CS 6170 Computational Topology: Topological Data Analysis **University of Utah**
Spring 2017 **School of Computing**

Lecture 3: Jan 17, 2017

Lecturer: Prof. Bei Wang <beiwang@sci.utah.edu> *Scribe: Jessica Lohse*

This lecture's notes defines topological space and looks at the union find algorithm and its usages.

3.1 Topological Space in Point Set Topology

Let:
$\mathbb{X}$: A point set, in its most simple version.
$\mathbb{U}$: A collection of subsets of $\mathbb{X}$. The elements of $\mathbb{U}$ are called open sets.

Definition 3.1. *$\mathbb{U}$ is a topology of $\mathbb{X}$ if:*

1. *$\mathbb{X}$, $\emptyset$ is in $\mathbb{U}$;*
2. *Any union of sets in $\mathbb{U}$ is in $\mathbb{U}$;*
3. *Any finite intersection of sets in $\mathbb{U}$ is in $\mathbb{U}$.*

Proof. Prove that $\mathbb{U}$ is a topology of $\mathbb{X}$, where $\mathbb{X} = \{1, 2, 3\}$ and $\mathbb{U} = \{\emptyset, \{1, 2, 3\}\}$.

Following the definition of a topology:

1. Both $\emptyset$ and $\mathbb{X}$ ($\{1, 2, 3\}$) are in $\mathbb{U}$;
2. $\emptyset \cup \{1, 2, 3\} = \{\emptyset, \{1, 2, 3\}\}$ is in $\mathbb{U}$;
3. $\emptyset \cap \{1, 2, 3\} = \{\emptyset\}$ is in $\mathbb{U}$.

In this case, $\mathbb{U}$ is a trivial topology on $\mathbb{X}$. □

3.1.1 Topological Space

Definition 3.2. *The* ($\mathbb{X}$, $\mathbb{U}$) *is a topological space.*

In some cases, the $\mathbb{U}$ is omitted as it is assumed that the $\mathbb{U}$ is understood. Saying that this space is a topological space attaches the relation from subsets to each set.

Exercise: Is $\mathbb{U} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\}$, also called the power set of $\mathbb{X}$, a topology of $\mathbb{X} = \{1, 2, 3\}$?
Seeing as any union of subsets in $\mathbb{U}$ is also in $\mathbb{U}$, it is a topology of $\mathbb{X}$. Because $\mathbb{U}$ contains all subsets of $\mathbb{X}$, it can be said that $\mathbb{U}$ is a discrete topology of $\mathbb{X}$.

Exercise: $\mathbb{R}^1$ is a real line and $\mathbb{B}$ is a collection of open sets in $\mathbb{R}$. ($\mathbb{R}, \mathbb{B}$) is a topological space and $\mathbb{B}$ is a topology of $\mathbb{R}$.

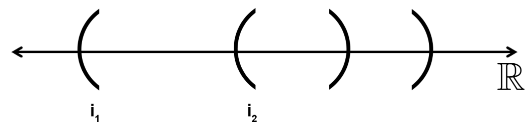

Figure 3.1: $i_1 \cup i_2$ is an open interval. Similarly, $i_1 \cap i_2$ is also an open interval.

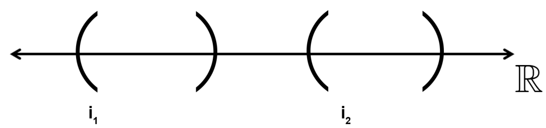

Figure 3.2: i_1 and i_2 also represent an open interval. The intersection of these sets is $\emptyset$, but their union is still an open interval.

3.2 Open Sets in Euclidean Space

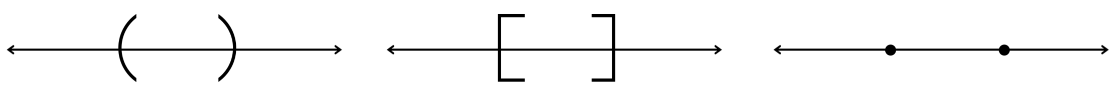

Figure 3.3: From left to right: An open set, closed set, and boundary set in $\mathbb{R}^1$.

