Report from Dagstuhl Seminar 17292

Topology, Computation and Data Analysis

Edited by

Hamish Carr[1], Michael Kerber[2], and Bei Wang[3]

1 **University of Leeds, GB,** h.carr@leeds.ac.uk
2 **TU Graz, AT,** kerber@tugraz.at
3 **University of Utah – Salt Lake City, US,** wang.bei@gmail.com

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 17292 "Topology, Computation and Data Analysis". This seminar was the first of its kind in bringing together researchers with mathematical and computational backgrounds in addressing emerging directions within computational topology for data analysis in practice. The seminar connected pure and applied mathematicians, with theoretical and applied computer scientists with an interest in computational topology. It helped to facilitate interactions among data theorist and data practitioners from several communities to address challenges in computational topology, topological data analysis, and topological visualization.

Seminar July 16–21, 2017 – http://www.dagstuhl.de/17292

1998 ACM Subject Classification D.2.11 Software Architectures – Data abstraction, F.2.2 Nonnumerical Algorithms and Problems – Computations on discrete structures, I 3.5: Computational Geometry and Object Modeling – Geometric algorithms, languages, and systems

Keywords and phrases computational topology, topological data analysis, Topological data visualization

Digital Object Identifier 10.4230/DagRep.7.7.88

1 Executive Summary

Hamish Carr
Michael Kerber
Bei Wang

The Dagstuhl Seminar titled *Topology, Computation and Data Analysis* has brought together researchers with mathematical and computational backgrounds in addressing emerging directions within computational topology for data analysis in practice. The seminar has contributed to the convergence between mathematical and computational thinking, in the development of mathematically rigorous theories and data-driven scalable algorithms.

Context

In the last two decades, considerable effort has been made in a number of research communities into computational applications of topology. Inherently, topology abstracts functions and graphs into simpler forms, and this has an obvious attraction for data analysis. This attraction is redoubled in the era of extreme data, in which humans increasingly rely on tools that extract mathematically well-founded abstractions that the human can examine and reason about. In effect, topology is applied as a form of data compression or reduction: topology is one of the most powerful forms of mathematical compression that we know how to apply to data.

Efforts to apply topology computationally to data, however, have largely been fragmented so far, with work progressing in a number of communities, principally computational topology, topological data analysis, and topological visualization. Of these, computational topology expands from computational geometry and algebraic topology to seek algorithmic approaches to topological problems, while topological data analysis and topological visualization seeks to apply topology to data analysis, of graphs and networks in the first case and of (usually) simulated volumetric data in the second. The research in these communities can roughly be clustered into theory (what are the underlying mathematical concepts), applications (how are they used for data analysis), and computation (how to compute abstractions for real datasets). It is crucial to advances in this area that these three branches go hand-in-hand, and communication between theoretical, applied, and computational researchers are therefore indispensable. On the other hand, there has been surprisingly little communication between the computational topology and topological visualization communities, mostly caused by the fact that each community has its own set of regular venues. As a consequence, the linkages in the two communities have been independent of each other, and results can take years to migrate from one community to the other.

Vision

Our goal was therefore to soften the aforementioned rather strict separation between computational topology and topological visualization by establishing new inter-community ties. The seminar aimed to bring together cross sections of both communities, including researchers with theoretical, applied, and computational backgrounds. By reducing redundancy and accelerating cross-communication, we expected a significant boost to both areas, perhaps even leading to a singular more dynamic community. As a side effect, we also wanted to provide a communication platform within each community between theory and application.

Topics

We identified specific research topics reflecting emerging trends in both communities. These topics were chosen to span the spectrum from the theoretical (category theory), to applicable theory (multidimensional persistent homology), and from applied theory (singularity theory and fiber topology) to the computational (scalable topological computation, applications) aspect.

Category theory: theory and applications. Category theory has recently gained momentum in computational topology, in particular through sheaves and cosheaves, which are extremely useful as an alternative foundation for level set persistence. Recent work has shown that the data of a Reeb graph can be stored in a category-theoretic object called a cosheaf, and this opens the way to define a metric for Reeb graphs known as the interleaving distance. Sheaves can also be used in deriving theoretical understandings between the Reeb space and its discrete approximations. Research into sheaves and their relationship with computation is, however, in its infancy, and would benefit from pooling the resources of experts in category theory and topological data analysis, to address questions such as how to simplify theories in computational topology, how to reinterpret persistent homology, or how to compare topological structures.

Multidimensional persistent homology. The second area of active research, both mathematically and computationally, is the extension of unidimensional persistence to multidimensional persistence. Mathematically, the lack of a complete discrete invariant for the multidimensional case raises the theoretical question of identifying meaningful topological invariants to compute. Some earlier proposals have been complemented by recent approaches and raise the immediate question of computability and applicability. Besides the invariants themselves, other questions such as the comparison of multidimensional data, or the efficient generation of cell complexes suitable for the multidimensional case are crucial, but hardly studied questions in this context. Computationally, existing algorithms for topological constructs rely on filtrations to encapsulate a sweep order through the data, thus serializing the problem for algorithmic implementation. For multidimensional data, this serialization is hard to achieve, and progress in this area is, therefore, crucial for computational advances in the topological analysis of data.

