Topological Data Analysis Made Easy with the Topology ToolKit

Guillaume Favelier — Sorbonne Universite
Charles Gueunet — Kitware, Sorbonne Universite
Attila Gyulassy — SCI Institute, University of Utah
Julien Jomier — Kitware
Joshua A. Levine — University of Arizona
Jonas Lukasczyk — TU Kaiserslautern
Daisuke Sakurai — Zuse Institute Berlin
Maxime Soler — Total, Sorbonne Universite
Julien Tierny — CNRS, Sorbonne Universite
Will Usher — SCI Institute, University of Utah
Qi Wu — SCI Institute, University of Utah & UC Davis

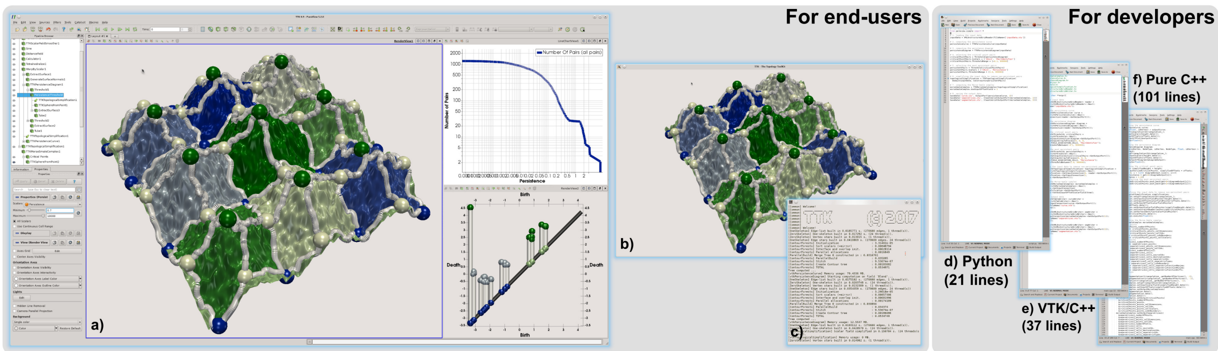

Figure 1: TTK is a software platform for topological data analysis in scientific visualization. It is both easily accessible to end users (ParaView plugins (a), VTK-based generic GUIs (b) or command-line programs (c)) and flexible for developers (Python (d), VTK/C++ (e) or dependence-free C++ (f) bindings). TTK provides an efficient and unified approach to topological data representation and simplification, which enables in this example a discrete Morse-Smale complex (a) to comply to the level of simplification dictated by a piecewise linear persistence diagram (bottom-right linked view, a). Code snippets are provided (d-f) to reproduce this pipeline.

1 Level of the Tutorial

This tutorial is targeted at a *Beginner* audience.

2 Abstract

This tutorial presents topological methods for the analysis and visualization of scientific data from a user's perspective, with the Topology ToolKit (TTK), a recently released open-source library for topological data analysis. Topological methods have gained considerably in popularity and maturity over the last twenty years and success stories of established methods have been documented in a wide range of applications (combustion, chemistry, astrophysics, material sciences, etc.) with both acquired and simulated data, in both post-hoc and in-situ contexts. While reference textbooks have been published on the topic, no tutorial at IEEE VIS has covered this area in recent years, and never at *a software level* and from *a user's point-of-view*. This tutorial fills this gap by providing a beginner's introduction to topological methods for practitioners, researchers, students, and lecturers. In particular, instead of focusing on theoretical aspects and algorithmic details, this tutorial focuses on how topological methods can be useful in practice for concrete data analysis tasks such as segmentation, feature extraction or tracking. The tutorial describes in detail how to achieve these tasks with TTK. First, after an introduction to topological methods and their application in data analysis, a brief overview of TTK's main entry point for end users, namely ParaView, will be presented. Second, an overview of TTK's main features will be given. A running example will be described in detail, showcasing how to access TTK's features via ParaView, Python, VTK/C++, and C++. Third, hands-on sessions will concretely show how to use TTK in ParaView for multiple, representative data analysis tasks. Fourth, the usage of TTK will be presented for developers, in particular by describing several examples of visualization and data analysis projects that were built on top of TTK. Finally, some feedback regarding the usage of TTK as a teaching platform for topological analysis will be given. Presenters of this tutorial include experts in topological methods, core authors of TTK as well as active users, coming from academia, labs, or industry. A large part of the tutorial will be dedicated to hands-on exercises and a rich material package (including TTK pre-installs in virtual machines, code, data, demos, video tutorials, etc.) will be provided to the participants. This tutorial mostly targets students, practitioners and researchers who are not experts in topological methods but who are interested in using them in their daily tasks. We also target researchers already familiar to topological methods and who are interested in using or contributing to TTK.

