5339 - Algorithms design under a geometric lens
Spring 2014, CSE, OSU
Lecture 2: Random partitions

Instructor: Anastasios Sidiropoulos

January 10, 2014

Metric embedding examples

Let K_n be the complete graph on n vertices.
Assume all edges have unit length.

Metric embedding examples

Let K_n be the complete graph on n vertices.
Assume all edges have unit length.

What is the best-possible distortion for embedding the shortest-path metric of K_n into the shortest-path metric of a tree?

Metric embedding examples

Let K_n be the complete graph on n vertices.
Assume all edges have unit length.

What is the best-possible distortion for embedding the shortest-path metric of K_n into the shortest-path metric of a tree?

Embedding into a star gives distortion $O(1)$.

Metric embedding examples

Let C_n be the cycle graph on n vertices.

Metric embedding examples

Let C_n be the cycle graph on n vertices.
What is the best-possible distortion for embedding the shortest-path metric of C_n into $\mathbb{R}^1$?

Metric embedding examples

Let C_n be the cycle graph on n vertices.
What is the best-possible distortion for embedding the shortest-path metric of C_n into $\mathbb{R}^1$?

Best-possible distortion $\Omega(n)$

Metric embedding examples

Let H be the $\sqrt{n} \times \sqrt{n}$ grid.

Metric embedding examples

Let H be the $\sqrt{n} \times \sqrt{n}$ grid.
Let X be the set of vertices on the boundary of H.

Metric embedding examples

Let H be the $\sqrt{n} \times \sqrt{n}$ grid.
Let X be the set of vertices on the boundary of H.

What is the best possible distortion for embedding $C_{\sqrt{n}}$ into the metric (X, d_H)?

Metric embedding examples

Let H be the $\sqrt{n} \times \sqrt{n}$ grid.
Let X be the set of vertices on the boundary of H.

What is the best possible distortion for embedding $C_{\sqrt{n}}$ into the metric (X, d_H)?

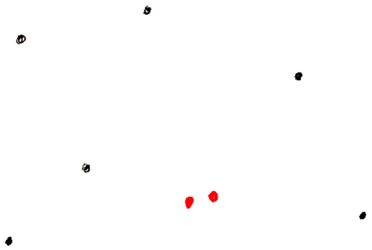
What is the best possible distortion for embedding H into $\mathbb{R}^1$?

The Closest Pair problem

Given a metric space (X, ρ), find a pair of minimum distance.

The Closest Pair problem

Given a metric space (X, ρ), find a pair of minimum distance.
I.e. find $x \neq y \in X$, minimizing $\rho(x, y)$.



An algorithm for the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$.
Assume X is in general position.

An algorithm for the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$.
Assume X is in general position.

Sort the points according to their x-coordinate.

An algorithm for the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$.
Assume X is in general position.

Sort the points according to their x-coordinate.
Let ℓ be a vertical line separating X into two sets $X = L \cup R$, each of size $n/2$.

An algorithm for the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$.
Assume X is in general position.

Sort the points according to their x-coordinate.
Let ℓ be a vertical line separating X into two sets $X = L \cup R$, each of size $n/2$.
Recurse on L, and on R.

An algorithm for the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$.
Assume X is in general position.

Sort the points according to their x-coordinate.
Let ℓ be a vertical line separating X into two sets $X = L \cup R$, each of size $n/2$.
Recurse on L, and on R.
Let α_L, α_R be the closest distance on L, R respectively.

An algorithm for the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$.
Assume X is in general position.

Sort the points according to their x-coordinate.
Let ℓ be a vertical line separating X into two sets $X = L \cup R$, each of size $n/2$.
Recurse on L, and on R.
Let α_L, α_R be the closest distance on L, R respectively.
Let

$$\alpha = \min\{\alpha_L, \alpha_R\}.$$

An algorithm for the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$.
Assume X is in general position.

Sort the points according to their x-coordinate.
Let ℓ be a vertical line separating X into two sets $X = L \cup R$, each of size $n/2$.
Recurse on L, and on R.
Let α_L, α_R be the closest distance on L, R respectively.
Let

$$\alpha = \min\{\alpha_L, \alpha_R\}.$$

For every point $x \in L$, all candidate points in $y \in R$ with $\|x - y\|_2 < \alpha$, lie inside a region S_x of diameter $O(\alpha)$. Why?

An algorithm for the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$.
Assume X is in general position.