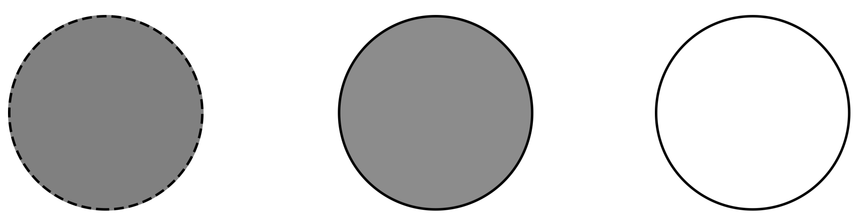

Figure 3.4: From left to right: An open set, closed set, and boundary set in $\mathbb{R}^2$.

Definition 3.3. *A subset of $\mathbb{R}^n$ is called open if given an point $x \in u$, $\exists$ a real number $\epsilon > 0$, such that for any point y in $\mathbb{R}^n$ whose distance from $x < \epsilon, y$ is in $\mathbb{U}$.*

In this definition, ϵ represents the neighborhood size. As a point gets closer to the boundary, ϵ shrinks, but never reaches 0.

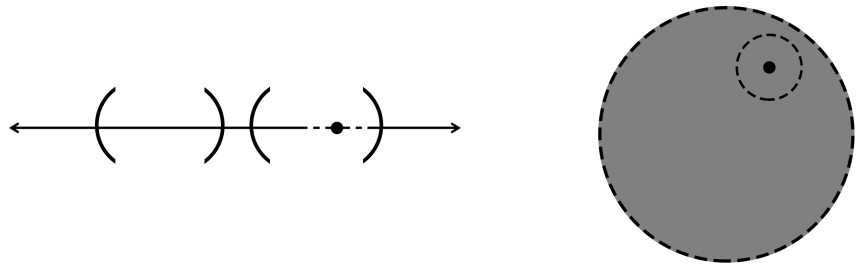

Figure 3.5: The dotted lines represent the possible values of a point ϵ away from a set point in $\mathbb{R}^1$ on the left and $\mathbb{R}^2$ on the right.

This is related to open sets in metric space, where a distance ϵ is also used.[BW12]

Definition 3.4. *A closed set is a set whose compliment is open.*

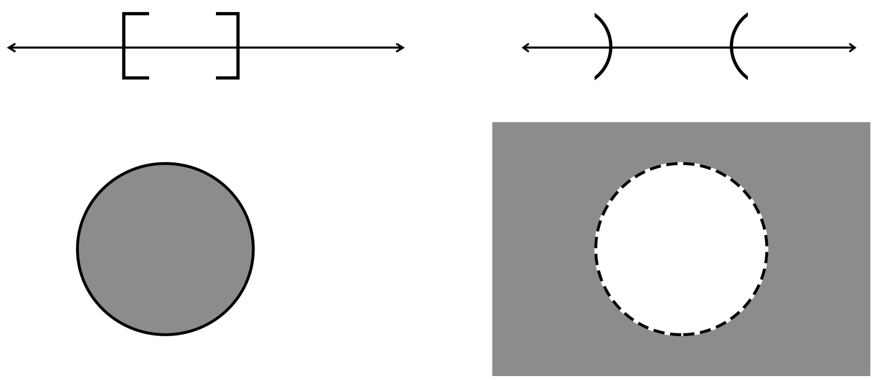

Figure 3.6: $\mathbb{R}^1$ and $\mathbb{R}^2$ representations of closed sets (left) and their compliments (right).

