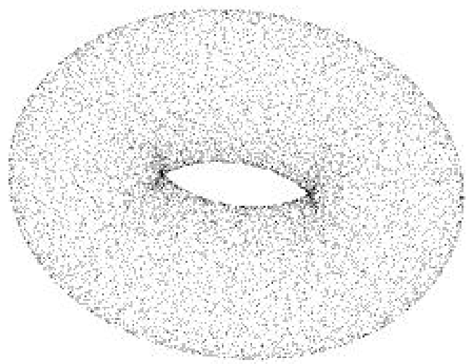
Seminar on Topological Data Analysis

Oihana Garaialde Ocaña and Marcus Zibrowius

Summer 2018



The aim of this seminar is to understand how topology can be used in data analysis. We will study *point clouds*, i. e. large (finite) sets of data points in, say, $\mathbb{R}^n$, and see how topological tools can be used to obtain qualitative and quantitative information in such diverse fields as image recognition, neuroscience, the study of diseases, etc. The main tool that we will study is *persistent homology*. The usual homology of a finite set of points is not very interesting, of course, but persistent homology is indeed able to detect the "shape" of point clouds.

The seminar is aimed at both **bachelor and master level students in pure or applied mathematics**. The course does not require any specific prior knowledge in topology as we shall introduce all the topological notions required throughout the course.

Each meeting will last 90 minutes. Each of you will give a talk of at most 75 minutes so that there is plenty of time for questions and feedback. All talks will be in English. Please prepare and hand in lecture notes for your talk at least one week before the talk.

Initial meeting:	**Wednesday 4th April, 14.30** in 25.22.00.81
Regular sessions:	**Wednesdays 14.00–15.30** in 25.22.U1.74 (starting 18th April)

Background

Studying topological spaces on a point-set level is inherently difficult. The aim of algebraic topology is therefore to translate questions about topological spaces into problems of a more algebraic nature. For example, the *fundamental group*, $\pi_1(X)$, of a topological space X and, more generally, *higher homotopy groups* $\{\pi_n(X)\}_{n\geq 1}$, provide a way of analysing topological spaces from the category of (abelian) groups [Hatch, Chapters 1, 4]. These groups can tell two non-homeomorphic as well as two non-homotopy equivalent topological spaces apart. However, these groups are extremely hard to compute. For instance, most of the homotopy groups of spheres are not known.

A more computable alternative to homotopy groups are *homology groups* [Hatch, Chapter 2]. Roughly speaking, the nth homology group counts the number of n-dimensional holes in a topological space. For example, a circle has exactly one one-dimensional hole, while a disk has no holes at all. There are different approaches or constructions to define homology groups: simplicial, singular and cellular homology, for instance.

Homology groups *per se* cannot possibly give any interesting information about point clouds, i. e. about discrete finite sets of points. In this seminar, we shall introduce *persistent homology groups* to study the shape of such point clouds. With a given point cloud, we will associate a sequence of nested topological objects, and study the homology of these associated objects instead of the homology of the initial point set. There will be homology classes that appear and disappear along the sequence of these topological objects, and we will be interested in the classes that *persist* (or 'last long') over the sequence.

Outline

There will be up to 15 talks in total. The talks with a $*$ are optional.

Talk 0: Introduction *(Oihana Garaialde Ocaña)* I will give an overview over the seminar and examine an example to illustrate how persistent homology can be applied to study a data set. More precisely, I will point out the steps that need to be carried out and connect them with the talks that you will be giving in future sessions.

Talk 1: Simplicial complexes *(Thomas Buchholz)* Define simplicial complexes and simplicial maps [Munk, Sections 1–3]. State the Simplicial Approximation Theorem and the Nerve Theorem, probably without proof [Munk, Sections 14 and 16].

Talk 2: Simplicial homology *(Alex and Karina)* Define the homology of chain complexes and introduce simplicial homology. Define Betti numbers. Explain the importance of homotopy invariance. Give lots of examples. [Munk] [Hatch, Theorem 2.44 and p. 130]

Talk 3: Simplicial complexes associated with point clouds *(Christof and Johannes)* Explain different simplicial complexes that can be defined from a discrete set of points: Vietoris-Rips complex, Čech complex, Delaunay and α-complexes. [EH10, pp. 59–74]

Talk 4*: Computability of homology Describe matrix reduction and give examples [Munk, Sections 10,11]. Also explain an algorithmic version of this method to compute homology with coefficients in a finite field [CZ05, Sections 4, 6].

