Topological Data Analysis

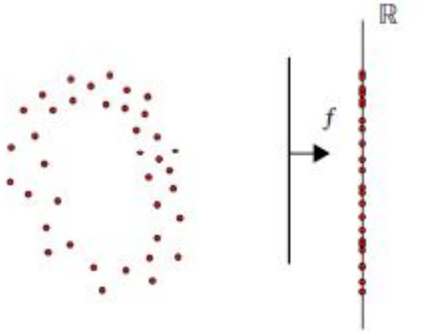
Tiffany Hu

what is computational geometry?????

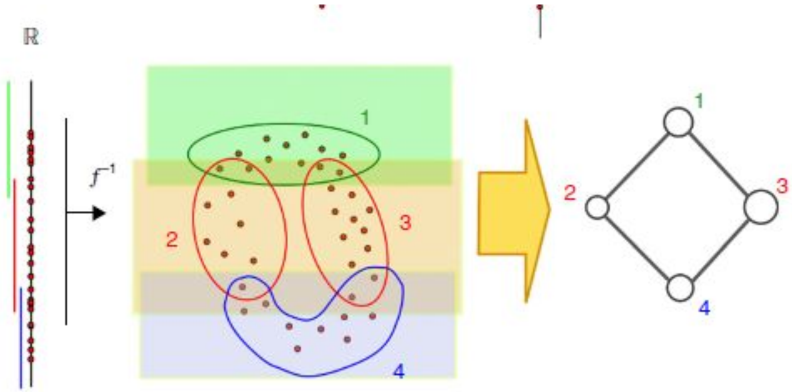
- geometry gives a concrete face to topological structures
  - points, connectedness
  - disk packings, sphere packings, shapes
  - triangulations, voronoi diagrams
- allows us to express spaces on a computer
  - techniques used to visualize data
  - difficult for algorithms to identify "holes"

points!!!!!

- you collect data, but you don't know the shape of the data
  - metric space with finitely many points
  - EX: finite set of genomic sequences
- translation of biological data into R^k



function f maps RNAseq point-cloud data to R^k



inverse function f^{-1} maps a covering of R^k to a covering of the point-cloud data

neighborhoods!!!!!

- to begin visualizing the data, we need to construct neighborhoods
  - we don't know the space of the points look like
- voronoi diagrams:

  Let S be a finite set of points in R^2 . Describing the elements of S as "sites", we want to find the region of points that are at least as close as to any other site using Euclidean distance.

  $$V = \{x \in R^2 \mid \| x - s \| \leq \| x - t \|, \forall t \in S\}$$

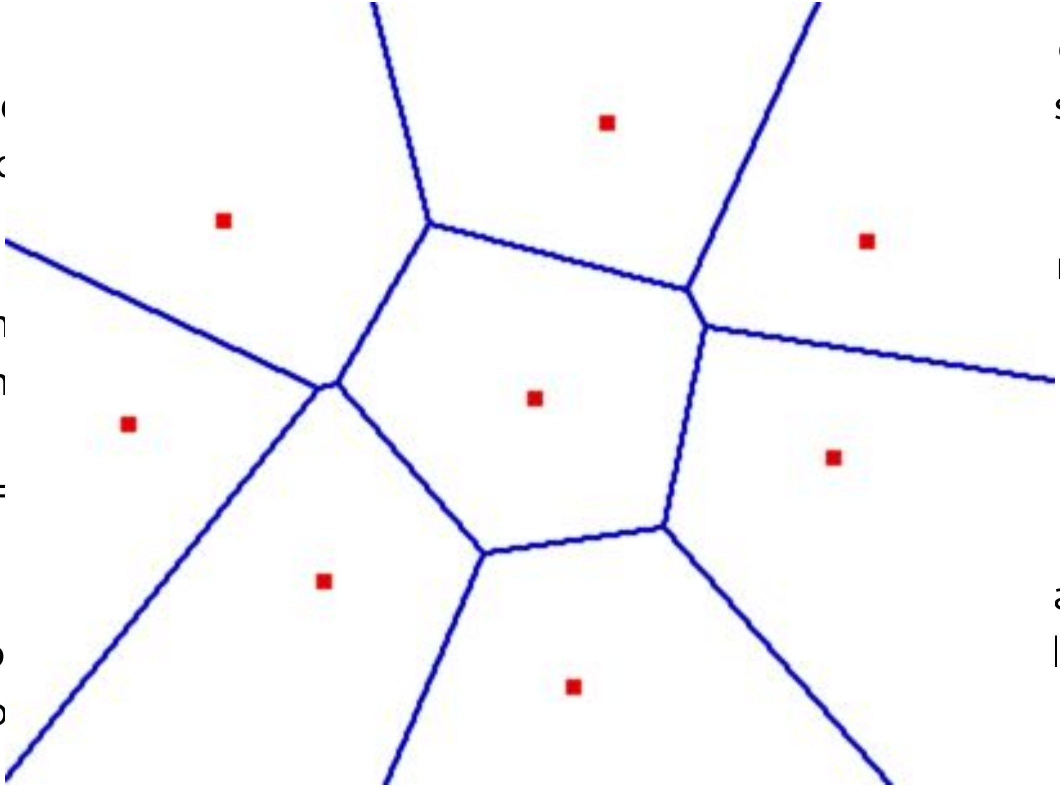
- trying to "scout out" the location of the points and form a visualization
- the set of points that satisfy the inequality form closed half-plane
  - a voronoi diagram is basically an intersection of lots of half-planes!! yay!!

