AMAT 584 TDA II

Instructor: Michael Lesnick mlesnick@albany.edu

Course Website: Google "lesnick 584" or similar.
Also serves as course syllabus!

Office hours: M,W 4:15-5:15 + by appointment

(also, please let me know in advance if you will come to office hours)

About this course: Continuation of my TDA I course from fall 19.

Topics: simplicial complexes
abstract linear algebra + quotient paces
simplicial homology (counting holes with linear algebra)
persistent homology (barcodes of data)
- construction
- computation
- stability
- applications
- statistical aspects
- multiparameter persistent homology
- Mapper (another TDA tool)

Official prereq: TDA I

Unofficial prereq: my fall 2019 TDA course

Topics I'll assume you know:

- metric spaces, abstract topological spaces
- homeomorphism
- homotopy equivalence
- path connectedness
- equivalence relations
- single linkage clustering
- graphs: connected components + cycles.

If you are unfamilar with some of this stuff, I discourage you from taking this course.

As with last semester, my handwritten notes will serve as the main course reference.

New for this semester: Better file names.

- Additional references will be suggested for each topic.

Homework:
- Assigned semi-regularly
- Always due at the <u>start</u> <u>of</u> <u>class</u>
- 5 minute grace period
- After that HW is late, will be accepted for %75 credit.
- After HW is delivered to the grader, late HW won't be accepted
- Two lowest HW scores will be dropped

Quizzes: May or may not be quizzes, worth the same as a hw.

Grading:
40%: HW + Quizzes
25%: In class midterm
30%: Takehome final
5%: Attendance
2% bonus: Class participation / engagement.

Exams may be curved, but not downward.

Regulations: University's Standards for Academic Integrity apply to this course.

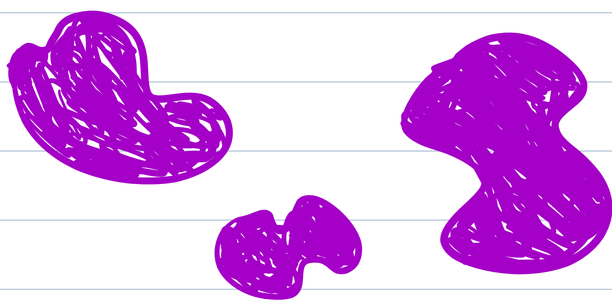
Feedback: I'm happy to recieve contructive, thoughtful feedback about the course, in person or by email.

Introduction to Persistent Homology

A major part of topology is the study of <u>holes</u> geometric objects.

In topology, we distinguish between holes of different dimension.

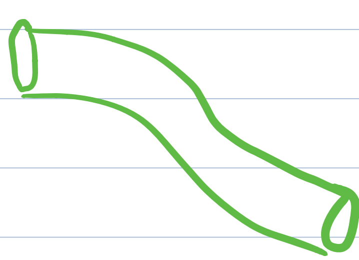
0-D holes are path components:



X

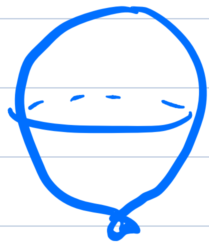
X has 3 0-D holes!

1-D holes in 3-D objects are tunnels (e.g. a hole you can see through)



A tube has a 1-D hole!

2-D holes in 3-D objects are hollow spaces

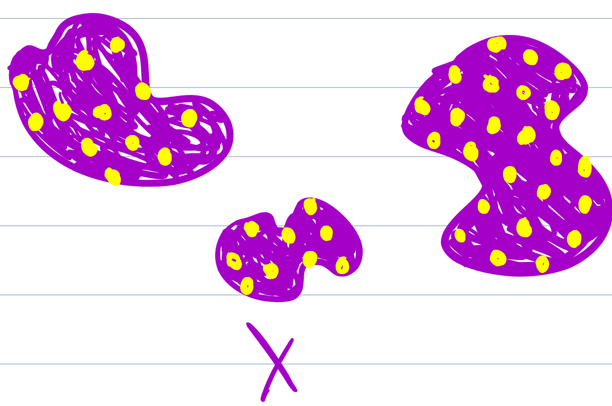


An inflated balloon has a 2-D hole!

