CS 6170: Computational Topology, Spring 2019
Lecture 01

Topological Data Analysis for Data Scientists

Dr. Bei Wang

School of Computing
Scientific Computing and Imaging Institute (SCI)
University of Utah
www.sci.utah.edu/~beiwang
beiwang@sci.utah.edu

Jan 8, 2019

Course Overview

- Course Syllabus, Final Project
- Course Schedule
- Course Webpage: for latest lectures, due dates, etc.
- Canvas: for assignments submission, announcements etc.
- Introduction: topological data analysis
- Examples of past projects

Course Webpage

- http://www.sci.utah.edu/~beiwang/teaching/cs6170-spring-2019.html
- http://www.sci.utah.edu/~beiwang/teaching/cs6170-spring-2019/syllabus.pdf
- http://www.sci.utah.edu/~beiwang/teaching/cs6170-spring-2019/schedule.html

Topological Data Analysis (TDA): a brief history

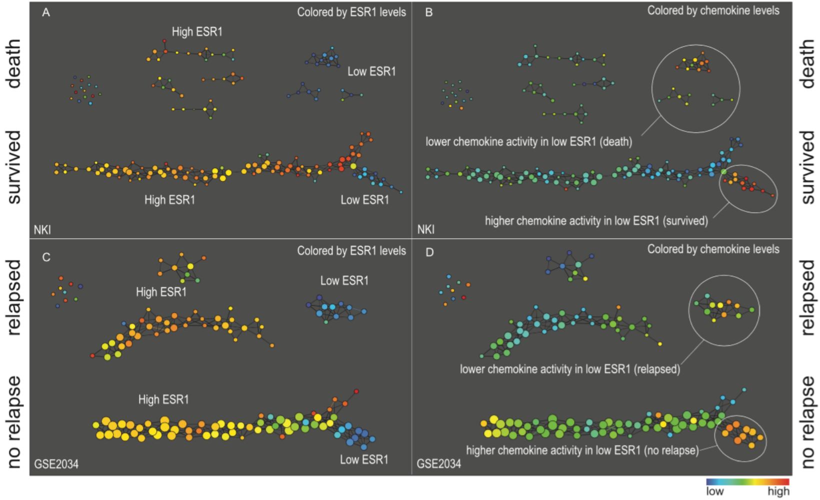
- A marriage: math and computer science
- Topological data analysis is cool: many data applications!
- Many great and fun people are players in this field: mathematicians, computer scientists, statisticians...
- Interdisciplinary: CS, math (algebraic topology, differential topology, i.e. Homology, Morse theory), statics (machine learning, manifold learning), electrical engineering (sensor networks), physics (universe)
- It is young (20 years), and a lot of open problems, that is, challenges and opportunities! (Imaging the field of computational geometry at its infancy...)
- The researchers are only in their 2nd generation (approximately): room to grow!
- Topological data analysis and visualization is inseparable

Three mutually inclusive modules

- TDA Foundations and Pipeline (FP)
- TDA, Machine Learning and Statistics (ML)
- TDA in Data Science (DS)

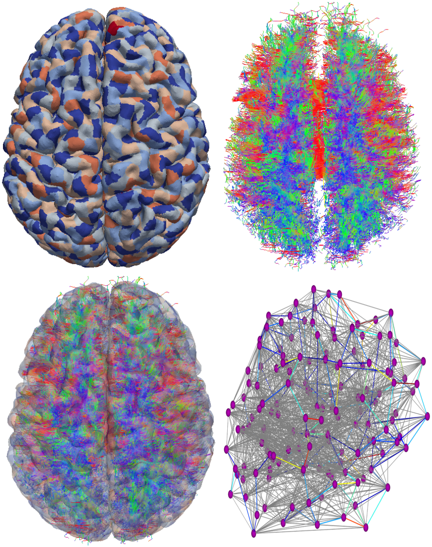
TDA in Data Science (DS)

Market/Gene Segmentation



Lum et al. (2013)

Brain Networks



Wong et al. (2016)

