Topological Data Analysis **Homework**
University of Fribourg Autumn Semester 2026
Instructor: Dr. Elena Xinyi Wang

Paper chase

Reading a research paper takes detours. A definition sends you to an earlier paper; a proof uses an unfamiliar idea; an experiment raises a new question. This assignment asks you to follow one of those trails and explain what you learn.

Choose a paper that interests you and use two further papers to work through questions that arise. This may be your first experience reading research papers in depth. You do not need experience writing papers, and I do not expect you to understand every proof. The aim is to explain what you have learned and how the three papers helped you get there.

What to write

Your report must fit within **three pages, including figures but excluding references**. Choosing what to explain, and explaining it concisely, is part of the assignment. Keep the template's 11-point font and margins. If you have a good reason to need more space, email me before submitting; I will decide whether to allow an exception.

Use the five short sections below. Write for a fellow master's student who knows the basic ideas of TDA but has not read your papers. Focus on a few ideas you can explain well.

1. **Paper 1: the main contribution.** In **one short paragraph**, explain the question, the main contribution, and why it matters. Be specific about what the authors prove, compute, or demonstrate. An application paper is just as welcome as a theoretical one.

2. **Paper 1: the mathematical or computational ingredients.** Explain one central method or result and the TDA tools behind it. For an application, describe the data, how they become a topological object, and how the output helps answer the original question. For a theoretical paper, explain the main construction and its assumptions. A small example or a well-chosen figure may help.

3. **Paper 1: where you got stuck.** Describe one difficulty that led you to look further. Locate it with a section, theorem, equation, figure, or page number, and explain what you tried. Go beyond naming an unfamiliar definition: what does it prevent you from understanding about the method, argument, or interpretation?

4. **Paper 2: follow the first lead.** Find a second research paper that helps with this difficulty. Briefly explain its relevant contribution, why you chose it, and what it clarified about Paper 1. Identify a new or remaining question to pursue. You do not need to summarise everything in the paper.

5. **Paper 3: follow the next lead.** Choose a third research paper to help with a question raised while reading Paper 2. Explain the connection and what you now understand better. End by looking back at your original question: what did you learn, and what remains unclear?

Choosing your papers

You are free to choose all three papers yourself; none has to come from the list below. Start with a question or an application that interests you. An application paper may use a tool introduced in one of the suggested papers, giving you a natural next step in the chase. Other TDA methods are welcome too.

- **Keep TDA central.** Paper 1 should develop, analyse, or make meaningful use of a topological method. Papers 2 and 3 may supply background from another area, but their role in your chase should be clear.
- **Choose something new to you.** Pick papers you have not already studied in depth. You should be able to grasp the main question while still having something to investigate.
- **Look beyond familiar work.** If you have a thesis or research supervisor, you may ask them for suggestions, but choose papers outside their own publications and mine. Their reference lists can be useful starting points.
- **Use three distinct research papers.** A conference paper and its expanded journal version count as one paper. Surveys, textbooks, lecture notes, and software documentation can help, but are additional sources rather than replacements for your three papers.

How to follow a useful trail

Start with the abstract, introduction, main results, and figures, then work carefully through one part that matters. Look in the references for the source of a method, or search for later papers that cite it. Keyword searches and recommendations can also help. You do not need to choose all three papers at the outset: let your questions guide the next choice.

For an application, follow the path from the data to the topological summary and then to the scientific claim. Ask what the evidence supports and what depends on modelling choices. You are welcome to discuss possible starting points with me.

The required project-proposal meeting

We will discuss your paper chase during the **required meeting about your project proposal**. Alongside your proposal, I will ask about what you learned from your chosen papers, the methods and results you describe, how the papers connect, and the parts you found difficult. Be ready to explain an example or a figure in your own words and to discuss how your reading might inform your project. You are not expected to know everything; you should be clear about what you understand and what you are still working out.

If this looks like an assignment you can hand over to AI, think again. A polished report cannot replace doing the reading and understanding it. Whatever tools you use, you are responsible for checking the claims and references, and you must be able to explain what you have written. The report and our conversation should reflect your own learning.

