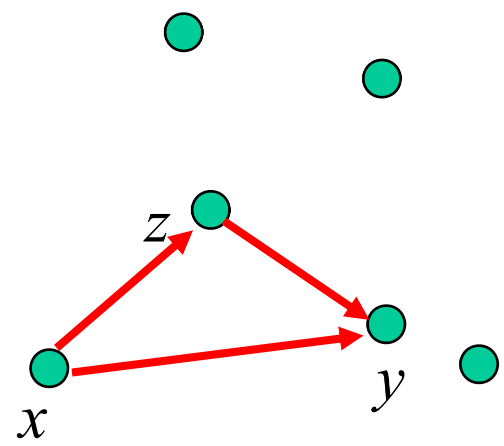
Approximating TSP with Neighborhoods in Doubling Metrics

Hubert Chan

University of Hong Kong

What is a Metric Space?

Points V with distance function d



Examples:

- Distances between cities
- Round trip delays between internet hosts
- Dissimilarity measures between documents

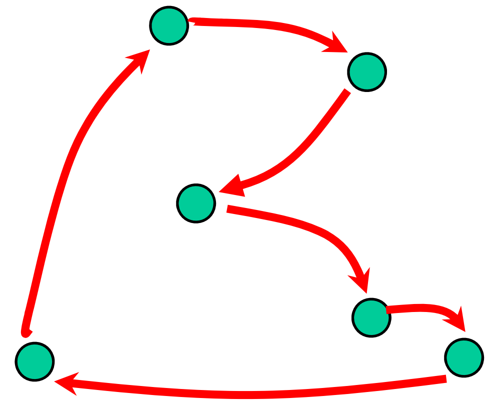
Simplifying Assumptions:

1. Triangle Inequality: $d(x,y) \leq d(x,z) + d(z,y)$

2. Symmetry: $d(x,y) = d(y,x)$

Traveling Salesman Problem

Traveling Salesman Problem: What is the shortest tour that visits each city once?



- Classical NP-complete Problem
- Application in circuit design, logistics
- Practical instances are solved routinely

Important Question:
Which metrics admit good algorithmic guarantees?

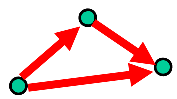
Approximating TSP on Different Metric Spaces

General

General distance function

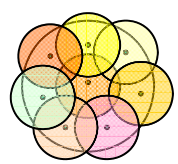
NP-hard to approx within any factor

Metrics



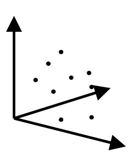
1.5-approx
NP-hard to approx better than 174/173

Doubling Metrics



$(1 + \varepsilon)$-approx in time [Talwar]
$\exp\{ (k/\varepsilon \log n)^{O(k)}\}$ (QPTAS)

k-Dim Euclidean Metrics



$(1 + \varepsilon)$-approx in time [Arora][Rao, Smith]
$n \exp\{(k/\varepsilon)^{O(k)}\} + O(kn \log n)$ (PTAS)

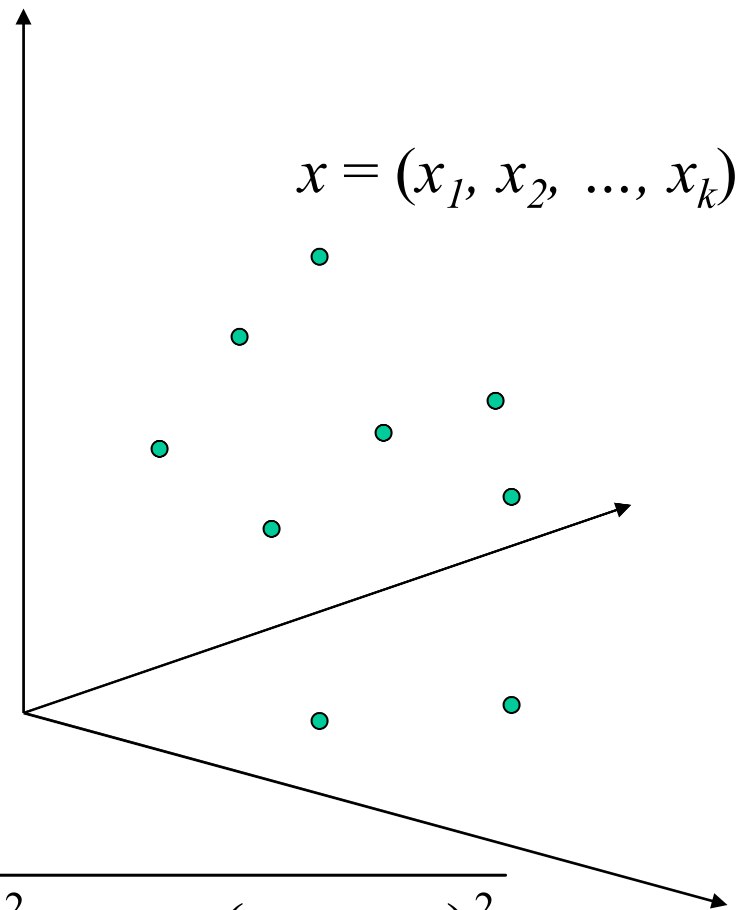
Specific

Roadmap

- ✓ TSP on Metric Spaces
  - ✓ Hardness and Approximation
- Special Classes of Metric Spaces
  - Euclidean Metrics and Doubling Dimension
- General Framework for Approximating TSP
  - Divide and Conquer
- TSP with Neighborhoods

Low Dim Euclidean Metrics

Nodes in *k*-dimensional space



Each node has *k* coordinates.

