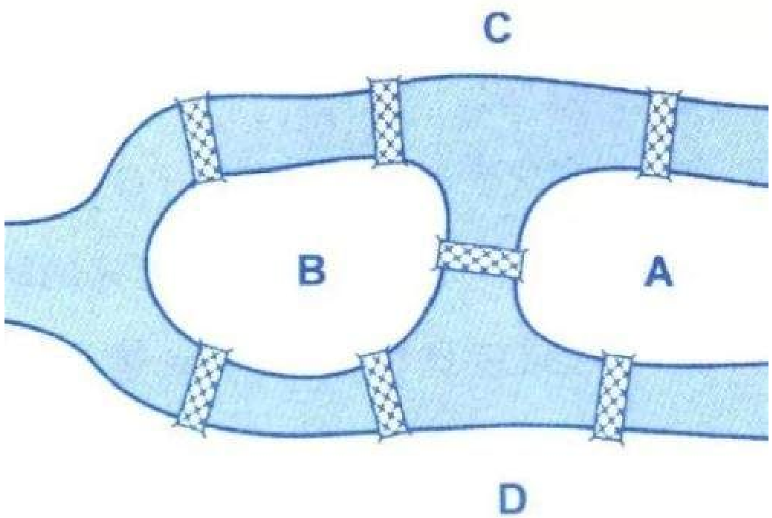
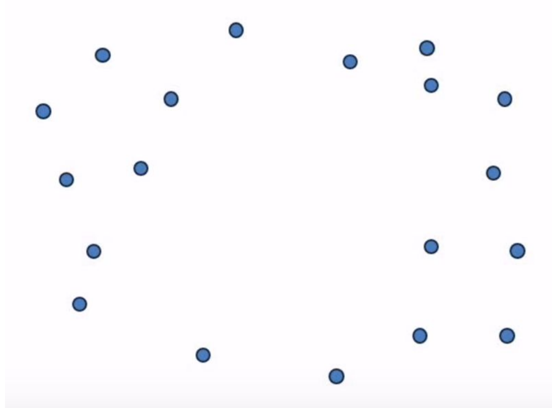
拓扑学（Topology），是一门研究拓扑空间的学科，主要研究空间内，在连续变化（如拉伸或弯曲，但不包括撕开或粘合）下维持不变的性质。拓扑学是从几何学与集合论里发展出来的学科，研究空间、维度与变换等概念。这些词汇的来源可追溯至哥特佛莱德·莱布尼兹，他在17世纪提出位置的几何学（Geometria Situs）和位相分析（Analysis Situs）的说法。莱昂哈德·欧拉（Leonhard Euler）的柯尼斯堡七桥问题与欧拉示性数（Euler Characteristic）被认为是该领域最初的定理。（from Wikipedia）



拓扑学虽然是数学的重要分支，但同样对计算机科学有很大影响。拓扑集合论分支在计算机领域的应用包括集合本身和机器语言等等，之后关于机器语言的介绍会涉及。作为一个图论笔记，这里先重点介绍一下拓扑几何学分支。

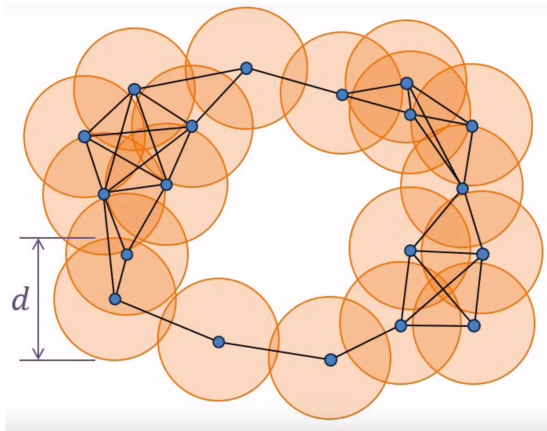
几何学分支在计算机领域的一大应用是利用持续同调（persistent homology）进行特征检测（feature ditection）。举个例子，如果问人们下图中19个点的分布显示了怎样的几何特征，大多数人可能会说它们形成了一个圆环的形状。但计算机如何得出这个推论呢？