Combustion simulation

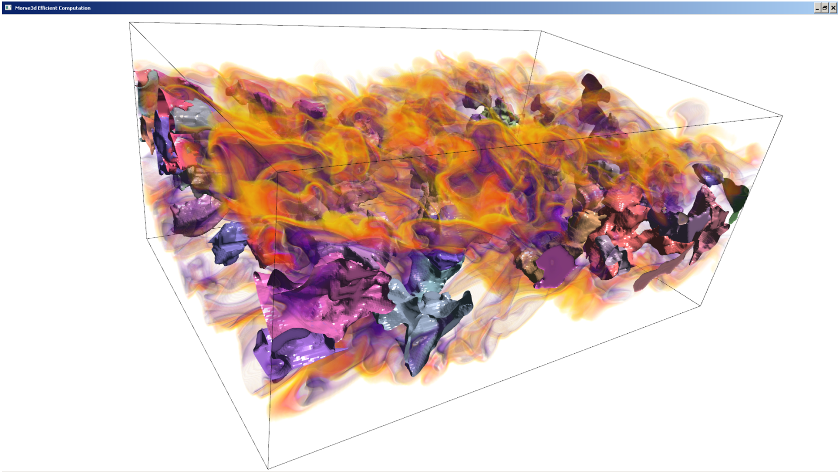


Image courtesy: V. Pascucci

Tracking 2D Combustion

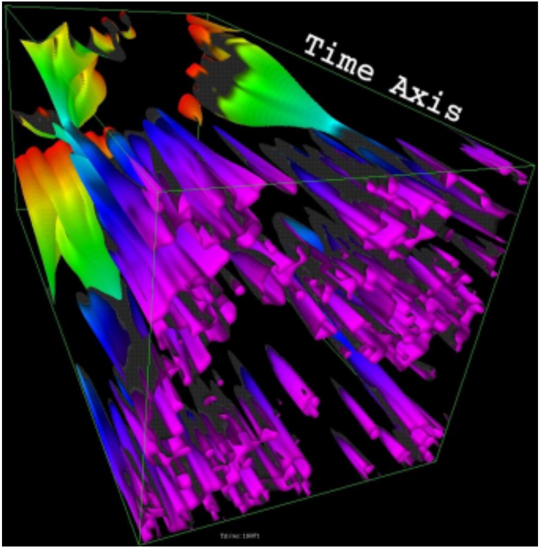
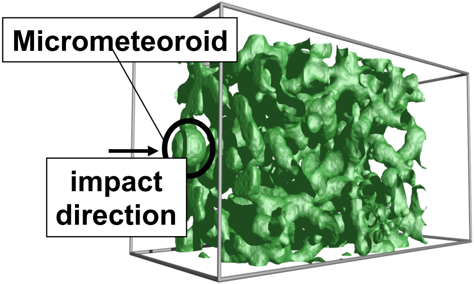
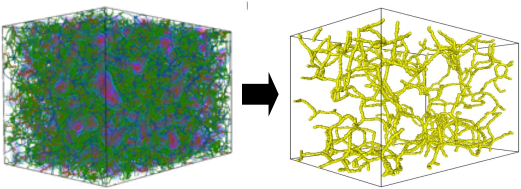


Image courtesy: V. Pascucci

Material science

Quantitative Analysis of the Impact of a Micrometeoroid in a Porous Medium; reconstruction the structure of porous medium

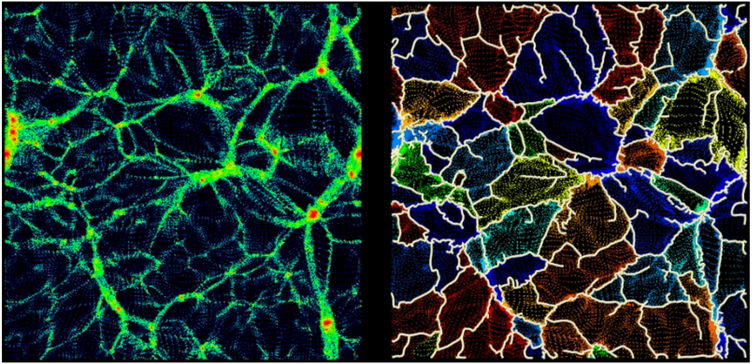




Gyulassy et al. (2007)

Astronomy: study the formation of filaments

TDA+ASTRONEMY
POTENTIALS



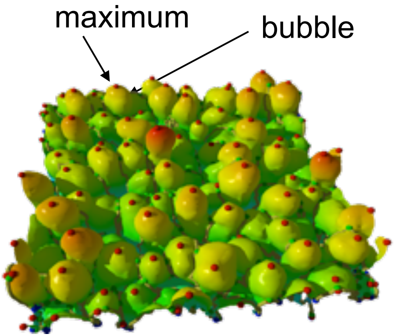
FILAMENTS STRUCTURE T. SOUSBIE, *DISPERSE*

http://www2.iap.fr/users/sousbie/disperse.html

TDA Foundations and Pipeline (FP)

Case study: feature definition

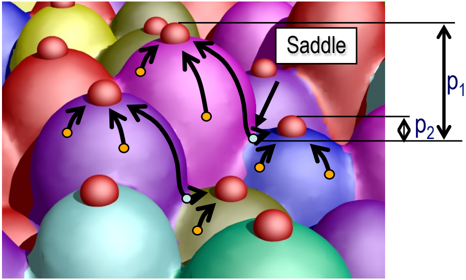
Analyze high-resolution Rayleigh Taylor instability simulations

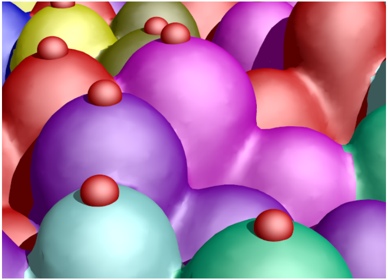


Laney et al. (2007)

Case study: persistence simplification

Analyze high-resolution Rayleigh Taylor instability simulations

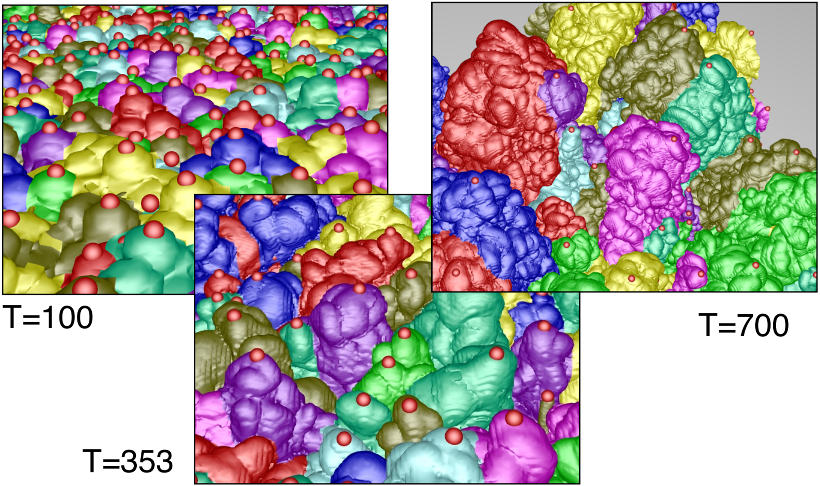




Laney et al. (2007)