Singularity theory and fiber topology in multivariate data analysis. Singularity theory and fiber topology both seek to extend Morse theory from scalar fields to multivariate data described as functions mapping $f : \mathbb{X} \to \mathbb{R}^d$. Since multivariate datasets are near-ubiquitous in scientific applications such as oceanography, astrophysics chemistry, meteorology, nuclear engineering and molecular dynamics, advances here are also crucial for topological data analysis and visualization. Methods from computational topology have been developed to support the analysis of scalar field data with widespread applicability. However, very few tools exist for studying multivariate data topologically: the most notable examples of these tools are the Jacobi set, the Reeb space, and its recent computational approximation, the Joint Contour Net. Here, we aim to bring together researchers in singularity theory, fiber topology and topological data analysis to develop new theory and algorithms driving a new generation of analytic tools.

Scalable computation. At the opposite pole from theory is the practical question: how do we apply topological analysis to ever-larger data sets? This question spans questions of algorithmic performance to the accuracy of representation: using the metaphor of compression, do we want lossy or lossless compression, how fast can we perform it, and what do we lose in the process? Moreover, the largest data sets are necessarily computed and stored on clusters, and scalability of topological computation therefore also depends on building distributed and parallel algorithms. For example, the standard algorithm for computing persistent homology is cubic in the number of simplices, but can be speeded up in theory and practice, and further improved by parallel computation. However, many challenges remain, including efficient generation, storage and management of simplicial complexes, streaming computation, I/O efficient computation, approximate computation, and non-simplicial complexes. Some of these approaches have already been applied in topological visualization, and cross-fertilization between the two communities is therefore of great interest.

Participants, Schedule, and Organization

The invitees were chosen according to the topics, bringing together enough expertise for each topic and resulting in a representative subset of both communities. Out of the 37 invited researchers in the first round, 28 accepted our invitation, pointing out the general interest for the seminar topic in both communities.

We decided for a mixed setup with introductory talks, contributed research talks and breakout sessions.

For the first day, we scheduled two overview talks per listed topic, which were delivered by Steve Oudot and Elizabeth Munch (Category theory), Michael Lesnick and Claudia Landi (Multidimensional persistent homology), Osamu Saeki and Julien Tierny (Singularity theory), and Yusu Wang and Valerio Pascucci (Scalable Computation). Further contributed talks by participants took place from Tuesday to Friday morning, resulting in a total of 19 contributed talks.

The afternoons of Tuesday and Thursday were used for breakout sessions. The format was different on the two days. Based on the discussions on Monday, we identified the topics "multivariate topology" and "scalable computation" as topics of general interest. We decided to let every participant discuss both topics, so we organized 4 discussion groups on multivariate topology in the early afternoon, and 3 discussion groups on scalable computation in the later afternoon (plus an alternative group with a different topic). We composed these groups mostly randomly, making sure that members of both communities are roughly balanced in each group. On Thursday afternoon, we let participants propose their topics of interest. 5 groups were formed discussing various aspects raised in contributed talks. On Wednesday and Friday morning, the outcomes of every discussion group were summarized and discussed in a plenary session.

Moreover, the majority of the participants joined an organized excursion to Trier on Wednesday afternoon.

Results and Reflection

The participants gave the unanimous feedback that the breakout sessions were a full success (and several proposed more time for such discussions in possible upcoming seminars). We first let people from a mixed background to discuss rather vague topics on Tuesday, and asked for specific topics on Thursday. Such an organizational plan led to a stimulating working environment, and helped to avoid idle breakout sessions.

We believe that we have fully achieved the goal of softening the separation between the two communities involved in this seminar. We expect visible evidence of newly formed inter-community ties fostered by the seminar, for instance through joint research projects and/or survey articles summarizing major open problems on the interface of both communities. To the best of our knowledge, 3 working groups are being formed and at least 1 position paper is underway that will combine expertise from both communities to tackle key research questions raised during the seminar.

2 Table of Contents

Working groups

3 Overview of Talks

3.1 Ripser: Efficient Computation of Vietoris-Rips Persistence Barcodes

Ulrich Bauer (TU München, DE)

I will discuss the efficient computation of the Vietoris-Rips persistence barcode for a finite metric space. The implementation in the newly developed C++ code "Ripser" focuses on memory and time efficiency, outperforming previous software by a factor of more than 40 in computation time and a factor of more than 15 in memory efficiency on typical benchmark examples. The improved computational efficiency is based on a close connection between persistent homology and discrete Morse theory, together with novel algorithmic design principles, avoiding the explicit construction of the filtration boundary matrix.

3.2 Interleaving Distance: Computational Complexity

Magnus Botnan (TU München, DE)

The computational complexity of computing the interleaving distance for multi-parameter persistent homology is not known. I will discuss a special instance of the problem which I believe is NP-Hard.

3.3 Dataflow EDSL: Parallel Topology Made Simple

Peer-Timo Bremer (LLNL – Livermore, US)

Efficient and scalable implementations, especially of more complex analysis approaches, require not only advanced algorithms but also an in-depth knowledge of the underlying runtime. Furthermore, different machine configurations and different applications may favor different runtimes, i.e., MPI vs. Charm++ vs Legion etc., and different hardware architectures. This diversity makes developing and maintaining a broadly applicable analysis software infrastructure challenging. We address some of these problems by explicitly splitting the definition and implementation of analysis and visualization algorithms. In particular, we present an embedded domain specific language (EDSL) to describe an algorithm as a generic task graph, that can be executed with different runtime backends (MPI, Charm++, Legion). We demonstrate the flexibility and performance of this approach using three different large-scale analysis and visualization use cases, i.e., topological analysis, rendering and compositing dataflow, and image registration of large microscopy scans. Despite the unavoidable overheads of a generic solution, our approach demonstrates performance portability at scale, and, in some cases, outperforms hand-optimized implementations.

