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

Lecture 2: Jan 12, 2017

Lecturer: Prof. Bei Wang <beiwang@sci.utah.edu> *Scribe: Tim Sodergren*

The basic theme we will explore is that all data have shape and that shape has meaning. That meaning is different for each scale.

2.1 Background

Our data motivation is a graph which we can generalize as simply a collection of nodes

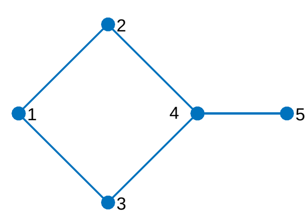
- All nodes are not necessarily of the same type.
- Nodes connect when there exists a relationship between them.
  - Abstract connections such as in a social network (e.g. Facebook, LinkedIn, etc.)
  - Concrete connections (e.g., transportation, biological networks) can have embedding such that the locations have meaning (e.g. roads).

More generally speaking, a graph is a combinatorial structure on a point cloud.

Let us consider a discrete object defined as:

$$G = (V, E)$$
$$E \subseteq VxV$$

Where V corresponds to the vertices and E corresponds to the edges (lines). In the following example:



these are given by:

$$V = \{1, 2, 3, 4, 5\}$$
$$E = \{12, 24, 13, 34, 45\}$$

Objects can be discrete or continuous. Topological Data Analysis (TDA) lives in between these two cases.

2.2 Connectivity

2.2.1 Graphs

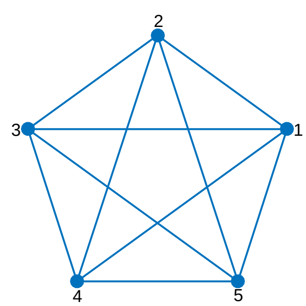
A graph, $G = (V, E)$ is called a **simple graph**, if the edge set, E, is a subset of the set of unordered pairs of vertices, V:

$$E \subseteq \binom{V}{2}$$

This means that no two edges connect the same two vertices and no edge joins a vertex to itself. The cardinalities of V and E are given by:

$$|V| = n$$
$$|E| = m \leq \binom{n}{2}$$

A **complete graph** (clique), K_n (where $n \geq 1$) , contains an edge for every pair of vertices (i.e., every node is connected to every other node). An example of a complete graph for K_5:



K_n represents the edge of the (n-1) simplex:

- $K_1 \rightarrow$ 0-simplex = single point
- $K_2 \rightarrow$ 1-simplex = line

- $K_3 \rightarrow$ 2-simplex = triangle
- $K_4 \rightarrow$ 3-simplex = tetrahedra
- $K_5 \rightarrow$ 4-simplex = 5-cell or pentatope
- ...

A **regular graph** is a graph where each vertex has the same degree.

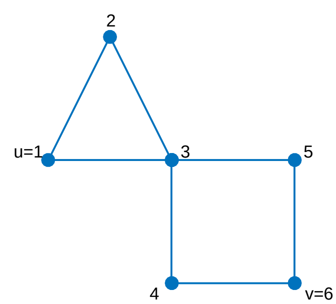
A **path**, γ, is the sequence of edges which connect a set of vertices:

$$\gamma(u, v) = \{u = u + 0, \dots, u + k = v\}$$
$$length(\gamma) = k$$

A path is considered simple if the vertices in the sequence are distinct:

$$u_i \neq u_j$$
$$i \neq j$$

So for the following example:



$$\gamma_1(1, 6) = \{1, 2, 3, 5, 6\}$$
$$\gamma_2(1, 6) = \{1, 3, 4, 6\}$$

are simple paths, and:

$$\gamma_3(1, 6) = \{1, 2, 1, 3, 4, 5, 6\}$$

is not a simple path as the vertex $u = 1$ repeats.

2.2.2 Definitions

Definition:A simple graph is connected if there exists a path between every pair of vertices.

Some algorithms for testing if a graph is connected:

1. DFS - depth first search

2. BFS = breadth first search
3. Union find

Definition: A connected component (CC) is a maximal subgraph that is connected.

Lemma 2.1. *A tree is the smallest connected graph (of a fixed number of nodes). If you remove any edge, it becomes disonnected.*

Definition: A spanning tree is the subset of G such that all vertices are covered by the minimum number of edges.

Lemma 2.2. *A graph is connected iff it has a spanning tree.*

Definition: A separation is a non-trivial partition of:

$$V = U \cup W$$
$$U, W \neq 0$$

such that no edge connects a vertex in U with a vertex in W. A simple graph is connected if it has no separation.

2.3 Topological Space

Topological spaces have some similarities to graphs but can be thought of as an abstraction of Euclidean space. It is a way to define when points are near each other without specifying how near. Concretely, a topology on a point set $\mathbb{X}$ is a collection U of subsets of $\mathbb{X}$, called open sets such that, $\mathbb{X}$ is open,the empty set $\emptyset$ is open, and if U_i is open for all i then the union of all U_i is open and the intersection of any two U_i is also open.

Definition: A topological space is path connected if every pair of points is connected by a path (e.g. a disk or cookie).

Definition: A separation of a topological space, $\mathbb{X}$, is a partition $U \cup W$ into non-empty, open subsets.

Definition: A topological space is connected if it has no separation.

Lemma 2.3. *Connectedness is weaker than path-connectedness (i.e., a connected space does not imply that the space is also path connected). There are, however, some pathological cases such as the topologist's sine curve.*