Last semester, we saw how to define path components, so we have a good definition of 0-D hole already.

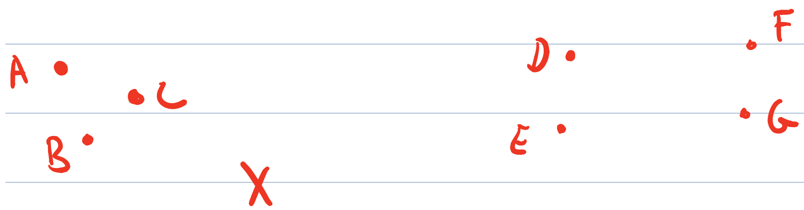
The language of homology will make precise the idea of higher-D holes.

Now, from a topological perspective, the clustering problem is a discrete version of the problem of finding connect components

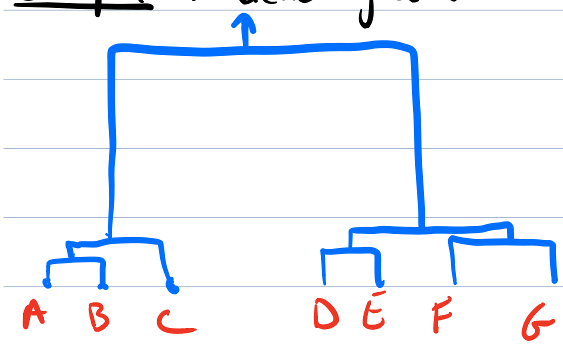


Last semester, we studied single linkage clustering:

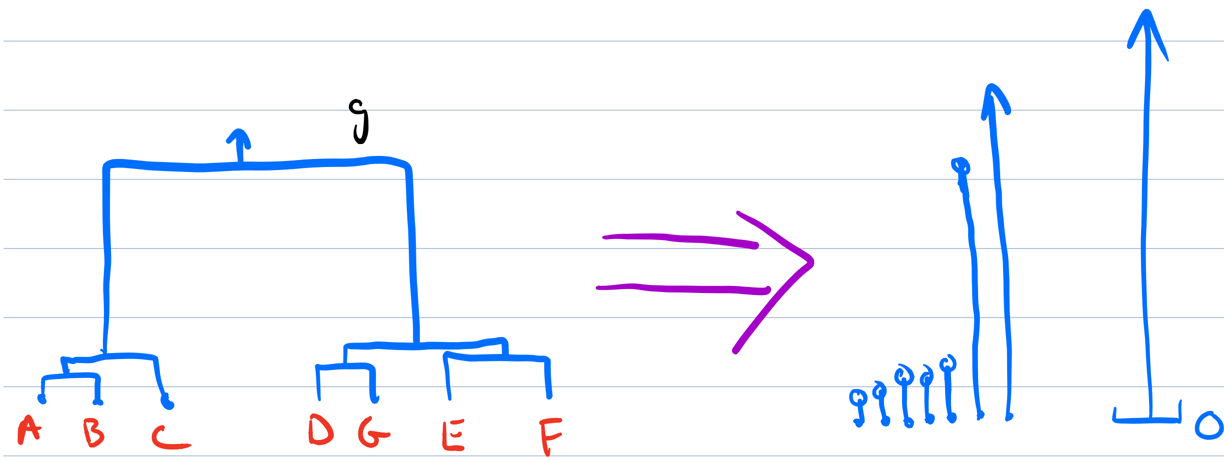
<u>Input</u>: $X \subset \mathbb{R}^n$ finite



<u>Output</u>: A dendrogram



We saw that there is a nice way to cut the dendrogram at its vertices to get a collection of intervals in $\mathbb{R}$, called a <u>barcode</u>.

