Topological Data Analysis

Fall 2025

Syllabus

Top. Review

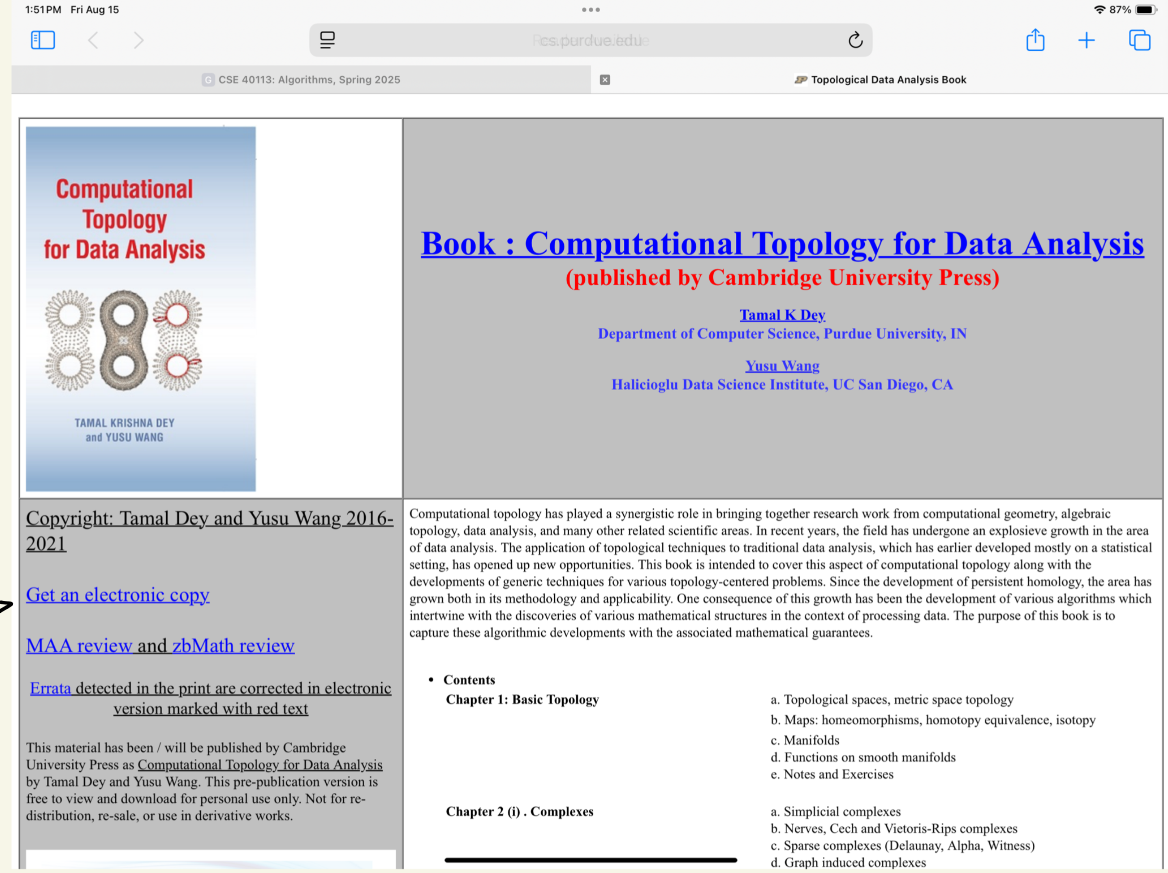
Course intro

- Syllabus + HW → main webpage
- HW submission → Canvas
  (+ gradebook)
- Prereqs: some linear algebra
  some programming background
- Slack?
- HW: Mix of pen/paper + coding
  ↳ flexible, so see me if you have any issues!

