A manual for MA2007B

Applications of Geometry and Topology for Data Science

Alejandro Ucan-Puc

January 25, 2023

Instituto Tecnólogico y de Estudios Superiores de Monterrey

Contents

List of Figures

List of Tables

Introduction to Topology | 1

Topology was first studied in...

1.1 Basic concepts and Examples

Topology is the study of properties that are preserved under continuous functions in a generic way, for this reason, we need to work with objects that are proper to the space and these objects are subsets. The set theory is needed to properly understand topology because it is the main language of the theory, if you need a refresh of set theory you can see the reference.

Definition 1.1.1 (Topological Space) *Let X be a set and τ a collection of subsets of X. We say that τ is a* topology *for X is it satisfies:*

1. *If the total and the empty set belong to τ.*

$$\varnothing, X \in \tau.$$

2. *For every subcollection of elements of τ, let say $\{U_\alpha\}_{\alpha \in A}$, we have that its union is an element of τ.*

$$\bigcup_{\alpha \in U} U_\alpha \in \tau.$$

3. *For every finite sub-collection of elements of τ, let say $\{U_j\}_{j=1}^n$, we have that its intersection is an element of τ.*

$$\bigcap_{j=1}^{n} U_j \in \tau.$$

The elements of τ are called open sets *of X and the pair (X, τ) is called a* topological space.

This is not the first time that we see the concept of openness, for example in calculus we already worked with open intervals, open boxes (the real plane or space), and open disks (complex numbers). Topology theory intends to generalize this concept to general sets.

Example 1.1.1 (Trivial topology) Let X be any set and consider the collection $\tau = \{\varnothing, X\}$. τ is a topology for X and it is called the trivial topology.

This is not the best topology to understand properties of X but it is a good counterexample when you want to generalize topological constructions.

Example 1.1.2 (Discrete topology) Let X be any set and consider the collection $2^X = \{\text{all possible subsets of } X\}$. τ is a topology for X and it is called the discrete topology.

Note that each element of X is an open set on this topology.

Example 1.1.3 (Co-finite topology) Let X be any set and consider the collection $\tau_{cf} = \{U \subset X : X \setminus U \text{ is finite or } \varnothing\}$, i.e., the complement of U in X is a finite set of points. τ is a topology for X and it is called the co-finite topology.

Example 1.1.4 (Spaces with different topologies) Let $X = \{1, 3, 5, 7\}$ and consider the collections

$$\begin{aligned}\tau_1 &= \{\varnothing, \{1\}, \{5\}, \{7\}, \{1,5\}, \{1,7\}, \{5,7\}, \{1,5,7\}, X\}.\\ \tau_2 &= \{\varnothing, \{1\}, \{3\}, \{7\}, \{1,3\}, \{1,7\}, \{3,7\}, \{1,3,7\}, X\}.\end{aligned}$$

Both (X, τ_1) and (X, τ_2) are topological spaces but they are not the same space. For example, the open set $\{1, 5\}$ is not part of (X, τ_2) and $\{1, 3\}$ is not part of (X, τ_1).

Therefore in order to obtain the same topological space we need to find exactly the same open sets in both topologies.

The next example states a definition that will be useful in future chapters (see Simplicial Complexes and Persistent homology).

Example 1.1.5 (Graphs topology) A *graph* is a combinatorial object that consists of two sets:

- V called vertices, and usually you can consider it as a set of points.
- E called edges, it is a subset of $V \times V$ (pairs of elements in V). This set describes how to vertices are related.

Usually, graphs are denoted as $\Gamma = (V, E)$. Given a graph, you can associate a geometric object as a set of points in $\mathbb{R}^n$ in identification with V and lines joining this set of points with the information of E. This geometric object is called the *geometric realization*. For example, we can consider the graph Γ given by $V = \{a, b, c, d\}$ and $E = \{(a, d), (b, d), (c, d)\}$. Its geometric realization is a tripod in the plane. If we consider the collections

$$\begin{aligned}T_1 &= \{\{(a,d)\}, \{(b,d)\}, \{(c,d)\}\}\\ T_2 &= \{\{(a,d), a\}, \{(b,d), b\}, \{(c,d), c\}, \{(a,d), (b,d), (c,d), d\}\}\end{aligned}$$

The collection $\tau = 2^{T_1 \cup T_2}$ is a topology for Γ. Later we will see that this topology is related to the metric space structure on the graph.