3 Tutorial Organization

Motivations As scientific datasets become more intricate and larger in size, advanced data analysis algorithms are needed for their efficient visualization and exploration. For scalar field visualization, topological analysis techniques [20, 26, 33] have shown to be practical solutions in various contexts by enabling the concise and complete capture of the structure of the input data into high-level *topological abstractions* such as contour trees [5, 13, 14], Reeb graphs [3, 24, 25, 36], or Morse-Smale complexes [8, 16, 17]. Such topological abstractions are fundamental data structures that enable advanced data analysis, exploration and visualization techniques, including for instance: small seed set extraction for fast isosurface traversal [6, 27, 40], feature tracking [29], transfer function design for volume rendering [43], data simplification [37] and compression [30], similarity estimation [32, 38], geometry processing [34, 41], or application-driven segmentation and analysis tasks. Successful applications in a variety of fields of science, including combustion [4, 15, 22], fluid dynamics [7, 21], material sciences [11, 18, 19, 23], chemistry [2, 12], and astrophysics [28, 31], have been documented, which further demonstrates the importance of these techniques.

While reference textbooks have been published [9] that present the fundamental aspects of these techniques, no tutorial has covered this area in recent years at IEEE VIS. The latest tutorial related to topology occurred nearly 10 years ago [42]. Moreover, despite their popularity and success in applications, topological methods have not yet been widely adopted as a standard data analysis tool for end users and developers. We believe one of the reasons for this is the lack of open-source software packages that implement these algorithms in a generic, user-friendly, and efficient way. Recently, the Topology ToolKit (TTK) [35], an open-source library for topological data analysis has been released (BSD license) to fill this gap. TTK is mostly written in C++ ($\sim$ 210k lines of code) and 8 institutions have contributed to its development so far, including 5 academic institutions (CNRS, Sorbonne Universite, University of Utah, Zuse Institute, University of Arizona) and 3 private companies (Kitware, Total, Caboma Inc.). TTK is currently supported under Linux, MacOS, and Windows. Since its initial release on Github a year ago, TTK's website collected more than 64k page-views, from more than 6.7k unique visitors. These statistics indicate that a user base for TTK exists and that further efforts towards the explanation of TTK's usage would be beneficial to the community.

The main motivation for this tutorial is therefore to introduce to beginners how topological methods can be useful for analyzing their data, and how to do it with TTK.

Target audience This tutorial mostly targets beginners, students, practitioners, and researchers who are not experts in topological methods but who are interested in using them in their daily tasks. It also targets researchers already familiar to topological methods and who are interested in using or contributing to TTK.

Tutorial goals The goals of this tutorial are to present the key tools in topological data analysis (the Persistence diagram, the Reeb graph and its variants, the Morse-Smale complex, etc.) and how they can be used in practice for precise data analysis tasks, including data segmentation and feature extraction. All examples will be illustrated with TTK. This tutorial also aims at presenting TTK and its different usage modalities (ParaView, Python, VTK/C++, C++). We expect participants to become capable of using TTK independently, at least with ParaView (possibly with Python), after attending the tutorial.

Hands-on material A large part of the tutorial will be dedicated to hands-on exercises with TTK and ParaView [1]. We will provide a rich material package including TTK pre-installs in virtual machines (to be used by attendees during the tutorial), code, data, demos, video tutorials, etc. Most of this material is already available on TTK's website [10]. Our idea is that participants who bring a laptop will be able to follow along, regardless of their native OS. Attendees who attend just to listen and learn will also benefit from the tutorial and receive sufficient material to try out our examples at home.

Qualification of the presenters Presenters include experts in topological methods, who have published many papers on the topic in premier venues as well as textbooks (Sect. 4). The list of presenters also includes the core authors of TTK and its most active users.

Proposal strengths In contrast to previous tutorials on topological methods [42], we believe this proposal to have a unique concrete and applicative appeal, by its focus on the *usage* of topological methods rather than on their *foundations*. Thus, we expect it to attract a larger audience than the specific subset of IEEE VIS attendees typically found in traditional topology sessions. Also, this tutorial is unique in the sense that TTK has never been presented before in a tutorial.