neighborhoods!!!!!

- to begin ... ods
    - we ... s look like
- voronoi ...

    Let ... nts of S as "sites", we
    wan ... as to any other site
    usin ...

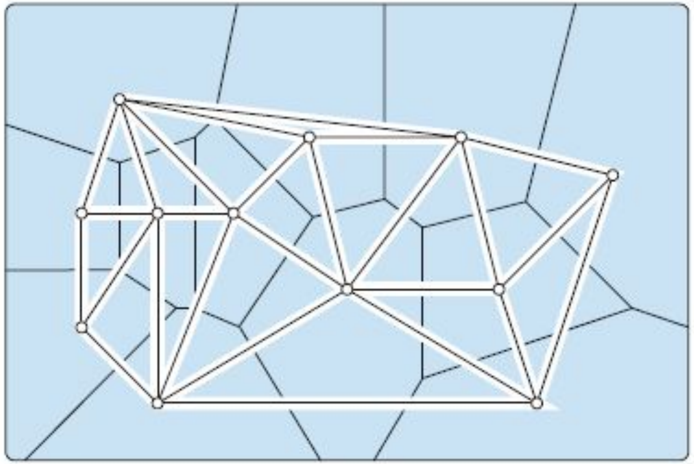
    Vs = ...



- trying to ... alization
- the set o ... lane
    - a vo ... lf-planes!! yay!!

triangles!!!!!

- from the voronoi diagram, we can construct the delaunay triangulation of the data
    - basically connecting two sites by a straight edge if two voronoi regions share an edge

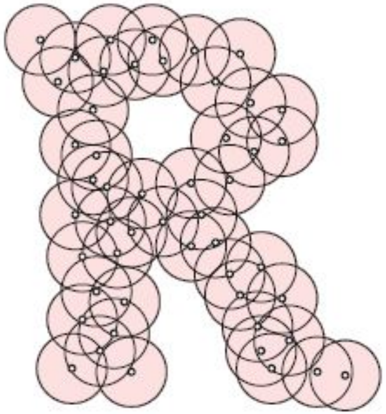


why do we care about triangles?!?!