Writing and references

Give complete, compact references and cite any additional sources. Define the notation you use, distinguish the authors' claims from your interpretation, and identify the paper and version when referring to a numbered result. **Please use the supplied `paper-chase-student.tex` template for your report.** It includes the preamble, course header, and report outline. Keep its font size and margins unchanged. Figures count towards the three-page limit; references do not.

Suggested papers

This is an optional menu, with applications first and the supporting methods and theory afterwards. You may start with a scientific question and follow its references to the TDA tools, or start with a tool and find an application you care about. You are equally welcome to choose papers outside this list.

Entry suggests a relatively gentle start after basic persistent homology; **Intermediate** calls for some new background; **Advanced** is best approached through one carefully chosen result. The notes suggest manageable reading goals for a first encounter with the literature.

A. Geometry, signals, and sensing

A1. On the Local Behavior of Spaces of Natural Images
Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian. *International Journal of Computer Vision* 76 (2008), 1–12. Publication; full text
Entry; point-cloud topology. A classic study connecting small image patches with a Klein-bottle model. Trace the preprocessing, density selection, and topological computations. Distinguish the evidence from sampled data from the mathematical model; this makes a useful study of what topology does and does not establish in an application.

A2. Persistent Cohomology and Circular Coordinates
Vin de Silva, Dmitriy Morozov, and Mikael Vejdemo-Johansson. *Discrete & Computational Geometry* 45 (2011), 737–759. Open article
Intermediate; cohomology and least squares. Turns persistent degree-one cohomology into circle-valued coordinates for data. Follow the lifting, smoothing, and integration steps, and explain one experiment. A natural question for the chase is why an integer lift is needed and how the resulting coordinates reflect the topology.

A3. Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis
Jose A. Perea and John Harer. *Foundations of Computational Mathematics* 15 (2015), 799–838. Publication; open preprint
Intermediate; Fourier analysis helps. Connects the geometry of sliding-window embeddings with periodicity and persistent homology. Begin with a sinusoid, explain the role of the window length, and follow one structural or convergence result. Keep the mathematical guarantees separate from the empirical performance of the periodicity score.

A4. Coverage in sensor networks via persistent homology
Vin de Silva and Robert Ghrist. *Algebraic & Geometric Topology* 7 (2007), 339–358. Publication; open article
Intermediate to advanced; relative homology. Gives a homological criterion guaranteeing coverage from limited proximity information. Draw the sensing region and its boundary, then identify the pair of complexes and the class used in the criterion. Pay close attention to the geometric and boundary assumptions that make the guarantee possible.

B. Neuroscience, biology, and materials

B1. Persistent homology analysis of brain artery trees
Paul Bendich, J. S. Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer. *The Annals of Applied Statistics* 10(1) (2016), 198–218. Publication; open article
Intermediate; basic statistics. A detailed example of turning geometric data into persistence diagrams and statistical summaries. Follow one association with age and explain the choices that produce the summary. Distinguish an observed association from a causal explanation; the loops studied arise in geometric thickenings of the embedded artery trees.

B2. Clique topology reveals intrinsic geometric structure in neural correlations
Chad Giusti, Eva Pastalkova, Carina Curto, and Vladimir Itskov. *Proceedings of the National Academy of Sciences* 112(44) (2015), 13455–13460. Article

Intermediate; correlation matrices and simplicial complexes. Studies geometric structure in correlations from neural recordings. Explain how ordering correlations produces a filtration of clique complexes, then interpret one comparison of Betti curves with random or geometric controls. Be clear about what the analysis reveals about correlations, rather than anatomical connections.

B3. Topology of viral evolution
Joseph Minhow Chan, Gunnar Carlsson, and Raul Rabadan. *Proceedings of the National Academy of Sciences* 110(46) (2013), 18566–18571. Publication; open article

Intermediate; some evolutionary biology. Investigates evolutionary relationships that a tree alone cannot capture. Follow one example from genetic distances to persistent homology and its biological interpretation. Ask how the distance model and evidence support a claim about recombination or reassortment; a persistent loop alone does not establish a particular event.

