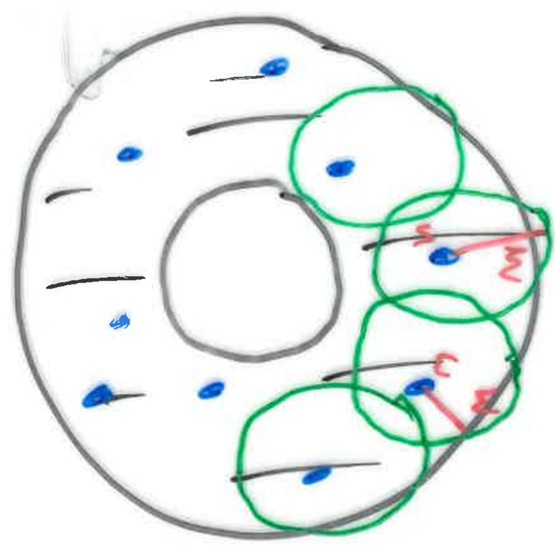
Suppose X is an unknown metric space. Suppose given a finite sample $S \subset X$ and the distance

$$d_X(x,y)$$

for all $x, y \in S$.

<u>Problem</u> We'd like to infer something about the topology of X using the distance $d_X(x,y)$ for $x,y \in S$.

<u>Example</u> A sample S is taken from an unknown metric space $X \subseteq \mathbb{R}^2$.

$X =$ 

S finite sample.

know

$d_X(x,y)$, $x,y \in S$

We could construct the ball

$$B(s,\varepsilon) = \{ x \in X : d_X(x,s) < \varepsilon \}.$$

We could hope that the collection

$$\mathcal{U}_\varepsilon = \{ B(s,\varepsilon) \}_{s \in X}$$

might be an open cover of X for suitable $\varepsilon > 0$.

If $\mathcal{U}_\varepsilon$ is an open cover of X then Leray's theorem says that the nerve $N\mathcal{U}_\varepsilon$ is homotopy equivalent to X.

The hypothesis of Leray's theorem is satisfied because a ball is contractible, and so too is any intersection of balls.

N.B. This means that $N\mathcal{U}_\varepsilon$ has the same Euler characteristic as X.

<u>However</u>, we can't construct the balls as we don't know X. Also, even if we could construct the balls, it would be expensive to compute their intersections.

Recall: $N\mathcal{U}_\varepsilon$ is a simplicial complex with one vertex s for each ball $B(s, \varepsilon)$, $s \in S$.

$N\mathcal{U}_\varepsilon$ has an edge

if $B(s, \varepsilon) \cap B(t, \varepsilon) \neq \emptyset$.

This will be non-empty if and only if $d_X(s, t) < 2\varepsilon$.

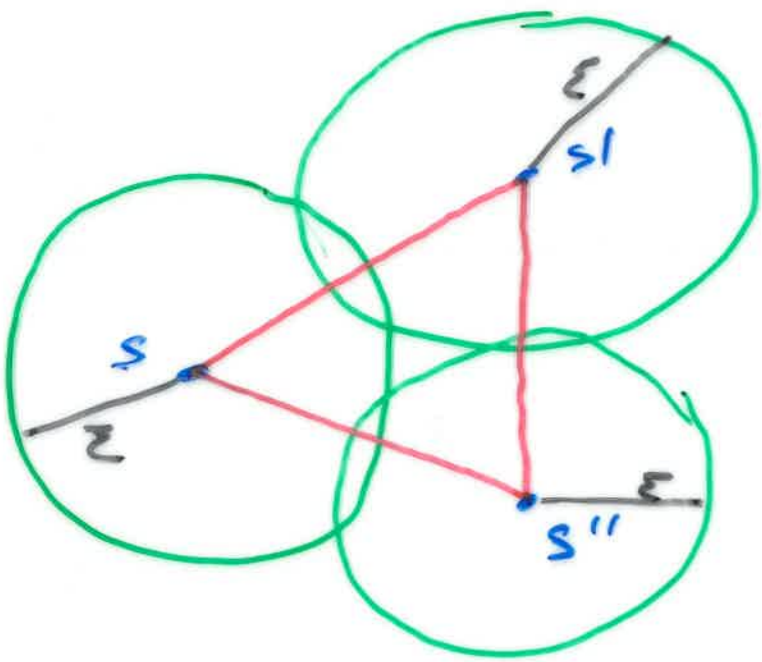
So we can construct the vertices and edges of $N\mathcal{U}_\varepsilon$ just from the distances between points in S.

$N\mathcal{U}_\varepsilon$ has a 2-simplex for

$$B(s, \varepsilon) \cap B(s', \varepsilon) \cap B(s'', \varepsilon) \neq \emptyset$$

$s, s', s'' \in S$.

If we happened to know X we could compute these balls and their intersections. However, this is a length computation, We could approximate $N\mathcal{U}_\varepsilon$ by considering the picture:



We could approximate the nerve $N\mathcal{U}_\varepsilon$ by the clique complex

$K_{2\varepsilon}$:

$K_{2\varepsilon}$ has a k-simplex

$$\sigma = \{s_0, s_1, \dots, s_k\}$$

whenever

$$d(s_i, s_j) < 2\varepsilon \text{ for all } s_i, s_j \in \sigma.$$

Proposition

$$N\mathcal{U}_{\varepsilon} \subseteq K_{2\varepsilon} \subseteq N\mathcal{U}_{2\varepsilon}$$

Conclusion:

K_{ε} is a good approximation to $N\mathcal{U}_{\varepsilon}$ since we consider a range of $\varepsilon > 0$.

The above ideas combine to form the <u>mapper clustering</u> algorithm.