3.4 What is Wrong with Time-Dependent Flow Topology?

Roxana Bujack (Los Alamos National Laboratory, US)

Vector field topology is a powerful visualization tool, because it can break down huge amounts of data into a compact, sparse, and easy-to-read description with little information loss. Visualization scientists struggle, because its generalization to time-dependent flow usually lacks a meaningful physical interpretation. We are looking for ways to overcome this problem.

3.5 Reeb Spaces, Fiber Surfaces and Joint Contour Nets

Hamish Carr (University of Leeds, GB)

Joint work of Hamish Carr, Julien Tierney, Aaron Knoll, David Duke, Amit Chattopadhyay, Zhao Geng, Osamu Saeki, Kui Wu, Pavol Klacansky, Valerio Pascucci
Main reference Kui Wu, Aaron Knoll, Benjamin J. Isaac, Hamish A. Carr, Valerio Pascucci: "Direct Multifield Volume Ray Casting of Fiber Surfaces", IEEE Trans. Vis. Comput. Graph., Vol. 23(1), pp. 941–949, 2017.
URL http://dx.doi.org/10.1109/TVCG.2016.2599040

Recent work in topological visualization has developed a set of tools for bivariate functions of the form $\mathbb{R}^3 \to \mathbb{R}^2$. Here, the analog of an isosurface is a fiber surface – the 2-manifold pre-image of a 1-manifold curve in the range. From this, the Reeb graph extends naturally to the Reeb space, which for bivariate functions is a 2-cell complex. This talk will give a summary of these recent developments, including variations on fiber surfaces and Reeb spaces, and some of the application-oriented results arising from the Joint Contour Net – a quantized approximation of the Reeb space.

3.6 A Discrete Gradient-Based Approach to Multivariate Data Analysis

Leila De Floriani (University of Maryland – College Park, US)

Joint work of Leila De Floriani, Federico Iuricich, Riccardo Fellegara, Sara Scaramuccia, Claudia Landi, Kenneth Weiss
Main reference Federico Iuricich, Sara Scaramuccia, Claudia Landi, Leila De Floriani: "A discrete morse-based approach to multivariate data analysis", in Proc. of the SIGGRAPH ASIA 2016, Macao, December 5-8, 2016 - Symposium on Visualization, pp. 5:1–5:8, ACM, 2016.
URL http://dx.doi.org/10.1145/3002151.3002166

In this talk, we present our recent work on topological analysis of big data based on discrete Morse theory, for applications to the efficient computation of (multi)persistent homology and to topology-based data visualization. In the first part of the talk, a new distributed data structure for simplicial complexes, the Stellar tree, is presented, which allows for an efficient generation and compact storage of the discrete Morse gradient and for an effective and efficient computation of the discrete Morse complex and its geometric embedding on very large data sets. Compactness and computational efficiency of the Stellar tree are demonstrated in comparison with state-of-the-art data structures for simplicial complexes. The second part of the talk has been focused on the case of multivariate data, i.e., data equipped with a vector-valued function. Such problem is especially relevant for computing multipersistent homology efficiently on large data sets and for investigating and extracting critical features of multivariate data, such as Pareto or Jacobi sets. Specifically, a new approach based on a discrete gradient compatible with the vector-valued function is presented, which has been proven to generate a chain complex which has the same persistent homology as the original input complex. This allows to drastically reduce the time and space required to compute the multipersistent module, as the results of our experiments with the public domain tool for multipersistent homology computation. Moreover, our preliminary results show theoretically anticipated connections between the critical simplices associated with the discrete gradient and Pareto sets, and form the basis for the current developments of this research.

3.7 Representations of Persistence and Time-Varying Persistence: Past, Present and Future

Pawel Dlotko (Swansea University, GB)

Standard computational topology pipeline barely considers the problem of post-processing of persistence diagrams. Yet, in data analysis, this is an important, if not essential step. Classical tools that allow for basic analysis of persistence diagrams are restricted to Wasserstein and Bottleneck distances. Yet, to use persistence as an input for standard statistics and machine learning algorithms, one requires more: in addition to be able to compute a distance between diagrams, one may need to average them, compute their scalar products, confidence bounds and similar. Some of those operations can be performed on persistence diagrams, but a lot of them are ambiguous on persistence diagrams. To address this issue, we will introduce various representations of persistence diagrams that implement all the mentioned operations. We will speculate on general, data-dependent representations and kernels, and discuss the existing implementations, including the implementation in Gudhi library. At the end, we will generalize all the introduced representations for time-varying persistence diagrams.

3.8 Topology-Guided Visual Exploratory Analysis

Harish Doraiswamy (New York University, US)

Joint work of Alex Bock, Theodoros Damoulas, Harish Doraiswamy, Nivan Ferreira, Juliana Freire, Claudio Silva, Adam Summers
Main reference Harish Doraiswamy, Nivan Ferreira, Theodoros Damoulas, Juliana Freire, Cláudio T. Silva: "Using Topological Analysis to Support Event-Guided Exploration in Urban Data", IEEE Trans. Vis. Comput. Graph., Vol. 20(12), pp. 2634–2643, 2014.
URL http://dx.doi.org/10.1109/TVCG.2014.2346449

Enormous amounts of data are being collected in different domains, from traditional ones such as biology to the more recent urban sciences. This has created new opportunities for using data-driven approaches to better support answering important questions that arise in these domains. Visualization and visual analytics systems have been successfully used to aid users obtain insight. However, manual (exhaustive) exploration of large data sets is not only time consuming, but often becomes impractical. It is therefore necessary to also guide users during this exploration process. Furthermore, it is also important that these tools be designed in a way that they are usable and within reach of domain experts who often lack computer science expertise. In this talk, I will present a few examples of how techniques from computational topology used in conjunction with visualization has been instrumental in guiding domain experts in their analysis process.

