AMAT 584 Lecture 9, February 10, 2019

Today: Simplicial Complexes from Data

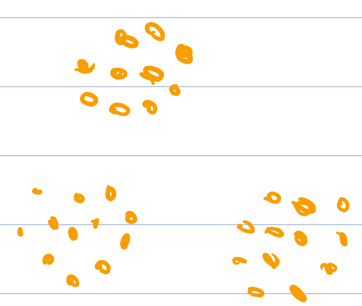
References: Edelsbrunner/Harer, Chazal/Boissonat/Yvinec (textbook)

So far in TDA II, we have focused on the study of simplicial complexes.

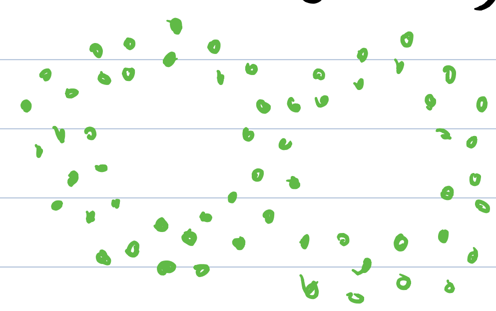
Toda[y], we'll dis[c]uss how simplicial complexes arise from data, in the context of persistent homology

Motivation: Use tools from topology to detect holes in data.

Data =
A finite subset of $\mathbb{R}^n$ (n possibly large) OR
A finite metric space (e.g. DNA sequences with the edit distance, or 3-D protein structures with RMSD metric)



0-D holes (clusters)



1-D holes (Loops)

We see that the data set on the left (above) has 3 clusters (0-D holes), while the data set on the right has 2 loops (1-D).

How do we use topology to detect this structure?

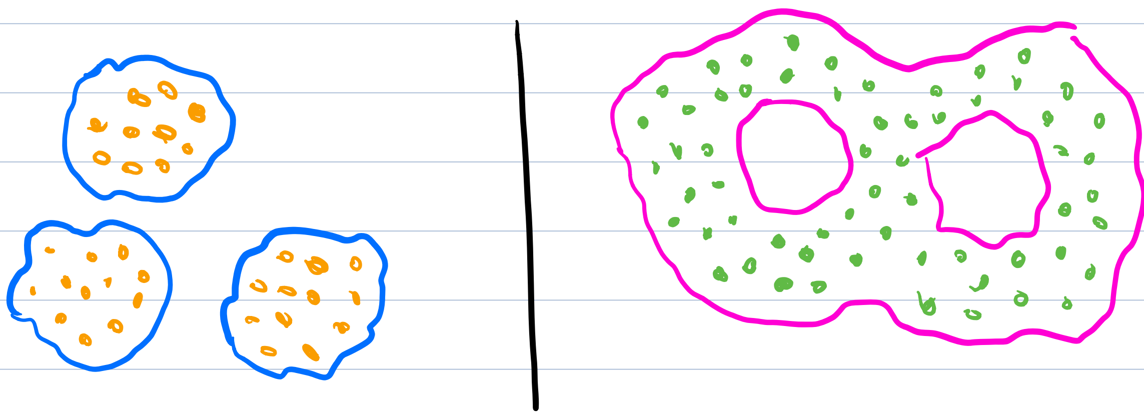
<u>Key Problem</u>: The data is <u>discrete</u>:

Strictly speaking:

- Every point is its own path component
- Neither data set has any holes.

} the topology of such objects is not interesting.

<u>Solution</u>: First <u>thicken</u> the data set
Then analyze the topology of the thickened data



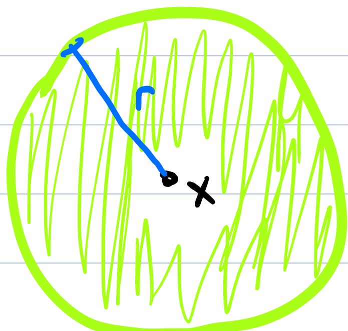
[We encountered this idea in TDA I when studying clustering]

How to thicken?

For $x \in \mathbb{R}^n$ and $r \geq 0$, let $B(x,r) = \{y \in \mathbb{R}^n \mid d_2(x,y) \leq r\}$

Euclidean distance.

$B(x,r)$ is called the closed ball of radius r centered at x.



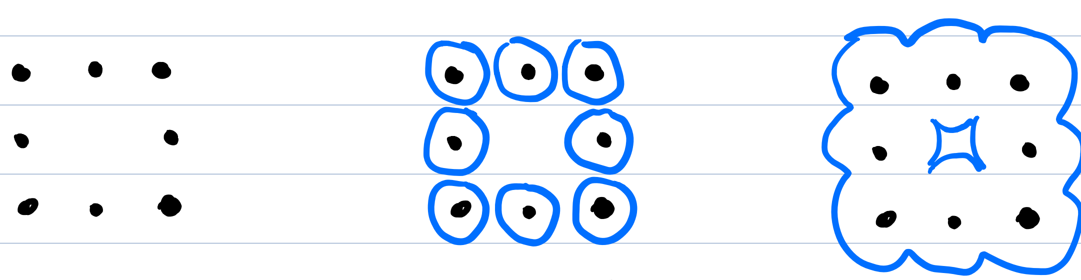
$B(x,r)$

Note: The boundary circle is included in $B(x,r)$.