B4. Hierarchical structures of amorphous solids characterized by persistent homology
Yasuaki Hiraoka, Takenobu Nakamura, Akihiko Hirata, Emerson G. Escolar, Kaname Matsue, and Yasumasa Nishiura. *Proceedings of the National Academy of Sciences* 113(26) (2016), 7035–7040. Article

Intermediate; geometric filtrations. Gives persistence diagrams a concrete interpretation in the structure of materials. Concentrate on silica glass and explain how simulated atomic configurations lead to diagrams and information about rings at different scales. One carefully explained figure is a good starting point for understanding the physical meaning of the topological features.

C. Mapper and exploratory data analysis

C1. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition
Gurjeet Singh, Facundo Mémoli, and Gunnar Carlsson. *Eurographics Symposium on Point-Based Graphics* (2007), 91–100. Open article

Entry; covers, clustering, and nerves. The original Mapper paper is short, concrete, and a natural place to start. Construct Mapper on a small example and explain the roles of the filter, cover, and clustering. Its construction and demonstrations invite further questions about stability and parameter choice.

C2. Extracting insights from the shape of complex data using topology
P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, and G. Carlsson. *Scientific Reports* 3 (2013), 1236. Open article

Entry; clustering and graph summaries. An accessible introduction to Mapper-style exploration through examples including voting records and basketball statistics. Choose one data set: explain what a point represents, how the network is constructed, and what a visible pattern means. Discuss the parameter choices and distinguish exploratory observations from independently validated conclusions.

C3. Statistical Analysis and Parameter Selection for Mapper
Mathieu Carrière, Bertrand Michel, and Steve Oudot. *Journal of Machine Learning Research* 19(12) (2018), 1–39. Open article

Intermediate to advanced; probability. Links practical choices in Mapper to approximation bounds and statistical guarantees under explicit geometric and sampling assumptions. Follow the construction and the main approximation bound, then choose either parameter selection or confidence regions. It is a useful follow-up to the original Mapper paper.

D. Representations, kernels, and learning

D1. Statistical Topological Data Analysis using Persistence Landscapes
Peter Bubenik. *Journal of Machine Learning Research* 16(3) (2015), 77–102. Open article

Intermediate; function spaces, with probability optional. Turns barcodes into functions that can be averaged and studied statistically. Construct a landscape for a small barcode, then choose either the stability results or the laws of large numbers and central limit theorems. A particularly good starting point for understanding why the representation of persistence matters.

D2. Persistence Images: A Stable Vector Representation of Persistent Homology
Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. *Journal of Machine Learning Research* 18(8) (2017), 1–35. Open article; erratum

Entry; integration and norm inequalities. A clear construction of finite vectors from diagrams, with a 1-Wasserstein stability theorem. Follow the weighting, smoothing, and pixel integration, especially the role of vanishing weights on the diagonal. The linked erratum corrects a formula in the dynamical-system example in Section 6.4.1.

D3. A Stable Multi-Scale Kernel for Topological Machine Learning
Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. *IEEE Conference on Computer Vision and Pattern Recognition* (2015), 4741–4748. Open article

Intermediate; inner products and Gaussian kernels. Uses a heat-equation construction to define a positive-definite kernel with 1-Wasserstein stability. Focus on reflection across the diagonal, the feature map, and the kernel formula. The short format makes this a useful comparison with landscapes or persistence images.

E. Algorithms and computation

E1. Dualities in Persistent (Co)Homology
Vin de Silva, Dmitriy Morozov, and Mikael Vejdemo-Johansson. *Inverse Problems* 27(12) (2011), 124003. Publication; open preprint

Intermediate; some cohomology. A compact bridge between algebra and algorithms, explaining the relationships among absolute and relative persistent homology and cohomology. Follow the running example and the matrix formulations. Ask why different computations recover related barcode information, and what changes when one needs actual cycles or cocycles.

