Reeb Graphs and Their Variants

Theory and Applications

Bei Wang Phillips

Publish under Bei Wang
Associate Professor, School of Computing
Faculty, Scientific Computing and Imaging Institute (SCI)
Adjunct Associate Professor, Department of Mathematics
University of Utah, USA
On Sabbatical at Leipzig University, MPI Leipzig, and ScaDS.AI Leipzig
www.sci.utah.edu/~beiwang
beiwang@sci.utah.edu

Dagstuhl Seminar on Combinatorial Topology
Feb 26, 2024

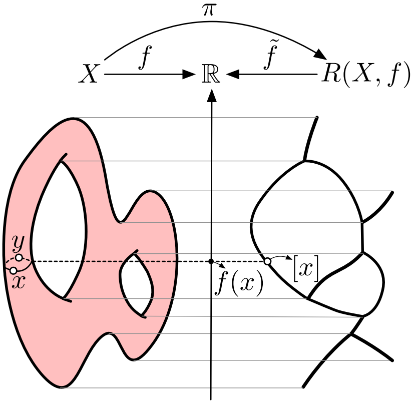
Overview

- Reeb graphs and their variants
- Applications
- Theoretical questions surrounding Reeb graphs and mapper graphs
- Measure theoretic Reeb graphs: robust topological data analysis

Reeb Graphs and Their Variants

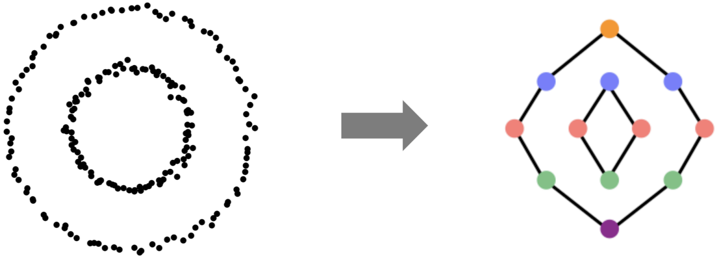
Reeb graph: a topological skeleton of data

- $(\mathbb{X}, f)$: a topological space with a continuous function $f : \mathbb{X} \to \mathbb{R}$.
- Reeb graph (Reeb [1946]): quotient space $R(\mathbb{X}, f) := \mathbb{X}/\!\sim_f$
- For $x, y \in \mathbb{X}$, $x \sim_f y$ iff x and y belong to the same connected component of the level set $f^{-1}(f(x))$.
- Reeb space: for $f : \mathbb{X} \to \mathbb{R}^d$, $R(\mathbb{X}, f) := \mathbb{X}/\!\sim_f$.

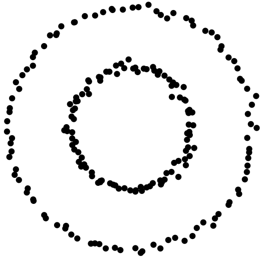


Mapper graph: a discrete approximation of a Reeb graph

- Singh et al. [2007]: Visualize the skeleton of (high-dimensional) point cloud data
- Main idea: turn point cloud data into a graph capturing both clusters and cluster relations.
- Parameters: number of cover elements of the range of the function, amount of overlaps, parameters associated with the clustering, etc.

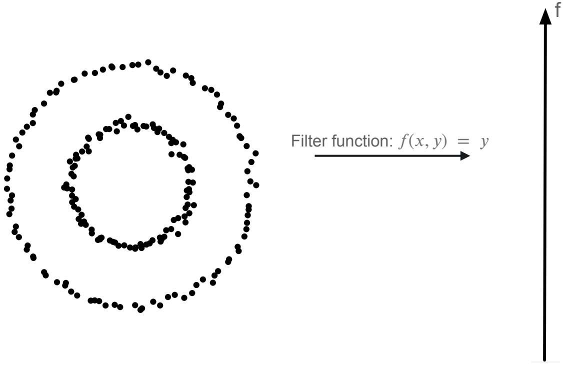


Mapper Graphs

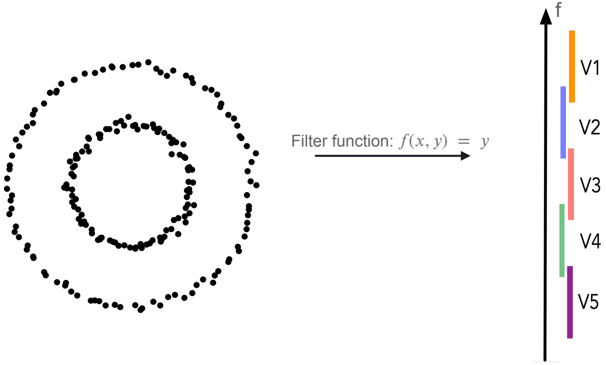


Filter function: $f(x, y) = y$

Mapper Graphs

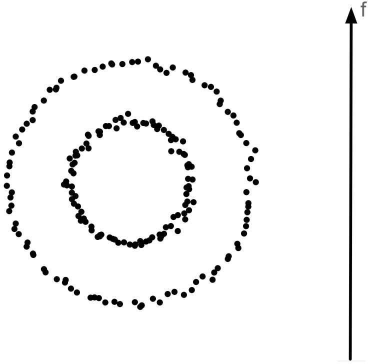


Mapper Graphs

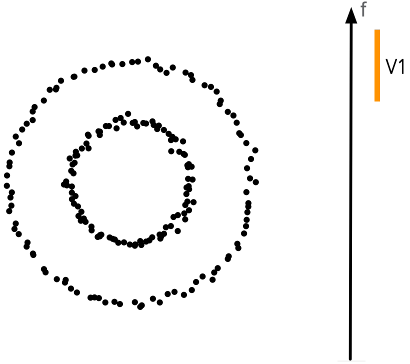


$\{V_\ell\}$: cover of image(f)