Sort the points according to their x-coordinate.
Let ℓ be a vertical line separating X into two sets $X = L \cup R$, each of size $n/2$.
Recurse on L, and on R.
Let α_L, α_R be the closest distance on L, R respectively.
Let

$$\alpha = \min\{\alpha_L, \alpha_R\}.$$

For every point $x \in L$, all candidate points in $y \in R$ with $\|x - y\|_2 < \alpha$, lie inside a region S_x of diameter $O(\alpha)$. Why?
Each region S_x contains at most $O(1)$ points from R. Why?

An algorithm for the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$.
Assume X is in general position.

Sort the points according to their x-coordinate.
Let ℓ be a vertical line separating X into two sets $X = L \cup R$, each of size $n/2$.
Recurse on L, and on R.
Let α_L, α_R be the closest distance on L, R respectively.
Let

$$\alpha = \min\{\alpha_L, \alpha_R\}.$$

For every point $x \in L$, all candidate points in $y \in R$ with $\|x - y\|_2 < \alpha$, lie inside a region S_x of diameter $O(\alpha)$. Why?
Each region S_x contains at most $O(1)$ points from R. Why?
For each $x \in L$, try all points in S_x.

An algorithm for the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$.
Assume X is in general position.

Sort the points according to their x-coordinate.
Let ℓ be a vertical line separating X into two sets $X = L \cup R$, each of size $n/2$.
Recurse on L, and on R.
Let α_L, α_R be the closest distance on L, R respectively.
Let

$$\alpha = \min\{\alpha_L, \alpha_R\}.$$

For every point $x \in L$, all candidate points in $y \in R$ with $\|x - y\|_2 < \alpha$, lie inside a region S_x of diameter $O(\alpha)$. Why?
Each region S_x contains at most $O(1)$ points from R. Why?
For each $x \in L$, try all points in S_x.
Proceed similarly for the points in R.

Running time

Total running time:

$$T(n) = 2 \cdot T(n/2) + O(n).$$

Therefore, $T(n) = O(n \log n)$.

Decision version of closest pair

Given a metric space (X, ρ), and some $r > 0$, decide whether the min-distance is at most r.

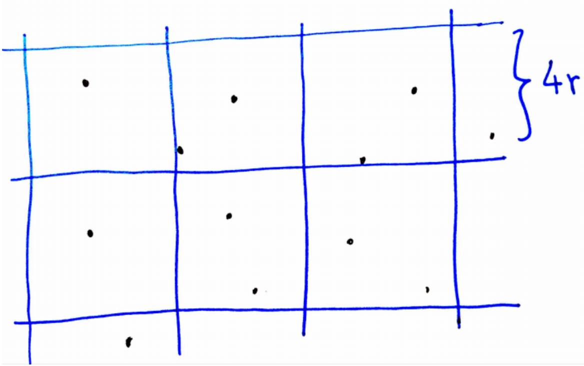
A simpler algorithm for the decision version in the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$, $r > 0$.

A simpler algorithm for the decision version in the Euclidean plane

Let $X \subset \mathbb{R}^2$, $|X| = n$, $r > 0$.

Impose a grid in $\mathbb{R}^2$, where each cell is a square of side length $4 \cdot r$.



A simpler algorithm (cont.)

Suppose that we *randomly shift* the grid.

A simpler algorithm (cont.)

Suppose that we *randomly shift* the grid.
I.e., we pick $t_x \in [0, 4r)$ and $t_y \in [0, 4r)$, uniformly and independently at random.

A simpler algorithm (cont.)

Suppose that we *randomly shift* the grid.
I.e., we pick $t_x \in [0, 4r)$ and $t_y \in [0, 4r)$, uniformly and independently at random.
We shift the grid horizontally by t_x, and vertically by t_y.

A simpler algorithm (cont.)

Suppose that we *randomly shift* the grid.
I.e., we pick $t_x \in [0, 4r)$ and $t_y \in [0, 4r)$, uniformly and independently at random.
We shift the grid horizontally by t_x, and vertically by t_y.

I.e. we get a partition of $\mathbb{R}^2$ into cells $\{C_{i,j}\}_{i,j \in \mathbb{Z}}$, where

$$C_{i,j} = [t_x + i4r, t_x + (i+1)4r) \times [t_y + j4r, t_y + (j+1)4r).$$

Separation probability

What is the probability that a pair of points $x, y \in X$ ends up in different cells?

Separation probability

What is the probability that a pair of points $x, y \in X$ ends up in different cells?