Distance function is the usual Euclidean distance.

$$x = (x_1, x_2, \ldots, x_k)$$

$$y = (y_1, y_2, \ldots, y_k)$$

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_k - y_k)^2}$$

Generalization: Metric Spaces with Low **Doubling Dimension**

Doubling Dimension

Generalization of Euclidean Metrics
A low-dim Euclidean metric has small doubling dim.

[Clarkson ’99] used the notion for nearest neighbor queries.

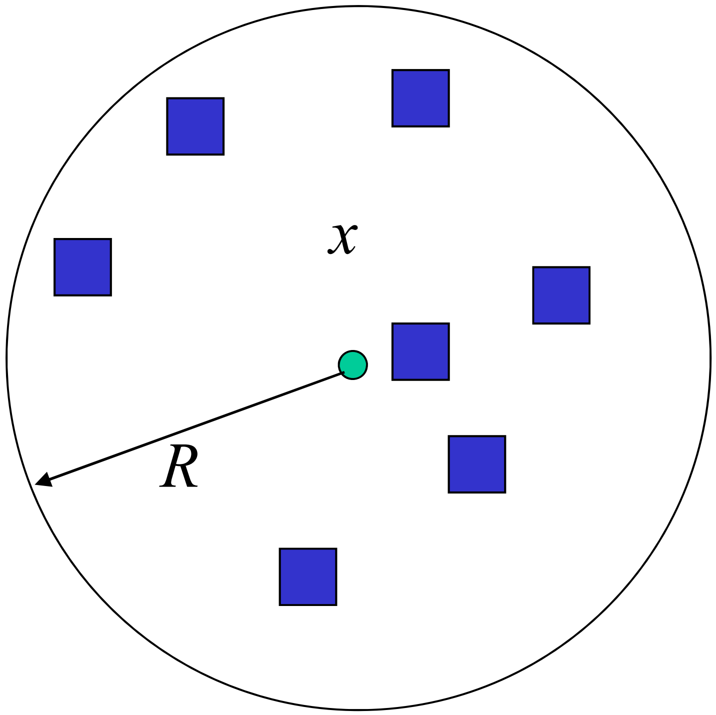
Received recent attention in CS community:
[Gupta, Krauthgamer, Lee 2003]

Hard problems more tractable: Quasi-polynomial time approximations for TSP, k-median, facility location
[Talwar 2004]

More good algorithms for near-neighbor
[Krauthgamer Lee 05] [Beygelzimer, Kakade, Langford ’06]

Ball $B(x, R)$

A ball $B(x, R)$ centered at x with radius R consists of points within distance R from x.

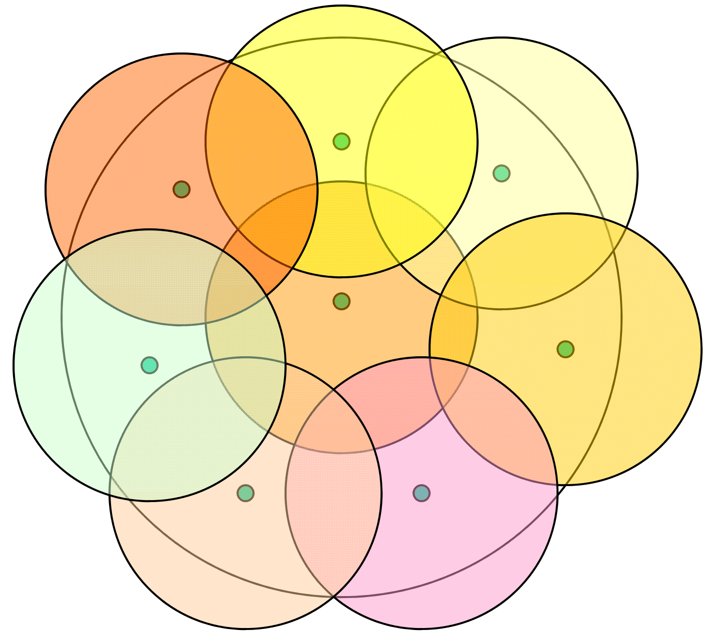


Doubling Dimension

A metric (V,d) has doubling dimension at most k if

for any $R > 0$, every ball of radius $2R$

is a union of at most 2^k balls of radius R.

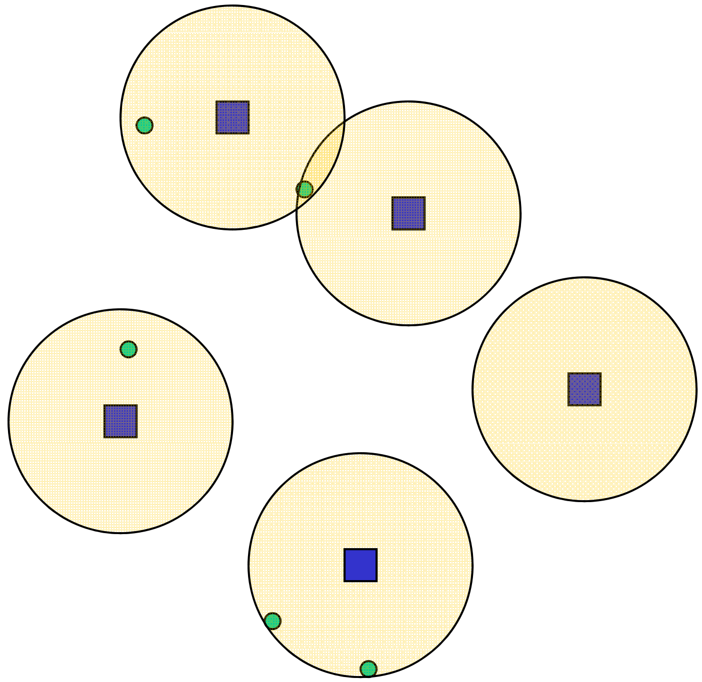


Examples:
A metric space in k-dim Euclidean space or k-dim manifold has doubling dim $O(k)$.