Talk 5: Persistent homology *(Alessa)* Define persistent homology groups and barcodes [CZ05]. Give examples. State the Fundamental Theorem of Persistent Homology and explain its proof [Bel15, Proof of Theorem 1.2.4], [Bel17, pp. 7–9] [CZ05].

Talk 6*: Software *(Alex)* Describe the two software packages *Ripser* [Rips] and *GUDHI* [GU]. Give a live demonstration, using examples from GHUDI's documentation and, as far as possible, examples from the seminar. Prior exposure to C++ and Python will be helpful, but keep in mind when planning your talk that half the audience will not be particularly familiar with either. Please speak to Marcus Zibrowius if you are interested in giving this talk.

Talk 7*: Morse Theory Define Morse functions and transversality conditions. Give an idea of the power of this theory. [EH10, Chapter VI] [Zo96, Chapter 5]

Talk 8: Stability Theorem *(Christof)* State the theorem, explain the motivation behind it and sketch its proof [CSEH] [EH08, Section 6].

Talk 9: Piecewise linear functions and Reeb graphs Morse functions can be too restricted. Define piecewise linear (PL) functions and PL Morse inequalities. Describe how to 'draw' PL functions via Reeb graphs. State analogous results to those about Morse functions for PL functions. [EH10] [CME+]

Talk 10: Application I: Statistics in persistent homology Summarize [FLR+]: The authors employ statistical methods to derive confidence sets that allow them to separate topological signals from topological noise, i. e. to isolate the significant parts of a given data set.

Talk 11*: An algorithm for Reeb graphs Summarize [DN09]: The article describe an efficient algorithm for computing Reeb graphs.

Talk 12: Mapper *(Alessa)* Summarize [SMG07]: The authors give a generalization of Reeb graphs to study high-dimensional data sets. This method is independent of the chosen cluster (a notion that may be examined in more detail in the next talk).

Talk 13*: An Impossibility Theorem *(Marcus Zibrowius)* Explain that there are no clustering algorithms that simultaneusly satisfy scale-invariance, richness and consistency. [Kl02]

Talk 14*: A_∞-structures for persistent homology Can usual Betti numbers distinguish two linked or unlinked circles? Give an idea of how one can equip homology groups with an A_∞-structure and of how the previous theory can be extended to A_∞-persistence. [Bel15, Chapters 2 and 3]

Talk 15: Application II: A subgroup of breast cancers Summarize [NLC11], one of the most famous applications of persistent homology to a real-world problem.

Talk 16*: Application III: Image recognition Summarize [LOC]: The authors explain a method that consists of computing the persistence diagrams of functions defined on different data modalities, including 2D shapes, textures and triangle meshes.

Talk 17*: Application IV: Neuronal morphologies Summarize [KDS+]: Nervous systems are characterized by neurons displaying a diversity of morphological shapes. The authors propose a stable topological measure as a standardized descriptor for any tree-like morphology, which encodes its skeletal branching anatomy.

Talk 18*: Application V: Sensor networks Summarize [SG07]: The authors consider coverage problems in sensor networks with minimal sensoring capabilities. In particular, they demonstrate that a stationary collection of sensor nodes with no localization can verify coverage in a bounded domain of unknown topological type, as long as the boundary is not too pinched.

Talk 19*: Application VI: Activity in the visual cortex Summarize [SMI+]: Information in the cortex is thought to be represented by the joint activity of neurons. The authors describe how fundamental questions about neural representation can be cast in terms of the topological structure of population activity. They find that the topological structures of activity patterns when the cortex is spontaneously active are similar to those evoked by natural image stimulation and consistent with the topology of a two-sphere.

Talk 20*: Your favourite application If you come across an interesting application of persistent homology that you would like to learn more about, different from the applications listed above, why not prepare a talk about it? Just speak to us, and we'll see how and where it can fit in.

Literature

For a quick introduction to persistent homology, see [Gh08] or [EH08]. General introductions to Algebraic Topology include [Hatch] and [Munk]. For a textbook geared specifically towards applications, see [Ghri].