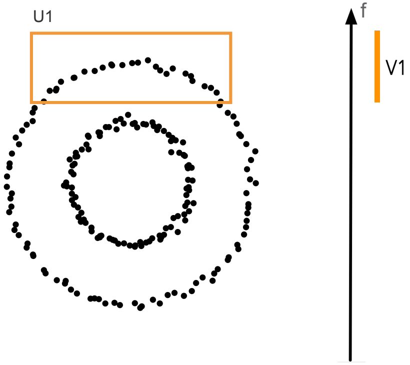
Mapper Graphs



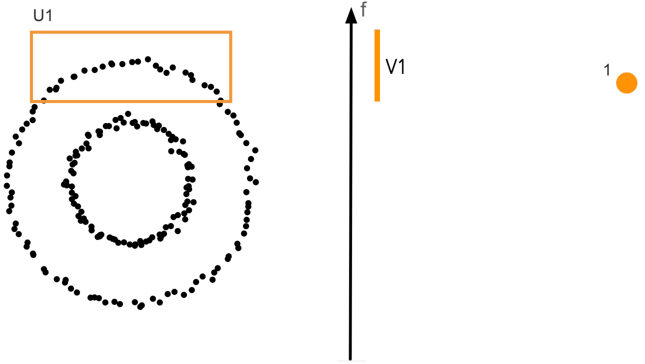
Mapper Graphs



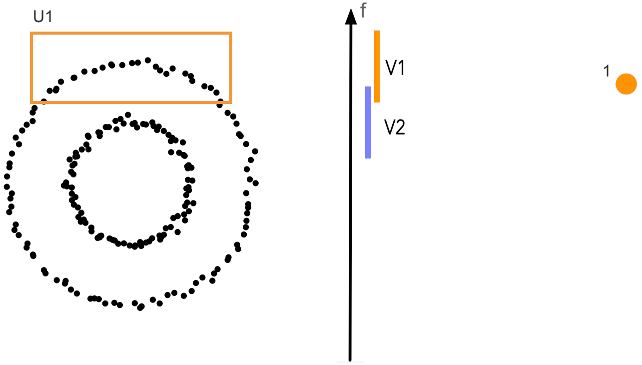
Mapper Graphs



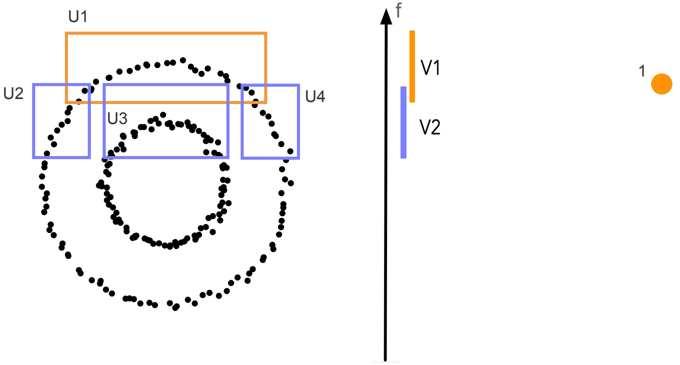
Mapper Graphs



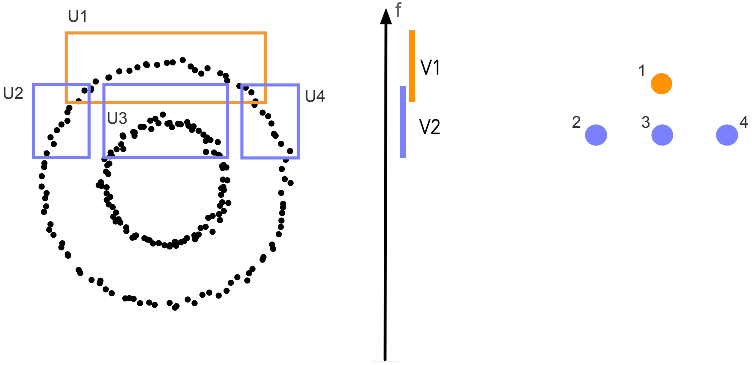
Mapper Graphs



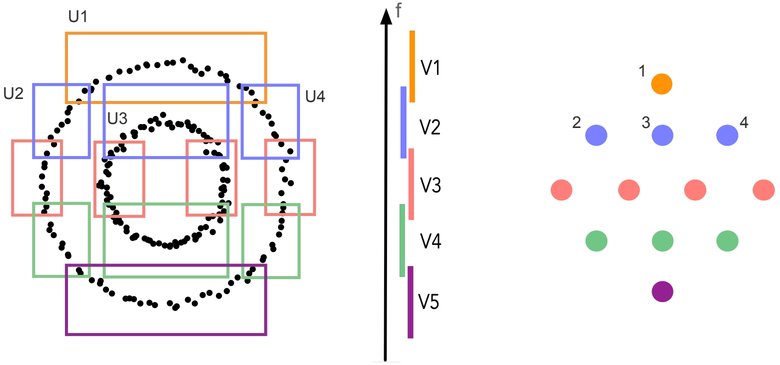
Mapper Graphs



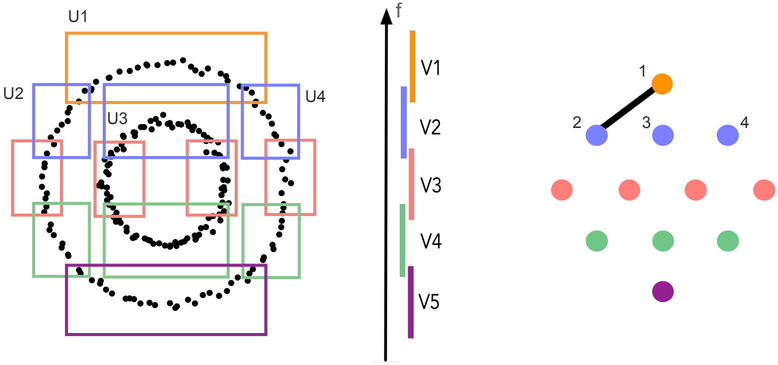
Mapper Graphs



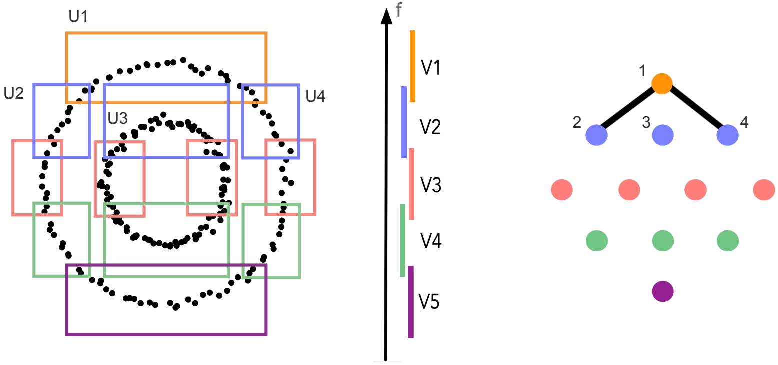
Mapper Graphs



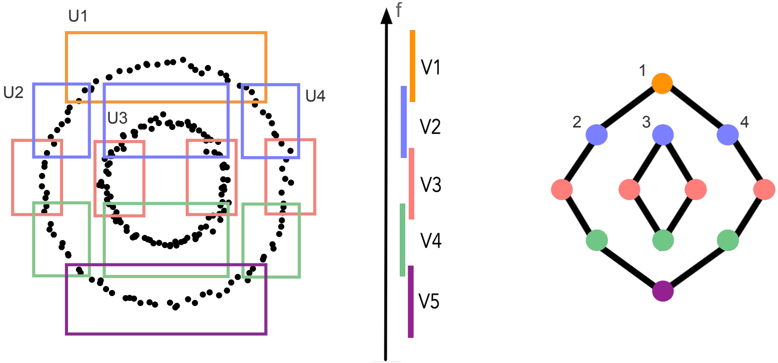
Mapper Graphs



Mapper Graphs



Mapper Graphs



Reeb graph: variants

- Mapper construction, mapper graph: Singh et al. [2007]
- α-Reeb graph: Chazal and Sun [2014]. Cover of range space with open intervals of length at most α.
- Multiscale mapper: Dey et al. [2016]. Consider a hierarchical family of covers and the maps between them.
- Multinerve mapper: Carriére and Oudot [2018]. Compute the multinerve of a connected cover.
- Joint Contour Net (JCN): Carr and Duke [2013], Geng et al. [2014]. A PL mapping over a simplicial mesh with multiple real-valued functions.
- Extended Reeb graph: Barral and Biasotti [2014]. Use cover elements from a partition of the domain without overlaps.
- Enhanced mapper graph: Brown et al. [2021]. Consider additionally the inverse map of intersections among cover elements.
- Ball mapper: Dlotko [2019]

