Math 412: Topology with Applications

March 18, 2016

- Instructor: Paul Bendich
- Day/Time: T/Th: 8:30-9:45 AM
- Room: Gross Hall 318
- Office Hours:
  - Mondays, 11-12, Math 210
  - Fridays, 11:45-1, Gross Hall 327

1 Course Summary

Introduction to the subject of Topological Data Analysis (TDA). We will learn some basic (and not-so-basic) topological, geometric, and algebraic tools, and see how to apply them in a variety of interesting situations. The course will begin by introducing a key TDA concept, the persistence diagram, and will quickly move on to several applications that use these diagrams. After the break, we will cover some more advanced topics from algebraic topology, and also see how they can be applied.

This course is centered around applications, and so the syllabus is designed to fit them in early and often. Along the way, you will also be exposed to some analytical techniques from statistics and machine-learning. Nonetheless, there will be quite a bit of traditional theorem-proof mathematics, both in the construction of TDA theory and in the production of algorithms fast and robust enough to meet the demands of actual data.

Each student will have the opportunity to participate in a course project, involving the application of TDA (and other) techniques to a dataset of their choosing.

2 Course Materials

We will use one textbook,

- *Computational Topology: An Introduction*, by Herbert Edelsbrunner and John Harer, American Mathematical Society,

but we will not go through it in a particularly linear fashion. In the Lecture Schedule below, all roman numerals refer to sections from this book. There will also be quite a lot of material taken from published journal articles and conference proceedings; these are indicated by bracketed numbers in the lecture schedule. Please do not feel obligated to read these extra papers in any detail, but they may well be useful for extra reading after lecture. All of these are publically available on the web.

3 Problem Sets

There will be homework problems, usually assigned after each lecture, and generally collected in bulk each Monday. Most of the exercises will be pencil-and-paper, but some of them may also involve some very basic computation. It is certainly permitted (and advisable!) to collaborate with others on these assignments. However, all work must be written up entirely on your own. **Late homework will not be accepted.**

4 Projects

Instead of a final exam, you will do a project. The requirements for this are quite open-ended: all that is needed is a real (or very compelling synthetic) dataset that you actually enjoy, and a question that can potentially be answered at least in part with TDA techniques. **Please ensure that you have full and legal access to the dataset, and that you have the right to perform the analyses you will perform.**

It is very likely that this will involve some form of coding, but all of the TDA packages out there (there is a Duke-based one called TDATools and several other freely available ones on the web) are very user-friendly, and I will be available to help you each step of the way. Given the size of the class, it'll be best if people are able to form groups of two or three, based on common interest. I'll facilitate this early in the semester.

More details about the project are given at the end of this document.

5 Grading

Half of your grade will be based on the problem sets, and half will be based on the project.

6 Lecture Schedule

Date	Material	Chapters/Papers
1/14	Course Overview	
1/19	Zero-dimensional Persistent Homology	VII.1, I.1
1/21	Metrics and Stability	VIII.2, VIII.4
1/26	Applications: Brain Arteries, Driver Behavior	[4], [10]
1/28	Homology (informal), Simplicial Complexes	
2/2	Simplicial Homology: Defintion, Examples	III.1, IV.1
2/4	Simplicial Homology: More Examples, Algorithm	IV.1, IV.2
2/9	Application: Neural Correlation in Clique Topology	[9]
2/11	Persistent Homology: Defintion, Examples	VII.1, VII.2
2/16	Persistent Homology: More Examples, Algorithm	VII.1, VII.2
2/18	Height Functions and Distance Functions	
2/23	Stability Theorems, Homology Inference	[5], [6]
2/25	Point-Cloud Triangulations	III.2, III.3
3/1	Data Expedition: Sliding-Windows and Weather Data	
3/3	Machine-Learning and Statistics Boot Camp	
3/8	blank	
3/10	blank	
3/22	Relative Homology	IV.3
3/24	Application: sensor networks	[8], [7]
3/29	Local Homology	[3]
3/31	Application: Road Network Reconstruction	[2]
4/5	Long Exact Sequence of the Pair	IV.3
4/7	Snake Lemma	IV.4
4/12	Mayer-Vietoris	IV.4
4/14	Rips-Collapse Algorithm	
4/19	Rips Complexes and Cover-Trees	[11]
4/21	Extended Persistence	VII.3
4/26	Application: Protein Docking	IX.2, [1]

7 Project Details

Here are the due dates associated to the project.