Textbook(s)

Main reference:

free pdf



1:51 PM Fri Aug 15

cs.purdue.edu

CSE 40113: Algorithms, Spring 2025

Topological Data Analysis Book

Computational Topology for Data Analysis

TAMAL KRISHNA DEY and YUSU WANG

Book : Computational Topology for Data Analysis

(published by Cambridge University Press)

Tamal K Dey
Department of Computer Science, Purdue University, IN

Yusu Wang
Halicioglu Data Science Institute, UC San Diego, CA

Get an electronic copy

MAA review and zbMath review

Errata detected in the print are corrected in electronic version marked with red text

Computational topology has played a synergistic role in bringing together research work from computational geometry, algebraic topology, data analysis, and many other related scientific areas. In recent years, the field has undergone an explosieve growth in the area of data analysis. The application of topological techniques to traditional data analysis, which has earlier developed mostly on a statistical setting, has opened up new opportunities. This book is intended to cover this aspect of computational topology along with the developments of generic techniques for various topology-centered problems. Since the development of persistent homology, the area has grown both in its methodology and applicability. One consequence of this growth has been the development of various algorithms which intertwine with the discoveries of various mathematical structures in the context of processing data. The purpose of this book is to capture these algorithmic developments with the associated mathematical guarantees.

- Contents

Others:

Listed on Syllabus

Most are available in the library, or come visit my office!

Project:

Ideally, this will connect to your research.

But - we'll have several assignments to help explore topics, if you're unsure of what direction to take.

Outline:

- talk summary → in Sept
- paper "chase"
- project proposal

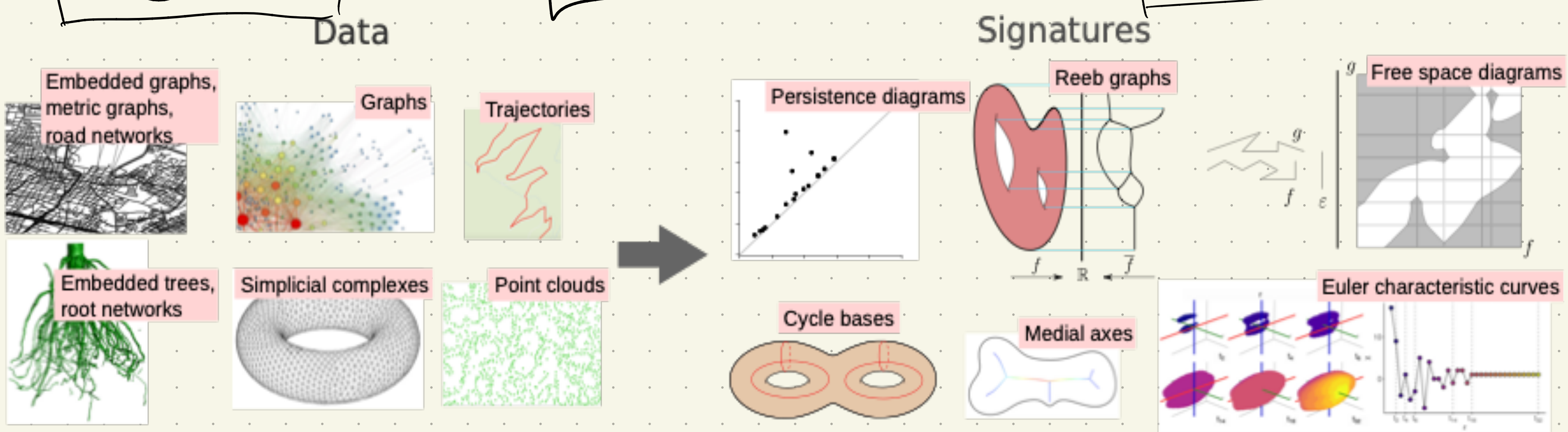
$\Big\}$ in October

- final presentation + submission

What is topological data analysis?