3.9 Persistence-Based Summaries for Metric Graphs

Ellen Gasparovic (Union College – Schenectady, US)

Joint work of Ellen Gasparovic, Maria Gommel, Emilie Purvine, Radmila Sazdanovic, Bei Wang, Yusu Wang, and Lori Ziegelmeier
Main reference Ellen Gasparovic, Maria Gommel, Emilie Purvine, Radmila Sazdanovic, Bei Wang, Yusu Wang, Lori Ziegelmeier, "A Complete Characterization of the 1-Dimensional Intrinsic Cech Persistence Diagrams for Metric Graphs", arXiv:1702.07379v2 [math.AT], 2017.
URL https://arxiv.org/abs/1702.07379

In this talk, we focus on giving a qualitative description of information that one can capture from metric graphs using certain topological summaries. In particular, we give a complete characterization of the persistence diagrams in dimension 1 for metric graphs under a particular intrinsic setting. We also look at two persistence-based distances that one may define for metric graphs and discuss progress toward establishing their discriminative capacities.

3.10 Robust Extraction and Simplification of 2D Tensor Field Topology

Ingrid Hotz (Linköping University, SE)

Joint work of Ingrid Hotz, Bei Wang, Jochen Jankowai

In this work, we propose a controlled simplification and smoothing strategy for symmetric 2D tensor fields that is based on the topological notion of robustness. Robustness measures the structural stability of the degenerate points with respect to variation of the underlying field. We consider an entire pipeline for the topological simplification of the tensor field by generating a hierarchical set of simplified fields based on varying the robustness values. Such a pipeline comprises of four steps: the stable exaction and classification of degenerate points, the computation and assignment of robustness values to the degenerate points, the construction of a simplification hierarchy, and finally the actual smoothing of the fields across multiple scales. We also discuss the challenges that arise from the discretization and interpolation of real world data.

3.11 The representation theorem of persistent homology revisited and generalized

Michael Kerber (TU Graz, AT)

Joint work of Rene Corbet, Michael Kerber
Main reference Rene Corbet, Michael Kerber, "The representation theorem of persistent homology revisited and generalized", arXiv:1707.08864v2 [math.AT], 2017.
URL https://arxiv.org/abs/1707.08864

The representation theorem by Zomorodian and Carlsson has been the starting point of the study of persistent homology under the lens of algebraic representation theory. In this work, we give a more accurate statement of the original theorem and provide a complete and self-contained proof. Furthermore, we generalize the statement from the case of linear sequences of R-modules to R-modules indexed over more general monoids. This generalization subsumes the representation theorem of multidimensional persistence as a special case.

3.12 Introduction to multidimensional persistent homology II

Claudia Landi (University of Modena, IT)

Joint work of Claudia Landi, Sara Scaramuccia, Federico Iuricich, Leila De Floriani

Many scientific fields need to study multivariate data. Multivariate data can be represented by multiple real-valued functions $f_1, f_2, \ldots, f_n : M \to \mathbb{R}$ defined on the same domain M, giving rise to a vector-valued function $f = (f_i) : M \to \mathbb{R}^n$. A sublevel set M^u of f at $u = (u_i) \in \mathbb{R}^n$ consists of those points p of M such that $f_i(p) \leq u_i$ for every $1 \leq i \leq n$. Varying u in $\mathbb{R}^n$ produces a multiparameter filtration of M by sublevel sets where $u_i \leq v_i$ for every $1 \leq i \leq n$ implies $M^u \subseteq M^v$. Multidimensional persistence detects the appearance and disappearance of homology features along this filtration with the multiparameter u varying in any increasing direction.

In the case when M is a smooth manifold and f is a smooth function, the values of the multiparameter u where homology features appear and disappear correspond to values taken at Pareto critical points. Intuitively, these are points where the gradients of the functions f_i disagree.

In the case when M is a simplicial complex and f is defined on its vertices and then extended to any simplex by taking the component-wise maximum over its vertices, a discrete gradient field compatible with the induced sublevel set filtration can be obtained by an algorithm based on homotopy expansion. The critical cells of such gradient field detect locations where homology classes are born and die along the filtration. In other words, they play, in the discrete setting, a role similar to that played by Pareto critical points in the smooth setting.

3.13 Topology Meets Machine Learning: How Both Fields Can Profit From Each Other

Heike Leitte (TU Kaiserslautern, DE)

Machine learning tries to reconstruct models from data, searching for underlying signals and patterns in it. A major hurdle is commonly noise and variations present in all real world data. Topological data analysis (TDA) provides a great set of tools to search for salient features in complex data, while filtering noise and short lived signals. In this talk, we will look at the major application fields of machine learning and see some examples of how they can be improved using modern TDA algorithms. We will also explore how visualisation can help to connect these two data analysis fields and make the results and the analysis process more easily accessible to the user.

3.14 An Introduction to Multidimensional Persistent Homology I

Michael Lesnick (Princeton University, US)

In topological data analysis, we often study data by associating to the data a filtered topological space, whose structure we can then examine using persistent homology. However, in many settings, a single filtered space is not a rich enough invariant to encode the interesting structure of our data. This motivates the study of multidimensional persistence, which associates to the data a topological space simultaneously equipped with two or more filtrations. The homological invariants of these "multi-filtered spaces," while much richer than their 1-dimensional counterparts, are also far more complicated. As such, adapting the usual 1-dimensional persistent homology methodology for data analysis to the multi-dimensional setting requires some new ideas. In this talk, I'll introduce multi-dimensional persistent homology and discuss some recent progress on this topic.