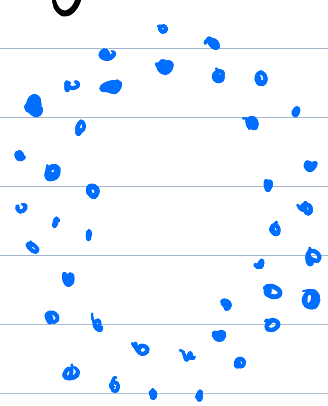


Intuition:

Each interval represents a cluster in the data, and the length of the interval is a measure of the "scale" of the cluster.

- Long intervals represent large-scale clusters
- Short intervals represent small-scale clusters.

We would like to extend the idea of barcodes to detect analogues of higher-D holes in data.

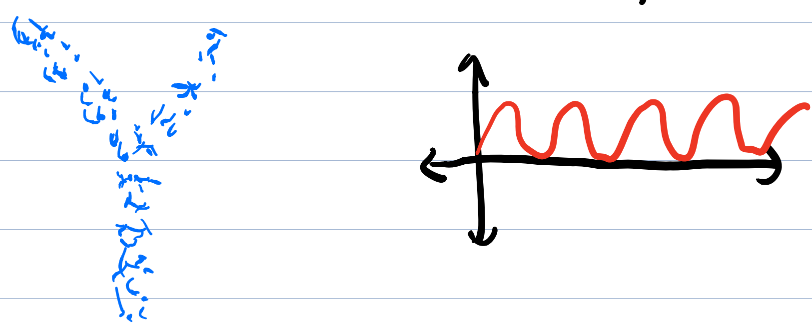


There are many applications where this is a very natural and useful thing, e.g. computational chemistry, drug discovery, materials science.

But persistent homology is not simply the business of detecting holes in data.

It can also be used to detect other shape features in data:

For example, tendrils, or periodicity in time series data:



The higher barcodes cannot be defined via a dendrogram; they must be defined algebraically. Homology will provide the link between topology and algebra.

Simplicial complexes

References: Munkres, Elements of Alg. Top.
: Blumberg, Rabadan

- A generalization of undirected graphs
- Contains not only vertices and edges, but also triangles, tetrahedra, and higher-dimensional analogues of these.

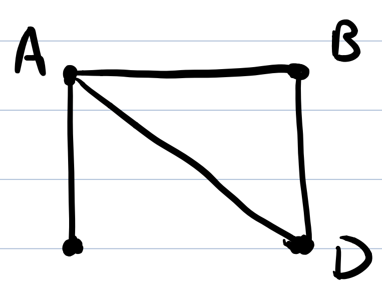
Recall: an (undirected) graph is a pair (V, E) where
- V is a set
- Elements of E are subsets of V w/ two elements.

We usually write $\{v, w\} \in E$ as $\underbrace{[v, w]}_{\text{order doesn't matter.}} = [w, v]$

We call: elements of V <u>vertices</u>
elements of W edges.

We draw a graph $G = (V, E)$ by plotting vertices as points in the plane and edges as lines connecting them

<u>Ex</u>: $G = (V, E)$ $V = \{A, B, C, D\}$
$E = \{[A, B], [A, C], [A, D], [B, D]\}$



<u>Def</u>: an abstract <u>simplicial</u> <u>complex</u> on a set S is a collection X of non-empty finite subsets of S such that if $\sigma \in X$ and $\phi \neq \tau \subset \sigma$, then $\tau \in S$.

The subsets are called simplices of X

The subsets of size $(k+1)$ are called k-simplices of X.

The simplex $\{a_0, \dots, a_k\}$ is written $[a_0, \dots, a_k]$.

Example: $S = \{A, B, C, D\}$

$X = \{[A], [B], [C], [D],$
$[A, B], [A, C], [A, D], [B, D],$
$[A, B, D]\}$