Case study: robust segmentation

The segmentation method is robust from early mixing to late turbulence



T=100

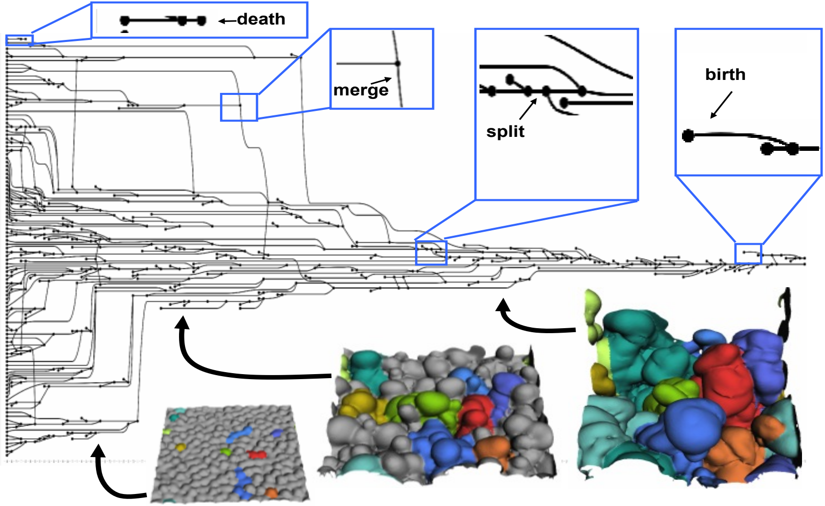
T=353

T=700

Laney et al. (2007)

Case study: event characterization

We characterize events that occur in the mixing process



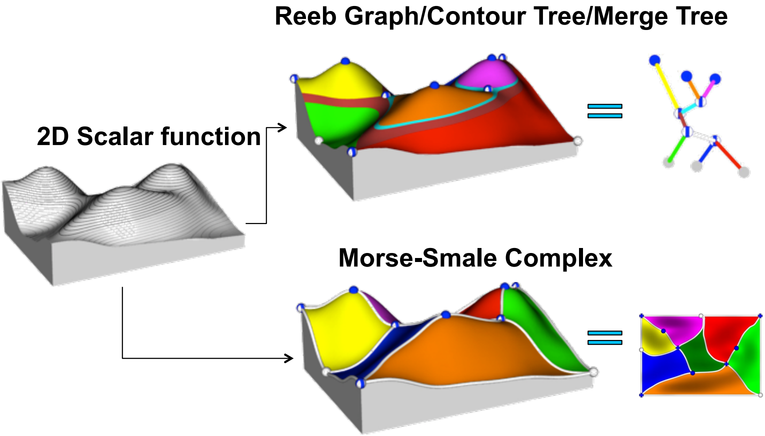
Laney et al. (2007)

A really old joke...

Who thinks the coffee mug and a donut is the same?

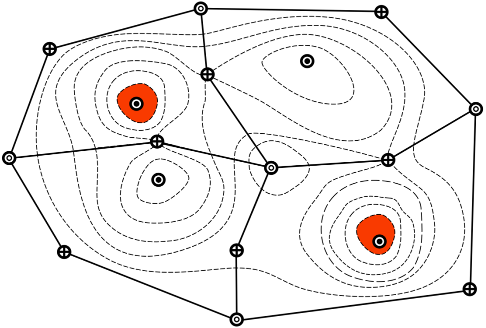
Key development in TDA

1. Abstraction of the data: topological structures
2. Separate features from noise: persistent homology



van Kreveld et al. (1997); Carr et al. (2003); Edelsbrunner et al. (2003a,b)