3.15 Introduction to categorical approaches in topological data analysis II

Elizabeth Munch (Michigan State University, US)

Arguably, the most beautiful mathematical idea coming from topological data analysis in the last decade is that of interleaving. An ϵ-interleaving can be thought of as an approximate isomorphism between two persistence modules with ϵ-allowed incorrectness. Using basic structures from category theory, one can think of a persistence module as a functor, then the ϵ-interleaving is a set of natural transformations between ϵ-shifted persistence modules which satisfy certain compatibility conditions. Once these ideas have been extended to category theory, we can then use the idea of interleaving for different choices of categories and functors to obtain known metrics, including bottleneck distance for persistence modules, Hausdorff distance for sets, and L_∞ for points or functions; as well as create new metrics for many disparate objects including Reeb graphs, mapper graphs, and multi-dimensional persistence modules.

3.16 Feature-Directed Visualization of Multifield Data

Vijay Natarajan (Indian Institute of Science – Bangalore, IN)

Main reference Vidya Narayanan, Dilip Mathew Thomas, Vijay Natarajan: "Distance between extremum graphs", in Proc. of the 2015 IEEE Pacific Visualization Symposium, PacificVis 2015, Hangzhou, China, April 14-17, 2015, pp. 263–270, IEEE Computer Society, 2015.
URL http://dx.doi.org/10.1109/PACIFICVIS.2015.7156386

Scientific phenomena are often studied through collections of related scalar fields generated from different observations of the same phenomenon. Exploration of such data requires a robust distance measure to compare scalar fields for tasks such as identifying key events and establishing a correspondence between features in the data. In this talk, I will pose the problem of designing appropriate distance measures to compare scalar fields in a feature-aware manner. Assuming that topological structures represent features in the data, what are good approaches towards the design of feature-aware distance measures between the scalar fields. In addition to provable properties, we will require the distance measure to be efficiently computable, and also interpretable.

3.17 Introduction to Categorical Approaches in Topological Data Analysis I

Steve Y. Oudot (INRIA Saclay – Île-de-France, FR)

The mathematical theory underlying topological data analysis, which is known as persistence theory, works at two different levels: the topological level, where it deals with nested families of topological spaces, as inspired from Morse theory; the algebraic level, where it deals with diagrams of vector spaces and linear maps, as inspired from quiver representation theory. While the objects involved in these two levels are very different in nature, they can be thought of as functors from partially ordered sets to some target categories. Category theory appears then as the right tool to build an abstraction of persistence, in which both levels can be cast and analyzed in tandem. This talk is therefore naturally divided into two parts: first, an introduction to the basics of category theory; second, an introduction to 1-dimensional persistence theory and its foundational results (decomposition, stability) from a categorical point of view.

3.18 A Stable Multi-Scale Kernel for Topological Machine Learning

Jan Reininghaus (Siemens Industry Software GmbH – Wien, AT)

Joint work of Jan Reininghaus, Stefan Huber, Ulrich Bauer, Roland Kwitt
Main reference Jan Reininghaus, Stefan Huber, Ulrich Bauer, Roland Kwitt: "A stable multi-scale kernel for topological machine learning", in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015, pp. 4741–4748, IEEE Computer Society, 2015.
URL http://dx.doi.org/10.1109/CVPR.2015.7299106

Topological data analysis offers a rich source of valuable information to study vision problems. Yet, so far we lack a theoretically sound connection to popular kernel based learning techniques, such as kernel SVMs or kernel PCA. In this work, we establish such a connection by designing a multi-scale kernel for persistence diagrams, a stable summary representation of topological features in data. We show that this kernel is positive definite and prove its stability with respect to the 1-Wasserstein distance. Experiments on two benchmark datasets for 3D shape classification and texture recognition show considerable performance gains of the proposed method compared to an alternative approach that is based on the recently introduced persistence landscapes.

3.19 Introduction to Singularity Theory and Fiber Topology in Multivariate Data Analysis

Osamu Saeki (Kyushu University – Fukuoka, JP)

Joint work of Yamamoto, Takahiro; Kawashima, Masayuki; Hiratuka, Jorge T.
Main reference Osamu Saeki, "Theory of singular fibers and Reeb spaces for visualization, Topological Methods in Data Analysis and Visualization IV – Theory, Algorithms, and Applications", Proc. Topology-Based Methods in Visualization 2015, pp. 3–33, Springer, 2017.
URL https://doi.org/10.1007/978-3-319-44684-4

In this talk, we consider generic differentiable maps between differentiable manifolds, and propose a mathematical formulation of fibers from the viewpoint of singularity theory. In fact, this formulation is shown to be essential also for visualization purposes. Then, classification results of fibers for certain dimension pairs are presented. We also present results on local characterizations of Reeb spaces. Our study of fibers for maps of 4-dimensional manifolds into surfaces indicates a (possibly new) concept of a Reeb diagram, which is expected to be a source of new problems. Some computational problems will also be presented from mathematical viewpoints.