R-Net

Radius $R > 0$

An R-net for V is a subset $N \subseteq V$ s.t.

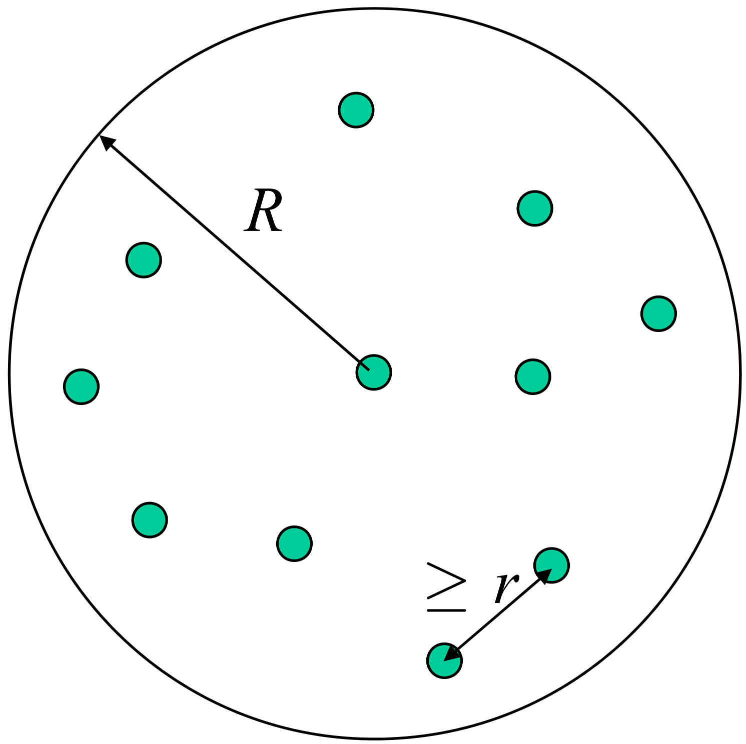


1. **Covering**: Every point in V is within distance R of some point in N.
2. **Packing**: Points in N are more than distance R away from one another.

R-nets & doubling dimension

Useful Property:

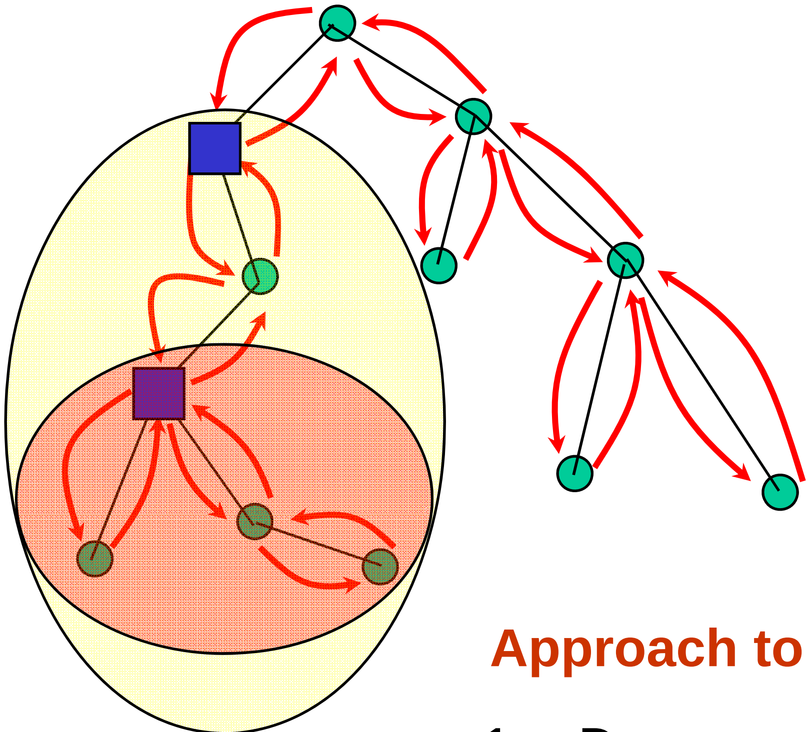
Given a metric (V, d) with **doubling dimension** k and any r-net N, any ball of radius R contains at most $(2R/r)^k$ net points in N.