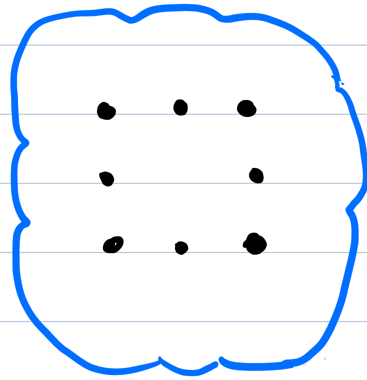
For $X = \{x_1, \dots, x_m\} \subset \mathbb{R}^n$ and $r \geq 0$, let

$$U(X,r) = B(x_1,r) \cup B(x_2,r) \cup \dots \cup B(x_m,r) = \bigcup_{i=1}^{m} B(x_i,r)$$

Thus, $U(X,r)$ is the union of the closed balls of radius r centered at the points of X.



$X = U(X,0)$ $\quad$ $U(X,r_1)$ $\quad$ $U(X,r_2)$



$U(X, r_3)$

This is a very sensible way to thicken a point cloud.

Questions: 1) How to choose r?

Answer: As in single linkage clustering, we consider all values of r, and study the topology of the of the family of spaces $\{U(X,r)\}_{r\in[0,\infty)}$.

This idea lies at the heart of persistent homology.

For now, we assume that $r\in[0,\infty)$ has been fixed.

2) How to work with $U(X,r)$ computationally?
For example, how do I count the holes in $U(X,r)$?

3) If X is a finite metric space not embedded in $\mathbb{R}^n$, how do I thicken?

Answer to both 2) and 3): Construct thickening using simplicial complexes.

Three constructions:

- Čech Complex
- Alpha complex (Delaunay Complex)
- Vietoris - Rips Complex

both are homotopy equivalent to $U(X,r)$

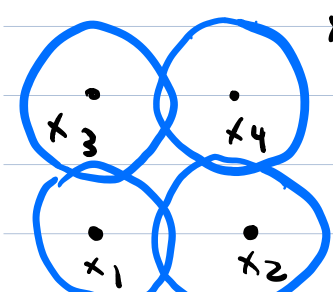
defined for finite metric spaces.

All are very closely related to the neighborhood graphs we studied in TDA I when discussing clustering.

Čech Complexes

Before giving the definition, let's start with two examples.

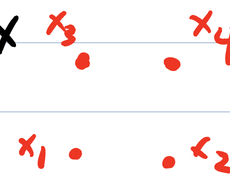
Let $X = \{\underset{x_1}{(0,0)}, \underset{x_2}{(1,0)}, \underset{x_3}{(0,1)}, \underset{x_4}{(1,1)}\}$.



Let $r = \frac{1}{2} + \delta$, δ small

I've drawn $B(x_i, r)$ for each $i \in \{1, \dots, 4\}$, in blue.

We construct an abstract simplicial complex

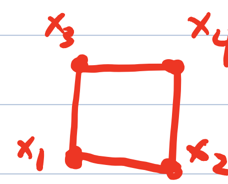


$\check{C}ech(X,r)$ with vertex set $\{x_1, \ldots, x_4\}$.

So the 0-simplices are of the form $[x_i]$, $i \in \{1, \ldots, 4\}$.

$\check{C}ech(X,r)$ contains the 1-simplex $[x_i, x_j]$ iff $B(x_i, r) \cap B(x_j, r) \neq \emptyset$.

Thus, the 1-simplices are $[x_1, x_3]$, $[x_1, x_2]$, $[x_2, x_4]$, $[x_3, x_4]$.



$\check{C}ech(X,r)$ contains the 2-simplex $[x_i, x_j, x_k]$ iff $B(x_i, r) \cap B(x_j, r) \cap B(x_k, r) \neq \emptyset$.

Thus there are no 2-simplices in $\check{C}ech(X,r)$.

If there are no 2-simplices, there are no higher simplices.

So $\check{C}ech(X,r) = \{[x_1], [x_2], [x_3], [x_4], [x_1, x_3], [x_1, x_2], [x_2, x_4], [x_3, x_4]\}$.

The geometric realization $|\check{C}ech(X,r)|$ is a square.

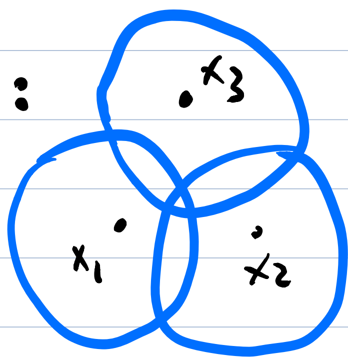
Note that $|\check{C}ech(X,r)|$ is homotopy equivalent to the union of balls $U(X,r)$. (But not homeomorphic.)

<u>General definition:</u>
For $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^n$ and $r > 0$, the Čech complex $\check{C}ech(X,r)$ is the abstract simplicial complex with vertex set $\{x_1, \dots, x_n\}$, such that

$$[x_{j_0}, \dots, x_{j_k}] \in \check{C}ech(X,r) \text{ iff}$$

$$B(x_{j_0}, r) \cap B(x_{j_1}, r) \cap \cdots \cap B(x_{j_k}, r) \neq \emptyset.$$

<u>Exercise</u>:



Consider $X = \{x_1, \dots, x_3\}$ as above, and r so that the balls $B(x_i, r)$ intersect as shown.

What is $Cech(X,r)$?