We believe that the list of presenters is also a strength of this proposal. First, it includes topology experts as well as core developers and users of TTK. More importantly, it includes researchers with a variety of experience profiles (Ph.D. students, post-docs, professors) and backgrounds (industry, labs, academia), which will ease interactions with a potentially heterogeneous audience. Moreover, the particularly large number of presenters (11) has two merits. First, it imposes a mini-symposium structure, where speakers will give presentation lasting between 10 and 20 minutes, which will result in a lively rhythm in the overall tutorial. Second, this large number of presenters will be instrumental during the hands-on exercises, as there will be enough presenters such that one presenter can assist a small group of attendees (typically 3 to 4).

Finally, we believe the detailed program of the tutorial (see Sect. 3) achieves a balance between concepts, usage descriptions and application examples.

Detailed content The tutorial is divided into three main parts (each part being subdivided into modules), for a target duration of approximately 3 hours. These three groups of modules can be organized differently to to fit any standard structure for breaks to match the tutorial schedule of IEEE VIS.

A. Preliminaries (60 minutes)

A1. General introduction *(5 minutes, by Julien Tierny)*

A2. Introduction to topological methods for data analysis *(30 minutes, by Attila Gyulassy)* This talk will present the core tools in topological data analysis (the Persistence diagram [9], the Reeb graph and its variants [3, 5, 14, 24, 25, 36], the Morse-Smale complex [8, 16, 17]). In particular, it will detail how these tools can be used for data segmentation and feature extraction.

A3. Quick introduction to ParaView's user interface *(25 minutes, by Julien Jomier)* This talk will provide a brief description of ParaView's main interface [1], in order to support its usage for beginners in the subsequent hands-on session. This will cover the usage of filters, pipeline design and view manipulation, state files backups and Python exports.

B. Hands-on exercises (70 minutes)

B1. General usage of TTK *(10 minutes, by Julien Tierny)* This talk will briefly describe TTK's usage philosophy. It will briefly present how TTK can be used from ParaView, Python, VTK/C++ or C++.

B2. Segmenting medical data with merge trees *(20 minutes, by Charles Gueunet)* This hands-on TTK/ParaView exercise will be a step-by-step tutorial showing how to extract individual bones in a medical CT scan interactively with merge trees.

B3. Extracting filament structures with the Morse-Smale complex *(20 minutes, by Guillaume Favelier)* This hands-on TTK/ParaView exercise will show step-by-step how to extract filament structures with the Morse-Smale complex on chemistry data.

B4. Topology-aware data compression *(20 minutes, by Maxime Soler)* This hands-on TTK/ParaView exercise will show step-by-step how to compress data while guaranteeing feature preservation.

C. Advanced usage (60 minutes)

C1. TTK's architecture and core data structures *(10 minutes, by Will Usher)* This talk will present TTK's architecture.

C2. Topology driven volume rendering with TTK *(10 minutes, by Qi Wu)* This talk will present topo-vol [39], an implementation of topology-driven volume rendering [43] built on top of TTK.

C3. Advanced data analysis with TTK *(10 minutes, by Jonas Lukasczyk)* This talk will describe the processing of the SciVis 2018 contest data with TTK.

C4. Bivariate data analysis with TTK *(10 minutes, by Daisuke Sakurai)* This talk will describe a user interface for bivariate data exploration built on top of TTK.

C5. TTK as a teaching platform *(15 minutes, by Joshua Levine)* This talk will provide feedback about our experience in using TTK in our topological data analysis classes.

C6. Concluding remarks *(5 minutes, by Julien Tierny)*

4 Background and Contact Information

Guillaume Favelier – *guillaume.favelier@sorbonne-universite.fr* – is a research engineer at Sorbonne Universite since late 2015. He received the master degree in Image Processing from Sorbonne Universite, also in 2015. His notable contributions to TTK include the implementation of the Morse-Smale complex and the implicit triangulation. Guillaume regularly demonstrates TTK capabilities through live demos and tutorials. His research interests include discrete geometry, rendering algorithms and visualization methods.

Charles Gueunet – *charles.gueunet@kitware.com* – is a Ph.D. student with Kitware and Sorbonne Universite. Charles received the engineering degree in 2015 from EISTI. His notable contributions to TTK include the implementation of the merge tree features.