Exercises:

1. Could the empty space be considered a topological space? In the case of a positive answer, describe a topology on it.
2. Consider $X = \{0, 2, 4, 6, 8\}$ and $\tau = \{\varnothing, \{0\}, \{0,2\}, \{0,4\}, \{0,2,6\}, \{0,4,6\}, X\}$. Is it τ a topology for X? Elaborate your answer (this means you prove all topology properties or give a counter-example for those properties)
3. With the same X as the previous exercise. Give a topology for X with no open sets with two elements.
4. Prove that the discrete topology and the co-finite topology on X are the same.

5. Prove that if X is any finite set, then the discrete topology and co-finite topology are the same.
6. Give a counter-example of the previous statement when X is an infinite set. *Consider an ordered set of numbers.*
7. Given the graph $\Gamma = (\{a, b, c\}, \{(a, b), (b, c), (c, a)\})$. Find the sets T_1 and T_2 that *generate* its topology.
8. Given a set X, a *basis* for a topology of X is a collection $\mathscr{B}$ of subsets of X (called base elements) such that:
   - For each $x \in X$, there is at least one base element B containing x.
   - If x belongs to the intersection of the base elements B_1 and B_2, then there is a base element B_3 containing x such that $B_3 \subset B_1 \cap B_2$.

   If $\mathscr{B}$ satisfies both conditions, we define the *topology generated by* $\mathscr{B}$ as: a subset U of X is said to be open in X (an element of τ) if for each $x \in U$ there is a base element $B \in \mathscr{B}$ such that $x \in B$ and $B \subset U$.
   Using the previous definition, prove that the open intervals in $\mathbb{R}$ are a basis for a topology on it. This topology is called the standard topology on $\mathbb{R}$. ⋆

1.2 Topological Properties

In mathematics, each research field is conformed by a pair: the objects and the transformations, so we are interested in the properties of the objects that are preserved by the transformations. In the previous section, we introduce topological spaces, that correspond to the objects of topology. Lately, we will introduce the transformations, but now we will present a set of properties of topological spaces that will be of our interest during the study of data.

Closedness

We mentioned that topology is the study of spaces using a collection of subsets called: open sets. The first property is to complement the open sets.

Definition 1.2.1 *Given a topological space (X, τ) and a subset V of X. We said that V is closed if its complement in X is an open set, i.e.,*

$$X \setminus V := \{x \in X : x \notin V\} \in \tau.$$

Closed sets on a topological space form a key part of topology, they are related to other topological properties and theorems. For example, closed sets are present in (topological) metric spaces to define continuity and compactness (we will see it later). Also, we can find closed sets in the definition of separation properties.

Example 1.2.1 Assume that $X = \{1, 3, 5, 7\}$ with the topology $\tau = \{\emptyset, \{1\}, \{5\}, \{1, 5\}, \{1, 5, 7\}, X\}$.

In this topological space, the sets $\varnothing, \{1, \}, \{3, 7\}, \{3, 5, 7\}, X$, are closed sets. Even more, we can enumerate all closed sets of X.

Note that, in a topological space, some sets could have open and closed properties. In the previous example, X is open and closed at the same time. In the literature, this sets are called *clopen* sets, but on this notes we will not give much attention on these.

Connectedness

As its name indicates, connectedness indicates if a set is made by part or it is a unit. Let detail on this.

Definition 1.2.2 *Given a set V of a topological space, a partition of V is a collection of open sets $\{A\}_j$ such that A_i*

Definition 1.2.3 *Given a set V of a topological space X, we said that V is* not connected *if there exists open sets A and B, such that:*

1. $V = A \cup B$,
2. $A \cap B = \varnothing$.

