Algorithmic Applications of Metrics Embeddings 1
Introduction to Spanners

R. Ravi

ravi@cmu.edu

Overview

- Metrics represent distances between nodes in a network
- Embeddings allow reducing problem to simpler metrics (e.g. trees)
- Informally, cost of embeddings is the amount of distortion in the pairwise distances => minimizing distortion is crucial
- Outline
  - Today: Introduction, Spanners (Embedding into subgraph metrics)
  - Tomorrow: Embedding into tree metrics
  - Day after: Tree covers for routing and approximate distance oracles

Definitions

- A metric is a distance function between a set of points X $d : X \times X \longrightarrow \mathbb{R}^+$

1. Identity $d(i,j) = \emptyset \iff i = j$
2. Symmetry $d(i,j) = d(j,i) \quad \forall\, i,j \in X$
3. Triangle inequality $d(i,j) \leq d(i,k) + d(k,j)$ $\forall\, i,j,k \in X$

Definitions

- Undirected graphs with nonnegative edge lengths naturally define a shortest path metric (a.k.a. the metric completion of the graph)

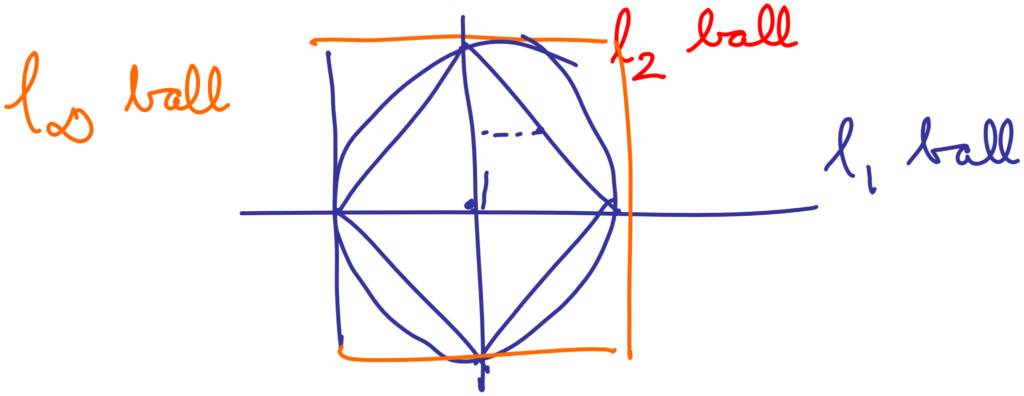
  - d(x,y) = length of the shortest path between x and y in the graph

Infinite metric spaces

- Real spaces R^k
- Minkowski norms l$_p$ for 1 · p · 1

$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{1/p}$$

- Picture of l$_p$ balls around the origin



Metric Embedding

$$d \xhookrightarrow{f} d'$$

- Map from one metric to another

$$f : (X, d) \longrightarrow (X', d')$$

- Contraction

$$\max_{x,y \in X} \frac{d(x,y)}{d'(f(x), f(y))}$$

- Expansion

$$\max_{x,y \in X} \frac{d'(f(x), f(y))}{d(x,y)}$$

- Distortion

$$\|f\| = \text{contraction}(f) \cdot \text{Expansion}(f).$$

- Alternate definition of distortion

Smallest D s.t.

$$\exists r: \quad r \cdot d(x,y) \leqslant d'(f(x), f(y)) \leqslant D\, r \cdot d(x,y)$$

Simplifying assumptions

- Most metrics (especially the guest metrics) will be finite n-point metrics
- Assume all distances are integers at least 1
- Mostly assume polynomial (in n) range of distance values

Algorithmic Applications of Metric Embeddings

- Problems where data is a metric
  - Embedding into a "simpler" metric where the problem may be tractable E.g., TSP, facility location
  - Distortion multiplies approximation or competitive ratio in approximation or online algorithms resply.
- Metric Relaxations
  - LP relaxation for a hard problem can be interpreted as a metric (triangle inequalities are linear); Embedding into simpler metric allows rounding
  - Distortion translates into approximation ratio

- Problems on metrics
  - Problem involves properties of metrics
  - E.g., Find a closest line/tree metric to given metric; Use properties of edit distances between sequences to find a good consensus sequence.

Important Results on Embeddings we will not cover

- [Bourgain '85, Marousek '95] Any n-point metric space embeds into l_p^k with distortion O(log n/p) [Or in O($\log^2$ n) dimensions with distortion O(log n)]
  - Proceeds via defining coordinates based on scaled distances from subsets of points (so-called Frechet-embeddings)
  - Applications to rounding LP relaxations for sparsest cut problems intrerpreted as metrics

- [Johnson-Lindenstrauss lemma ’84] Any n-point subset in (arbitrarily high) d-dimensional Euclidean space can be embedded into a Euclidean subspace of dimension O(log n/ε^2) with distortion (1 + ε).
  - Projection into random subspace of this dimension works
  - Applications to nearest neighbor searching, learning mixtures of Gaussians

Important Results on Embeddings we will cover

- Spanners or subgraph embeddings: Embed a weighted graph into its own subgraph to
  - Approximately preserve distances
  - Keep the subgraph sparse/light