Attila Gyulassy – *jediati@sci.utah.edu* – received the bachelors of Arts in computer science and applied mathematics from the University of California, Berkeley in 2003 and the PhD degree in computer science from the University of California, Davis in 2009. His research interests as a research scientist at the Scientific Computing and Imaging (SCI) Institute, University of Utah, include topology-based data analysis and visualization.

Julien Jomier – *julien.jomier@kitware.com* – is currently directing Kitware's European subsidiary in Lyon, France. Julien received both his B.S. and M.S in Electrical Engineering and Information Processing in 2002 from the ESCPE-Lyon (France) and an M.S. in Computer Science from The University of North Carolina at Chapel Hill (UNC) in 2003. He worked on a variety of projects in the areas of parallel and distributed computing, mobile computing, image processing, and visualization. Prior to joining Kitware, Mr. Jomier was a Faculty Research Lecturer of Radiology at UNC and a member of the Computer-Aided Diagnosis and Display Laboratory.

Joshua A. Levine – *josh@email.arizona.edu* – is an assistant professor in the Department of Computer Science at University of Arizona. Prior to starting at Arizona, he was an assistant professor at Clemson University, and before that a postdoctoral research associate at the University of Utah's SCI Institute. He received his PhD from The Ohio State University. His research interests include visualization, geometric modeling, topological analysis, mesh generation, vector fields, performance analysis, and computer graphics.

Jonas Lukasczyk – *jl@jluk.de* – is currently a research associate and Ph.D. student in the Visual Information Analysis Group at the Technical University of Kaiserslautern (TU-KL). He received his Bachelor and Master degree in Computer Science at the TU-KL in 2012 and 2015. His work focuses on Topology-Based Visual Analytics for Extreme Scale Simulations. His research interests include topological analysis, HPCP, visualization, and web technologies. Several of his approaches use TTK as a backbone for reliable and reproducible data analysis. For more information on his current work visit www.jluk.de.

Daisuke Sakurai – *sakurai@zib.be* – is a postdoctoral researcher at Zuse Institute Berlin (ZIB). He received his Ph.D. at the University of Tokyo in 2015, at which time he was also a special student at the Japan Atomic Energy Agency (JAEA). After graduation he worked at JAEA, Institute of Physical and Chemical Research (aka RIKEN), and the Computer Science Department (LIP6) of Sorbonne Universite. From April 2017 he is at the current position working for the project *High Definition Clouds and Precipitation for advancing Climate Prediction* (HD(CP)^2). His research interests include multi-field visualization, atmospheric sciences, topological analysis, and mathematical visualization.

Maxime Soler – *soler.maxime@total.com* – is a Ph.D. student with Total and Sorbonne Universite. Maxime received the engineering degree in 2015 from EISTI. His notable contributions to TTK include the implementation of topology-aware compression and distances between persistence diagrams.

Julien Tierny – *julien.tierny@sorbonne-universite.fr* – received the Ph.D. degree in Computer Science from Lille 1 University in 2008 and the Habilitation degree (HDR) from Sorbonne Universite in 2016. He is currently a CNRS permanent research scientist, affiliated with Sorbonne Universite since September 2014 and with Telecom ParisTech from 2010 to 2014. Prior to his CNRS tenure, he held a Fulbright fellowship (U.S. Department of State) and was a post-doctoral research associate at the Scientific Computing and Imaging Institute at the University of Utah. His research expertise includes topological data analysis for scientific visualization. He is the lead developer of the Topology ToolKit (TTK).

Will Usher – *will@sci.utah.edu* – is a PhD student at the SCI Institute at the University of Utah. His research interests include interactive ray tracing, virtual reality, and distributed rendering. He has used TTK to develop a topology guided volume visualization tool, and recently rewritten TTK's build system to follow a modern CMake approach, making it easier to use.

Qi Wu – *qadwu@ucdavis.edu* – is a PhD candidate at the Computer Science Department of the University of California, Davis. Prior to UC Davis, he obtained his Master degree from the SCI institute, University of Utah. His research interests include scientific visualization, ray tracing and distributed rendering on large scale systems. Together with Will Usher, he has used TTK to develop TopoVol, which is a topology guided volume exploration tool enabling sophisticated volume classification and rendering.

ACKNOWLEDGMENTS

This work is partially supported by the European Commission grant H2020-FETHPC-2017 "VESTEC" (ref. 800904).

