TDA Seminar

In recent years, the idea to use topological methods to analyze data has gained some traction. This is mainly due to the development of *persistent homology*, which generally works as follows:

We start with some data P that has a notion of distance, usually a finite metric space. Using the points in P as vertices, one then builds a filtration of simplicial complexes $K_\bullet(P)$ by successively connecting vertices, edges, etc. via higher dimensional simplices depending on their distances to each other. Afterwards, we apply homology with field coefficients to each complex in the filtration and the inclusion maps. This yields a family of vector spaces $H(K_\bullet(P))$ that are connected by linear maps. The isomorphism type of such a diagram of vector spaces can be completely described by a discrete invariant called a *barcode* that records the dimension of each vector space and all the ranks of the maps between spaces. The barcode $B(H(K_\bullet(P)))$ can be efficiently computed by performing column operations on a certain matrix and is then used to infer properties of the initial data P. Importantly, this inference is stable because the mapping $P \mapsto B(H(K_\bullet(P)))$ can be shown to be 1-Lipschitz with respect to appropriate metrics.

The rough plan for the seminar is to first go through all of the above steps in more detail and answer the following questions: What are the standard ways of constructing simplicial complexes from finite metric spaces and how are the different constructions related? How are barcodes defined formally? How does one compute the barcode of filtrations of simplicial complexes? What can one learn about the data from the barcode? What are the metrics for the stability theorem and how could one prove it?

After that, there are several possible topics that could be chosen according to the interests of the participants.

1. Homology computation and standard complex constructions (Maurice)

Give a very brief review of simplicial complexes and their homology (any standard algebraic topology reference you like will do). Explain how to compute simplicial homology via matrix reduction/smith normal form, maybe compute a small example ([5, Section 3.1], [12, §1.11]. Define the standard Rips, Čech, and Delaunay/alpha complexes to construct simplicial complexes from point clouds/finite metric spaces, explain that they form filtrations, maybe draw some examples ([5, Section 1.5], [13, Definition 3, Sections 5.2.1-5.2.3]).

2. More on the standard complex constructions (Jiajun)

State the most important results pertaining to the previously introduced constructions: Inclusions between Rips and Čech ([5, Corollary 1.12]), homotopy equivalence between Delaunay and Čech ([2, Čech-Delaunay Collapsing Theorem] without the part about Wrap complexes), nerve theorem ([5, Section 4.2]), Nyiogi–Smale–Weinberger theorem ([5, Theorem 5.4]). Draw pictures and present proof ides as you see fit. If time permits you can also present [5, Theorem 5.5].

3. Persistence modules and their structure theory (Merik)

Define persistent Betti numbers (parts of [5, Section 5.2]), persistence modules as poset representations, interval modules, barcodes, and persistence diagrams [5, Section 6] . Explain barcode uniqueness via the Krull–Remak–Schmidt–Azumaya theorem ([9, Theorem 2.7]). State Crawley-Boevey's Theorem on existence of barcodes for PFD persistence modules with totally ordered index set ([5, Theorem 6.18]). If time permits maybe prove it for the special case of a finite index set ([5, Theorem 6.16]).

4. Persistent homology and barcode computation (Dani)

Explain that the matrix reduction algorithm from the first talk not only computes Betti numbers of a single complex, but actually yields the barcode of a whole filtration if the rows and columns of the matrix are ordered correctly [5, Section 7.1], [1, Section 3.1]. Talk about how to speed this computation up by using clearing [5, Section 7.2], [1, Section 3.2] and cohomology-based algorithms [5, Section 8.3], [1, Section 3.3]).

5. Stability theorem (Kevin W.)

Define the bottleneck [5, Definition 9.4] and interleaving distances [5, Definition 9.12], explain that they are extended pseudo-metrics, state the algebraic stability (or isometry) theorem [5, Theorem 9.13]. State the stability theorem for sublevel set persistence [5, Theorem 9.6], state stability theorems for the standard simplicial filtrations [5, Theorems 9.7 and 9.8] and prove all these theorems assuming algebraic stability [5, page 57]. If time permits: prove the converse algebraic stability part of the isometry theorem, i.e. $d_I \leq d_B$ [5, Section 13.1].