Contour tree



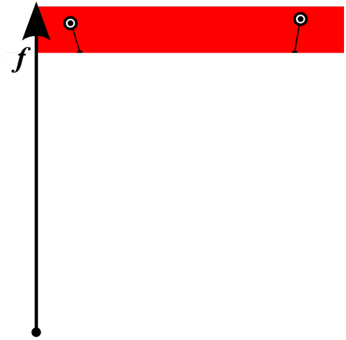


Image courtesy: V. Pascucci

Contour tree

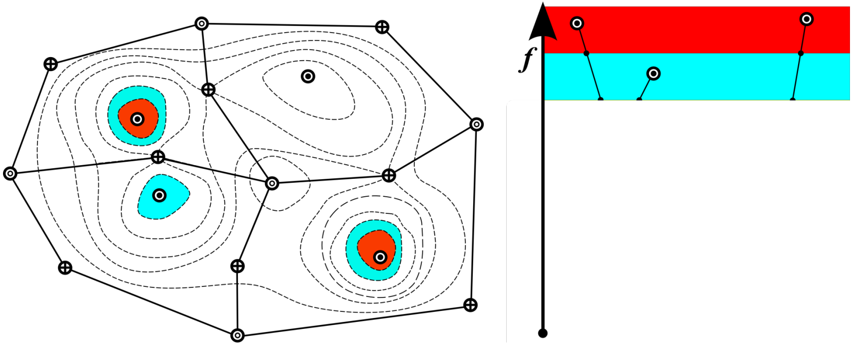


Image courtesy: V. Pascucci

Contour tree

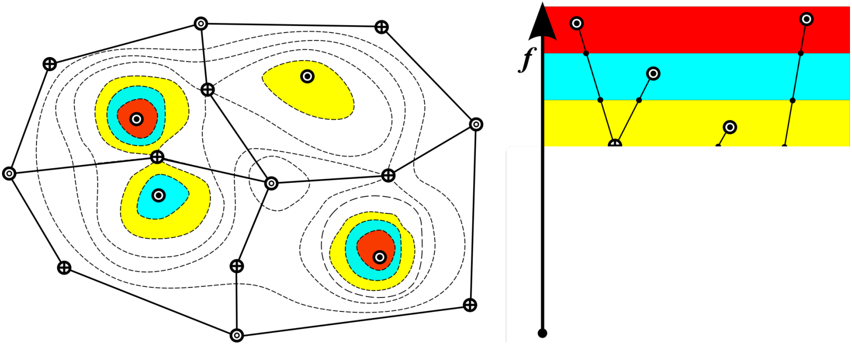


Image courtesy: V. Pascucci

Contour tree

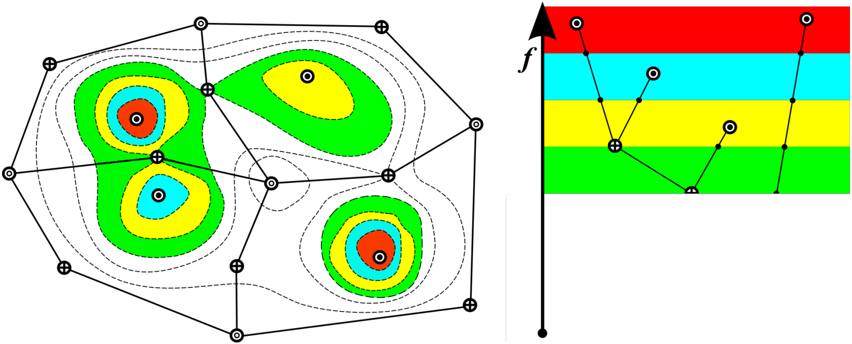


Image courtesy: V. Pascucci

Contour tree

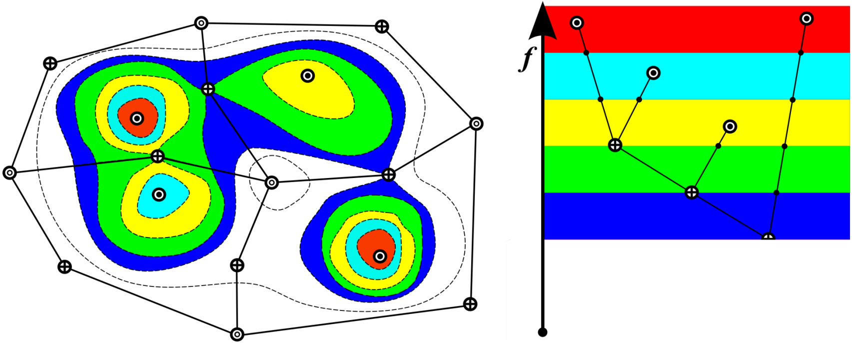


Image courtesy: V. Pascucci

Contour tree

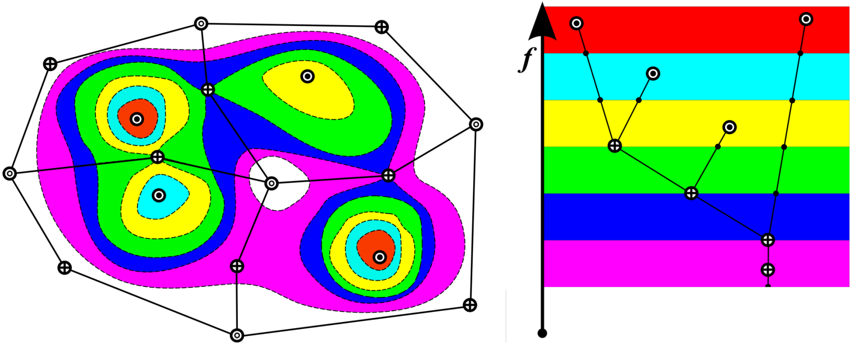
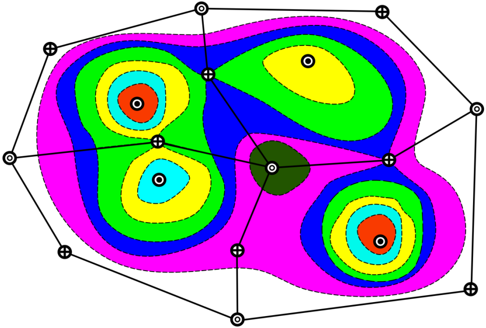


Image courtesy: V. Pascucci

Contour tree



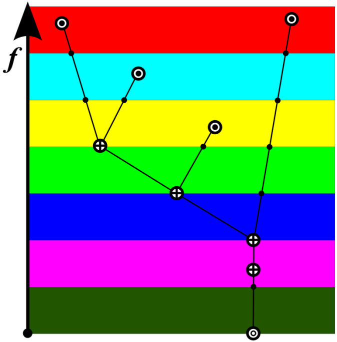
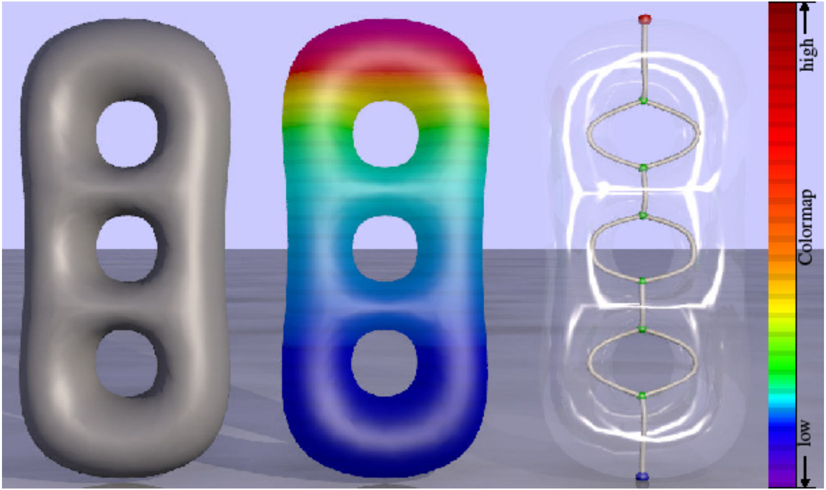