3.2.1 Continuity

Definition 3.5. *A function $f : \mathbb{X} \to \mathbb{Y}$, where $\mathbb{X}$ and $\mathbb{Y}$ are both topological spaces, is continuous if the preimage of every open set is open. $\forall$ open sets $V \sqsubseteq \mathbb{Y}, f^{-1}(V) = \{x \in \mathbb{X} \mid f(x) \in V\}$ is an open set of $\mathbb{X}$.*

Exercise: When we think of a continuous function, we think of a continuous curve. The figure below shows a function that is not continuous at 0.

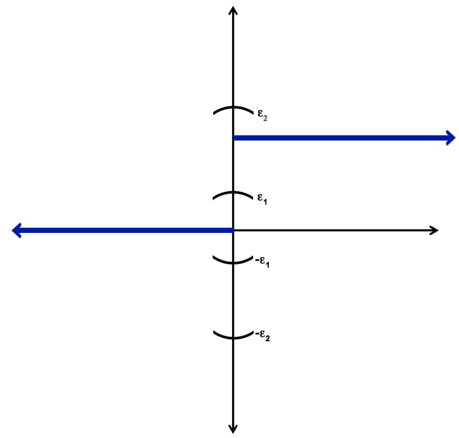

Figure 3.7: $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$.

For any interval $(-\epsilon, \epsilon)$, where $|\epsilon| < 1$, $f^{-1}(-\epsilon, \epsilon)$ is not an open set.

Proof. Given the above function, $f(x) = 0$ for $(-\infty, 0]$ and $f(x) = 1$ for $(0, \infty)$. The union of these sets is the entire x-axis. For simplicity, assume $f : \mathbb{X} \to \mathbb{Y}$, where $\mathbb{X} = \mathbb{Y} = \mathbb{R}$. To be continuous, the pre-image in $\mathbb{X}$ of every open set in $\mathbb{Y}$ must be open.
Consider an open set $(-\epsilon, \epsilon) \in \mathbb{Y}$. Its pre-image is the set of all points $x \in \mathbb{X}$ such that $f(x)$ is in the open set $(-\epsilon, \epsilon)$. Look at any such open set in $\mathbb{Y}$ where $|\epsilon| > 1$, eg. the interval $(-\epsilon_2, \epsilon_2)$ formed by parentheses at the top and bottom. This set contains both 0 and 1. Since the function f maps all $x \in \mathbb{X}$ to either 0 or 1 in $\mathbb{Y}$, the pre-image of any such set is the entire x-axis (the open interval $(-\infty, \infty) = \mathbb{X} = \mathbb{R}$.

However, when $|\epsilon| < 1$, as in the case of $(-\epsilon_1, \epsilon_1)$ - the interval formed by middle two parentheses - the set contains 0 but not 1. Since the set of points $x \in \mathbb{X}$ that map to 0 is $(-\infty, 0]$ which is not an open set in $\mathbb{X}$, the pre-image of an open set in $\mathbb{Y}$ is not open and thus the function is not continuous. □

If we allow an infinite number of open sets of any size, the intersection of those sets will be a single point. This makes every point an open set and continuous, which is not what we want.

Definition 3.6. *A path in topological space is a continuous function from [0, 1]* $\to \mathbb{X}$.

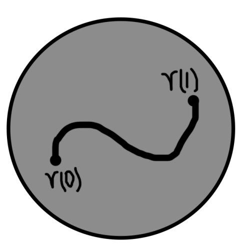

Figure 3.8: A simple path γ from $\gamma(0)$ to $\gamma(1)$.

Definition 3.7. *The topological space is path-connected if every pair of points is connected by a path.*

Definition 3.8. *Separation is when a path is partitioned into two nonempty, open subsets.*

Definition 3.9. *If a path has no separation it is connected. This is a weaker relationship than path-connected.*

Exercise: The Topologist's Sine Curve is an example of a function that is connected, but not path connected. The Curve is modeled by, $f(x) = \begin{cases} 0 & \text{if } x = 0 \\ \sin(1/x) & \text{if } x > 0 \end{cases}$. As you approach 0 from the right, the $\sin(1/x)$ function oscillates so much, you never actually reach the (0,0) point. Therefore, the function is not path-connected because a path does not exist between (0,0) and the rest of the curve. [EL15]

3.3 Union Find Algorithm or the Disjoint Sets Data Structure

The union find algorithm is an algorithm that decides connectedness.

