Approximate Near Neighbor (ANN)

- Preprocess: a dataset $P \subset \mathbb{R}^d$ on $n$ points.
- Query: $q \in \mathbb{R}^d$ such that $\exists p \in P$, $\text{dist}(q, p) \leq r$
- Output: $p' \in P$, such that $\text{dist}(q, p') \leq c \cdot r$
- High dimensional $\rightarrow \omega(\log n) \leq d \leq n^{o(1)}$

F.A.Q

- What parameters do we care about?
  - Approximation, space and query time.
- We want polynomial dependence on the dimension of the metric space.
  - For a normed space $(\mathbb{R}^d, \| \cdot \|)$, dimension is $d$.
  - For a metric space $(X, d_X)$, dimension is $\log |X|$, or doubling dimension.

The Question

- Characterize metric spaces $(X, d_X)$ which admit efficient ANN data structures.

Cutting Modulus of Metric Space

- **Definition (Cutting Modulus)**
  The cutting modulus $\Xi(X, \epsilon)$ of a metric space $(X, d_X)$ is the smallest $R > 0$ such that for every $G = (V, E)$ and any $1$-Lipschitz map $f : G \rightarrow X$:
  - Either there exists a ball of radius $R$ in $X$ containing $\Omega(|V|)$ points from $f(G)$, or
  - $G$ contains a cut $S \subset V$ with conductance at most $\epsilon$.

Main Result

- **Theorem**
  For any $\epsilon > 0$, there exists a cell-probe data structure for ANN achieving:
  - Approximation: $O(\Xi(X, \epsilon))$.
  - Space: $\text{poly}(d) \cdot n^{1+O(\epsilon)}$.
  - Query time: $\text{poly}(d) \cdot n^{O(\epsilon)}$.

- For several important cases, obtain time efficient data structures:
  - $\ell_p$, $S_p$, and general norms*.

- **Theorem (Partitioning metric spaces)**
  For a metric space $(X, d_X)$ and $\epsilon > 0$, there exists a collection $C$ of subsets of $X$ with $|C| \leq |X|^{O(\log \frac{|X|}{\epsilon})}$ such that for any subset $P \subset X$ of $n$ points:
  - Either there exists a ball of radius $\Xi(X, \epsilon)$ containing $\Omega(n)$ points from $P$, or
  - There exists a distribution $S$ on $C$ such that for all $p, q \in X$ with $d_X(p, q) \leq 1$, $\Pr_{S \sim S}[S(p) \neq S(q)] \leq O(\epsilon)$, and for all $S \in \text{supp}(S)$, $\Omega(1) \leq \frac{|S \cap P|}{n} \leq 1 - \Omega(1)$.