Roadmap

- ✓ TSP on Metric Spaces
  - ✓ Hardness and Approximation
- ✓ Special Classes of Metric Spaces
  - ✓ Euclidean Metrics and Doubling Dimension
- General Framework for Approximating TSP
  - Divide and Conquer
- TSP with Neighborhoods

Easy Instances of TSP



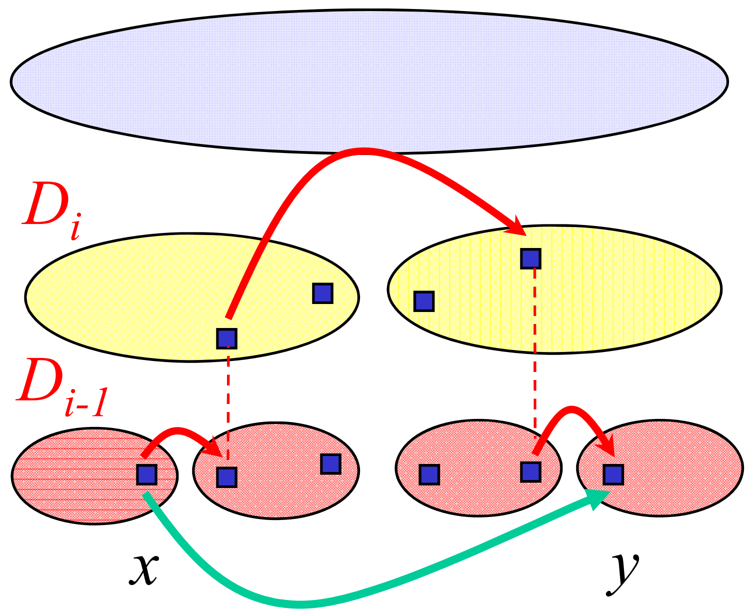
Optimal Tour for Tree Metric

- Tour enters and leaves subtree through a single point
- True for smaller subtrees too.

Approach to approximate TSP in general:

1. Decompose metric recursively into clusters
2. Assign few points in each cluster as portals
3. Restrict to tour that enters and leaves clusters via portals (“portal respecting”)

General Framework for TSP [Arora, Talwar]



1. Randomized Hierarchical decomposition of metric into "clusters"

   Level i cluster diameter D_i such that $D_{i-1} \leq D_i/4$

2. Assign portals to each cluster (some appropriate net)

3. Show existence of a "good" portal respecting tour

4. Dynamic Program to find best portal-respecting tour.
   B = # portals in child clusters $\Rightarrow$ Run-time = $2^{O(B \log B)}$

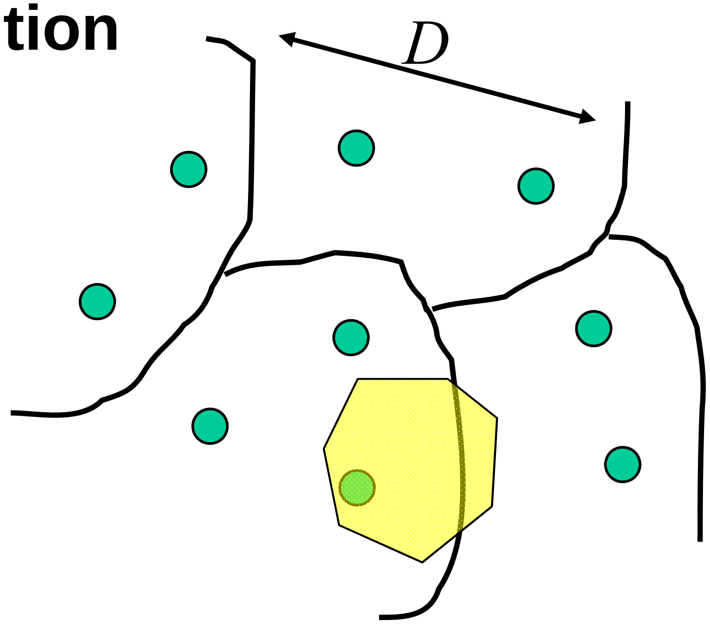
Doubling metric:
$B = (\log n)^{O(k)}$

How to Divide? - Padded Decomposition

***D*-Bounded *β*-Padded Decomposition**

Random partition of (V,d) s.t.

1. Each cluster has diameter at most D.
2. If a set S has diameter δ, Pr[S separated] $\leq \beta \, \delta \, / \, D$.



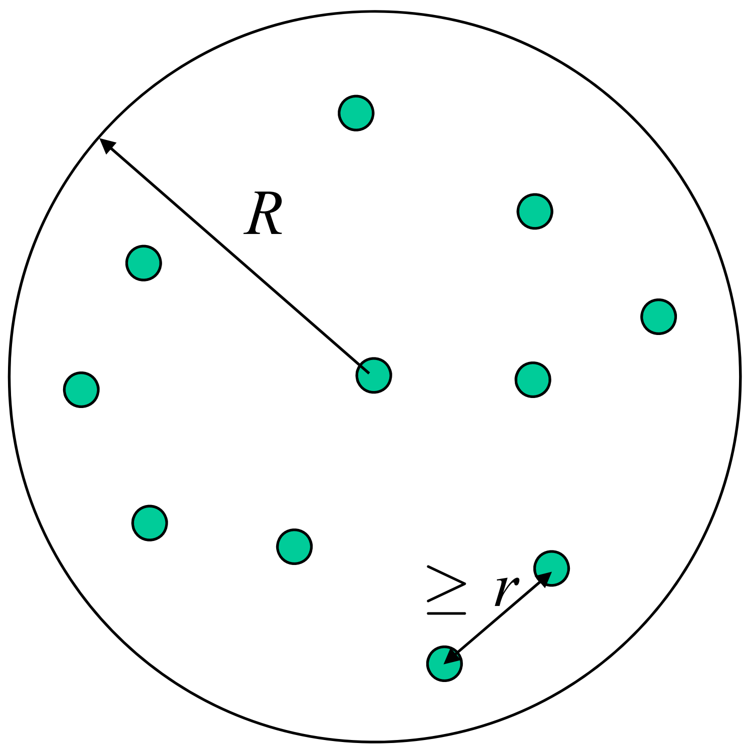
Theorem

For any D, a metric with doubling dimension k has D-bounded O(k)-padded decomposition.