On the contrary, if there is no such partition, then we say that V is connected

Example 1.2.2 Assume that $X = \{1, 3, 5, 7\}$ with the topology $\tau = \{\varnothing, \{1\}, \{5\}, \{1, 5\}, \{1, 5, 7\}, X\}$.

The set $\{1, 5\}$ is not connected, because $\{1, 5\} = \{1\} \cup \{5\}$. The sets $\{1\}, \{5\}, \{1, 5, 7\}, X$ are connected.

Example 1.2.3 Assume that $X = \{1, 3, 5, 7\}$ with the discrete topology $\tau = 2^X$. On this topology, X is no longer connected, because:

$$X = \{1\} \cup \{3, 5, 7\} = \{1, 3\} \cup \{5, 7\}.$$

In fact, the only connected sets are $\{1\}, \{3\}, \{5\}$ and $\{7\}$.

So connectedness of a space depends on the topology of the ambient space.

Connected compontents

Between the connected subspaces of a topological space, there are some that are characteristic of the space. Before we state the definition of connected components, we need to introduce the concept of equivalence relations.

In mathematics, we can relate objects by equivalence relations. In mathematics, a *relation* between two sets X and Y is just a set of ordered pairs (x, y), for example a function is a relation. The "lower or equal" ($\leq$) is a relation between real numbers.

Definition 1.2.4 *An equivalence relation on a set X, denoted by $\sim$, is a mathematical relation that satisfies:*

1. *Reflexive: each element is related to itself*

$$a \sim a.$$

2. *Symmetric: if a is related to b, then b is related to a.*

$$a \sim b \Rightarrow b \sim a.$$

3. *Transitivity: if a is related to b, and b is related to c, then a is related to c.*

$$a \sim b \text{ and } b \sim c \Rightarrow a \sim c.$$

What does an equivalence relation imply on a set? Well, given an equivalence relation on a set, we can "collect" all interrelated elements in a subset of the ambient set. These collections of interrelated elements are called *equivalence classes* and they produce a partition on the ambient set.

Definition 1.2.5 *Let X be a set with an equivalence relation $\sim$. We define the equivalence class of x as the set*

$$[x] := \{y \in X : y \sim x\}.$$

Given the equivalence relation $[y]$, we say that y is a class representative.

Examples of equivalence relations can be found in. We will define an equivalence relation on topological spaces given by connected sets.

Definition 1.2.6 *Let (X, τ) be a topological space. Let $\sim$ defined as follow, x and y are related ($x \sim y$) if and only if there exists a connected subspace of X containing both points. The equivalence classes for this equivalence relation are called* Connected components *of X.*

Example 1.2.4 Assume that $X = \{1, 3, 5, 7\}$ with the topology $\tau = \{\emptyset, \{1\}, \{5\}, \{1, 5\}, \{1, 5, 7\}, X\}$.

The unique connected component of X is X itself, and this follows because the space is connected on this topology.

Example 1.2.5 Assume that $X = \{1, 3, 5, 7\}$ with the discrete topology $\tau = 2^X$. On this topology, X is no longer connected, and the connected components are $\{1\}$, $\{3\}$, $\{5\}$ and $\{7\}$.

Connectedness on Graphs

As graphs are an essential part of TDA, we need to understand how it works connectedness over graphs.

Let $\Gamma = (V, E)$ be a graph, and we will think that there is no distinction between the combinatorial object and the geometric realization.

Definition 1.2.7 *We say that Γ is* path-connected *if every two points can be joined by a trajectory of continuous edges.*

Proposition 1.2.1 *Let $\Gamma = (V, E)$ be a path-connected graph, then Γ is connected in the graph topology.*

Proof. Assume that Γ is not path-connected. Then there exists at least a pair of points that are not connected by a path. □

For graphs, the connected components look like subgraphs (subsets that are graphs) that are path-connected.