References

1 O. Saeki, Topology of singular fibers of differentiable maps, Lecture Notes in Math., Vol. 1854, Springer Verlag, 2004.
2 O. Saeki, Theory of singular fibers and Reeb spaces for visualization, Topological Methods in Data Analysis and Visualization IV – Theory, Algorithms, and Applications, H. Carr, C. Garth, T. Weinkauf (Eds.), Proc. Topology-Based Methods in Visualization 2015, pp. 3–33, Springer, 2017.
3 O. Saeki and T. Yamamoto, Singular fibers of stable maps of 3–manifolds with boundary into surfaces and their applications, Algebraic and Geometric Topology 16, 1379–1402, 2016.

4 D. Sakurai, O. Saeki, H. Carr, Hsiang-Yun Wu, T. Yamamoto, D. Duke, and S. Takahashi, Interactive visualization for singular fibers of functions $f : \mathbb{R}^3 \to \mathbb{R}^2$, IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 945–954, 2016.

3.20 A Topological Visualization Approach to Combinatorial Optimization

Gerik Scheuermann (Universität Leipzig, DE)

Topological data visualization has been applied to many different domain. In this talk, we take a look at a discipline that has rarely been studied by topological visualization – despite a very close relation to topology. We look at combinatorial optimization. In this discipline, there are many topological considerations, but hardly any topological visualization approaches. Obviously, optimization is hard if there are many local extrema, otherwise it is easy. Therefore, a topological study of the problem can provide insight. We show that for enumerable problems (even with millions or billions of points), topological visualization allows to visually study algorithmic behavior which is not possible with typically used visualizations. Thus, while these problems are still very small, parameter tuning and testing of optimization algorithms is simplified. For larger instances, we show that sampling of the landscape is a promising direction to go.

3.21 Noise Systems and Multidimensional Persistence

Martina Scolamiero (EPFL – Lausanne, CH)

In this talk I will introduce a framework that allows to compute a new class of stable discrete invariants for multidimensional persistence. In doing this, we generalise the notion of interleaving topology on multidimensional persistence modules by using noise systems. A filter function is usually chosen to highlight properties we want to examine from a dataset. Similarly, our new topology allows some features of datasets to be considered as noise.

3.22 Discrete Morse Theory and Simplicial Map Persistence

Donald Sheehy (University of Connecticut – Storrs, US)

One efficient way to compute the persistent homologous of simplicial maps involves converting the sequence of complexes into a proper filtration. In this talk, I will show that this approach follows naturally from discrete Morse theory on the mapping telescope.

3.23 Spectral Sequences for Parallel Computation

Primoz Skraba (Jozef Stefan Institute – Ljubljana, SI)

Spectral sequences represent a family of incremental algorithms for computing topological invariants. They are a fundamental tool used by both algebraic topologists and homological algebraists, most often to compute (co)homology, although in some cases also more difficult invariants such as homotopy groups. Often they are difficult to follow due to extensive notation and because they are generally applied to difficult problems (making the examples themselves difficult), In our case, however, the Mayer-Vietoris spectral sequence, a special case of the Leray spectral sequence (which is itself a special case of the Grothendieck spectral sequence), provides an algorithm for computing (co)homology. In this talk, we introduced what the spectral sequence is from an algorithmic point of view. We showed how to set it up and the operations which must be efficiently implemented in order to make the algorithm as a whole efficient. We concentrated on the structure which indicates how much parallelization can be achieved with this approach as well as discuss the obstacles which remain in order to extend this to persistence.

3.24 Topological Analysis of Bivariate Data

Julien Tierny (CNRS-UPMC – Paris, FR)

Main reference Julien Tierny, Guillaume Favelier, Joshua A. Levine, Charles Gueunet, Michael Michaux: "The Topology ToolKit", IEEE Trans. Vis. Comput. Graph., Vol. 24(1), pp. 832–842, 2018.
URL http://dx.doi.org/10.1109/TVCG.2017.2743938

Multivariate scalar data sets are becoming increasingly popular in scientific visualization applications, since modern numerical simulations and acquisition devices have now the ability to simultaneously track a large number of variables on a single geometrical domain. Thus, the topological methods developed over the last twenty years for the analysis and visualization of univariate scalar data need to be completely revisited in that setting. The bivariate case is an appealing first step in this generalization effort, in particular since users often tend to project multivariate functions to the (two-dimensional) screen in the form of 2D scatterplots for visualization purposes. This talk reviews recent algorithms for the extension to the bivariate case of the notions of level sets (to fibers), critical points (to Jacobi sets) and Reeb graphs (to bivariate Reeb spaces). Applications to continuous scatterplot peeling, silhouette simplification, medial structure computation and feature similarity estimation are discussed. Finally, I will present current research problems, including the question of Jacobi set simplification, for which solutions are expected to eventually enable a wide adoption of bivariate topological data analysis in scientific visualization applications.

3.25 Generalizations of the Rips Filtration for Quasi-Metric Spaces with Corresponding Stability Results

Katharine Turner (Australian National University – Canberra, AU)

Rips filtrations over a finite metric space and their corresponding persistent homology are prominent methods in Topological Data Analysis to summarize the "shape" of data. For finite metric space X and distance r the traditional Rips complex with parameter r is the flag complex whose vertices are the points in X and whose edges are $\{[x, y] : d(x, y) \leq r\}$. From considering how the homology of these complexes evolves we can create persistence modules (and their associated barcodes and persistence diagrams). Crucial to their use is the stability result that says if X and Y are finite metric space then the bottleneck distance between persistence modules constructed by the Rips filtration is bounded by $2d_{GH}(X, Y)$ (where d_{GH} is the Gromov-Hausdorff distance). Using the asymmetry of the distance function we construct four different constructions analogous to the persistent homology of the Rips filtration and show they also are stable with respect to the Gromov-Hausdorff distance. These different constructions involve ordered-tuple homology, symmetric functions of the distance function, strongly connected components and poset topology.