How to choose portals? - R-nets

Useful Property:

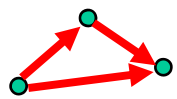
Given a metric (V, d) with **doubling dimension** k and any r-net N, any ball of radius R contains at most $(2R/r)^k$ net points in N.



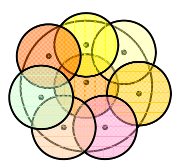
Approximating TSP on Different Metric Spaces

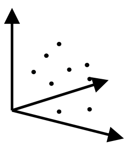
General

General distance function — NP-hard to approx within any factor

Metrics  — 1.5-approx; NP-hard to approx better than 174/173

[SODA'08 C., Gupta] $(1+\varepsilon)$-approx in sub-exp time $\cap_{\delta>0} \exp\{n^{\delta}\}$

Doubling Metrics  — $(1+\varepsilon)$-approx in time $\exp\{(k/\varepsilon \log n)^{O(k)}\}$ [Talwar] (QPTAS)

k-Dim Euclidean Metrics  — $(1+\varepsilon)$-approx in time $n \exp\{(k/\varepsilon)^{O(k)}\} + O(kn \log n)$ [Arora][Rao, Smith] (PTAS)

Specific

Roadmap

✓ TSP on Metric Spaces

  ✓ Hardness and Approximation

✓ Special Classes of Metric Spaces

  ✓ Euclidean Metrics and Doubling Dimension

✓ General Framework for Approximating TSP

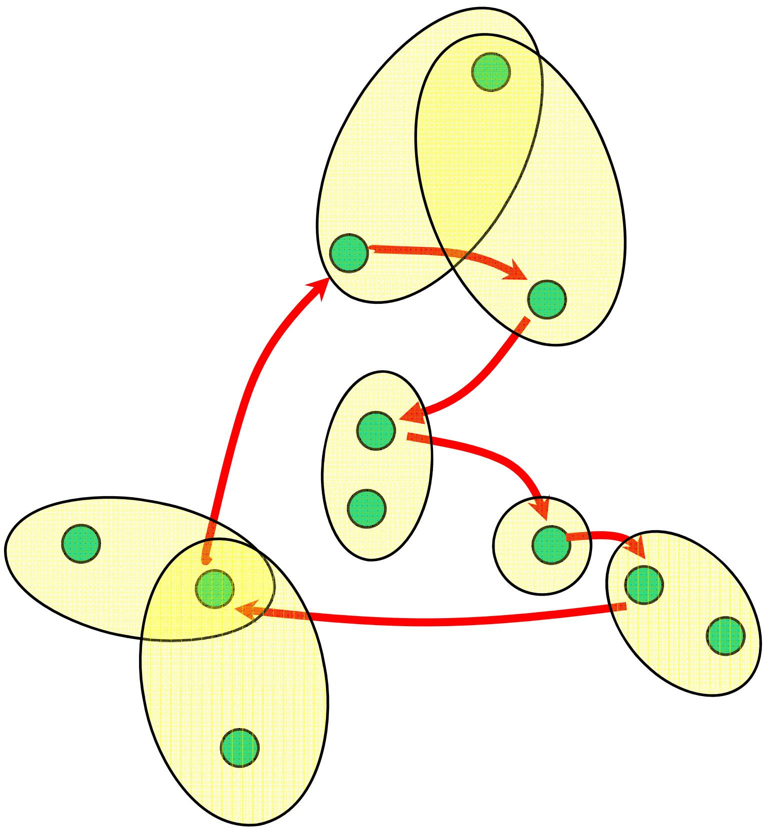
  ✓ Divide and Conquer

- TSP with Neighborhoods

Motivation

1. You have a list of items and the shops where each item can be found. What is the shortest tour for buying every item?

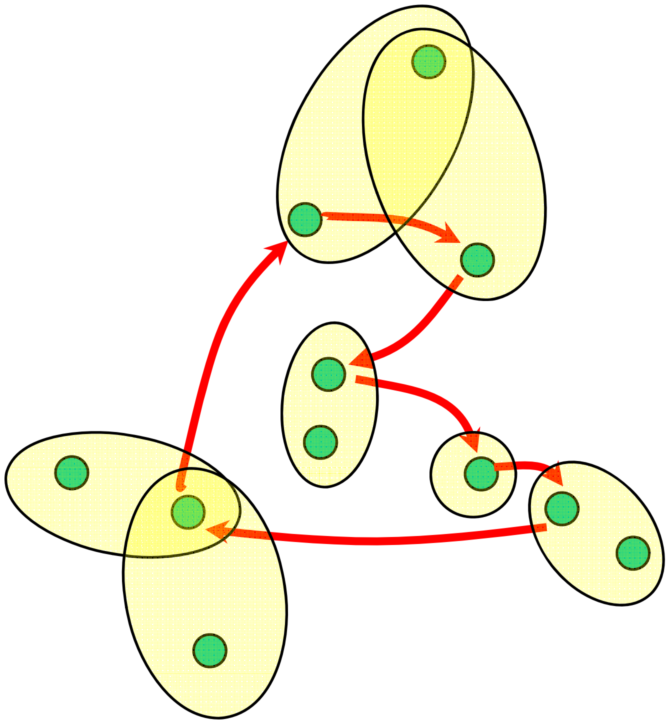
2. There are outbreaks of several viruses. What is the shortest tour to collect a sample for each virus?