Example 1.2.6 Consider the graph K_n given by $V = \{1, 2, \cdots, n\}$ and $E = V \times V$. By definition, this graph is path-connected for every n.

Example 1.2.7 Consider the graph given by $V = \{1, 2, \cdots, n\}$ and E given by these conditions:

- $(v, u) \in E$ if u and v are even.
- $(v, u) \in E$ if u and v are odd.

By its definition, this graph is not path-connected and is composed of two connected components.

Compactness

As you can imagine, compactness observes if the set's elements are sort of close together.

Definition 1.2.8 *Let (X, τ) be a topological space and E be a subset. An open cover for E is a collection of open sets of X, $\{U_\alpha\}_{\alpha \in A}$, such that*

$$E \subset \bigcup_{\alpha \in A} U_\alpha .$$

Definition 1.2.9 *For (X, τ) and E be a subset. We said that E is* compact *if for every open covering for E there exists a finite cover (finite number of elements) that covers E.*

Example 1.2.8 Consider the set of positive integers $X = \mathbb{N}$ with the discrete topology $\tau = 2^{\mathbb{N}}$. Every finite set is compact.

Also, if X has the co-finite topology, then also every finite set is compact.

Now, the compactness property seems artificial and that maybe all subsets are compact but this does not happen. In the next section, we will introduce more examples of all properties on metric spaces.

Exercises:

1. Let X be a non-empty set with the discrete topology. Prove that for the topological space $(X, 2^X)$, the only connected components of X are singletons (sets with a unique element). ★
2. Is it true that for any non-empty set X, with the trivial topology, is connected? Elaborate on your answer.
3. Construct an example of a disconnected X with only three connected components.
4. Construct an example of a graph with n vertices that is connected but it is not the graph K_n.
5. Let $X = \mathbb{Z}$ give a topology on X such that has two connected components.

6. Is it true that for any non-empty set X, with the trivial topology, is compact? Elaborate on your answer.
7. Prove that $X = \mathbb{Z}$ with the discrete topology is not compact. (*Hint: Try to construct an open covering on* $\mathbb{Z}$) ⋆

1.3 Metric Spaces and Metric Topology

Metric spaces are an important set of examples for topology, this follows because their metric structure induces a natural topology on them.

Definition 1.3.1 *Let X be a non-empty set. A* metric *on X is defined as a function $d : X \times X \to \mathbb{R}$ that satisfies:*

1. *Non degeneracy: $d(p, q) = 0$ if and only if $p = q$.*
2. *Symmetry: $d(p, q) = d(q, p)$.*
3. *Triangle inequality: for every $p, q, r \in X$,*

$$d(p, q) \leq d(p, r) + d(r, q).$$

A metric space *is a pair (X, d) where d is a metric for X.*

Example 1.3.1 Let $X = \mathbb{R}$ and $d(x, y) = |x - y|$, the function given by the absolute value. Then (X, d) is a metric space. The metric d is known as the standard metric on the reals.

We can consider the functions:

- $d_1(x, y) = \left|\ln\left(\frac{x}{y}\right)\right|$.
- $d_2(x, y) = \frac{|x-y|}{1+|x-y|}$.

Both functions are metrics for X.

Example 1.3.2 Let $X = \mathbb{R}^n$ be the set of ordered $n-$tuples of reals. If $x = (x_1, \cdots, x_n)$ and $(y_1, \cdots, y_n)$, then let

$$d(x, y) = \sqrt{\sum_{j=1}^{n}(x_j - y_j)^2}$$

this function makes X a metric space, and it is known as the ℓ_2-metric.