Reeb graph: variants

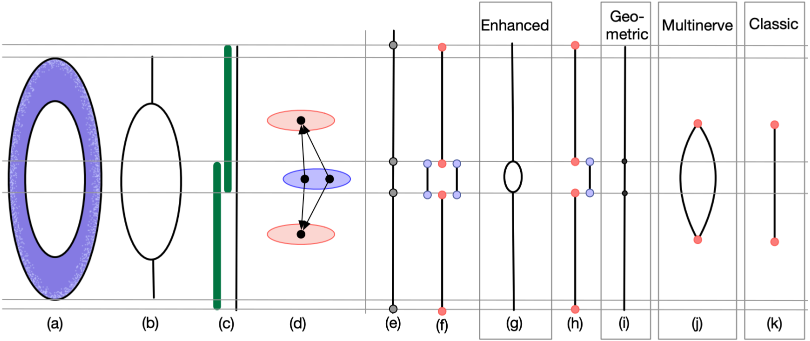


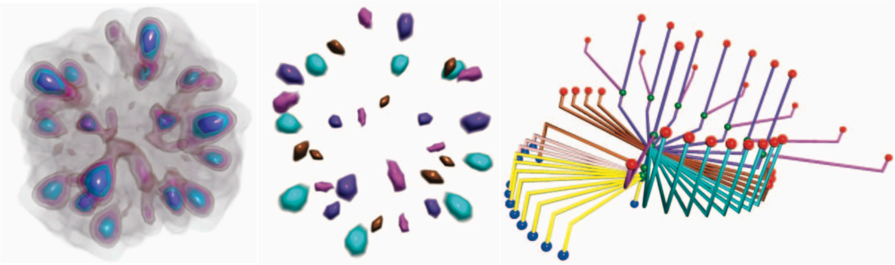
Fig 6. from Brown et al. [2021]
(a) Torus with a height function. (b) Reeb graph. (c) Nice cover.
(d) Visualization of the mapper cosheaf. (e) Stratification of $\mathbb{R}$.
(f) Disjoint union of closed intervals with quotient isomorphic to the enhanced mapper graph. (g) Enhanced mapper graph: Brown et al. [2021].
(i) Geometric mapper graph: Munch and Wang [2016].
(j) Multinerve mapper graph: Carriére and Oudot [2018].
(k) Classic mapper graph: Singh et al. [2007].

Applications

Application: shape skeletonization

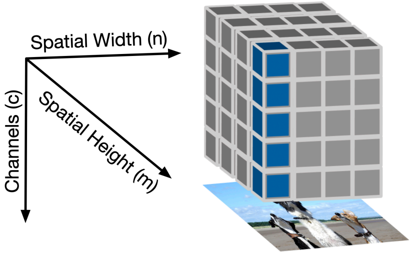
Pascucci et al. [2007]

Application: symmetry detection



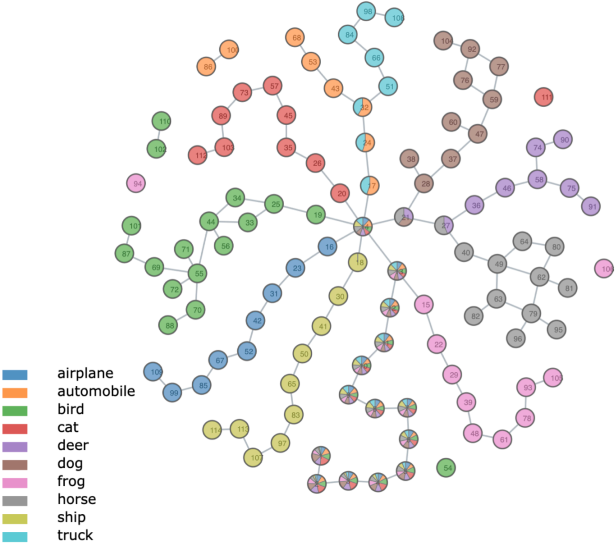
Symmetric patterns identified using contour trees
in electron microscopy data of a protein molecule
Thomas and Natarajan [2011].
See Yan et al. [2021] for a survey.

Application: deep learning



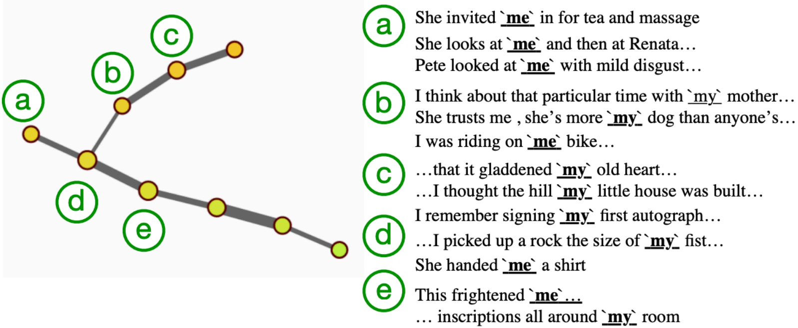
Neuron activations: neuron outputs from a convolutional neural network (CNN) form a high-dimensional point cloud Purvine et al. [2023].

Application: deep learning



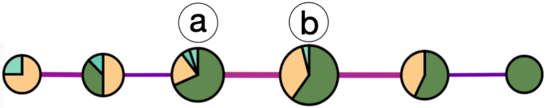
Topological summary of neuron activations using mapper graphs from a ResNet-18 model with CIFAR10 dataset, Zhou et al. [2021].

Application: topology of word embeddings



Pronoun differentiation. BERT Configuration: layer 12, Euclidean norm, 80 intervals, 30% overlap. Rathore et al. [2021].

Application: topology of word embeddings



Cost	I expected to pay for this service , but imagine my surprise when I received a bill **for** MORE than what I paid to have the original return prepared .
Cost	Also , a week after the work , Phet called me up to see how my car was running and to let me know that they had accidentally overcharged me for part of the work and wanted to give me a refund **for** that amount .
Possession	I paid 2 k cash **for** a truck with a blown motor .
Possession	Well - when you pay over $ 1000 **for** something you want it to hold up and look good !!!
Possession	$ 80 **for** a dish that has about one small lobster tail and is full of filler vegetables !
Theme	And another $ 100 **for** wrapping the furniture .
Theme	The dealer wanted $ 1300 to fix that and another $ 1500 **to** fix some other things .
Theme	Plus they will overcharge you **for** just about everything , and smile while doing it .

Capturing model confusion. BERT, Layer 9, fine-tuned on Supersense-Role during batch update 50. Rathore et al. [2023].

Theoretical Properties of Reeb Graphs and Mapper Graphs: An Incomplete Review

Theoretical questions on Reeb graphs

- **Comparison**: What is a reasonable distance or similarity measure $d(R_1, R_2)$ between a pair of Reeb graphs? With desirable properties?
  - Metric or Pseudometric
  - Discriminative
  - Computational complexity: easily implementable
  - **Stability**: w.r.t. simplification or perturbation of the underlying function?
- **Information content**: What information is encoded by the Reeb graph? How much information can we recover about the original data from the Reeb graph by solving an inverse problem?
- See Yan et al. [2021] and Bollen et al. [2022] for surveys.