- By 3/21: project selection (subject to approval by me!)
- By 4/4: Annotated Bibliography, pdf to me by 5 pm.
- 4/25: Presentation, uploaded by 5 pm.
- 5/7: paper due, pdf to me by 10 pm.

I expect your project to produce two end-products: a presentation and a paper.

Presentation The presentation should be about 20 minutes long, and do an excellent job of summarizing your work at a high-level that is broadly understandable by all students in the class. Its main purpose will be to generate feedback from the other students, and myself, which you can then incorporate into your final paper. A secondary purpose might be to produce something that you can use to show off your work in later contexts.

Given the large size of the class, we won't use class time for presentations. Instead, I will require everyone to record themselves giving the presentation (having a slide show with your voice recorded is perfectly fine) and to upload this recording to a website (to be arranged). I'll then break the class into smaller groups, ask everyone in a group to watch all the presentations in their group and leave comments. Of course, I will watch all the presentations as well!

Paper The paper should be a maximum of six pages, including figures and references, and must be written in LaTex (I will provide a template). References should be in the form seen in this document.

Your paper should provide detailed and coherent answers to the following questions:

- What is your dataset and why do people care about it?
- What type of insight do you hope to gain using TDA methods?
- What other methods have people used to gain similar insight on this type of data?
- How did you actually use TDA methods on this data (give both conceptual and implementation-oriented details)?
- What did you find?
- What else would you try, given more time?

In addition, I expect you to incorporate the feedback from your presentation in a sensible way.

References

[1] Pankaj K. Agarwal, Herbert Edelsbrunner, John Harer, and Yusu Wang. Extreme elevation on a 2-manifold. *Discrete & Computational Geometry*, 36:553–572, 2006.

[2] Mahmuda Ahmed, Brittany Terese Fasy, and Carola Wenk. Local persistent homology based distance between maps. In *SIGSPATIAL*. ACM, Nov. 2014.

[3] P. Bendich, D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and D. Morozov. Inferring local homology from sampled stratified spaces. In *Proceedings 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 536–546, 2007.

[4] Paul Bendich, J.S. Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer. Persistent homology analysis of brain artery trees. *Annals of Applied Statistics*, 2015. to appear.

[5] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, January 2007.

[6] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have l_p-stable persistence. *Found. Comput. Math.*, 10(2):127–139, February 2010.

[7] V. de Silva and R. Ghrist. Coordinate-free coverage in sensor networks with controlled boundaries via homology. *The International Journal of Robotics Research*, 25(12):1205–1222, 2006.

[8] Robert Ghrist and Abubakr Muhammad. Coverage and hole-detection in sensor networks via homology. *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005.*, (1):254–260, 2005.

[9] Chad Giusti, Eva Pastalkova, Carina Curto, and Vladimir Itskov. Clique topology reveals intrinsic geometric structure in neural correlations. *PNAS*, 112(44):13455–13460, 2015.

[10] David Rouse, Adam Watkins, David Porter, John Harer, Paul Bendich, Nate Strawn, Elizabeth Munch, Jonathan DeSena, Jesse Clarke, Jeffrey Gilbert, Peter Chin, and Andrew Newman. Feature-aided multiple hypothesis tracking using topological and statistical behavior classifiers. volume 9474, pages 94740L–94740L–12, 2015.

[11] Donald R. Sheehy. Linear-size approximations to the Vietoris-Rips filtration. *Discrete & Computational Geometry*, 49(4):778–796, 2013.