Lemma

$\Pr[C(x) \neq C(y)] \leq \frac{\|x-y\|_2}{2r}$.

Separation probability

What is the probability that a pair of points $x, y \in X$ ends up in different cells?

Lemma

$\Pr[C(x) \neq C(y)] \leq \frac{\|x-y\|_2}{2r}$.

Proof.

What is the probability that x and y are separated by a vertical line of the grid?

Separation probability

What is the probability that a pair of points $x, y \in X$ ends up in different cells?

Lemma

$\Pr[C(x) \neq C(y)] \leq \frac{\|x-y\|_2}{2r}$.

Proof.

What is the probability that x and y are separated by a vertical line of the grid?

The probability that x and y are separated by a vertical line of the grid, is at most $\frac{\|x-y\|_2}{4r}$.

Separation probability

What is the probability that a pair of points $x, y \in X$ ends up in different cells?

Lemma

$\Pr[C(x) \neq C(y)] \leq \frac{\|x-y\|_2}{2r}$.

Proof.

What is the probability that x and y are separated by a vertical line of the grid?
The probability that x and y are separated by a vertical line of the grid, is at most $\frac{\|x-y\|_2}{4r}$.
Why?

Separation probability

What is the probability that a pair of points $x, y \in X$ ends up in different cells?

Lemma

$\Pr[C(x) \neq C(y)] \leq \frac{\|x-y\|_2}{2r}$.

Proof.

What is the probability that x and y are separated by a vertical line of the grid?
The probability that x and y are separated by a vertical line of the grid, is at most $\frac{\|x-y\|_2}{4r}$.
Why?
The probability that x and y are separated by a horizontal line of the grid, is at most $\frac{\|x-y\|_2}{4r}$.

Separation probability

What is the probability that a pair of points $x, y \in X$ ends up in different cells?

Lemma

$\Pr[C(x) \neq C(y)] \leq \frac{\|x-y\|_2}{2r}$.

Proof.

What is the probability that x and y are separated by a vertical line of the grid?
The probability that x and y are separated by a vertical line of the grid, is at most $\frac{\|x-y\|_2}{4r}$.
Why?
The probability that x and y are separated by a horizontal line of the grid, is at most $\frac{\|x-y\|_2}{4r}$.
By the union bound, the probability that x and y are separated is at most $\frac{\|x-y\|_2}{2r}$. □

The algorithm

Repeat the following $O(\log n)$ times:

The algorithm

Repeat the following $O(\log n)$ times:
Compute a random partition using a randomly shifted grid.

The algorithm

Repeat the following $O(\log n)$ times:
Compute a random partition using a randomly shifted grid.
There are at most n non-empty cells (since there are n points).

The algorithm

Repeat the following $O(\log n)$ times:
Compute a random partition using a randomly shifted grid.
There are at most n non-empty cells (since there are n points).
If there exists a non-empty cell with at least 100 points, then it must contain a pair at distance at most r.

The algorithm

Repeat the following $O(\log n)$ times:
Compute a random partition using a randomly shifted grid.
There are at most n non-empty cells (since there are n points).
If there exists a non-empty cell with at least 100 points, then it must contain a pair at distance at most r.
Why?

The algorithm

Repeat the following $O(\log n)$ times:
Compute a random partition using a randomly shifted grid.
There are at most n non-empty cells (since there are n points).
If there exists a non-empty cell with at least 100 points, then it must contain a pair at distance at most r.
Why?
Otherwise, for every non-empty cell C, check all pairs of points in C.

The algorithm

Repeat the following $O(\log n)$ times:
Compute a random partition using a randomly shifted grid.
There are at most n non-empty cells (since there are n points).
If there exists a non-empty cell with at least 100 points, then it must contain a pair at distance at most r.
Why?
Otherwise, for every non-empty cell C, check all pairs of points in C.
This takes time $O(n)$, since every cell has $O(1)$ points.

The algorithm

Repeat the following $O(\log n)$ times:
Compute a random partition using a randomly shifted grid.
There are at most n non-empty cells (since there are n points).
If there exists a non-empty cell with at least 100 points, then it must contain a pair at distance at most r.
Why?
Otherwise, for every non-empty cell C, check all pairs of points in C.
This takes time $O(n)$, since every cell has $O(1)$ points.
If after $O(\log n)$ repetitions no pair is found, then output NO.

Analysis

Suppose that there exist $x \neq y \in X$, with $\|x - y\|_2 \leq r$.