Theoretical questions on mapper graphs

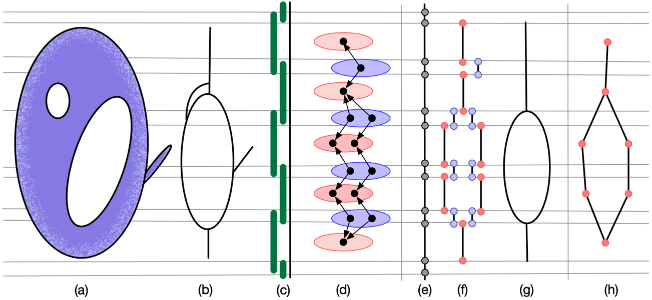
- **Information content**: What information is encoded by the mapper? How much information can we recover about the original data from the mapper by solving an inverse problem?
- **Stability:** What is the structural stability of the mapper with respect to perturbations of its function, domain and cover?
- **Convergence:** What is an appropriate metric under which the mapper converges to the Reeb graph as the number of sampled points goes to infinity and the granularity of the cover goes to zero?
- **Parameter tuning**: How to effectively and automatically tune the parameters that best capture the topology of the underlying data?
- See Brown et al. [2021] for a discussion and a number of references.

Distances between Reeb graphs

- Interleaving distance: Chazal et al. [2009], Morozov et al. [2013], De Silva et al. [2016b], Munch and Stefanou [2019].
- Labeled interleaving distance: Lan et al. [2023]
- Function distortion distance: Bauer et al. [2014, 2015]
- Edit distance: Bauer et al. [2016, 2020], Sridharamurthy et al. [2020]
- Distances based on branch decompositions and matching: Beketayev et al. [2014], Saikia et al. [2014]
- And many more, including those for merge trees and contour trees.
- Coming up: optimal transport distance by extending Li et al. [2023]

Information content

- Dey et al. [2017] studied topological information encoded by Reeb spaces, mappers and multi-scale mappers. 1-dimensional homology of the mapper was shown to be no richer than the domain itself.
- Carriére and Oudot [2018] characterized the information encoded in the mapper using the extended persistence diagram of its corresponding Reeb graph.
- Brown et al. [2021]: enhanced mapper graph reduces the information loss during summarization.



Convergence

- Babu [2013]: showed that the mapper converges to the Reeb graph in the bottleneck distance, using levelset zigzag persistence.
- Dey et al. [2017]: a convergence result between the mapper and the domain under a Gromov-Hausdorff metric.
- Carriére and Oudot [2018]: convergence between the (multinerve) mapper and the Reeb graph using the functional distortion distance.
- Carriére et al. [2018]:
    - Proved convergence.
    - Provided a confidence set for the mapper using a bottleneck distance on extended persistence diagrams.
    - The mapper is an optimal estimator of the Reeb graph.
    - Provided a statistical method for automatic parameter tuning using the rate of convergence.

Convergence

- Munch and Wang [2016]: characterized the mapper using constructible cosheaves. Proved the convergence between the (classic) mapper and the Reeb space in interleaving distance.
- Brown et al. [2021]: Probabilistic convergence between the mapper graph and the Reeb graph.
  - Recover (a variant of) the mapper graph using the theory of cosheaves in a probabilistic setting.
  - Assume points randomly sampled from a probability density function concentrated on the space.
  - With high probability, the distance between an enhanced mapper graph and the Reeb graph is upper bounded by the resolution of the cover as the number of samples goes to infinity.

Parameter tuning for mapper graphs

- Carriére et al. [2018]: a statistical procedure for selecting the most stable parameters for the mapper graph.
- Chalapathi et al. [2021]: adaptive covers using information criteria: multi-pass AIC and BIC. Based on X-means clustering algorithms.
- Alvarado et al. [2023]: G-Mapper. Splitting a cover repeatedly according to a statistical test for normality. Based on G-means clustering.
- Bui et al. [2020]: F-Mapper. Based on the Fuzzy C-means algorithm, a centroid-based overlapping clustering method.

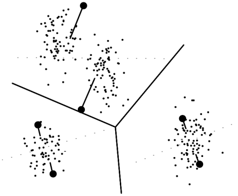


Figure 3: The first step of parallel local 2-means. The line coming out of each centroid shows where it moves to.

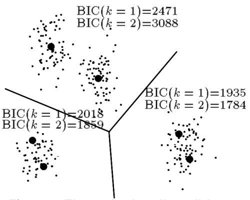


Figure 4: The result after all parallel 2-means have terminated.

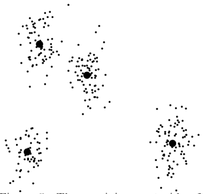


Figure 5: The surviving centroids after all the local model scoring tests.

Pelleg and Moore [2000]

Stability of mapper graphs

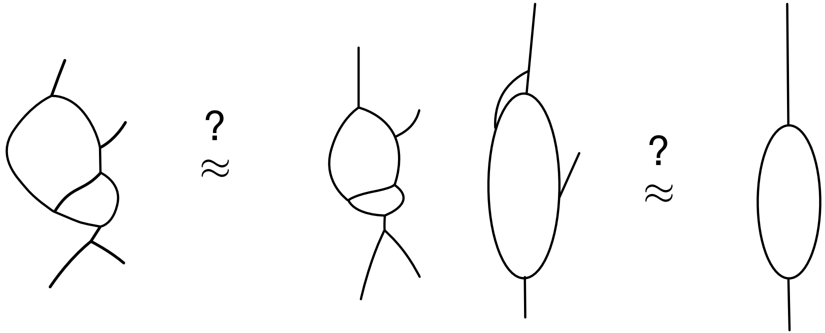
- Carriére and Oudot [2018]: stability for the mapper graph using the stability of extended persistence diagrams equipped with the bottleneck distance under Hausdorff or Wasserstein perturbations of the data.
- Brown et al. [2021]:
  - Stability for the mapper cosheaf in the interleaving distance.
  - Stability of the enhanced mapper graph with respect to perturbation of the data $(\mathbb{X}, f)$
  - The local stability depends on how the cover is positioned in relation to the critical values of f.

Stability of Reeb graphs

Theorem ([De Silva et al., 2016a, Theorem 4.4])

Let $R(\mathbb{X}, f)$ and $R(\mathbb{X}, g)$ be two Reeb graphs built from the same ambient space $\mathbb{X}$. Then the interleaving distance satisfies

$$d_I(R(\mathbb{X}, f), R(\mathbb{X}, g)) \leq \|f - g\|_\infty.$$



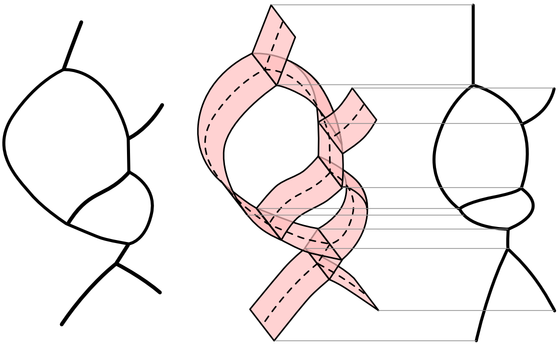
Smoothing of Reeb graph

- Given a Reeb graph G, let G_ε denote the space $G \times [-\varepsilon, \varepsilon]$ and define the ε-smoothing of G as the Reeb graph of the function

$$\begin{aligned} f_\varepsilon : \quad G_\varepsilon &\longrightarrow \mathbb{R} \\ (x, t) &\longmapsto f(x) + t. \end{aligned}$$