（from Youtube）

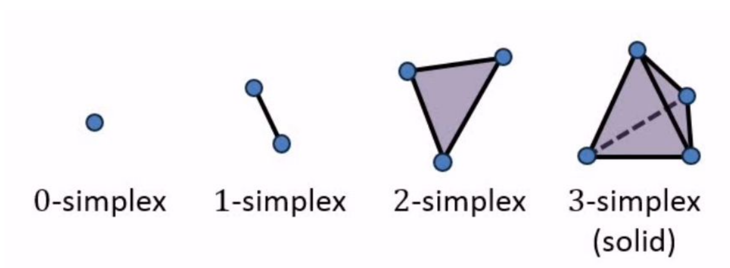
这时候就可以用到持续同调了。对于上面这张图，我们进行如下操作：

(1) 选取一个直径d，如下图所示
(2) 连接每对距离小于等于d的点



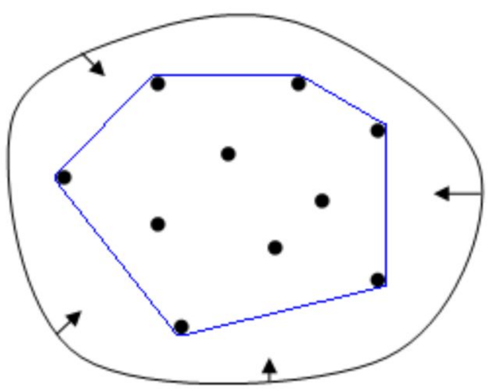
（from Youtube）

于是我们得到了上面这张图。它能说明什么呢？容我先介绍一下n维单型（n-simplex）。n维单型是对三角形在不同维度的泛化，下图展示了0维到3维的单型。



（from Youtube）

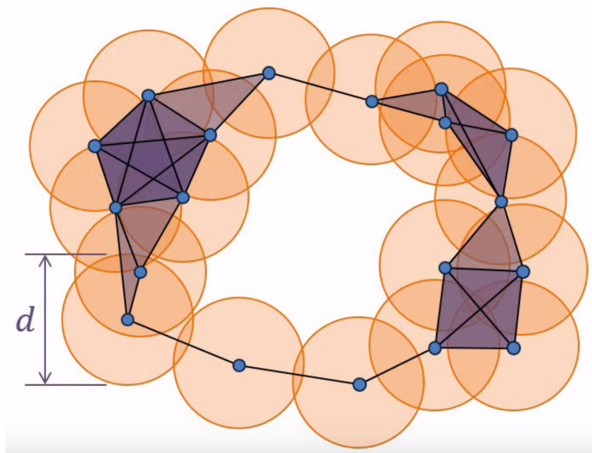
具体来说，n维单型是一个n维的多胞形（polytope，即一类由平的边界构成的几何结构）上n+1个顶点的凸包（convex hull，即所有包含这些顶点的凸集的交集，下图是一个二维凸包的例子）。作为一个三维生物我也不知道怎么脑补4维及以上的单型 :(



（from Wikipedia）

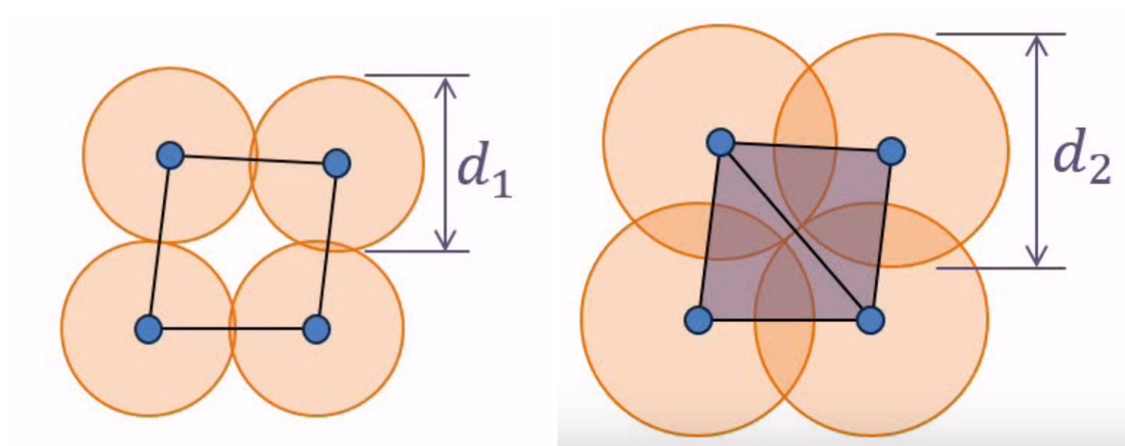
回到之前的问题，在连接每对距离小于直径d的点之后，我们可以进行如下操作：

(3) 填充出现的所有n维单型，如下图所示的紫色区域



（from Youtube）

我们可以看到，当多条边连成一个封闭的多边形时，橙色的圆也相应围出了空隙（hole，比如橙色中间的白色部分）；当多条边围出的多边形被紫色填充时，空隙也相应消失。举个例子，对于下图的四个点，如果我们不断增加直径d，我们可以看到空隙从出现到消失的过程，也是直径从d1增长到d2空隙存在的过程。