Problem Definition

Input: a metric space (V, d) and a collection of subsets (a.k.a regions or neighborhoods) R_1, R_2, …, R_n in V.

Output: a tour with shortest length that visits each neighborhood R_i at least once.



General Version is Hard

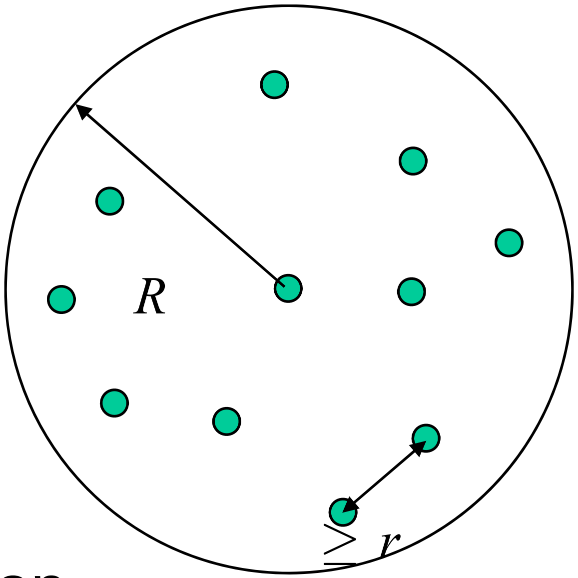
1. As hard as the classical **Traveling Salesman Problem** (TSP), which is **APX**-hard for general Euclidean metrics.
2. Generalizes **Set Cover** and **Hitting Set**, which is Θ**(log n)**-hard to approximate.

Lower Bound [Halperin, Krauthgamer '03]
Inapproximability threshold: $\Omega(\log^{2-\epsilon} n)$

Upper Bound [GKR00 + FRT04]
$O(\log N \log k \log n)$
n = # of regions
N = # of points
k = # of points in a region

Special Cases (1)

The underlying metric has bounded ***doubling dimension***: a packing inside a bounded subset has a limited number of points.

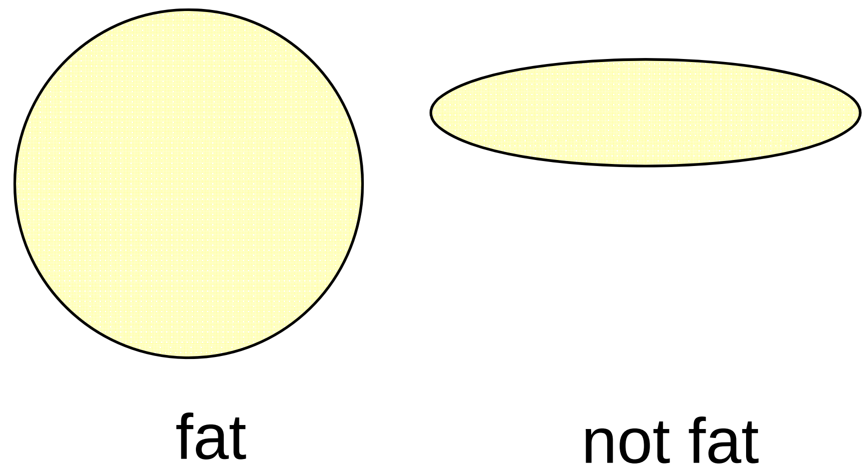


TSP is APX-hard without this assumption.

The very particular case of Euclidean plane is often considered.

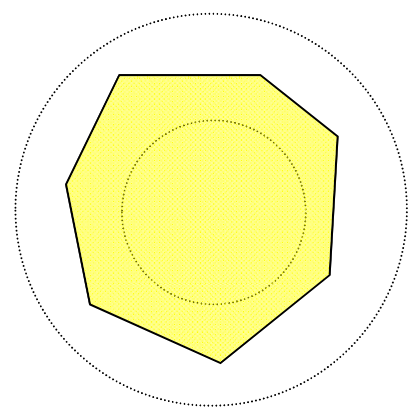
Special Cases (2)

The regions are "fat".



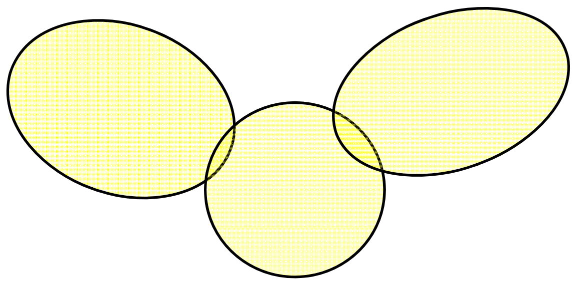
For $\alpha \geq 1$, region R is α-fat if there exist a point x and $r > 0$ s.t.

$$B(x, r) \subseteq R \subseteq B(x, \alpha\, r)$$

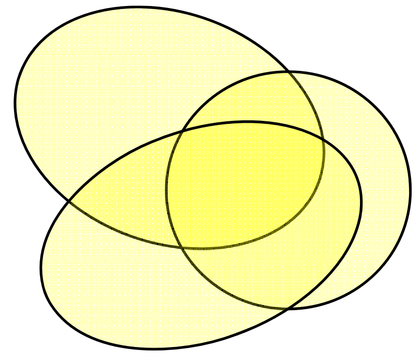


Special Cases (3)

The regions have limited intersection.