Analysis

Suppose that there exist $x \neq y \in X$, with $\|x - y\|_2 \leq r$.
Then, at every iteration, there exists a cell that contains both x and y with probability at least $1/2$.

Analysis

Suppose that there exist $x \neq y \in X$, with $\|x - y\|_2 \leq r$.
Then, at every iteration, there exists a cell that contains both x and y with probability at least $1/2$.
Therefore, after k iterations, we find the pair x, y with probability at least $1 - 2^{-k}$.

Analysis

Suppose that there exist $x \neq y \in X$, with $\|x - y\|_2 \leq r$.
Then, at every iteration, there exists a cell that contains both x and y with probability at least $1/2$.
Therefore, after k iterations, we find the pair x, y with probability at least $1 - 2^{-k}$.
Thus, after $O(\log n)$ iterations, we find x, y with high probability.

Analysis

Suppose that there exist $x \neq y \in X$, with $\|x - y\|_2 \leq r$.
Then, at every iteration, there exists a cell that contains both x and y with probability at least $1/2$.
Therefore, after k iterations, we find the pair x, y with probability at least $1 - 2^{-k}$.
Thus, after $O(\log n)$ iterations, we find x, y with high probability.

Running time: $O(n \log n)$.

Random partitions

Let (X, ρ) be a metric space, and let $r > 0$.

Random partitions

Let (X, ρ) be a metric space, and let $r > 0$.

A *r-partition* of (X, ρ) is a partition P of X into *clusters* (i.e. subsets) of diameter at most r.

Random partitions

Let (X, ρ) be a metric space, and let $r > 0$.

A *r-partition* of (X, ρ) is a partition P of X into *clusters* (i.e. subsets) of diameter at most r.
For any $x \in X$, we denote by $P(x)$ the cluster containing x.

Random partitions

Let (X, ρ) be a metric space, and let $r > 0$.

A *r-partition* of (X, ρ) is a partition P of X into *clusters* (i.e. subsets) of diameter at most r.
For any $x \in X$, we denote by $P(x)$ the cluster containing x.

Let $\beta > 0$.
A *(β, r)-Lipschitz partition* of (X, ρ) is a distribution $\mathcal{D}$ over r-partitions of (X, ρ), such that for any $x, y \in X$:

$$\Pr_{P \in \mathcal{D}}[P(x) \neq P(y)] \leq \beta \cdot \frac{\rho(x, y)}{r}$$

Examples of random partitions

For any $r > 0$, the space $\mathbb{R}^1$ admits a $(1, r)$-Lipschitz partition.

Examples of random partitions

For any $r > 0$, the space $\mathbb{R}^1$ admits a $(1, r)$-Lipschitz partition.

For any $r > 0$, the space $\mathbb{R}^2$ admits a $(O(1), r)$-Lipschitz partition.

Examples of random partitions

For any $r > 0$, the space $\mathbb{R}^1$ admits a $(1, r)$-Lipschitz partition.

For any $r > 0$, the space $\mathbb{R}^2$ admits a $(O(1), r)$-Lipschitz partition.

For any $r > 0$, the space $\mathbb{R}^d$ admits a $(O(\sqrt{d}), r)$-Lipschitz partition.

Examples of random partitions

For any $r > 0$, the space $\mathbb{R}^1$ admits a $(1, r)$-Lipschitz partition.

For any $r > 0$, the space $\mathbb{R}^2$ admits a $(O(1), r)$-Lipschitz partition.

For any $r > 0$, the space $\mathbb{R}^d$ admits a $(O(\sqrt{d}), r)$-Lipschitz partition.

What about general metric spaces?

Random partitions for general metric spaces

Let (X, ρ) be a metric space.

Random partitions for general metric spaces

Let (X, ρ) be a metric space.
Let $X = \{x_1, \ldots, x_n\}$.

Random partitions for general metric spaces

Let (X, ρ) be a metric space.
Let $X = \{x_1, \ldots, x_n\}$.
Pick a random permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$.

Random partitions for general metric spaces

Let (X, ρ) be a metric space.
Let $X = \{x_1, \ldots, x_n\}$.
Pick a random permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$.
Pick $\alpha \in [1/2, 1)$, uniformly at random.