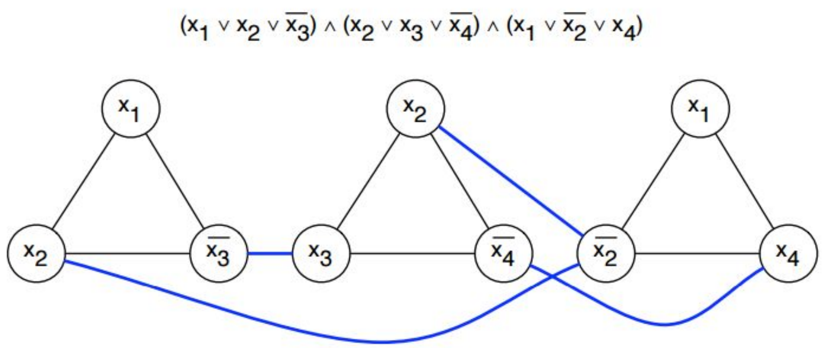


（from Youtube）

这一过程可以用下图直观地表示，每个空隙的存在可以用区间（d1, d2）来表示，这个区间（如下图的蓝条）被称之为barcode或persistance，即持续同调中的“持续”。对于例子中的19个点，当我们从零开始不断增加直径d时，不同的空隙会产生不同的barcode。这些barcode有长有短，短的代表出现的空隙较小或狭长，通常是噪音（noise）；长的代表出现的空隙较大或包含较大的圆形区域，通常就是我们需要的特征（feature）。

（from Youtube）

除了持续同调之外，拓扑几何学分支衍生出的概念比如完美图（perfect graph），弦图（chordal graph），区间图（interval graph），在图论中也同样存在，它的应用包括资源分配（resource allocation）。资源可以用顶点表示，资源之间的限制条件可以用边表示。这一描述很容易让人联想起Boolean Satisfiability Problem（SAT），提到NP-Complete最经常想到的3-SAT问题就在其中（NP-Complete以后会讲，这里就不赘述了）。下图是一个将SAT问题的逻辑形式转化成图的例子。不仅是逻辑学中的SAT问题，一些其它领域的问题也可以转化为图论问题进行解决。



（from StackExchange）

由于内容较多，我们先正式介绍一下几个概念，下一篇文章会讲拓扑排序算法和应用。

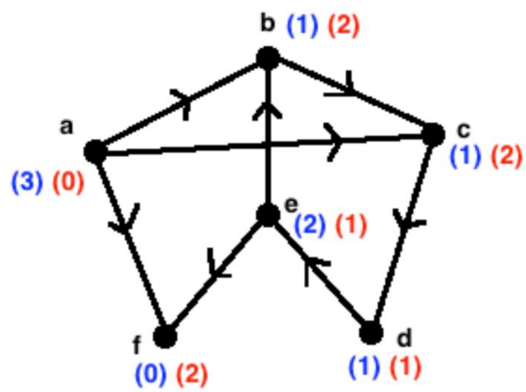
1. 顶点/节点（vertex/node），边（edge），图（graph）
   (1) 顶点/节点
   有的定义会说顶点是包含坐标的点，但在这个系列里顶点和节点是一个意思，不一定包含坐标，但包含点边关系。一个顶点可以是多条边的起点或终点，通常用V表示。
   (2) 边

有的定义会说边是包含数学表达式的线段，但在这个系列里边不一定包含数学表达式，但包含边点关系。一条边连接两个顶点，通常用E =（v1，v2）表示，v1即起点，v2即终点。

(3) 图

图的定义就多了，在这个系列里图包含一系列顶点和边，通常用G =（V，E）表示。图分有向图（directed graph，简称digraph）和无向图（undirected graph）。在有向图中每条边只能从起点到终点，不能从终点到起点；在无向图中起点和终点是双向的，相当于复制边相同的有向图中的所有边并交换起点和终点。因此，无向图是有向图的子集，这一点在论证算法的限制时很重要。

2. 度（degree），入度（in-degree），出度（out-degree）
   度是关于顶点的一个概念，分为入度和出度。一个顶点的入度是以该顶点为终点的边的数量；一个顶点的出度是以该顶点为起点的边的数量。举个例子，下图中红色括号内为该顶点的入度，蓝色括号内为该顶点的出度。一个顶点的度为入度和出度之和。图关于入度和出度的性质有很多也十分重要，比如在任何图中度为奇数的顶点数量总是偶数，其中一些性质和开头提到的七桥问题密切相关。