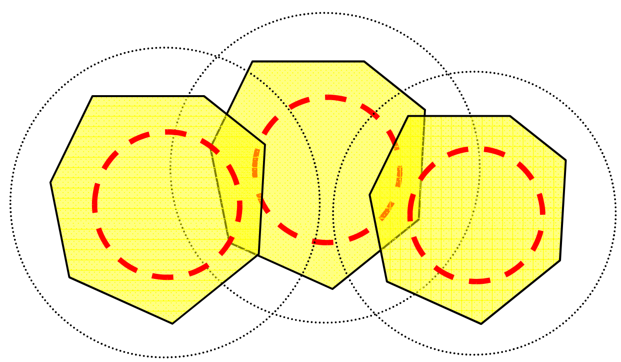


weakly disjoint



arbitrary intersection

Formally, related to α-fat regions. The “cores” do not intersect.



Some Results

(1) Euclidean Plane
(2) Fat Regions
(3) Weakly Disjoint Regions
(4) Regions of Similar Size

	Assumptions	Approx Ratio
DM03	(1)-(4)	PTAS
dBGK05	(1),(2),(3)	O(1)
EFS06	(1),(2),(4)	O(1)

Best Previous Result

Mitchell (SODA ’07)
PTAS for Euclidean plane, fat and weakly disjoint regions (assumptions (1)-(3))

Techniques

1. Guillotine subdivision
2. Only applies to Euclidean plane, would not work even for 3 dimensions.

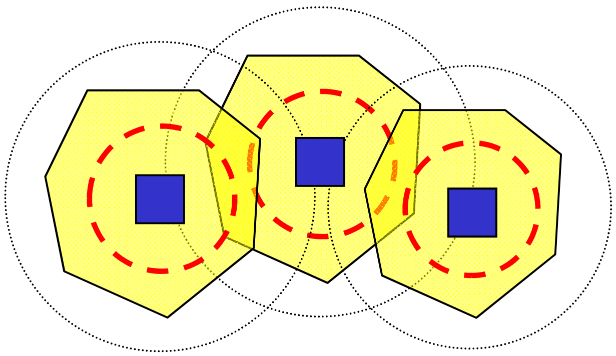
Our Contribution

- More general underlying metric space (with bounded doubling dimension)
- Combining and generalizing the notion of fatness and disjointness

A group of regions $\{R_j\}$ is α-fat weakly disjoint if there exist $r > 0$ and for each R_j, a point z_j s.t.

(1) $\{z_j\}$ is an r-packing, i.e., any 2 such points are at least distance r apart.

(2) Each R_j is contained in $B(z_j, \alpha\, r)$.



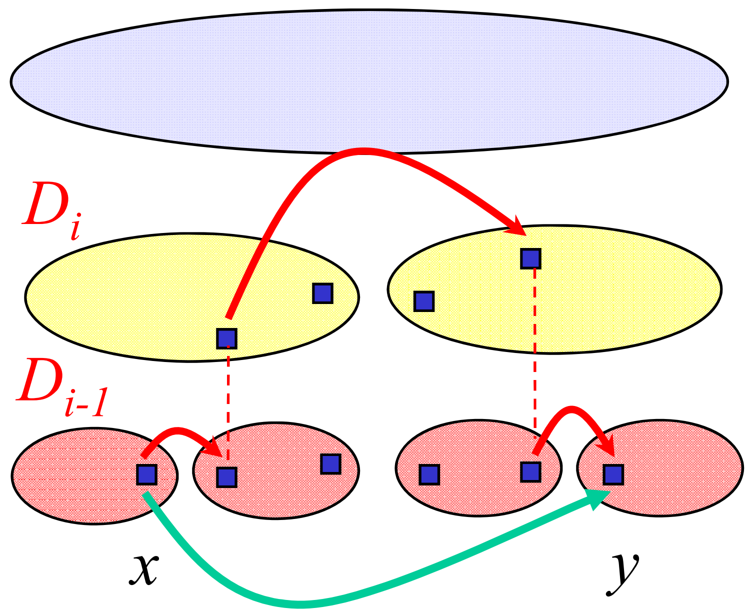
Our Result

Theorem **[QPTAS for TSPN. C., Elbassioni SODA'10]**
For metric space with doubling dimension k,
Δ groups of α-fat weakly disjoint regions, we have (1+ϵ)-approx in time $\exp((\Delta/\epsilon)^k \, O(\alpha)^{k^2} \log^k n)$, where n is the total number of regions.

Remark
We have weakened assumptions (1)-(4).
If we do not bound the number Δ of groups, the problem remains APX-hard.