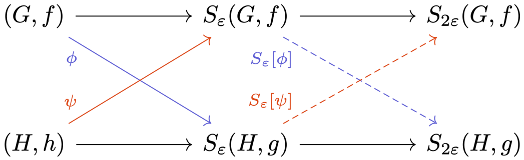
- De Silva et al. [2016a]: *ε-smoothing of a Reeb graph*

$$S_\varepsilon(G, f) := G_\varepsilon / \sim_{f_\varepsilon}$$



Interleaving distance

- For any $\epsilon > 0$, an ε-interleaving between two Reeb graphs (G, f) and (H, g) is a pair of maps, φ and ψ such that the following diagram commutes,



  $S_\varepsilon[\phi]$ is the map induced by $\Phi : G \times [-\varepsilon, \varepsilon] \to S_\varepsilon(H, g) \times [-\varepsilon, \varepsilon]$, $\Phi(x, t) = (\phi(x), t)$. See De Silva et al. [2016a] for details.

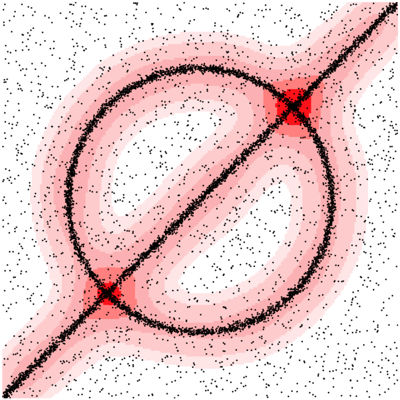
- The interleaving distance $d_I((G, f), (H, h))$ is defined as

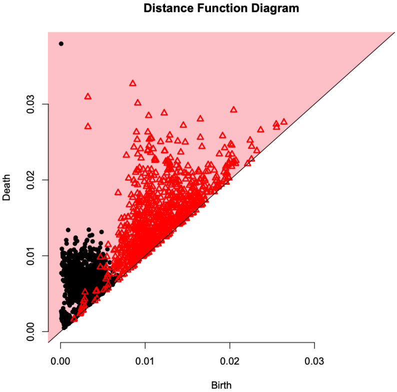
$$d_I((G, f), (H, h)) = \inf_{\varepsilon}\{\exists\ \varepsilon\text{-interleaving of } (G, f) \text{ and } (H, h)\}.$$

- **Smallest amount of thickening that gives rise to an ε-interleaving.**

Measure Theoretic Reeb Graphs

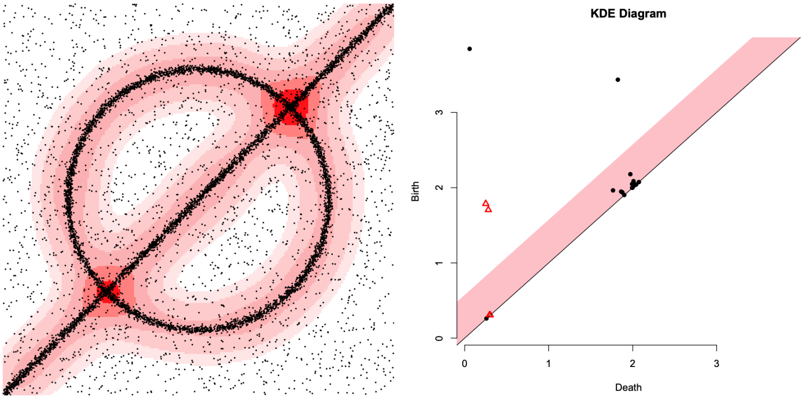
Robust topological descriptors and structural inference





Examine superlevel sets of kernel density estimate (sublevel sets of the kernel distance) for robust topological inference. Phillips et al. [2015]

Robust topological descriptors and structural inference

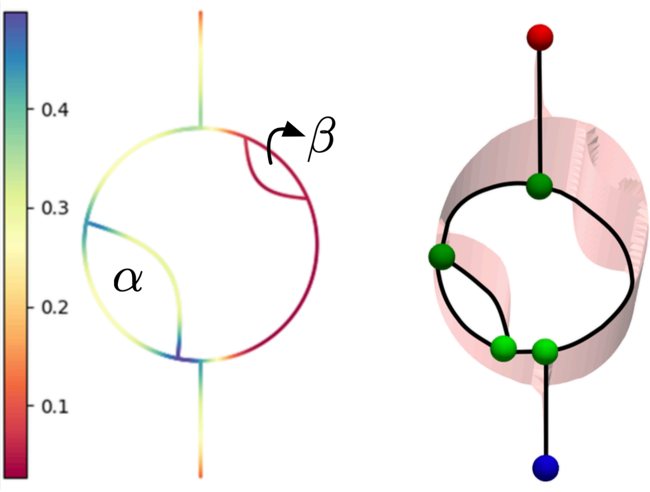


Examine superlevel sets of kernel density estimate (sublevel sets of the kernel distance) for robust topological inference. Phillips et al. [2015]

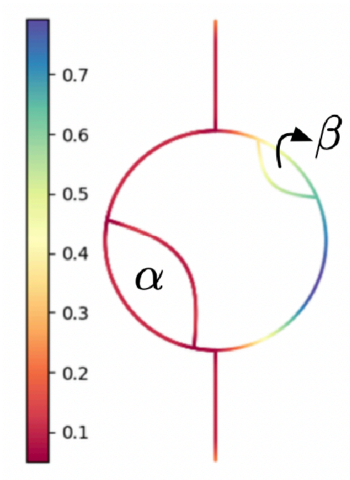
Robust topological descriptors

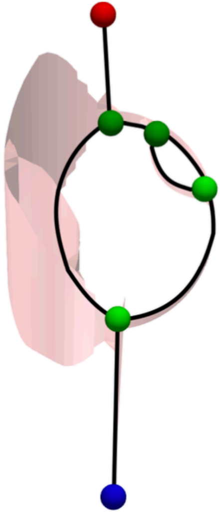
- Data science applications: associate weights to data points, which represent how much we trust these data points, or how important their corresponding features are.
- Conventional Reeb graphs do not take into consideration the data distributions and (possibly) non-uniform importance of data points, leading to discrepancies between the represented and actual topologies of the data.
- E.g., a significant loop in the Reeb graph might be caused by a sparse set of data points or lies in regions of low importance in function values.
- **Capture robust topology in data**.

An illustrative example

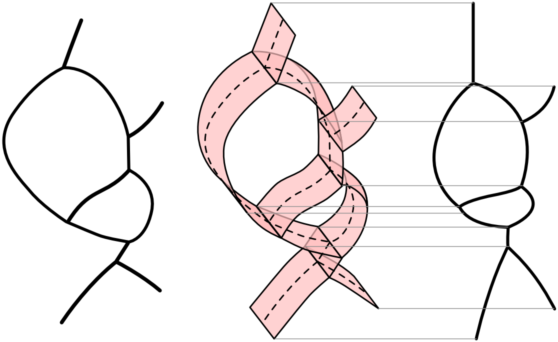


An illustrative example





Locally smoothed Reeb graph



What if we vary the thickening parameter?