Image courtesy: V. Pascucci

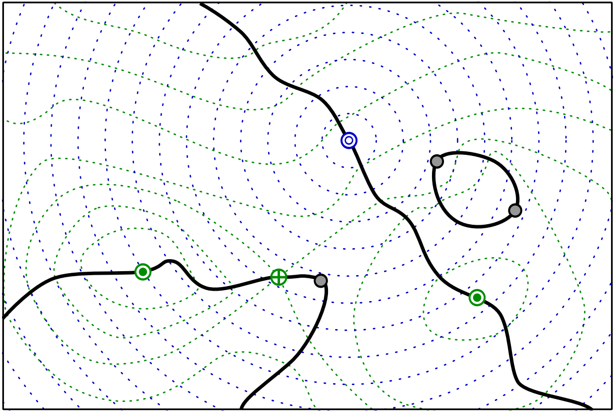
Reeb graph

Graph obtained by continuos contraction of all the contours in a scalar field, where each contour is collapsed to a distinct point.



Cole-McLaughlin et al. (2003)

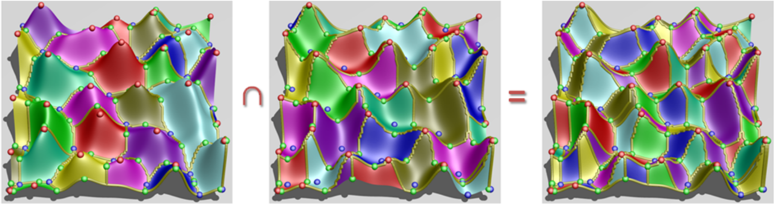
Jacobi Set



Edelsbrunner and Harer (2004)

Morse-Smale complex

A partition of the data into monotonic regions

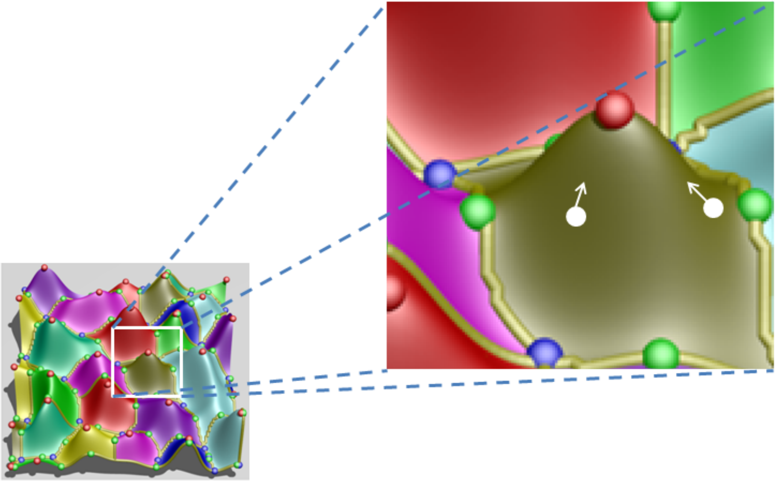


Ascending Manifolds ∩ Descending Manifolds = Morse-Smale Complex

Edelsbrunner et al. (2003a,b)