- we can now describe data as a shape using concept of α-shapes!

  constructing the α-shape:

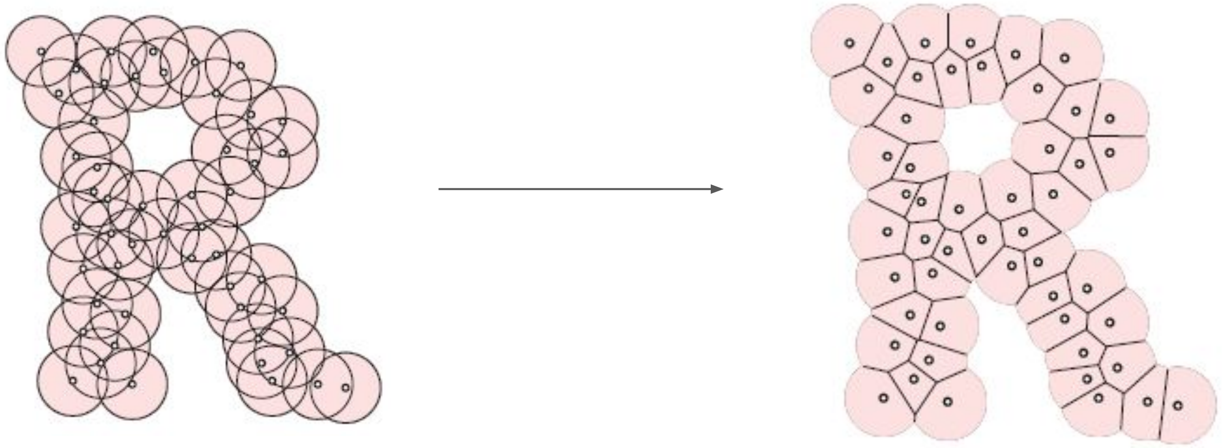
  1) Let α be a fixed radius. Let Dx(α) be the closed disk with center x and radius α.



why do we care about triangles?!?!

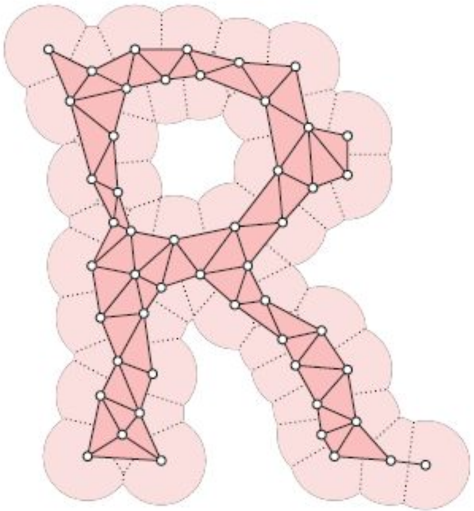
2) overlay voronoi diagram with the union of the disks

- decomposing the triangulation



why do we care about triangles?!?!

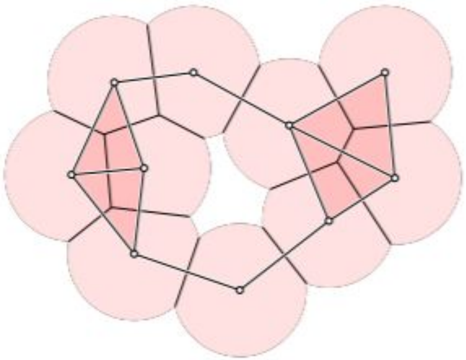
3) triangulation = α−complex

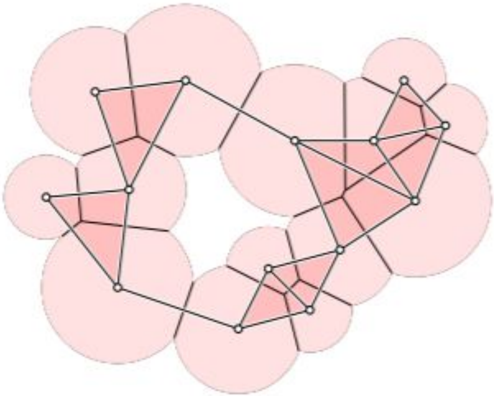


A(0) = set of sites
A(∞) = Delaunay triangulation

okay... so?????

- the takeaway is that the value of α determines the derived shape of the data
  - you can have weighted diagrams, for example
  - Applications: protein models, space filling models of molecules





- filtration works in determining threshold value
  - determines which shape you should be looking at

triangulation of a space = **simplicial complex** (a data structure in defining topological spaces)

A set of k +1 points, {u0, u1, . . . , uk}, is affinely independent if the k vectors {u1−u0, u2−u0, . . . , uk−u0} are linearly independent. A k−simplex is the convex hull of k+1 affinely independent points.