Locally smoothed Reeb graph

- Smoothing with variable thickness.
- Let $\mathbb{X}$ be a topological space and $f : \mathbb{X} \to \mathbb{R}$ be a function on $\mathbb{X}$.
- Let $r : \mathbb{X} \to \mathbb{R}$ be a bounded positive function on $\mathbb{X}$ with

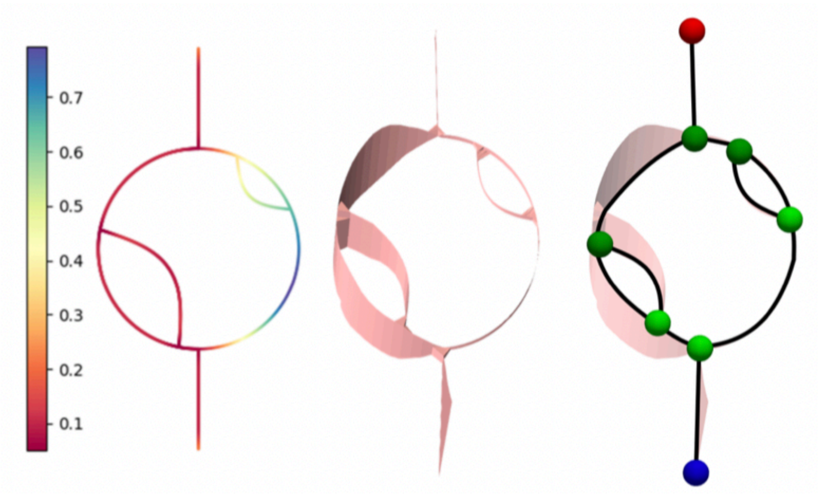
$$M := \sup_{x \in \mathbb{X}} r(x).$$

- The function r is viewed as a local smoothing factor.
- Let $\mathbb{X}_r$ denote the space

$$\mathbb{X}_r = \{(x, t) \in \mathbb{X} \times [-M, M] \mid |t| \leq r(x)\}.$$

- Then the function f naturally extends to a function f_r on $\mathbb{X}_r$ by $f_r(x, t) = f(x) + t$.
- The r-*smoothed Reeb graph* of $(\mathbb{X}, f)$ is the Reeb graph $R(\mathbb{X}_r, f_r)$.

Locally smoothed Reeb graph



Stability of locally smoothed Reeb graph

Lemma (Stability of locally smoothed Reeb graph)

Let $\mathbb{X}$ be a topological space and f be a function on $\mathbb{X}$. Additionally, let r_1 and r_2 be two bounded positive function on $\mathbb{X}$ with

$$\varepsilon := \sup_{x \in \mathbb{X}} |r_1(x) - r_2(x)|.$$

Then the r_1-smoothed Reeb graph $R(\mathbb{X}_{r_1}, f_{r_1})$ and the r_2-smoothed Reeb graph $R(\mathbb{X}_{r_2}, f_{r_2})$ are ε-interleaved.

Metric measure spaces

- A *metric measure space* is a triple $(\mathbb{X}, d_{\mathbb{X}}, \mu)$ where $(\mathbb{X}, d_{\mathbb{X}})$ is a metric space and μ is a probability measure on (the Borel σ-algebra of) $\mathbb{X}$.
- What measure do we use to model the data importance?
- Use distance to a measure $d_{\mu,m}$ or Kernel distance to a measure $D_{\mu,K}$ as local smoothing parameters:
  - Distance to a measure: $d_{\mu,m}$, Buchet et al. [2016]
  - Kernel distance to a measure: $D_{\mu,K}$ Phillips et al. [2015]

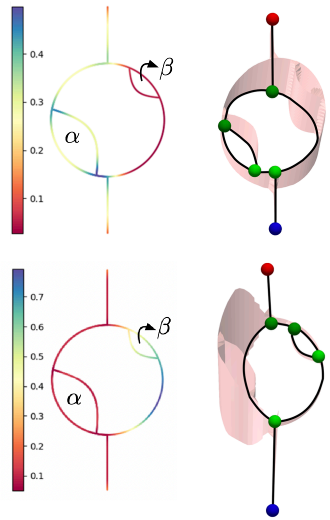
Distance between measures

- Let $(\mathbb{X}, d_{\mathbb{X}})$ be a metric space and μ, ν be two probability measures on $\mathbb{X}$. The 2-Wasserstein distance between μ and ν is defined as

$$W_2(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \left(\int_{\mathbb{X}\times\mathbb{X}} d_{\mathbb{X}}(x,y)^2 d\pi(x,y) \right)^{1/2},$$

  where $\Pi(\mu,\nu)$ is the set of all probability measures on $\mathbb{X} \times \mathbb{X}$ with marginals μ and ν.

Locally smoothed Reeb graphs



Locally smoothed Reeb graphs based on distance to a measure (top) and kernel distance to a measure (bottom).

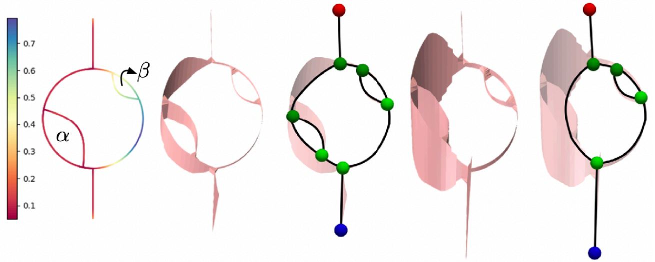
Stability of measure theoretic Reeb graphs

Theorem (Stability of $d_{\mu,m}$-smoothed Reeb graph)

Let $(\mathbb{X}, d_{\mathbb{X}}, \mu)$ *and* $(\mathbb{X}, d_{\mathbb{X}}, \nu)$ *be two metric measure spaces and* $f, g : \mathbb{X} \to \mathbb{R}$ *be two continuous functions. Let* $m \in (0, 1]$ *be a mass parameter. Then we have*

$$d_I(R(\mathbb{X}_{d_{\mu,m}}, f_{d_{\mu,m}}), R(\mathbb{X}_{d_{\nu,m}}, g_{d_{\nu,m}})) \le \|f - g\|_\infty + \frac{1}{\sqrt{m}} W_2(\mu, \nu).$$

Stability w.r.t. perturbations of the function and the measure.



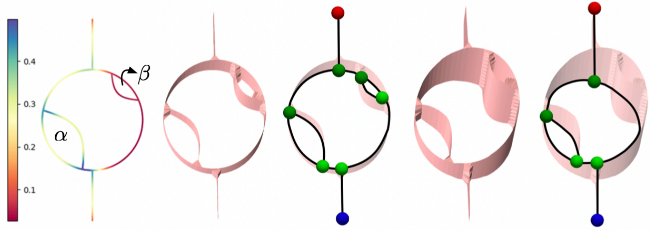
Stability of measure theoretic Reeb graphs

Theorem (Stability of $D_{\mu,K}$-smoothed Reeb graph)

Let $\mathbb{X}$ be a topological space. Let μ and ν be two probability measures on $\mathbb{X}$. Let K be an integrally strictly positive definite kernel function on $\mathbb{X}$. Consider two continuous functions $f, g : \mathbb{X} \to \mathbb{R}$. Then we have

$$d_I(R(\mathbb{X}_{D_{\mu,K}}, f_{D_{\mu,K}}), R(\mathbb{X}_{D_{\nu,K}}, g_{D_{\nu,K}})) \leq \|f - g\|_\infty + D_K(\mu, \nu).$$

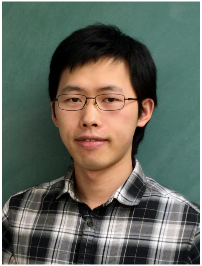
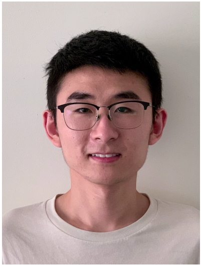
Stability w.r.t. perturbations of the function and the measure.