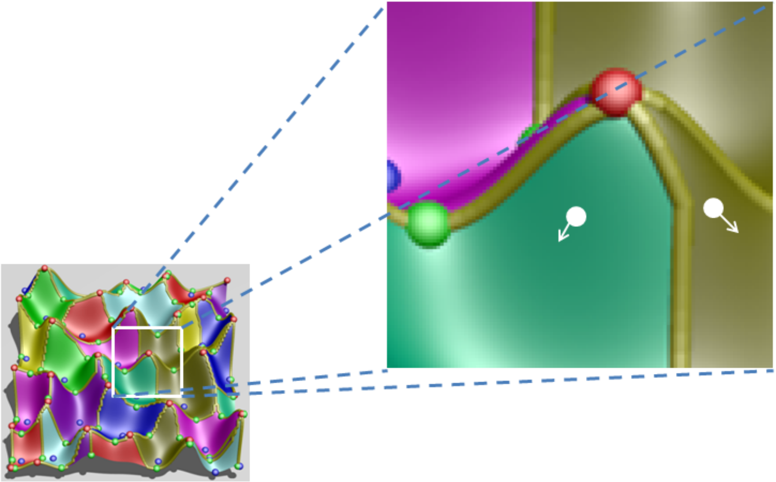
Ascending Manifolds

Compute steepest ascent gradient from each point in dataset



Descending Manifolds

Compute steepest descent gradient from each point in dataset



Terrain simplification

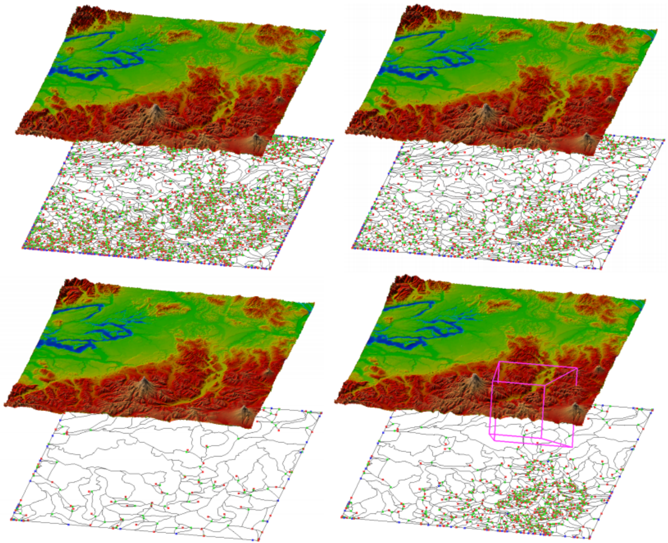


Figure 11: (Upper-left) Puget Sound data after topological noise removal. (Upper-right) Data at persistence of 1.2% of the maximum height. (Lower-left) Data at persistence 20% of the maximum height. (Lower-right) View-dependent re□nement (purple: view frustum).

Bremer et al. (2003)

Persistent homology

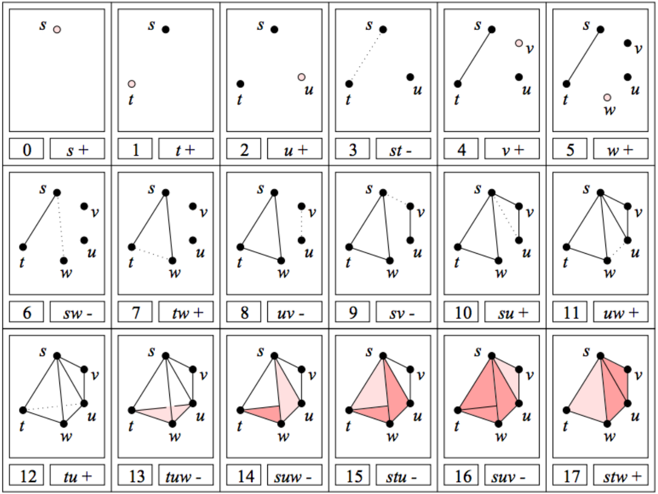
When data is corrupted by noise, how can we tell features from noise?

"The eye, or the brain, performs the marvelous task of taking the sense data of individual points and assembling them into a coherent image of a continuum infers the continuous from the discrete."

Figure: The Seine at La Grande Jatte by Georges Seurat

Weinberger (2011)

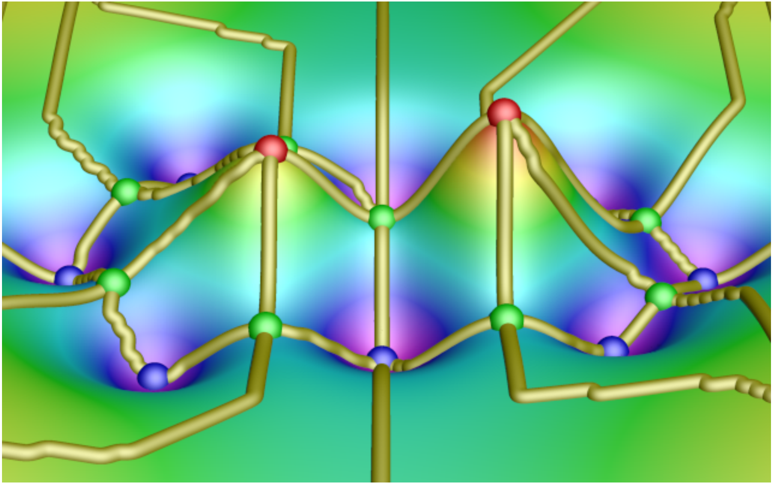
Persistent homology: computation



Edelsbrunner et al. (2002)