6. Mapper (Fernando)

Define the Reeb graph of a real-valued function on a manifold, show that it is a topological graph [5, Section 11.1]. Define the continuous mapper associated to a function and a cover of the reals [5, Section 11.2]. Define the mapper of a filtered point cloud, briefly talk about possible clustering methods to use [5, Sections 10.3 and 11.3]. Show pictures of examples [16], show pictures detailing how the output depends on the choices involved [8].

7. More representation theory and multiparameter persistence (Julia)

Talk about quiver representations, Gabriel's theorem, wild/tame representation type [14, Section 1.1]. Explain why multiparameter persistence modules (i.e. ones indexed by $\mathbb{R}^d$) do not admit barcode decompositions in general [5, Section 12.2]. Sketch a possible remedy through signed barcodes [6, Corollary 5.1].

8. Clustering (Dia)

Present some classical clustering methods, interpret hierarchical clustering as H_0 persistence, show non-existence of ideal clustering, ToMaTo ([5, Section 10.1-10.2]).

9. Covid phylogenetic tree (Kevin K.)

Explain what a phylogenetic tree is and how persistent homology can potentially be used on such trees to find adaptive mutations of viruses following [4].

10. Cosmic microwave background (Enya)

Explain what the cosmic microwave background is and how persistent homology can be used to detect possible non-Gaussianity in it following [15].

References

[1] U. Bauer. Ripser: efficient computation of Vietoris-Rips persistence barcodes. *J. Appl. Comput. Topol.*, 5(3):391–423, 2021.

[2] U. Bauer and H. Edelsbrunner. The Morse theory of Čech and Delaunay complexes. *Trans. Amer. Math. Soc.*, 369(5):3741–3762, 2017.

[3] U. Bauer and M. Lesnick. Induced matchings and the algebraic stability of persistence barcodes. *J. Comput. Geom.*, 6(2):162–191, 2015.

[4] M. Bleher, L. Hahn, J. A. Patino-Galindo, M. Carriere, U. Bauer, R. Rabadan, and A. Ott. Topological data analysis identifies emerging adaptive mutations in sars-cov-2, 2021. Preprint.

[5] M. B. Botnan. Topological data analysis, 2020. Lecture Notes.

[6] M. B. Botnan, S. Oppermann, and S. Oudot. Signed barcodes for multi-parameter persistence via rank decompositions and rank-exact resolutions. Preprint, 2021.

[7] P. Bubenik, M. Hull, D. Patel, and B. Whittle. Persistent homology detects curvature. *Inverse Problems*, 36(2):025008, 2020.

[8] M. Carrière, B. Michel, and S. Oudot. Statistical analysis and parameter selection for Mapper. *J. Mach. Learn. Res.*, 19:Paper No. 12, 39, 2018.

[9] F. Chazal, V. de Silva, M. Glisse, and S. Oudot. *The structure and stability of persistence modules*. SpringerBriefs in Mathematics. Springer, 2016.

[10] B. Giunti. TDA applications. Zotero Library.

[11] Inria. GUDHI tutorials. Github repository.

[12] J. R. Munkres. *Elements of algebraic topology*. Addison-Wesley Publishing Company, Menlo Park, CA, 1984.

[13] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6:1–38, 2017.

[14] S. Y. Oudot. *Persistence theory: from quiver representations to data analysis*, volume 209 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015.

[15] P. Pranav, R. J. Adler, T. Buchert, H. Edelsbrunner, B. J. T. Jones, A. Schwartzman, H. Wagner, and R. van de Weygaert. Unexpected topology of the temperature fluctuations in the cosmic microwave background. *A&A*, 627:A163, 2019.

[16] G. Singh, F. Mémoli, G. E. Carlsson, et al. Topological methods for the analysis of high dimensional data sets and 3d object recognition. *PBG@ Eurographics*, 2, 2007.