Discussion: robust topological descriptors

- What is the robust topology of data?
- Data importance captured by a measure.
- Our stability results hold for Reeb spaces, $R(\mathbb{X}, f), f : \mathbb{X} \to \mathbb{R}^d$, multivariate data analysis.
- Our findings demonstrate the stability of both Reeb graph and Reeb space constructions against perturbations of the function and the measure, thereby offering robust improvements for these topological descriptors.
- More: range-integrated reeb graphs, with a measure on the range of a function. Stability results.
- Ongoing: applications in topological data analysis and visualization.

Acknowledgement

- Measure Theoretic Reeb Graphs and Reeb Spaces. **Qingsong Wang**, Guanqun Ma, Raghavendra Sridharamurthy, Bei Wang. SoCG 2024. arXiv preprint arXiv:2401.06748, 2024.
- This work was partially supported by NSF DMS-2301361, IIS-2145499, and DOE DE-SC0021015.

Enrique Alvarado, Robin Belton, Emily Fischer, Kang-Ju Lee, Sourabh Palande, Sarah Percival, and Emilie Purvine. G-Mapper: Learning a cover in the mapper construction. arXiv preprint arXiv:2309.06634, 2023.

Aravindakshan Babu. *Zigzag Coarsenings, Mapper Stability and Gene-network Analyses*. PhD thesis, Stanford University, 2013.

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

Ulrich Bauer, Xiaoyin Ge, and Yusu Wang. Measuring distance between Reeb graphs. In *Proceedings of the 30th International Symposium on Computational Geometry*, pages 464–474, 2014.

Ulrich Bauer, Elizabeth Munch, and Yusu Wang. Strong equivalence of the interleaving and functional distortion metrics for reeb graphs. In *31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34, pages 461–475. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015.

References II

Ulrich Bauer, Barbara Di Fabio, and Claudia Landi. An edit distance for Reeb graphs. In A. Ferreira, A. Giachetti, and D. Giorgi, editors, *Eurographics Workshop on 3D Object Retrieval*, Eindhoven, The Netherlands, 2016. The Eurographics Association. doi: 10.2312/3dor.20161084.

Ulrich Bauer, Claudia Landi, and Facundo Memoli. The Reeb graph edit distance is universal. *Foundations of Computational Mathematics*, 21 (5):1441–1464, 2020. doi: https://doi.org/10.1007/s10208-020-09488-3.

Kenes Beketayev, Damir Yeliussizov, Dmitriy Morozov, Gunther Weber, and Bernd Hamann. Measuring the distance between merge trees. *Topological Methods in Data Analysis and Visualization III*, pages 151–165, 2014. doi: https://doi.org/10.1007/978-3-319-04099-8_10.

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

Adam Brown, Omer Bobrowski, Elizabeth Munch, and Bei Wang. Probabilistic convergence and stability of random mapper graphs. *Journal of Applied and Computational Topology*, 5:99–140, 2021. doi: 10.1007/s41468-020-00063-x.

Mickaël Buchet. *Topological inference from measures*. Theses, Université Paris Sud - Paris XI, December 2014.

Mickaël Buchet, Frédéric Chazal, Steve Y Oudot, and Donald R Sheehy. Efficient and robust persistent homology for measures. *Computational Geometry*, 58:70–96, 2016.

Quang-Thinh Bui, Bay Vo, Hoang-Anh Nguyen Do, Nguyen Quoc Viet Hung, and Vaclav Snasel. F-Mapper: A fuzzy mapper clustering algorithm. *Knowledge-Based Systems*, 189(105107), 2020.

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

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

Mathieu Carriére, Bertrand Michel, and Steve Oudot. Statistical analysis and parameter selection for mapper. *Journal of Machine Learning Research*, 19:1–39, 2018.

Nithin Chalapathi, Youjia Zhou, and Bei Wang. Adaptive covers for mapper graphs using information criteria. *IEEE International Conference on Big Data (IEEE BigData)*, 2021.

Frédéric Chazal and Jian Sun. Gromov-hausdorff approximation of filament structure using reeb-type graph. In *Proceedings of the thirtieth annual symposium on Computational geometry*, pages 491–500, 2014.

Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the 25th Annual Symposium on Computational Geometry*, pages 237–246, New York, NY, USA, 2009. ACM.

Vin De Silva, Elizabeth Munch, and Amit Patel. Categorified Reeb graphs. *Discrete & Computational Geometry*, 55(4):854–906, 2016a. doi: https://doi.org/10.1007/s00454-016-9763-9.

Vin De Silva, Elizabeth Munch, and Amit Patel. Categorified Reeb graphs. *Discrete & Computational Geometry*, 55(4):854–906, 2016b.

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

Tamal K. Dey, Facundo Mémoli, and Yusu Wang. Topological analysis of nerves, Reeb spaces, mappers, and multiscale mappers. In Boris Aronov and Matthew J. Katz, editors, *33rd International Symposium on Computational Geometry*, volume 77 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 36:1–36:16, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Pawel Dlotko. Ball mapper: a shape summary for topological data analysis. *arXiv: 1901.07410*, 2019.

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

Fangfei Lan, Salman Parsa, and Bei Wang. Labeled interleaving distance for Reeb graphs. arXiv preprint arXiv:2306.01186, 2023.

Mingzhe Li, Xinyuan Yan, Lin Yan, Tom Needham, and Bei Wang. Flexible and probabilistic topology tracking with partial optimal transport. arXiv preprint arXiv:2302.02895, 2023.

Dmitriy Morozov, Kenes Beketayev, and Gunther Weber. Interleaving distance between merge trees. *Proceedings of Topology-Based Methods in Visualization (TopoInVis)*, 2013.

Elizabeth Munch and Anastasios Stefanou. The ℓ^∞-cophenetic metric for phylogenetic trees as an interleaving distance. In *Research in Data Science*, Association for Women in Mathematics Series, pages 109–127. Springer International Publishing, 2019.

Elizabeth Munch and Bei Wang. Convergence between categorical representations of reeb space and mapper. In *32nd International Symposium on Computational Geometry (SoCG 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

Valerio Pascucci, Giorgio Scorzelli, Peer-Timo Bremer, and Ajith Mascarenhas. Robust on-line computation of reeb graphs: Simplicity and speed. *ACM Trans. Graph.*, 26(3):58–es, jul 2007. ISSN 0730-0301. doi: 10.1145/1276377.1276449. URL `https://doi.org/10.1145/1276377.1276449`.

Dau Pelleg and Andrew Moore. X-means: Extending k-means with efficient estimation of the number of clusters. *Proceedings of the 17th international conference on machine learning*, pages 727–734, 2000.