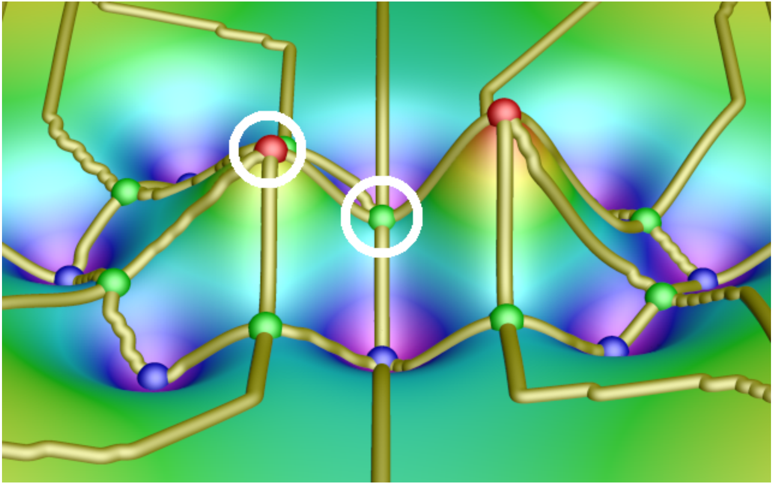
Persistent homology

Simplifying topological features



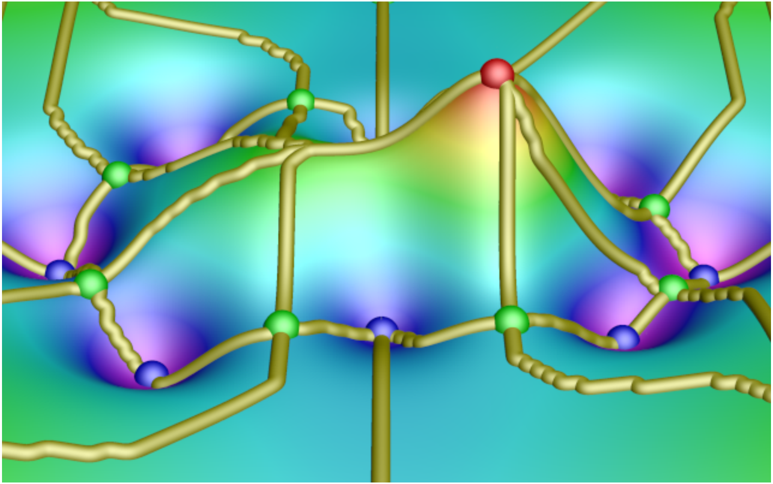
Persistent homology

Simplifying topological features



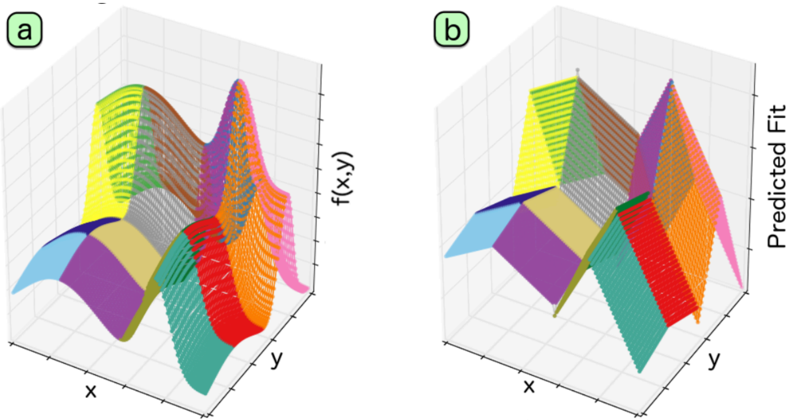
Persistent homology

Simplifying topological features



TDA, Machine Learning and Statistics (ML)

TDA with regression: topological partition

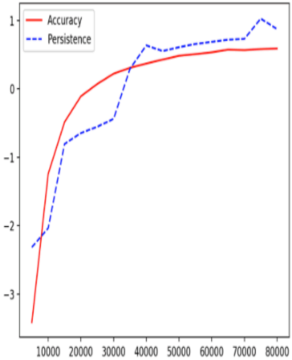


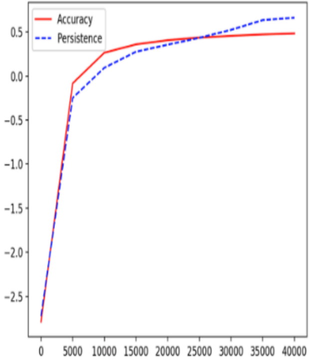
Fit linear models to each partition.

Gerber et al. (2010); Maljovec et al. (2016)

Understanding how CNN learn using TDA

CNN: convolutional neural networks

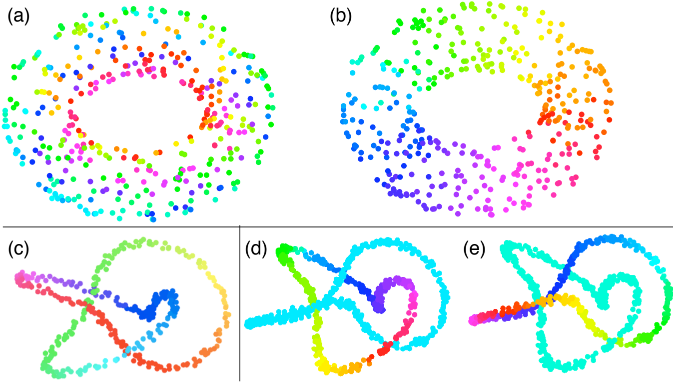




`https://www.ayasdi.com/blog/artificial-intelligence/going-deeper-understanding-convolutional-neural-networks-learn-using-tda/`

TDA with dimensionality reduction: circular structures

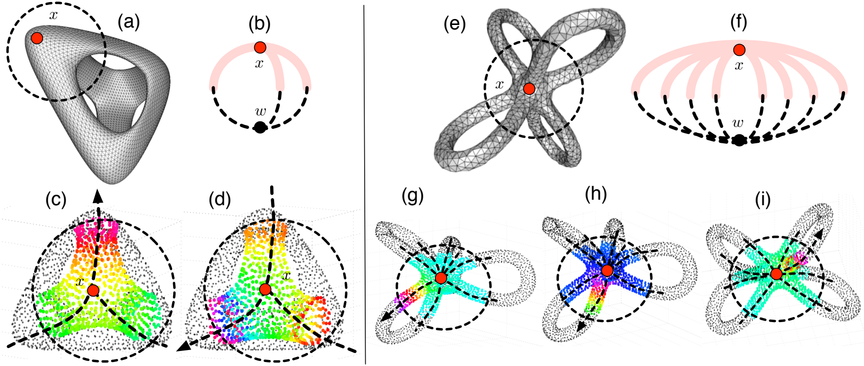
Parametrizing data (for circular features) in high-dimensions.



de Silva et al. (2009)

Detect branching features in high-dim data

Parametrizing data in high-dimensions.