REFERENCES

[1] J. Ahrens, B. Geveci, and C. Law. Paraview: An end-user tool for large-data visualization. *The Visualization Handbook*, pp. 717–731, 2005.
[2] H. Bhatia, A. G. Gyulassy, V. Lordi, J. E. Pask, V. Pascucci, and P.-T. Bremer. Topoms: Comprehensive topological exploration for molecular and condensed-matter systems. *J. of Comp. Chem.*, 2018.
[3] S. Biasotti, D. Giorgio, M. Spagnuolo, and B. Falcidieno. Reeb graphs for shape analysis and applications. *TCS*, 2008.
[4] P. Bremer, G. Weber, J. Tierny, V. Pascucci, M. Day, and J. Bell. Interactive exploration and analysis of large scale simulations using topology-based data segmentation. *IEEE TVCG*, 2011.
[5] H. Carr, J. Snoeyink, and U. Axen. Computing contour trees in all dimensions. In *Symp. on Dis. Alg.*, pp. 918–926, 2000.
[6] H. Carr, J. Snoeyink, and M. van de Panne. Simplifying flexible isosurfaces using local geometric measures. In *IEEE VIS*, 2004.
[7] F. Chen, H. Obermaier, H. Hagen, B. Hamann, J. Tierny, and V. Pascucci. Topology analysis of time-dependent multi-fluid data using the reeb graph. *Computer Aided Geometric Design*, 2013.
[8] L. De Floriani, U. Fugacci, F. Iuricich, and P. Magillo. Morse complexes for shape segmentation and homological analysis: discrete models and algorithms. *Comp. Grap. For.*, 2015.
[9] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2009.
[10] G. Favelier, C. Gueunet, A. Gyulassy, J. Jomier, J. Levine, J. Lukasczyk, D. Sakurai, M. Soler, J. Tierny, W. Usher, and Q. Wu. Topological data analysis made easy with the Topology Toolkit. In *IEEE VIS Tutorials*, 2018. `https://topology-tool-kit.github.io/ieeeVisTutorial.html`.
[11] G. Favelier, C. Gueunet, and J. Tierny. Visualizing ensembles of viscous fingers. In *IEEE SciVis Contest*, 2016.
[12] D. Guenther, R. Alvarez-Boto, J. Contreras-Garcia, J.-P. Piquemal, and J. Tierny. Characterizing molecular interactions in chemical systems. *IEEE TVCG*, 2014.
[13] C. Gueunet, P. Fortin, J. Jomier, and J. Tierny. Contour forests: Fast multi-threaded augmented contour trees. In *IEEE LDAV*, 2016.
[14] C. Gueunet, P. Fortin, J. Jomier, and J. Tierny. Task-based Augmented Merge Trees with Fibonacci Heaps,. In *IEEE LDAV*, 2017.
[15] A. Gyulassy, P. Bremer, R. Grout, H. Kolla, J. Chen, and V. Pascucci. Stability of dissipation elements: A case study in combustion. *Comp. Graph. For.*, 2014.
[16] A. Gyulassy, P. T. Bremer, B. Hamann, and V. Pascucci. A practical approach to morse-smale complex computation: Scalability and generality. *IEEE TVCG*, 2008.
[17] A. Gyulassy, D. Guenther, J. A. Levine, J. Tierny, and V. Pascucci. Conforming morse-smale complexes. *IEEE TVCG*, 2014.
[18] A. Gyulassy, A. Knoll, K. Lau, B. Wang, P. Bremer, M. Papka, L. A. Curtiss, and V. Pascucci. Interstitial and interlayer ion diffusion geometry extraction in graphitic nanosphere battery materials. *IEEE TVCG*, 2015.
[19] A. Gyulassy, V. Natarajan, M. Duchaineau, V. Pascucci, E. Bringa, A. Higginbotham, and B. Hamann. Topologically Clean Distance Fields. *IEEE TVCG*, 2007.
[20] C. Heine, H. Leitte, M. Hlawitschka, F. Iuricich, L. De Floriani, G. Scheuermann, H. Hagen, and C. Garth. A survey of topology-based methods in visualization. *Comp. Grap. For.*, 2016.
[21] J. Kasten, J. Reininghaus, I. Hotz, and H. Hege. Two-dimensional time-dependent vortex regions based on the acceleration magnitude. *IEEE TVCG*, 2011.