Similarly to the dimension one ($\mathbb{R}$), there are more than one metric on $\mathbb{R}^n$, for example:

- The ℓ_∞ norm:
$$d(x, y) = \max_{j=1,\cdots,n}\{|x_j - y_j|\}.$$
- The ℓ_1 norm:
$$d(x, y) = \sum_{j=1}^{n}|x_j - y_j|.$$
- The discrete metric:
$$d(x, y) = \delta_{xy}.$$

The geometry of $\mathbb{R}^n$ changes for different metrics. For example, given a metric space (X, d), a point $x \in X$ and a positive real number r, we can define a special subset of X, this subset is a *ball* centered at x of radius r:

$$B_d(x, r) := \{y \in X : d(x, y) < r\}.$$

You can look for example, at different balls plotted on each geometry in the next figure.

In the next example, we will assume that a graph Γ means also its geometric representation in some $\mathbb{R}^n$.

Example 1.3.3 Given a graph $\Gamma = (V, E)$. We define the following metric on it. Let $d : V \times V \longrightarrow \mathbb{R}$ given as follows: for any pair (u, v), the number $d_\Gamma(u, v)$ is the length of the shortest path that joins u and v, and by length we mean the sum of the distance of between the points in a path for a given metric in $\mathbb{R}^n$.

In the literature, previous metrics on the graph can be uniformized by taking the distance between the vertices to be always one.

Metric Topology

The metric space structure on a set X induces a natural topology on it, this topology is known as the *metric topology*. Before we state the description of open sets, we need to understand the concept of the basis for a topology.

Definition 1.3.2 *Let X be a set, a* basis *for a topology on X is a collection $\mathcal{B}$ of subsets of X (called* base elements*) that satisfies:*

1. *For each $x \in X$, there is at least one base element of B such that $x \in B$.*
2. *If x belongs to the intersection of two base elements B_1 and B_2, then there is a base element B_3 containing x such that*

$$B_3 \subset B_1 \cap B_2.$$

If $\mathcal{B}$ satisfies these two conditions, then we define the topology $\tau_{\mathcal{B}}$ generated by $\mathcal{B}$ as follows: A subset $U \subset X$ belongs to $\tau_{\mathcal{B}}$ if for each $x \in U$ there is a base element $B \in \mathcal{B}$ such that $x \in B \subset U$.

The previous definition is nothing new, you have seen a similar concept in Complex Calculus when we talked about open subsets and neighborhoods of $\mathbb{C}$.

Definition 1.3.3 *Let (X, d) be a metric space. Consider the collection $\mathcal{B}_d$ as the set of all balls centered on points of x of any radius.*

The topology on X induced by the basis $\mathcal{B}_d$ is called the metric topology *on X.*

Example 1.3.4 Consider $\mathbb{R}^2$ with the euclidean metric (or standard metric), the basis of the topology $\mathcal{B}_d$ consists of all circular regions on $\mathbb{R}^2$.

The metric topology is very special because topological properties have a more detailed description here. In what follows, we will describe all topological properties for the metric topology. From now on, when we think of a metric space (X, d) we will think it with the metric topology.

Closed sets

Definition 1.3.4 *Let (X, d) be a metric space and $Q \in X$ a point set. We say that $q \in X$ is a* limit point *(or* accumulation point*) of Q if for any real number ε (no matter how small) there exists a point $p \in Q$ different from q such that*

$$d(q, p) < \varepsilon.$$

The previous definition, in topological terms, means that for a limit point q, any ball centered at q always intersects Q.

Example 1.3.5 Consider the set $(\mathbb{R}, d)$ with the standard metric. The set $Q = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$, the point $0 \in \mathbb{R}$ is a limit point for Q. You can see in the next figure you can see a sketch of why.

Example 1.3.6 Consider the set $(\mathbb{R}^2, d)$ with the standard metric. The set $Q = \left\{\left(x, \frac{\sin(x)}{x}\right) : x \in (0, \infty)\right\}$. The point $(0, 1) \in \mathbb{R}^2$ is a limit point for Q.