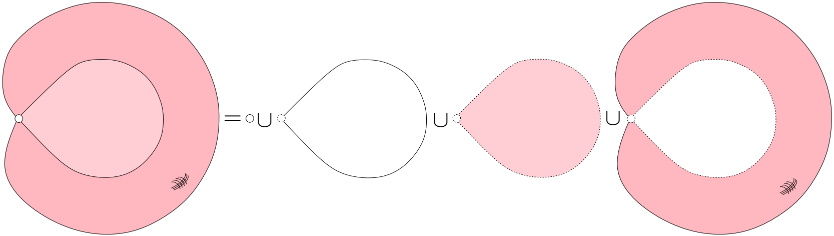


Wang et al. (2011)

Stratification learning in high dimensions

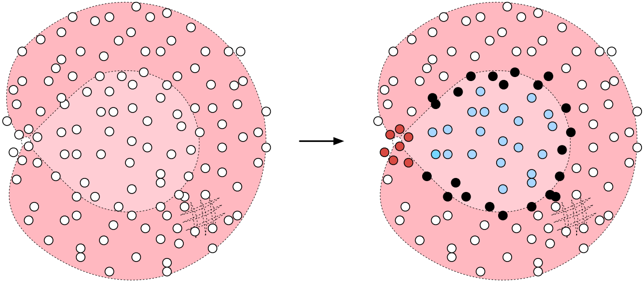
The coarsest stratification of a pinched torus

1. Decompose into manifold pieces (strata). 2. Pieces fit “nicely”.



Bendich et al. (2012)

Stratification learning in high dimensions



Bendich et al. (2012)

Challenges and Opportunities

Challenges and opportunities for data science

- Robustness of topological structures
- Scalability, approximation
- High-dimensional data
- Integration with statistics and machine learning
- Integration with visualization
- **Many applications!** Usability, interpretability

References I

Bendich, P., Mukherjee, S., and Wang, B. (2012). Local homology transfer and stratification learning. *Proceedings 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1355–1370.

Bremer, P.-T., Edelsbrunner, H., Hamann, B., and Pascucci, V. (2003). A multi-resolution data structure for two-dimensional Morse-Smale functions. *IEEE Visualization*.

Carr, H., Snoeyink, J., and Axen, U. (2003). Computing contour trees in all dimensions. *Computational Geometry: Theory and Applications*, 24(3):75–94.

Cole-McLaughlin, K., Edelsbrunner, H., Harer, J., Natarajan, V., and Pascucci, V. (2003). Loops in Reeb graphs of 2-manifolds. *Proceedings of the nineteenth annual symposium on Computational geometry*, pages 344–350.

de Silva, V., Morozov, D., and Vejdemo-Johansson, M. (2009). Persistent cohomology and circular coordinates. *Proceedings 25th Annual Symposium on Computational Geometry*, pages 227–236.

Edelsbrunner, H. and Harer, J. (2004). Jacobi sets of multiple Morse functions. *Foundations of Computational Mathematics*, pages 37–57.

Edelsbrunner, H., Harer, J., Natarajan, V., and Pascucci, V. (2003a). Morse-Smale complexes for piece-wise linear 3-manifolds. *Proceedings 19th Annual symposium on Computational geometry*, pages 361–370.

Edelsbrunner, H., Harer, J., and Zomorodian, A. J. (2003b). Hierarchical Morse-Smale complexes for piecewise linear 2-manifolds. *Discrete and Computational Geometry*, 30(87-107).

Edelsbrunner, H., Letscher, D., and Zomorodian, A. J. (2002). Topological persistence and simplification. *Discrete & Computational Geometry*, 28:511–533.

Gerber, S., Bremer, P.-T., Pascucci, V., and Whitaker, R. (2010). Visual exploration of high dimensional scalar functions. *IEEE Transactions on Visualization and Computer Graphics*, 16:1271 – 1280.

Gyulassy, A., Natarajan, V., Hamann, B., Duchaineau, M., Pascucci, V., Bringa, E., and Higginbotham, A. (2007). Topologically clean distance fields. *IEEE Transactions on Visualization and Computer Graphics*.

References III

Laney, D., Bremer, P.-T., Mascarenhas, A., Miller, P., and Pascucci, V. (2007). Understanding the structure of the turbulent mixing layer in hydrodynamic instabilities. *IEEE Transactions on Visualization and Computer Graphics*, 13(1):1053–1060.

Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., and Carlsson, G. (2013). Extracting insights from the shape of complex data using topology. *Scientific Reports*, 3.

Maljovec, D., Wang, B., Rosen, P., Alfonsi, A., Pastore, G., Rabiti, C., and Pascucci, V. (2016). Topology-inspired partition-based sensitivity analysis and visualization of nuclear simulations. *Proceedings IEEE Pacific Visualization Symposium (PacificVis)*.

van Kreveld, M., van Oostrum, R., Bajaj, C., Pascucci, V., and Schikore, D. (1997). Contour trees and small seed sets for isosurface traversal. *Proceedings 13th Annual Symposium on Computational Geometry*, pages 212–220.

References IV

Wang, B., Summa, B., Pascucci, V., and Vejdemo-Johansson, M. (2011). Branching and circular features in high dimensional data. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 17(12):1902–1911.

Weinberger, S. (2011). What is persistent homology? *Notices of the AMS*, 58(1).

Wong, E., Palande, S., Wang, B., Zielinski, B., Anderson, J., and Fletcher, P. T. (2016). Kernel partial least squares regression for relating functional brain network topology to clinical measures of behavior. *International Symposium on Biomedical Imaging (ISBI)*.