CS 6170: Computational Topology, Spring 2019 Lecture 03

Topological Data Analysis for Data Scientists

Dr. Bei Wang

School of Computing
Scientific Computing and Imaging Institute (SCI)
University of Utah
www.sci.utah.edu/~beiwang
beiwang@sci.utah.edu

Jan 15, 2019

Topological Spaces

Book Chapter A.I.

Topological spaces

Definition (Topological space)

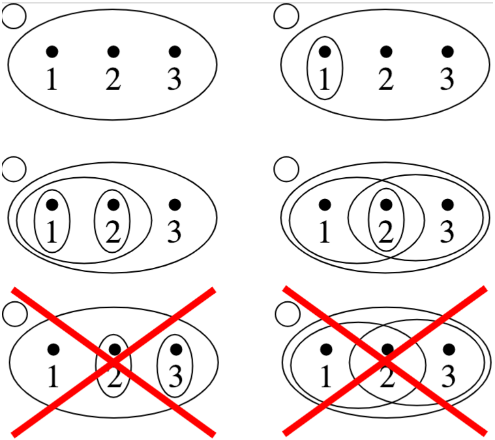
A *topological space* is an ordered pair $(\mathbb{X}, \mathcal{U})$, where $\mathbb{X}$ is a set, $\mathcal{U}$ is a collection of subsets of $\mathbb{X}$, such that:

- The empty set and $\mathbb{X}$ itself belong to $\mathcal{U}$.
- Any arbitrary (finite or infinite) union of members of $\mathcal{U}$ still belongs to $\mathcal{U}$.
- The intersection of any finite number of members of $\mathcal{U}$ still belongs to $\mathcal{U}$.

The elements of $\mathcal{U}$ are called *open sets* and the collection $\mathcal{U}$ is called a *topology* on $\mathbb{X}$.

- The pair $(\mathbb{X}, \mathcal{U})$ is called a *topological space*.
- We usually assume that $\mathcal{U}$ is understood, and refer to $\mathbb{X}$ as a topological space.

Topological spaces

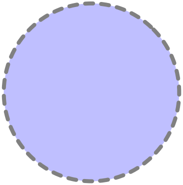


Topologies on a three-point set.

https://commons.wikimedia.org/wiki/File:Topological_space_examples.svg

Open set

- An open set is an abstract concept generalizing the idea of an open interval in the real line.
- A set is open if it doesn't contain any of its boundary points.
- Any collection of sets can be called open, as long as the union of an arbitrary number of open sets is open, the intersection of a finite number of open sets is open, and the space itself is open.
- A subset $U \subset \mathbb{R}^n$ is called *open* if given any point $x \in U$, there exists a real number $\epsilon > 0$ such that for all points $y \in \mathbb{R}^n$ such that $\mathrm{dist}(x, y) < \epsilon$, then $y \in U$.
- A *closed set* if a set whose complement is an open set.



https://en.wikipedia.org/wiki/Open_set

Topological space (point set topology)

Examples of topological spaces $(\mathbb{X}, \mathcal{U})$:

- $\mathbb{X} = \{1, 2, 3\}$, $\mathcal{U} = \{\emptyset, \{1, 2, 3\}\}$. $\mathcal{U}$ is a (trivial) topology on $\mathbb{X}$.
- $\mathbb{X} = \{1, 2, 3\}$, $\mathcal{U}$ is the power set of $\mathbb{X}$.
- $\mathbb{X}$ is the real line $\mathbb{R}^1$. $\mathcal{U}$ is the set of all open intervals on the real line $\mathbb{R}^1$.
- We refer to $\mathbb{X}$ as a topological space assuming $\mathcal{U}$ is understood.

Continuous function

- A continuous function is a function for which sufficiently small changes in the input result in arbitrarily small changes in the output [wikipedia].
- A function $f : \mathbb{X} \to \mathbb{Y}$ is *continuous* if the pre-image of every open set is open.
- Example of a discontinuous function, $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \begin{cases} 0, & \text{if } x \leq 0; \\ 1, & \text{otherwise.} \end{cases}$$

For any $0 < \epsilon < 1$, $(-\epsilon, +\epsilon)$ is open, $f^{-1}((-\epsilon, +\epsilon))$ is not.

Connectivity of a topological space

Definition (Path connectivity)

A topological space is *path-connected* if every pair of points is connected by a path.

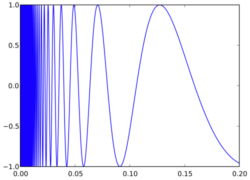
- A function between topological spaces $f : \mathbb{X} \to \mathbb{Y}$ is *continuous* if the presage of every open set is open.
- A *path* is continuous function from the unit interval $\gamma : [0, 1] \to \mathbb{X}$.

Separation of a topological space

Definition (Separation)

A *separation* of a topological space $\mathbb{X}$ is a partition $\mathbb{X} = U \cup W$ into two non-empty, open subsets that are disjoint. A topological space is *connected* if it has no separation.

- Connectedness is strictly weaker than path-connectedness.
- A topological space that is connected but is not path-connected: topologist's sine curve: $T = \{(x, \sin(\frac{1}{x})) : x \in (0, 1]\} \cup \{(0, 0)\}$



https://commons.wikimedia.org/wiki/File:Topologist%27s_sine_curve.svg

Union-Find

Book Chapter A.I.

Union-find on a high level

- Given a set of elements which are partitioned into subsets, you have to keep track of the connectivity of each element in a particular subset or connectivity of subsets with each other.
- To do this operation efficiently, you can use union-find data structure.

Galler and Fischer (1964); Galil and Italiano (1991)

https://www.hackerearth.com/practice/notes/disjoint-set-union-union-find/

Algorithmic question of deciding connectedness

- Union-find: a disjoint-set data structure.
- A data structure that tracks a set of elements partitioned into a number of disjoint (non-overlapping) subsets.
- Represent each set as a tree of elements (thinking grapes).
- Maintain a collection of sets under three operations (book notations):
  - **MakeSet(x)**: Create a set containing a single element x.
  - **Find(x)**: Return the root of the tree containing x.
  - **Union(x,y)**: Make the root of tree containing x to also be the root of tree containing y.

A social network example (slightly different perspective)

- There are 5 people A, B, C, D and E.
- A is a friend of B, B is a friend of C, D is a friend of E.
- As we can see:
  - A, B and C are connected to each other.
  - D and E are connected to each other.
- Use union-find to check whether one friend is connected to another in a direct or indirect way or not.
- We can also determine the two different disconnected subsets: $\{A, B, C\}$ and $\{D, E\}$.

You have to perform two types of operations here:

- **Union (A, B)**: connect two elements A and B.
- **FindPath(C, D)**: find if there is any path connecting two elements C and D.

https://www.hackerearth.com/practice/notes/disjoint-set-union-union-find/

References I

Galil, Z. and Italiano, G. (1991). Data structures and algorithms for disjoint set union problems. *ACM Computing Surveys*, 23.

Galler, B. A. and Fischer, M. J. (1964). An improved equivalence algorithm. *Communications of the ACM*, 7:301–303.