Definition 1.3.5 *Let (X, d) be a metric space and $Q \in X$ a point set. We call* boundary *of Q to the set of all its limit points, denoted by ∂Q. The* closure *of Q is defined to be the union of Q and its boundary,* i.e.,

$$\overline{Q} = Q \cup \partial Q.$$

We say that Q is closed in X if it is equal to its closure:

$$Q = \overline{Q}.$$

Compact Sets

In the metric topology, compact sets have a more intuitive description. On metric spaces, any compact set has to be closed and bounded, if one of these two properties fails the set cannot be compact.

Definition 1.3.6 *Let Q be a subset of a metric space (X, d). The diameter of Q, denoted by* $\operatorname{diam}(Q)$, *is the minimal possible upper-bound for all distances of points of Q,* i.e.,

$$\operatorname{diam}(Q) := \sup_{x,y \in Q} d(x, y).$$

The diameter can be a real number or ∞.

Definition 1.3.7 *Let Q be a subset of a metric space (X, d). We say that Q is* bounded *if its diameter is finite.*

$$\operatorname{diam}(Q) < \infty.$$

Example 1.3.7 Let $(\mathbb{R}, d)$ where d is the euclidean metric. The set $Q = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$ is not compact.

Note that even if the diameter of Q is 1, the set is not closed because does not contain its limit point.

Proposition 1.3.1 *Let $X = \mathbb{R}^n$ and d the standard metric. A subset $Q \subset \mathbb{R}^n$ is compact if and only if it is bounded and closed.*

Example 1.3.8 Let $Q = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$. In the previous example, we said that Q is bounded but it is not closed. Therefore, from the previous proposition, Q cannot be compact of $(\mathbb{R}, d)$.

Therefore, Proposition 1.3.1 proposes a way to check if a set can be compact. For general metric spaces, the other direction of the proposition does not follow so easily, we need extra conditions on the metric space and these conditions are beyond the intentions of these notes. In the particular case of $\mathbb{R}^n$ with the euclidean metric, closed and bounded means compact.

Theorem 1.3.2 *Let $Q \subset \mathbb{R}^n$ where $\mathbb{R}^n$ has the euclidean metric. Then Q is compact if and only if it is closed and bounded.*

The previous Theorem implies that whenever we look at a data set as a finite collection of points in some $\mathbb{R}^n$, these sets are compact because they are bounded and closed.

You may ask, what happens with compactness in the case of metric graphs? Well, if the graph has finite vertices and its geometrization belongs to some $\mathbb{R}^n$, then the graph is compact for the metric graph. But if the graph has infinite vertices, then this is another story, we need to be careful about how its metric is defined to determine if it is compact or not.

How we can infer the topological properties of a data universe via a data set? Well, a data set is a sample of a bigger space of information. If we look, for example, that in our sample (observations) there is some accumulation, then we can infer that close in that region we have accumulation points and the data universe could contain a closed subset which is not a finite set of points.

Connected sets

There is no particular change in the definition of connected sets in metric spaces, but limit points can be helpful when we look for a partition.

Definition 1.3.8 *A point set $Q \in (X, d)$ is said to be disconnected if Q admits a partition into two disjoint non-empty sets U and V so that there is no point in U that is a limit point of V and no point in V that is a limit point of U.*

Example 1.3.9 The set $Q = \{\frac{1}{n} : n \in \mathbb{N}\}$ is disconnected in $(\mathbb{R}, d)$.

Example 1.3.10 The following sets are connected in $(\mathbb{R}^2, d)$:

- A punctured ball centered at cero: $B_d(0, r) \setminus \{0\}$.
- An anular region centered at cero: $B_d(0, r_1) \setminus B_d(0, r_2)$.
- An ball region without several regions removed:

Theorem 1.3.3 *Let* $(\mathbb{R}, d)$ *where* d *is the euclidean metric. Then* $[0, 1]$ *is compact and connected, and* $(0, 1)$ *is connected.*

Proof. That $[0, 1]$ is compact follows from the fact that $[0, 1]$ is closed and bounded. The prove that $[0, 1]$ and $(0, 1)$ are connected are similar. We will do it for $(0, 1)$.

Let $\square$

1.4 Special Subsets of $\mathbb{R}^n$