[Bel15] F. Belchí, *A_∞-persistence*, Ph.D. Thesis, 2015.

[Bel17] F. Belchí, *Optimazing the topological information of the A_∞-persitence groups*. https://arxiv.org/abs/1706.06019

[BK07] P. Bubenik, P. T. Kim, *A statistical approach to persistent homology*, Homology, Homotopy and Applications 9 (2), 337–362, 2007.

[CZ05] G. Carlsson, Zomorodian *Computing persistent homology*, Disc. Computational Geometry 33 (2), 249–274, 2005.

[CME+] K. Cole-McLaughlin, H. Edelsbrunner, J. Harer, V. Natarajan, V. Pascucci, *Loops in Reeb Graphs of 2-Manifolds*, http://www.pascucci.org/pdf-papers/reeb-loops.pdf

[DN09] H. Doraiswamya, V. Natarajanb *Efficient algorithms for computing Reeb graphs*, Computational Geometry 42, 67, 606–616, 2009.

[CSEH] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, *Stability of Persistence Diagrams*, Disc. Comp. Geom. 37: 103–120, 2007.

[EH08] H. Edelsbrunner, J. L. Harer, *Persistent homology – a Survey*, Surveys on discrete and computational geometry, 257–282, Contemp. Math., 453, Amer. Math. Soc., Providence, RI, 2008.

[EH10] H. Edelsbrunner, J. L. Harer, *Computational topology. An introduction.* American Mathematical Society, Providence, RI, 2010.

[FLR+] B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, A. Singh, *Confidence sets for persistence diagrams.* https://arxiv.org/abs/1303.7117

[Ghri] R. Ghrist, *Elementary Applied Topology*, 2014.

[Gh08] R. Ghrist, *Barcodes: the persistent topology of data.* Bull. Amer. Math. Soc. (N.S.) 45 (2008), no. 1. https://doi.org/10.1090/S0273-0979-07-01191-3

[GU] *GUDHI.* http://gudhi.gforge.inria.fr

[Hatch] A. Hatcher, *Algebraic Topology.* https://www.math.cornell.edu/~hatcher/AT/AT.pdf

[KDS+] L. Kanari, P. Dlotko, M. Scolamiero, R. Levi, J. Shillcock, K. Hess, H. Markram, *Quantifying topological invariants of neuronal morphologies.* https://arxiv.org/abs/1603.08432

[Kl02] J. Kleinberg, *An Impossibility Theorem for Clustering*, NIPS 2002 Proceedings of the 15th International Conference on Neural Information Processing Systems, 463–470, 2002.

[LOC] C. Li, M. Ovsjanikov, F. Chazal, *Persistence-based Structural Recognition*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1995–2002, 2014.

[Munk] J. R. Munkres, *Elements of Algebraic Topology*, Addison-Wesley Publishing Company, 1984.

[NLC11] M. Nicolaua, A. J. Levineb, G. Carlsson, *Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival*, PNAS, 108 (17) 7265–7270, 2011.

[PWT+] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer and F. Vaccarino, *Homological scaffolds of brain functional networks*, J. R. Soc. Interface, 11: 20140873, 2014. http://dx.doi.org/10.1098/rsif.2014.0873

[Rips] *Ripser.* https://github.com/Ripser/ripser

[SC04] V. de Silva, G. Carlsson, *Topological estimation using witness complexes*, Eurographics Symposium on Point-Based Graphics, 2004.

[SG07] V. de Silva, R. Ghrist, *Coverage in sensor networks via persistent homology*, Algebraic and Geometric Topology 7, 339–358, 2007.

[SMG07] G. Singh, F. Mémoli, G. Carlsson, *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*, Eurographics Symposium on Point-Based Graphics 2007.

[SMI+] G. Singh, F. Memoli, T. Ishkhanov, G. Sapiro, G. Carlsson, D. L. Ringach, *Topological analysis of population activity in* *visual cortex*, PMC (8): 11.1–1118 (2010). http://doi.org/10.1167/8.8.11

[Zo96] A. J. Zomorodian, *Computing and comprehending topology: persistence and hierarchical Morse complexes*, Ph.D. Thesis, 1996.