[22] D. E. Laney, P. Bremer, A. Mascarenhas, P. Miller, and V. Pascucci. Understanding the structure of the turbulent mixing layer in hydrodynamic instabilities. *IEEE TVCG*, 2006.
[23] J. Lukasczyk, G. Aldrich, M. Steptoe, G. Favelier, C. Gueunet, J. Tierny, R. Maciejewski, B. Hamann, and H. Leitte. Viscous fingering: A topological visual analytic approach. In *Physical Modeling for Virtual Manufacturing Systems and Processes*, 2017.
[24] S. Parsa. A deterministic o(m log m) time algorithm for the reeb graph. In *Symp. on Comp. Geom.*, 2012.
[25] V. Pascucci, G. Scorzelli, P. T. Bremer, and A. Mascarenhas. Robust on-line computation of Reeb graphs: simplicity and speed. *ACM Trans. on Graph.*, 2007.
[26] V. Pascucci, X. Tricoche, H. Hagen, and J. Tierny. *Topological Methods in Data Analysis and Visualization: Theory, Algorithms and Applications*. Springer, 2010.
[27] E. Santos, J. Tierny, A. Khan, B. Grimm, L. Lins, J. Freire, V. Pascucci, C. Silva, S. Klasky, R. Barreto, and N. Podhorszki. Enabling advanced visualization tools in a web-based simulation monitoring system. In *Proc. of IEEE eScience*, 2009.
[28] N. Shivashankar, P. Pranav, V. Natarajan, R. van de Weygaert, E. P. Bos, and S. Rieder. Felix: A topology based framework for visual exploration of cosmic filaments. *IEEE TVCG*, 2016. `http://vgl.serc.iisc.ernet.in/felix/index.html`.
[29] B. S. Sohn and C. L. Bajaj. Time varying contour topology. *IEEE TVCG*, 2006.
[30] M. Soler, M. Plainchault, B. Conche, and J. Tierny. Topologically controlled lossy compression. In *Proc. of IEEE PacificVis*, 2018.
[31] T. Sousbie. The persistent cosmic web and its filamentary structure: Theory and implementations. *Royal Astronomical Society*, 2011. `http://www2.iap.fr/users/sousbie/web/html/indexd41d.html`.
[32] D. M. Thomas and V. Natarajan. Multiscale symmetry detection in scalar fields by clustering contours. *IEEE TVCG*, 2014.
[33] J. Tierny. *Topological Data Analysis for Scientific Visualization*. Springer, 2018.
[34] J. Tierny, J. Daniels, L. G. Nonato, V. Pascucci, and C. Silva. Interactive quadrangulation with Reeb atlases and connectivity textures. *IEEE TVCG*, 2012.
[35] J. Tierny, G. Favelier, J. A. Levine, C. Gueunet, and M. Michaux. The Topology ToolKit. *IEEE TVCG*, 2017. `https://topology-tool-kit.github.io/`.
[36] J. Tierny, A. Gyulassy, E. Simon, and V. Pascucci. Loop surgery for volumetric meshes: Reeb graphs reduced to contour trees. *IEEE TVCG*, 2009.
[37] J. Tierny and V. Pascucci. Generalized topological simplification of scalar fields on surfaces. *IEEE TVCG*, 2012.
[38] J. Tierny, J.-P. Vandeborre, and M. Daoudi. Partial 3D shape retrieval by reeb pattern unfolding. *Comp. Grap. For.*, 28:41–55, 2009.
[39] W. Usher and Q. Wu. Topo-Vol: Topology guided volume exploration. `https://github.com/Twinklebear/topo-vol`, 2017.
[40] M. van Kreveld, R. van Oostrum, C. Bajaj, V. Pasucci, and D. Schikore. Contour trees and small seed sets for isosurface traversal. In *Symp. on Comp. Geom.*, 1997.
[41] A. Vintescu, F. Dupont, G. Lavoué, P. Memari, and J. Tierny. Conformal factor persistence for fast hierarchical cone extraction. In *Eurographics (short papers)*, 2017.
[42] G. Weber, P.-T. Bremer, H. Carr, and A. Gyulassy. Scalar topology in visual data analysis. In *IEEE VIS Tutorials*, 2009.
[43] G. Weber, S. E. Dillard, H. Carr, V. Pascucci, and B. Hamann. Topology-controlled volume rendering. *IEEE TVCG*, 2007.