Techniques

General Framework for TSP [Arora, Talwar]



1. Randomized Hierarchical decomposition of metric into "clusters"

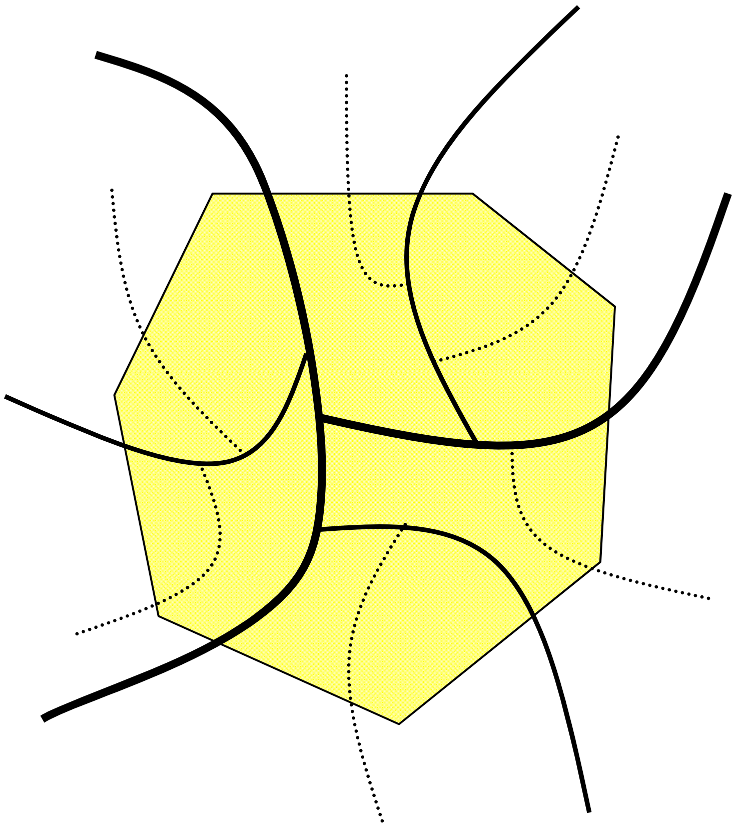
   Level i cluster diameter D_i such that $D_{i-1} \leq D_i/4$

2. Assign portals to each cluster (some appropriate net)
3. Show existence of a "good" portal respecting tour
4. Dynamic Program to find best portal-respecting tour.
   B = # portals in child clusters $\Rightarrow$ Run-time = $2^{O(B \log B)}$

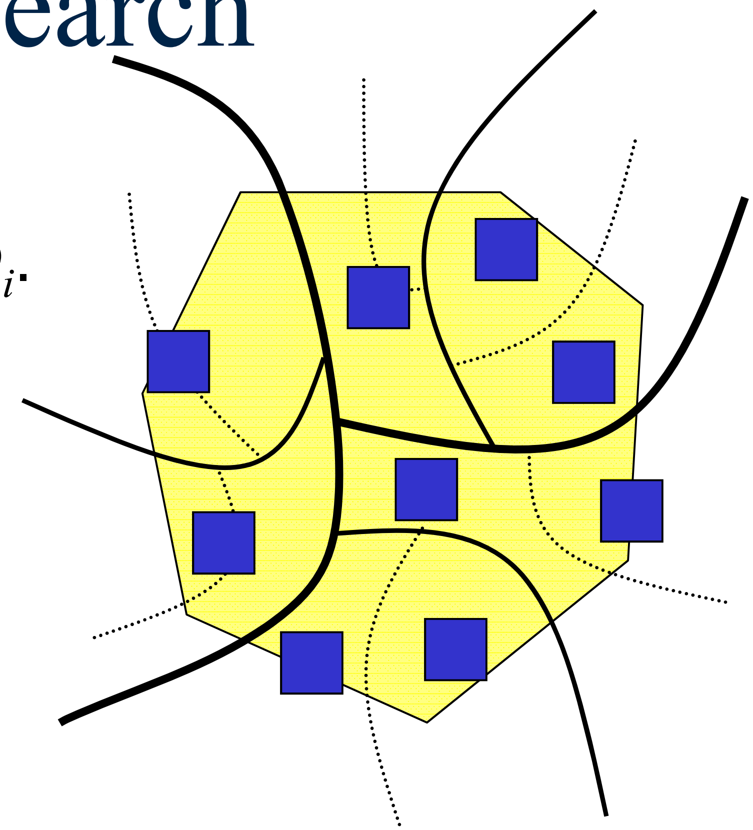
Technical Hurdle

When a region is divided, which part should be visited?

Each part is further subdivided recursively, leading to exponential number of cases to be considered.



Pruning the Search



If a set S has diameter δ,
$\Pr[S$ first sep. at level $i] \le \beta\, \delta / D_i$.

For some small $\gamma > 0$, when the sub-region gets smaller than $\gamma\, D_i$, pick any point and stop further partitioning the subregion.

If there are L levels, the expected increase in cost is at most $\sum_i \beta\, \delta / D_i \cdot \gamma\, D_i = L\, \beta\, \gamma\, \delta$

Summing over all regions, we have $L\beta\gamma \times$ Sum of Diameters

Structural Lemma

Lemma

If there are Δ groups of regions, then

Sum of Diameters $\leq \Delta \, O(\alpha)^k$ OPT

Picking the pruning parameter γ appropriately, we can show the pruning procedure increases the cost by at most ϵ OPT.

***Theorem* [QPTAS for TSPN]**

For metric space with doubling dimension k, Δ groups of α-fat weakly disjoint regions, we have (1+ϵ)-approx in time $\exp((\Delta/\epsilon)^k \, O(\alpha)^{k^2} \log^k n)$, where n is the total number of regions.

Open Problems

Is there a PTAS for the case when the underlying metric is Euclidean (with appropriate assumptions on the regions)?

Note that a PTAS is not known for TSP on doubling metrics.