E2. Ripser: efficient computation of Vietoris–Rips persistence barcodes
Ulrich Bauer. *Journal of Applied and Computational Topology* 5 (2021), 391–423. Open article

Intermediate; boundary-matrix reduction. Explains the mathematical and algorithmic choices behind an influential implementation. Choose one or two ideas, such as implicit coboundaries, clearing, or apparent pairs, and explain why they save work. The experiments give a useful opportunity to distinguish practical performance from worst-case guarantees.

F. Foundations and statistical guarantees

F1. Computing Persistent Homology
Afra Zomorodian and Gunnar Carlsson. *Discrete & Computational Geometry* 33 (2005), 249–274. Publication

Intermediate; some module theory. A foundational account of the algebra behind barcodes and their computation over a field. Follow the passage from a filtration to a graded module, and explain why the resulting interval description is possible. This leads naturally to asking what changes with more than one filtration parameter.

F2. Stability of Persistence Diagrams
David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. *Discrete & Computational Geometry* 37 (2007), 103–120. Publication

Intermediate; the original proof is demanding. The basic robustness theorem bounds bottleneck distance by the uniform change in suitably tame functions. Begin with the hypotheses and a perturbed-function example, then follow one part of the proof. A useful next question is how to express the same stability principle purely in terms of persistence modules.

F3. Geometric Inference for Probability Measures
Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. *Foundations of Computational Mathematics* 11 (2011), 733–751. Publication; author's copy

Advanced; measures and Wasserstein distance. Introduces distance to a measure to make geometric inference less sensitive to outliers. The empirical nearest-neighbour interpretation gives a concrete starting point. Focus on the definition and stability theorem before tackling reconstruction, and explain why ordinary distance to a point cloud can be fragile.

F4. Confidence Sets for Persistence Diagrams
Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. *The Annals of Statistics* 42(6) (2014), 2301–2339. Publication; open manuscript

Intermediate; probability and persistence stability. Connects stability bounds with statistical confidence sets. Choose one construction, such as subsampling, and explain exactly what its confidence statement guarantees and under which assumptions. This is a good route from an observed persistence diagram to the question of what can be inferred from it.

G. Further theory

These are options for students who would like to explore the algebra further. They are not prerequisites for choosing an application paper.

G1. Zigzag Persistence
Gunnar Carlsson and Vin de Silva. *Foundations of Computational Mathematics* 10 (2010), 367–405. Publication; open preprint

Intermediate; linear algebra and persistence modules. Allows arrows to point in either direction while retaining interval decompositions. Begin with an example that an ordinary filtration cannot describe, then compute a small zigzag barcode. The full algorithmic development is a more demanding second step.

G2. The Theory of Multidimensional Persistence
Gunnar Carlsson and Afra Zomorodian. *Discrete & Computational Geometry* 42 (2009), 71–93. Publication; full text

Advanced; multigraded modules. Explains why several filtration parameters lead to a substantially different classification problem and develops the rank invariant. Start with a bifiltration and ask what the invariant records and what it forgets. The classification arguments require more algebra than the one-parameter theory.

A few possible first steps

Follow a question from your reading rather than treating these as fixed routes. All three papers may come from outside the list.

- **Neural data:** B2 (neural correlations) can lead to E2 (Ripser) if you want to understand how persistence is computed for a clique filtration.
- **Exploratory analysis:** C2 (voting or basketball data) can lead to C1 (Mapper), then to C3 if parameter selection becomes your next question.
- **Statistical evidence:** B1 (artery trees) may lead you to D1 (landscapes) to investigate another way to summarise and average persistence.

Two useful background surveys

You are welcome to consult these while learning the background. They are additional reading aids, rather than substitutes for the three research papers:

Gunnar Carlsson, *Topology and data*, *Bulletin of the American Mathematical Society* 46(2) (2009), 255–308.

Frédéric Chazal and Bertrand Michel, *An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists*, *Frontiers in Artificial Intelligence* 4 (2021), 667963.