Random partitions for general metric spaces

Let (X, ρ) be a metric space.
Let $X = \{x_1, \ldots, x_n\}$.
Pick a random permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$.
Pick $\alpha \in [1/2, 1)$, uniformly at random.
For any $i \in \{1, \ldots, n\}$, let

$$C_i = \mathrm{Ball}(x_{\sigma(i)}, \alpha \cdot r/2) \setminus \left(\bigcup_{j=1}^{i-1} C_j \right),$$

where

$$\mathrm{Ball}(x, t) = \{y \in X : \|x - y\|_2 \leq t\}.$$

Random partitions for general metric spaces

Let (X, ρ) be a metric space.
Let $X = \{x_1, \ldots, x_n\}$.
Pick a random permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$.
Pick $\alpha \in [1/2, 1)$, uniformly at random.
For any $i \in \{1, \ldots, n\}$, let

$$C_i = \mathrm{Ball}(x_{\sigma(i)}, \alpha \cdot r/2) \setminus \left(\bigcup_{j=1}^{i-1} C_j \right),$$

where

$$\mathrm{Ball}(x, t) = \{y \in X : \|x - y\|_2 \leq t\}.$$

Let $P = \{C_1, \ldots, C_n\}$ be the resulting random partition of X.

Analysis

Lemma

P is an r-partition (with probability 1).

Analysis

Lemma

P is an r-partition (with probability 1).

Proof.

Each cluster is contained inside some $\mathrm{Ball}(x_i, r/2)$. Therefore, it has diameter at most r. □

Analysis

Lemma

The resulting distribution is a $(O(\log n), r)$*-Lipschitz partition.*

Analysis

Lemma

The resulting distribution is a $(O(\log n), r)$*-Lipschitz partition.*

Proof: Fix $x, y \in X$.

Analysis

Lemma

The resulting distribution is a $(O(\log n), r)$*-Lipschitz partition.*

Proof: Fix $x, y \in X$.
We say that x_i *settles* $\{x, y\}$ if x_i is the first point w.r.to σ such that $\text{Ball}(x_i, r/2) \cap \{x, y\} \neq \emptyset$.

Analysis

Lemma

The resulting distribution is a $(O(\log n), r)$-Lipschitz partition.

Proof: Fix $x, y \in X$.
We say that x_i *settles* $\{x, y\}$ if x_i is the first point w.r.to σ such that $\text{Ball}(x_i, r/2) \cap \{x, y\} \neq \emptyset$.
Assume after reordering X, that

$$\rho(x_1, \{x, y\}) \leq \rho(x_2, \{x, y\}) \leq \ldots \leq \rho(x_n, \{x, y\}),$$

where $\rho(z, \{x, y\}) = \min\{\rho(z, x), \rho(z, y)\}$.

Proof (cont.)

Let $s \in \{1, \ldots, n\}$.

Proof (cont.)

Let $s \in \{1, \ldots, n\}$.
Let

$$I_s = [\min\{\rho(x_s, x), \rho(x_s, y)\}, \max\{\rho(x_s, x), \rho(x_s, y)\}).$$

Proof (cont.)

Let $s \in \{1, \ldots, n\}$.
Let

$$I_s = [\min\{\rho(x_s, x), \rho(x_s, y)\}, \max\{\rho(x_s, x), \rho(x_s, y)\}).$$

In order for $P(x) \neq P(y)$ when x_s settles $\{x, y\}$, it must be that

$$\alpha \cdot r/2 \in I_s.$$

Proof (cont.)

$$
\begin{aligned}
\Pr[P(x) \neq P(y)] &\leq \sum_{s=1}^{n} \Pr[P(x) \neq P(y) \text{ and } x_s \text{ settles } \{x, y\}] \\
&\leq \sum_{s=1}^{n} \Pr[\alpha \cdot \rho/2 \in I_s \text{ and } x_s \text{ settles } \{x, y\}] \\
&\leq \sum_{s=1}^{n} \Pr[x_s \text{ settles } \{x, y\} | \alpha \cdot \rho/2 \in I_s] \cdot \Pr[\alpha \cdot \rho/2 \in I_s] \\
&\leq \sum_{s=1}^{n} \frac{1}{s} \cdot \frac{4 \cdot |\rho(x_s, x) - \rho(x_s, y)|}{r} \\
&\leq \sum_{s=1}^{n} \frac{1}{s} \cdot \frac{4\rho(x, y)}{r} \\
&\leq O(\log n) \cdot \frac{\rho(x, y)}{r}
\end{aligned}
$$

We obtain the following:

Theorem ([Bartal '96])

For any $r > 0$, any n-point metric space admits a $(O(\log n), r)$-Lipschitz partition.