This algorithm represents each set as a tree element.

Exercise:

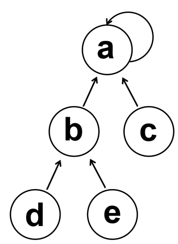

Figure 3.9: A possible tree of the set {a, b, c, d, e}.

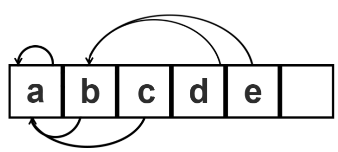

Figure 3.10: An array representation of the same tree.

The benefit of this algorithm over depth first search or breadth first search is that it stores the graph as a data structure containing a collection of sets. This works as a reversed tree, since instead of traversing from the root to the children, it traverses from the children to the root.

The union find algorithm has three main operations:

1. *MakeSet(x)*: Create a set that contains the single element *x*. This creates a single node that has a pointer to itself.
2. *Find(x)*: Find the root of the tree containing *x*. For the tree in Figure 3.9, *Find*(e) = *Find*(b) = a. This works it way from *x* to its parent, and that parent's parent, until the root is reached.

3. *Union(x,y)*: Make root of one tree containing *x* to be root of another tree containing *y*.

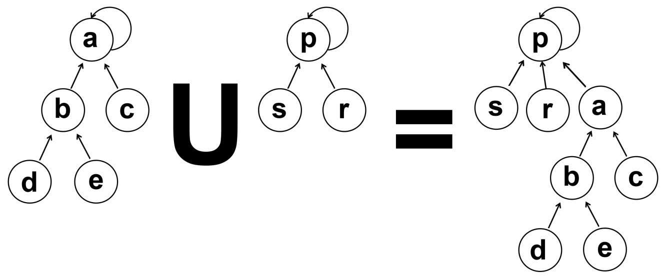

Figure 3.11: The *Union(a,p)*.

3.3.1 Run Time

With *Union(x,y)*, the run time is affected based on if larger trees are being attached to smaller tree or vice versa. If singletons are unioned together, creating a tree like the one below, *Find(e)* will operate at the worst run time possible.

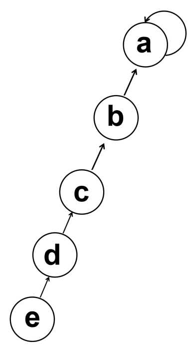

Figure 3.12: When singletons are unioned together, the run time for *Find(e)* will be $O(n)$.

The run times for the union find operations in O-notation, where α is a very slow growing function that functions as a constant and it assumed that the root is already known for *Union(x,y)* [B09]:

	MakeSet(x)	*Find(x)*	*Union(x,y)*
Worst Case	$O(1)$	$O(1)$	$O(\log n)$
Amortized	$O(1)$	$O(\alpha(n))$	$O(\alpha(n))$

3.3.2 Reducing Run Time

There are two ”hacks” that can be used to reduce run time with union find operations:

1. Union By Rank: Always hang the smaller tree on the larger tree. This requires extra storage for the rank/depth of tree.
   Instead of *Union(a,p)* in Figure 10, the union by rank hack knows that the tree containing p is smaller. Therefore, it will instead perform *Union(p,a)*.