Jeff M Phillips, Bei Wang, and Yan Zheng. Geometric inference on kernel density estimates. In *Proceedings of the 31st International Symposium on Computational Geometry*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.

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

References VIII

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

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

Georges Reeb. Sur les points singuliers d'une forme de pfaff completement integrable ou d'une fonction numerique [on the singular points of a completely integrable pfaff form or of a numerical function]. *Comptes Rendus Acad. Sciences Paris*, 222:847–849, 1946.

Himangshu Saikia, Hans Peter Seidel, and Tino Weinkauf. Extended branch decomposition graphs: Structural comparison of scalar data. *Computer Graphics Forum (CGF)*, 33(3):41–50, 2014. doi: https://doi.org/10.1111/cgf.12360.

Gurjeet Singh, Facundo Mémoli, Gunnar E Carlsson, et al. Topological methods for the analysis of high dimensional data sets and 3d object recognition. *PBG@ Eurographics*, 2:091–100, 2007.

Raghavendra Sridharamurthy, Talha Bin Masood, Adhitya Kamakshidasan, and Vijay Natarajan. Edit distance between merge trees. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 26(3):1518–1531, 2020. doi: 10.1109/TVCG.2018.2873612.

Bharath K Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert RG Lanckriet. Hilbert space embeddings and metrics on probability measures. *The Journal of Machine Learning Research*, 11:1517–1561, 2010.

Dilip Mathew Thomas and Vijay Natarajan. Symmetry in scalar field topology. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 17(12):2035–2044, 2011. doi: 10.1109/TVCG.2011.236.

Lin Yan, Talha Bin Masood, Raghavendra Sridharamurthy, Farhan Rasheed, Vijay Natarajan, Ingrid Hotz, and Bei Wang. Scalar field comparison with topological descriptors: properties and applications for scientific visualization. *Computer Graphics Forum (CGF)*, 40(3): 599–633, 2021. doi: https://doi.org/10.1111/cgf.14331.

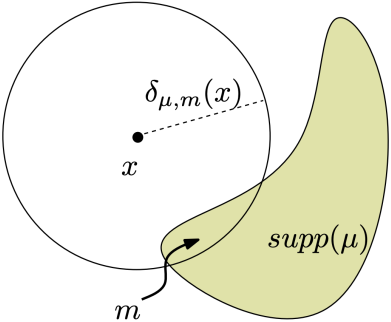
References X

Youjia Zhou, Nithin Chalapathi, Archit Rathore, Yaodong Zhao, and Bei Wang. Mapper Interactive: A scalable, extendable, and interactive toolbox for the visual exploration of high-dimensional data. *IEEE 14th Pacific Visualization Symposium (PacificVis)*, 2021.

Extra

Distance to a measure: pointwise importance or robustness

- $(\mathbb{X}, \mu)$: a metric measure space; $m \in (0, 1]$: a mass parameter.
- Define $\delta_{\mu,s}$ as $\delta_{\mu,s} : x \in \mathbb{X} \mapsto \inf\{r > 0 \mid \mu(\bar{B}(x, r)) > s\}$.
- $\bar{B}(x, r)$ denotes the closed ball of radius r centered at x.
- $\delta_{\mu,s}$: smallest distance needed to capture a mass of at least s.



Chazal 2010[1]

[1] https://atmcs4.appliedtopology.org/talks/Chazal/geomInference.pdf

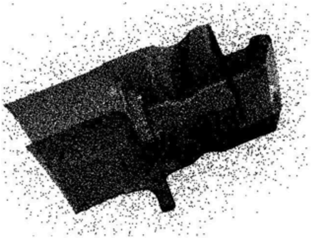
Distance to a measure $d_{\mu,m}$

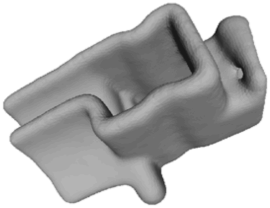
Definition ([Buchet et al., 2016, Definition 1.1])

Let $(\mathbb{X}, \mu)$ be a metric measure space and let $m \in (0, 1]$ be a mass parameter. We define the *distance to a measure* function $d_{\mu,m} : \mathbb{X} \to \mathbb{R}$ as

$$d_{\mu,m} : x \in \mathbb{X} \mapsto \sqrt{\frac{1}{m} \int_0^m \delta_{\mu,s}^2(x) ds},$$

where $\delta_{\mu,s} : x \in \mathbb{X} \mapsto \inf\{r > 0 \mid \mu(\bar{B}(x, r)) > s\}$. $\bar{B}(x, r)$ denotes the closed ball of radius r centered at x.





Stability of distance to a measure $d_{\mu,m}$

Theorem ([Buchet et al., 2016, Theorem 3.3] for $\mathbb{R}^n$; [Buchet, 2014, Proposition 3.14] for general metric spaces)

Let μ and ν be two probability measures on a metric space $(\mathbb{X}, d_{\mathbb{X}})$ and let $m \in (0, 1]$ be a mass parameter. Then:

$$\|d_{\mu,m} - d_{\nu,m}\|_\infty \leq \frac{1}{\sqrt{m}} W_2(\mu, \nu),$$

where $W_2(\mu, \nu)$ is the 2-Wasserstein distance between μ and ν.

Integrally strictly positive definite kernel function

Definition (Sriperumbudur et al. [2010])

Let $\mathbb{X}$ be topological space. A (Borel) measurable function $K : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ is called an *integrally strictly positive definite* kernel function if for all finite signed Borel measures μ on $\mathbb{X}$, there is

$$\int_{\mathbb{X}\times\mathbb{X}} K(x, x')d\mu(x)d\mu(x') > 0.$$

E.g. a Gaussian kernel function

Kernel distance

Definition (Sriperumbudur et al. [2010], Phillips et al. [2015])

Let $\mathbb{X}$ be a topological space. Let μ and ν be two probability measures on $\mathbb{X}$. Let K be an integrally strictly positive definite kernel function. Then the kernel distance D_K between μ and ν is defined as

$$D_K(\mu, \nu) := \sqrt{\kappa(\mu, \mu) + \kappa(\nu, \nu) - 2\kappa(\mu, \nu)},$$

where $\kappa(\mu, \nu)$ is defined as $\kappa(\mu, \nu) := \int_{\mathbb{X}\times\mathbb{X}} K(x, x')d\mu(x)d\nu(x')$.

Kernel distance to a measure $D_{\mu,K}$

Definition (Phillips et al. [2015])

Let μ be a probability measure on a topological space $\mathbb{X}$. Let K be an integrally strictly positive definite kernel function. Then the kernel distance $D_{\mu,K}$ with respect to μ is a function $D_{\mu,K} : \mathbb{X} \to \mathbb{R}$ defined as

$$D_{\mu,K}(x) = D_K(\mu, \delta_x),$$

where δ_x is the Dirac delta measure at x.

Stability of a kernel distance to a measure

Theorem (Phillips et al. [2015])

Let μ and ν be two probability measures on a topological space $\mathbb{X}$. Let K be an integrally strictly positive definite kernel function. Then

$$\|D_{\mu,K} - D_{\nu,K}\|_{\infty} \leq D_K(\mu, \nu),$$

where $D_K(\mu, \nu)$ is the kernel distance between μ and ν.