Therefore, we see how we get the shape of the data and from that, we can get the actual topological space we are working in.

DATA SET > TOPOLOGICAL SPACE

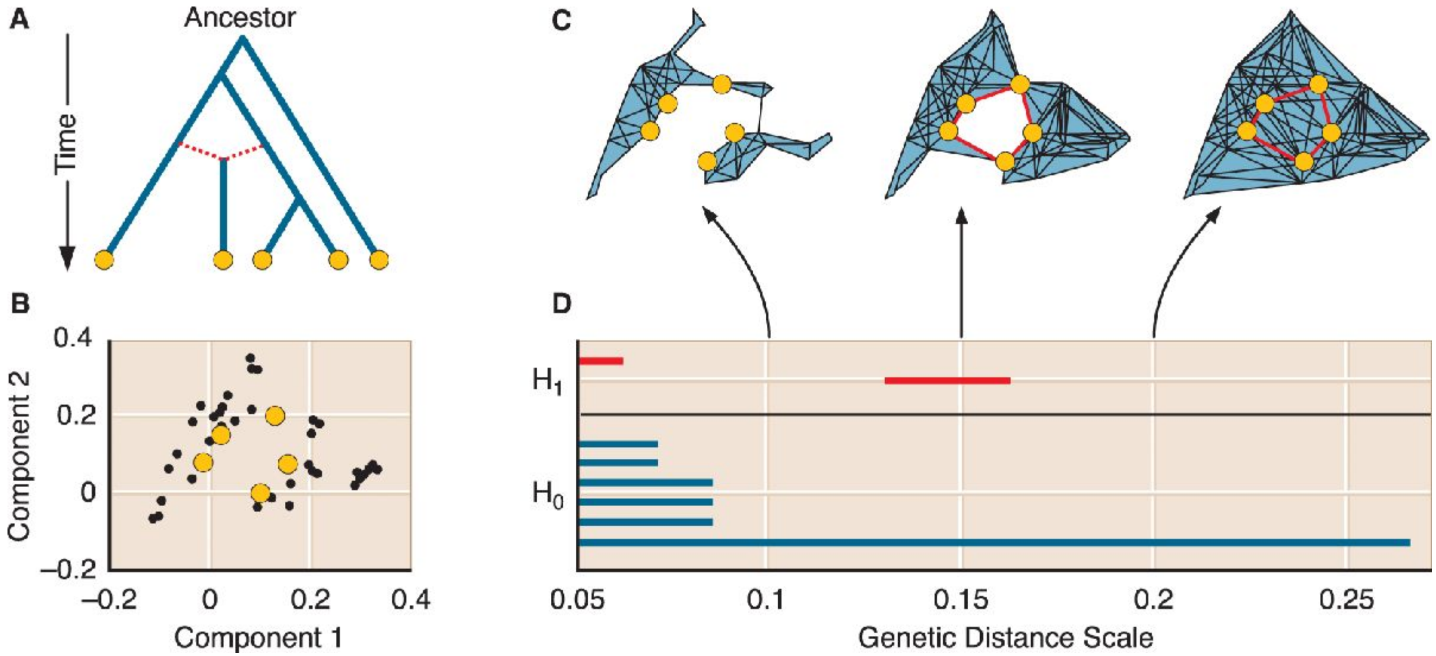
homology!!

- trying to calculate homology is the reason that we're doing all of this
- chain groups & boundary of a p-simplex -> homology
  - p-chain is a formal sum of p-simplices in a simplicial complex
  - boundary of a p-simplex is the set of (p-1)-faces
    - p-boundary is the boundary of a (p+1)-chain

Hp = Zp/Bp

Zp: subgroup of p-chains
Bp: subgroup of p-boundaries

- output: betti number that tells how many holes of each dimension you have
  - rank of the homology group
  - number of independent components of different dimensions that are in the space


A
Ancestor
Time
B
Component 2
0.4
0.2
0
−0.2
−0.2
0
0.2
0.4
Component 1
C
D
H1
H0
0.05
0.1
0.15
0.2
0.25
Genetic Distance Scale