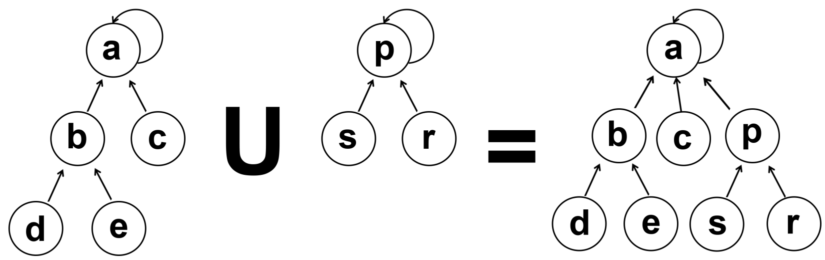

Figure 3.13: Performing *Union(a,p)* with the Union By Rank hack.

2. Path Compression: When using *Find(x)*, connect all nodes on the path from x to the root directly to the root. This shrinks the height of the tree and increases the efficiency of *Find(x)*.

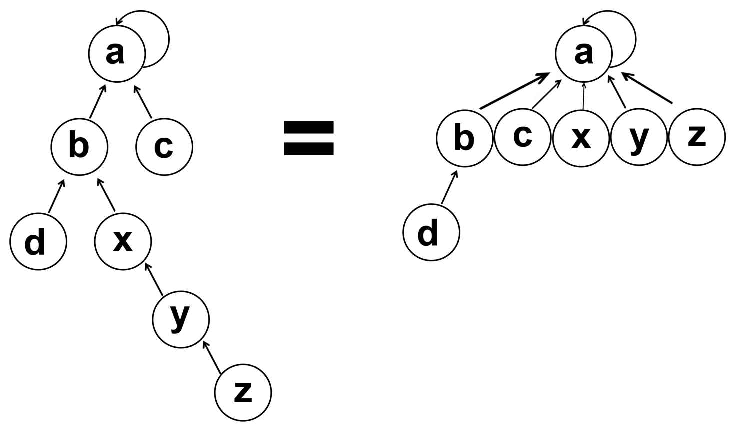

Figure 3.14: Performing *Find(z)* with the Path Compression hack.

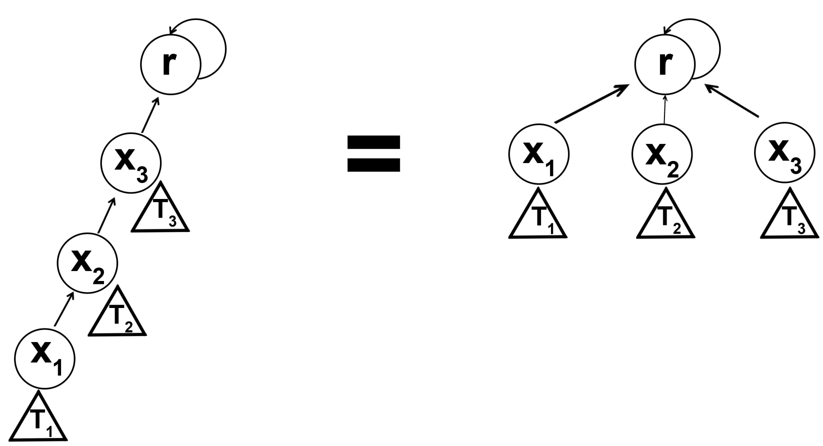

Figure 3.15: *Find*(x_1) moves all nodes or trees below the node and its parents with it when the Path Compression hack is used.

References

[B] M. BARILE "Topologist's Sine Curve," *MathWorld- A Wolfram Web Resource, http://mathworld.wolfram.com/Topolog*

[BW12] J.G. BROIDA and S.G. WILLIAMSON, *Comprehensive Introduction to Linear Algebra: Appendix A, http://cseweb.ucsd.edu/ gill/CILASite/Resources/17AppABCbib.pdf*, 2012

[EL15] E. LAMB "A Few of My Favorite Spaces: The Topologist's Sine Curve," *https://blogs.scientificamerican.com/roots-of-unity/a-few-of-my-favorite-spaces-the-topologist-s-sine-curve/*, 2015