Traditional pipeline:

geometric data → topological signature → statistics ML visualization



Goals:

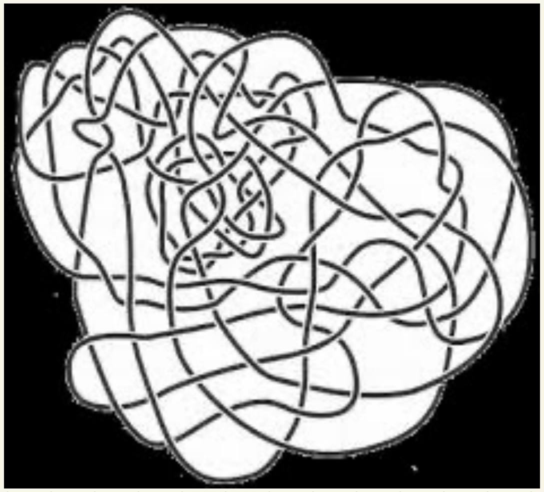
Some history:

Recognizing exact topology can be hard.

How so?

- Deciding if 2 4-manifolds are homeomorphic is undecidable [Markov 1960, van Meter 2005]

- Deciding if you can "unknot" a curve using a fixed number of moves is NP-Hard [de Mesmay, Sedgewick, & Tancer 2021]



An approach

Since we can't solve the problem exactly, focus on invariants and ways to simplify the data.

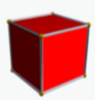
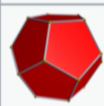
This is not new!

Examples:

- Knot invariants
- Curve skeletons
- Manifold approximation
  - ↳ ie meshes

A first example: Euler characteristic

Introduced first by Maurolico in 1537.
(This is known as
Then published by Euler in 1758:

Name	Image	Vertices V	Edges E	Faces F	Euler characteristic: $\chi = V - E + F$
Tetrahedron		4	6	4	**2**
Hexahedron or cube		8	12	6	**2**
Octahedron		6	12	8	**2**
Dodecahedron		20	30	12	**2**
Icosahedron		12	30	20	**2**

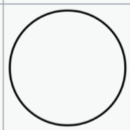
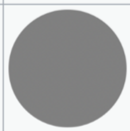
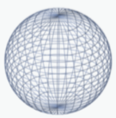
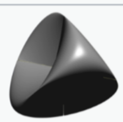
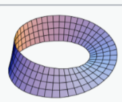
For any <u>embedded</u> planar graph $G=(V,E)$ with F faces,

$$V-E+F=2$$

Note: planar $\Rightarrow$

More generally:

Euler characteristic: $V - E + F = \chi$

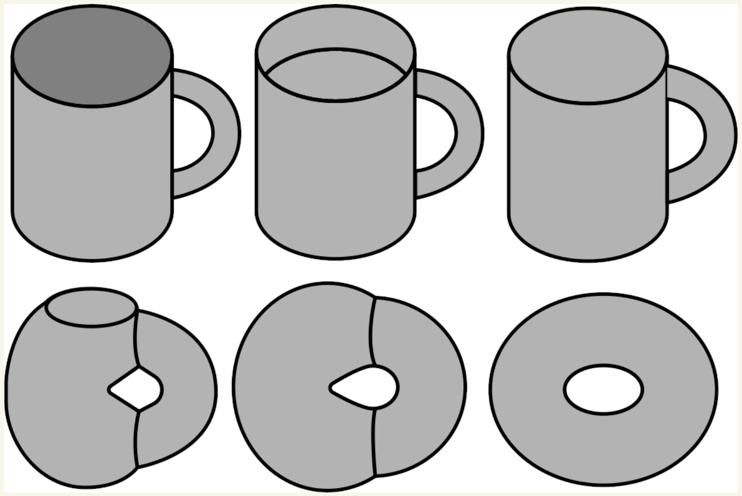
Name	Image	χ
Interval		1
Circle		0
Disk		1
Sphere		2
Torus *(Product of two circles)*		0
Double torus		−2
Triple torus		−4
Real projective plane		1
Möbius strip		0

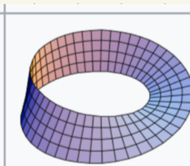
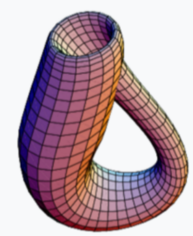
Ideal for computers:

- requires a discrete representation
- Given a data structure encoding V, E, + F, easy to calculate

Topological signatures: invariants

Different Euler characteristic $\Rightarrow$ different space



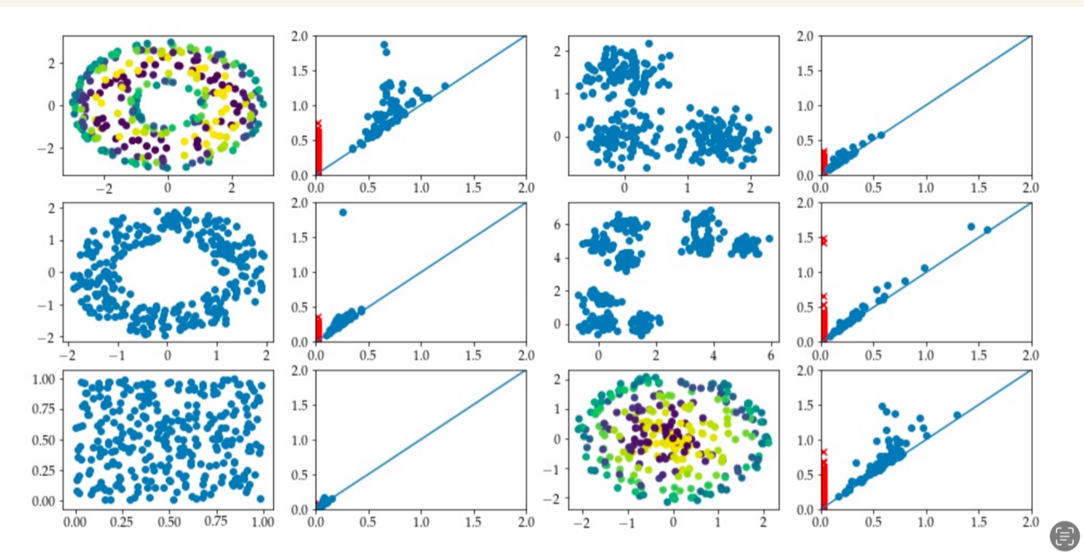
Möbius strip		0
Klein bottle		0

But different spaces might have the same Euler characteristic (as well as very different geometry)

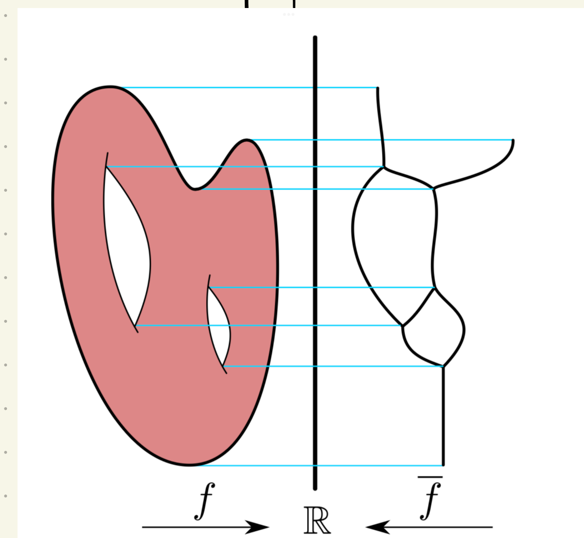
Back to signatures

We'll cover a range of possible choices, on a sliding scale of complexity & discriminativity.