3.26 Scalable Computation in Computational Topology

Yusu Wang (Ohio State University – Columbus, US)

Recent years have witnessed a tremendous amount of growth in the field of computational topology. In addition to significant theoretical and algorithmic developments, topological methods have been used in various application domains, including in visualization. With the rapid increase in the number of applications and in the scale of data sizes, it is important that topological methods are scalable and can handle the challenge of mass data sizes. In this talk, I will survey some of the algorithmic efforts for topological methods, with a special focus on the computation of persistent homology (in various settings). I will also briefly touch on the computation of the Reeb graph and related structures.

4 Working groups

4.1 Discussion group on "Mean Reeb Graphs"

Ellen Gasparovic (Union College – Schenectady, US), Peer-Timo Bremer (LLNL – Livermore, US), Ingrid Hotz (Linköping University, SE), Elizabeth Munch (Michigan State University, US), Vijay Natarajan (Indian Institute of Science – Bangalore, IN), Steve Y. Oudot (INRIA Saclay – Île-de-France, FR), Julien Tierny (CNRS-UPMC – Paris, FR), Katharine Turner (Australian National University – Canberra, AU), and Bei Wang (University of Utah – Salt Lake City, US)

The goal of our working group was to make precise the notion of a *mean Reeb graph* and discuss how to compute it. We agreed that such a mean should be a descriptive statistic that lends itself readily to topological and geometric interpretation.

We began by asking ourselves many questions, including:

- If we obtain the same Reeb graph from two different functions, should the mean be the same independent of the functions giving rise to them?
- Do we want the metric we use to depend on the original functions?
- Should the mean depend on the application?
- Should we use a feature-based distance and/or a spatial distance?
- Would some sort of *augmented* or *labeled* Reeb graph be desirable?
- Should we consider a whole set of Fréchet means from different metrics?
- Is this related to the notion of "tensor swelling"? Or perhaps uncertainty visualization for contour trees?

We then discussed possible metrics that one can define on the space of Reeb graphs, so that the associated Fréchet means of sets of Reeb graphs have nice properties. Possibilities included the functional Gromov-Hausdorff distance, interleaving and functional distortion (FD) distances [1, 2], as well as the persistence distortion distance [4]. We decided to first look at several simple examples involving pairs of Reeb graphs, figure out what we thought the means should be in those instances, and then find a distance that would yield the desired means.

Later in the week, we focused on the induced intrinsic bottleneck distance $\hat{d}_B$ of [3], i.e., given Reeb graphs R_f and R_g, we have

$$\hat{d}_B(R_f, R_g) := \inf_{\gamma} |\gamma|_B$$

where γ ranges over all paths $\gamma : [0,1] \to \mathit{Reeb}$ ($\gamma(0) = R_f$ and $\gamma(1) = R_g$) that are continuous in d_{FD}, $|\gamma|_B = \sup_{n,\Sigma} \sum_{i=1}^{n-1} d_B(\gamma(t_i), \gamma(t_{i+1}))$ ($n \in \mathbb{N}$ and Σ ranges over all partitions $0 = t_0 \leq t_1 \leq \ldots \leq t_n = 1$ of $[0,1]$), and d_B is the usual bottleneck distance. This distance has many nice properties, including the fact that it is globally equivalent to the similarly defined intrinsic version of the functional distortion distance, $\hat{d}_{FD}$, which implies that they both induce the same topology on *Reeb* [3].

The next step is to prove that the bottleneck distance between Reeb graphs is locally intrinsic in a certain sense, in the same way as Carriére and Oudot showed that it is in the space of persistence diagrams.

References

1. Ulrich Bauer, Xiaoyin Ge, and Yusu Wang. *Measuring distance between Reeb graphs*. Proc. of the 30th Symposium on Computational Geometry, 2014.
2. Ulrich Bauer, Elizabeth Munch, and Yusu Wang. *Strong Equivalence of the Interleaving and Functional Distortion Metrics for Reeb Graphs*. Proc. of the 31st Symposium on Computational Geometry, 2015.
3. Mathieu Carrière and Steve Oudot. *Local Equivalence and Intrinsic Metrics between Reeb Graphs*. Proc. of the 33rd Symposium on Computational Geometry, 2017.
4. Tamal K. Dey, Dayu Shi, and Yusu Wang. *Comparing Graphs via Persistence Distortion*. Proc. of the 31st Symposium on Computational Geometry, 2015.

4.2 Discussion Group on "Multivariate Topology"

Michael Kerber (TU Graz, AT), Harish Doraiswamy (New York University, US), Christoph Garth (TU Kaiserslautern, DE), Claudia Landi (University of Modena, IT), Michael Lesnick (Princeton University, US), Jan Reininghaus (Siemens Industry Software GmbH – Wien, AT), and Julien Tierny (CNRS-UPMC – Paris, FR)

The discussion group started by brainstorming a list of possible questions to address. The list of topics includes:

- Computing the matching distance exactly or finding better ways of approximating it
- Simplification of Jacobi sets/Reeb spaces (in a PL setting)
- Good ways to depict Reeb spaces
- Stable kernels for multi-dimensional persistence (with connection to machine learning)
- Approximate Reeb space in non-manifold PL-setting
- Definition of a critical point in a multi-filtration that works in a PL-domain/combinatorially
- More (convincing) application scenarios (e.g., ensemble classifications in weather simulations)

Since this list was impossible to cover in the short time, the discussion focussed on specific aspects related to these questions:

- For the visualization community, the denoising aspect of persistent homology is important, but there are two issues: measuring importance and localization. Moreover, the point was raised that localization is not stable, as the pairing of critical points obtained from persistent homology can change a lot after small perturbations (even though the persistence diagram is stable). Do such effects also occur in the case of Reeb spaces?
- For the case of time-varying data, the common approach is to simplify each time step separately, but the major problem is to link these time frames. A mapper-style approach for time-varying data has been used in visualization in the past.
- A problem of current approaches is that they depend on or favor the chosen coordinate directions. A point was raised with respect to whether randomized constructions could help to eliminate such effects.
- For 1-dimensional data sets, a simplicial simplification based on persistence has been proposed by Bauer et al. To extend this to higher dimensions, the question is whether one can pick compatible slices. This relates to the talk by Claudia Landi, and the discussion reviewed some aspects of her talk.

- While in Claudia Landi's talk, the setup was with two functions and two dimensions, a question came up as what happens if there are more functions than dimensions. Based on Osamu's talk, it seems that everything becomes more difficult in that case. From the perspective of a researcher in visualization, it is a strange phenomenon that adding one additional function removes the "niceness" from the problem.
- A question was discussed as what complexes can arise as 2-dimensional Reeb spaces (with a generic manifold input). There seemed to be an agreement that it is not a manifold in general. It was discussed whether it is a pure complex (without any definite answer).

As a wrap-up, it was commonly agreed that being able to compare multi-dimensional modules is essential to visualization applications and to many of the discussed questions. A kernel for multi-dimensional persistence would add to the range of applications. Moreover, a first implementation for the (approximate) matching distance is currently in preparation.

4.3 Discussion Group on "Scalable Computation"

Michael Kerber (TU Graz, AT), Peer-Timo Bremer (LLNL – Livermore, US), Leila De Floriani (University of Maryland – College Park, US), Michael Lesnick (Princeton University, US), Vijay Natarajan (Indian Institute of Science – Bangalore, IN), and Primoz Skraba (Jozef Stefan Institute – Ljubljana, SI)

The group started with a discussion on parallelizing the computation of persistent homology. First, it was discussed whether existing approaches for merge trees extend to higher-dimensional homology in a simple way. The conclusion was that this is not the case, mostly because it appears hard to obtain global information (like homology) based on local computations.

For the existing approaches to parallel computation, it is usually the case that one node does a substantial amount of work in the end. The question is whether this can be avoided. Again, the fact that persistence computation is, in fact, a linear algebra problem prevents an easy application of local computation.

Moreover, the possibility of a GPU implementation for persistent homology was briefly discussed.

It was established that "scalable" has different meanings in different communities: for the visualization community, it mostly means parallelizable algorithms (which were also the main topic of discussion in this group), but in the field of algorithm design, it also means asymptotically faster algorithms. It seems that for point cloud data, the latter aspect should currently be in focus because current approaches are too slow even if they would be parallelized. For the case of cubical data, however, the algorithms appear optimal from an asymptotic point of view, and hence parallelization becomes important for performance.

Another point of discussion was the computation of high-dimensional persistent homology. Indeed, there are application domains whether high-dimensional simplices arise quite naturally, for instance in robotics. The question of whether the homology information in dimension 4 (or higher) is useful without interpretability was a controversial point of discussion. Some participants referred to machine learning, where sometimes the learned features can also not be interpreted in many cases. It was questioned by others whether this is the right way to go.

The group agreed that more work should go into the practical efficiency of low-dimensional Rips complexes with a low distance threshold. This includes topics such as the distributed creation of the Rips complex and distributed nearest-neighbor queries.

Finally, there was a request about a state of the art report on the sizes of data sets that can be handled by current approaches.

Participants

- Ulrich Bauer
TU München, DE
- Magnus Botnan
TU München, DE
- Peer-Timo Bremer
LLNL – Livermore, US
- Roxana Bujack
Los Alamos National Laboratory, US
- Hamish Carr
University of Leeds, GB
- Leila De Floriani
University of Maryland – College Park, US
- Pawel Dlotko
Swansea University, GB
- Harish Doraiswamy
New York University, US
- Brittany Terese Fasy
Montana State University – Bozeman, US
- Christoph Garth
TU Kaiserslautern, DE
- Ellen Gasparovic
Union College – Schenectady, US
- Hans Hagen
TU Kaiserslautern, DE
- Ingrid Hotz
Linköping University, SE
- Michael Kerber
TU Graz, AT
- Claudia Landi
University of Modena, IT
- Heike Leitte
TU Kaiserslautern, DE
- Michael Lesnick
Princeton University, US
- Elizabeth Munch
Michigan State University, US
- Vijay Natarajan
Indian Institute of Science – Bangalore, IN
- Steve Y. Oudot
INRIA Saclay – Île-de-France, FR
- Valerio Pascucci
University of Utah – Salt Lake City, US
- Jan Reininghaus
Siemens Industry Software GmbH – Wien, AT
- Osamu Saeki
Kyushu University – Fukuoka, JP
- Gerik Scheuermann
Universität Leipzig, DE
- Martina Scolamiero
EPFL – Lausanne, CH
- Donald Sheehy
University of Connecticut – Storrs, US
- Primoz Skraba
Jozef Stefan Institute – Ljubljana, SI
- Julien Tierny
CNRS-UPMC – Paris, FR
- Katharine Turner
Australian National University – Canberra, AU
- Bei Wang
University of Utah – Salt Lake City, US
- Yusu Wang
Ohio State University – Columbus, US