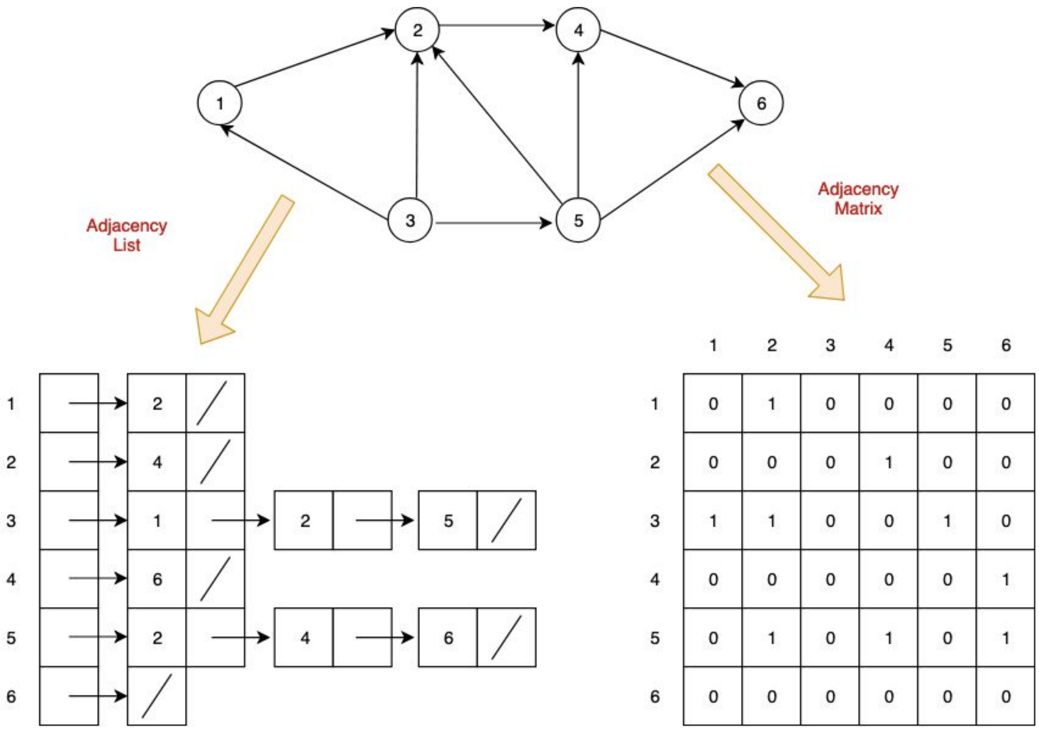


（from Mathonline）

3. 拓扑次序（topological order）
   拓扑次序仅适用于有向无环图（directed acyclic graph），且每个有向无环图至少有一个拓扑次序。它对于一个图内的所有顶点进行线性排序，使得对于任意一条边e = (v1, v2)来说，v1总是排在v2前面。

4. 邻接链表（adjacency list）和邻接矩阵（adjacency matrix）
   邻接链表和邻接矩阵是图的两种常见表达方式。邻接链表将图中每条边按起点划分，下图中的邻接链表是包含每个顶点的集合，每个顶点又指向所有可以达到的相邻顶点的链表；邻接矩阵是一个|V| * |V|的二维矩阵，下图中第i行第j列为1代表图中存在从顶点i到顶点j的边，为0则代表不存在从顶点i到顶点j的边。通常邻接链表在

稀疏图（sparse graph，即边数量较少的图）中占用较少的空间（邻接矩阵中会存在大量0）。



（from AlgorithmTutor）

相关材料：

https://en.wikipedia.org/wiki/Simplex
https://en.wikipedia.org/wiki/Polytope
https://en.wikipedia.org/wiki/Convex_hull
https://en.wikipedia.org/wiki/Persistent_homology
https://www.youtube.com/watch?v=h0bnG1Wavag
https://en.wikipedia.org/wiki/Boolean_satisfiability_problem
https://www.tutorialspoint.com/graph_theory/graph_theory_fundamentals.htm
https://en.wikipedia.org/wiki/Directed_acyclic_graph
http://mathonline.wikidot.com/out-degree-sequence-and-in-degree-sequence
https://en.wikipedia.org/wiki/Topological_sorting
https://algorithmtutor.com/Data-Structures/Graph/Graph-Representation-Adjacency-List-and-Matrix/