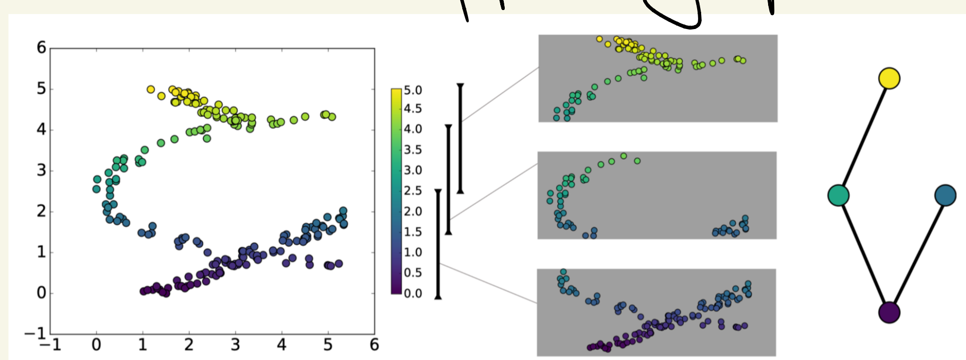
Examples:



Persistent homology



Reeb & Mapper graphs



Active research directions

This is a fairly young & vibrent area.

Emerging directions:

- Machine Leering with TDA
- Time series & dynamic systems
- Parellelization
- Vizualization
- Algebraic methods & multi-dimensional persistence

- Many applications: atmospheric data, image processing, biomedical, neuroscience, vision, etc.

Our goals

Understand the computation + interpretation of several commonly used tools in TDA:

- Euler curves
- Persistent homology
- Mapper + Reeb graphs
- Morse-Smale Complexes

In the end, understand what types of signatures are likely to be both useful + practical on data sets (as well as which open source tools exist).

But first – topology!

Chapter 1 covers an intro to topology.

Depending on math, might seem obvious, or might seem very hard!

Either is ok.

Worth reading textbook to be sure you get main definitions in context, & come see me if you have questions.

Topology

Topological space: a set $\mathbb{T}$ with elements (called points) + a set of subsets T, such that

- $\emptyset, \mathbb{T} \in T$
- $\forall\, U \subseteq T$, union of sets in U is in T
- $\forall$ finite $U \subseteq T$, intersection of sets in U is also in T

Ex: $\mathbb{T} = \{a, b, c\}$, $T = \{\emptyset, \{a\}, \{b\}, \{a,b\}, \{a,b,c\}\}$

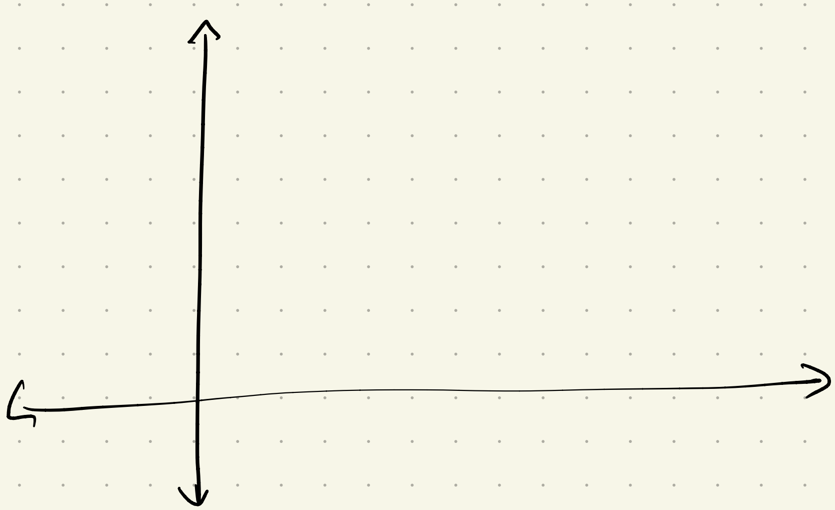
Check:

Metric space:

a pair $(\mathbb{T}, d)$, where $\mathbb{T}$ is a set and $d : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ satisfies

- $d(p,q) = 0 \iff p = q$
- $d(p,q) = d(q,p) \quad \forall p,q \in \mathbb{T}$
- $d(p,q) \leq d(p,r) + d(r,q) \quad \forall p,q,r \in \mathbb{T}$

Example: $\mathbb{T} = \mathbb{R}^2$, $d((u_1,u_2),(v_1,v_2)) = \sqrt{(u_1-v_1)^2 + (u_2-v_2)^2}$



Metric topology

Given a metric space $(\mathbb{T}, d)$, an open metric ball is

$$B_o(c, r) = \{ p \in \mathbb{T} \mid d(p, c) < r \}$$

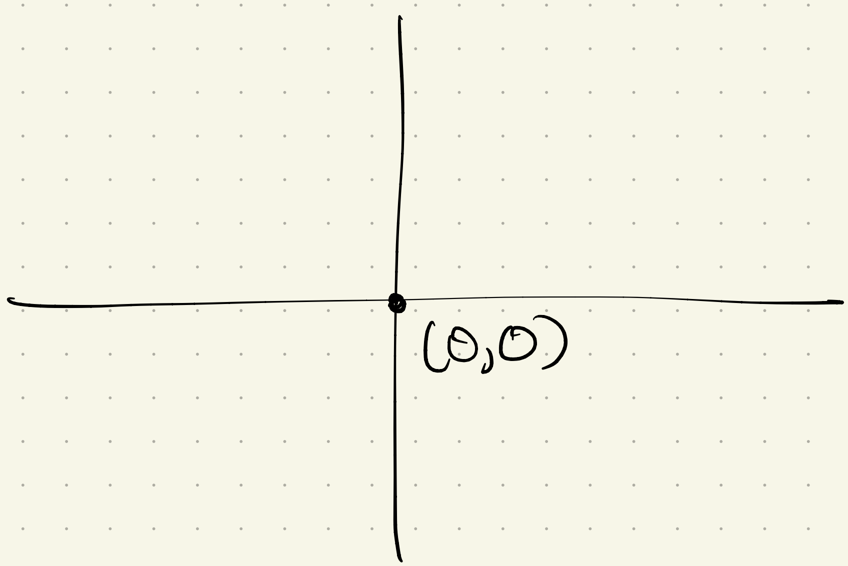
The metric topology is the set of all metric balls.

<u>Ex</u>: $\mathbb{R}^2$ again:

Many different metric topologies!

Fix $\mathbb{R}^2$, & let's try $B_0(0,1)$ for:

- $\|u-v\|_1 = |u_1 - v_1| + |u_2 - v_2|$
- $\|u-v\|_2 = \sqrt{(u_1-v_1)^2 + (u_2-v_2)^2}$
- $\|u-v\|_\infty = \max\{|u_1-v_1|, |u_2-v_2|\}$



Open & closed sets

Fixing a topology T, U is open if $U \in T$.
We say U is closed if $\mathbb{T} \setminus U$ is open.

Back to first example:

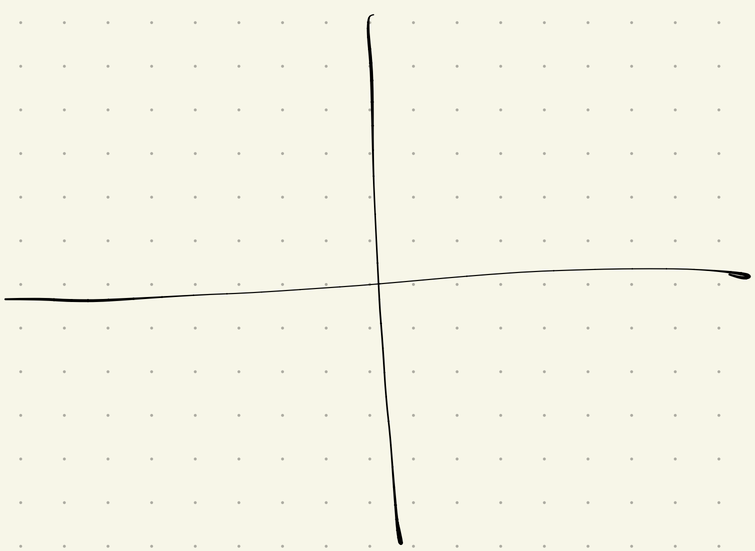
$\mathbb{T} = \{a, b, c\}$, $T = \{\phi, \{a\}, \{b\}, \{a,b\}, \{a,b,c\}\}$

Closed sets:

In a metric space, can get some alternate definitions:

Consider $Q \subseteq \mathbb{T}$. A point $p \in \mathbb{T}$ is a **limit point** of Q if $\forall \varepsilon > 0$, Q contains some point $q \neq p$ with $d(p,q) < \varepsilon$

Example: $\mathbb{R}^2$ & $B_0(p, 1)$:

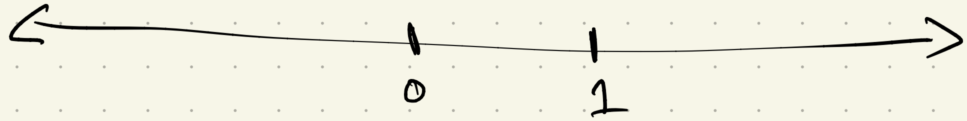


The closure of a point set $Q \subseteq \mathbb{T}$ is the set containing Q and all of its limit points.

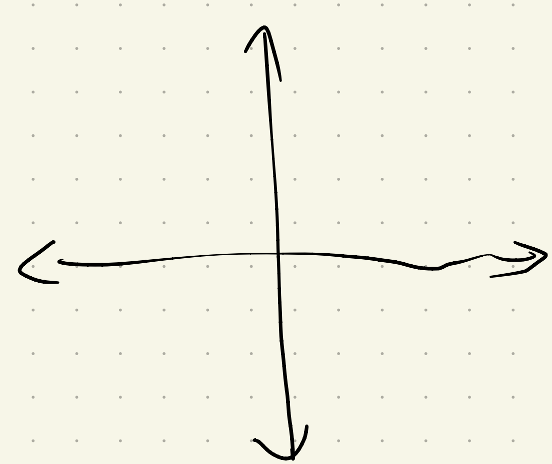
↳ written $Cl(Q)$, or $\overline{Q}$

We say Q is closed if $Q = Cl(Q)$.

Example: $(0,1) \subseteq \mathbb{R}$ is an open ball



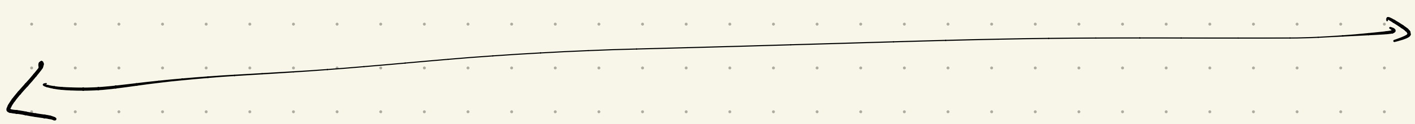
Example: $\mathbb{R}^2$ + $B_0(0,1)$



A open (resp, closed) cover of a topological space $(\mathbb{T}, \mathcal{T})$ is a collection C of open (resp. closed) sets s.t.

$$\mathbb{T} = \bigcup_{c \in C} c$$

Example: $\mathbb{R}$, $C = \{ (n-1, n+1) \mid n \in \mathbb{Z} \}$



A topological space is disconnected if $\exists$ 2 disjoint nonempty open sets $U, V \in \mathcal{T}$ s.t. $\mathbb{T} = U \cup V$.

(The space is connected if it is not disconnected.)

Ex: $A = (1,2) \cup (3,4) \subset \mathbb{R}$

Note: subspace topology: Given $\mathbb{U} \subseteq \mathbb{T}$, $\mathbb{U}$ can inherit topology from $\mathcal{T}$ via

$$\{X \cap U \mid X \in \mathcal{T}\}$$

Next time:

Maps, homeomorphisms, & homotopies!

(See remainder of Chapter 1)

Overall goal: understand enough about maps to get to "nice" functions, & start Morse theory