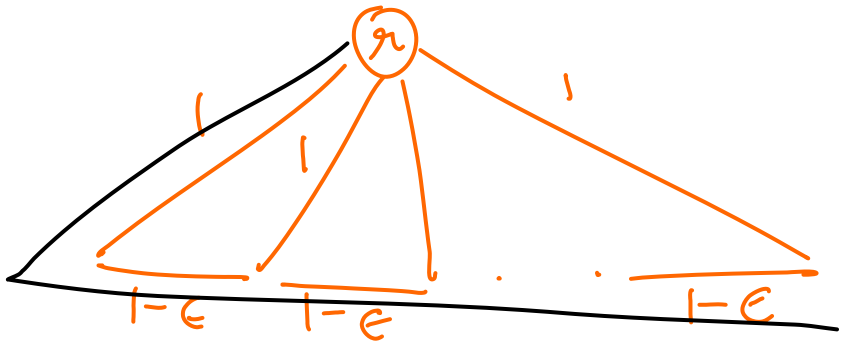
  We will cover spanner constructions with and without a root
- Embeddings into a distribution of trees
  - Any n-point metric embeds into a polynomially computable distribution of tree metrics with expected distortion Θ(log n)
  - Applications to online and approximation algorithms

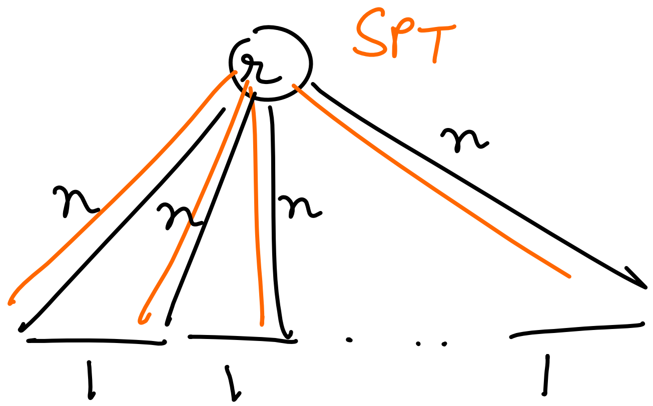
  We will discuss this construction tomorrow
- Distance oracles and compact routing schemes
  - Can associate O(kn$^{1/k}$) labels with every node that can be used to recover paths with distortion at most (2k -1) between all pairs
  - Applications to constructing tree covers for network routing

  We will discuss this the day after tomorrow

Spanners: Motivation

- Given an undirected graph with edge lengths, a spanner is a lightweight, distance preserving subgraph
- Rooted case: Given a root r, preserving shortest distances from it (SPT) and finding a minimum length tree (MST) are often conflicting goals





LASTs

- Given an undirected graph with edge lengths and a root r, an (a,b)-LAST or Light Approximate Shortest path Tree is a tree T such that
  - (distance guarantee) $\forall v,$ $d_T(r,v) \leqslant a \cdot d_G(r,v)$
  - (weight guarantee) $\sum_{e \in T} d_e \leqslant b \cdot d(MST)$
- Theorem [KRY'93 foll. ABP '91]

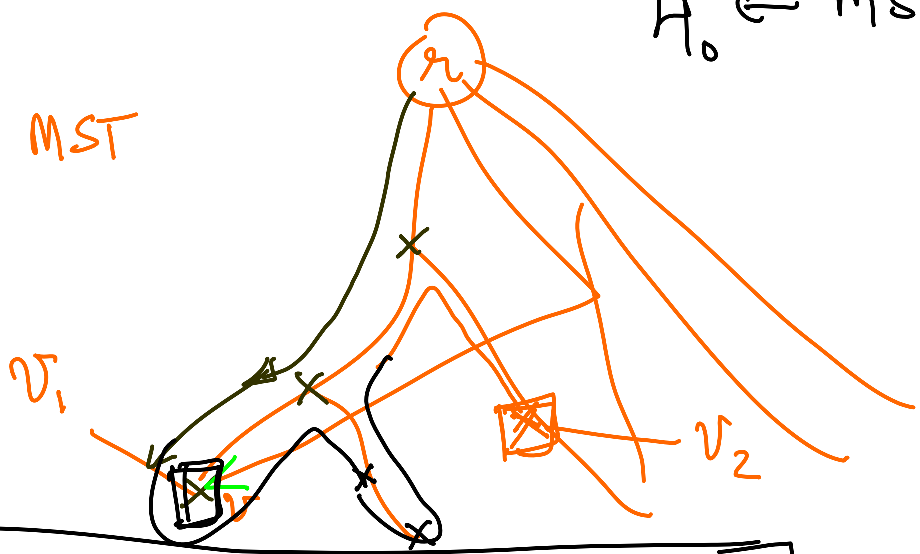
$$\forall \alpha > 1, \; \exists \; \left(\alpha, 1 + \frac{2}{\alpha - 1}\right)\text{-LAST}$$

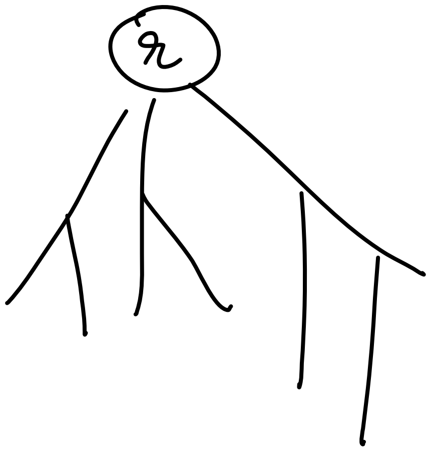
LAST Construction

- Greedy Addition Algorithm

LAST Construction Example

$H_0 \leftarrow MST$





If $d_{curr}(r,v) > \alpha\, d_G(r,v)$

add $SP(r,v)$

Curr graph $= H_i$

milestones v_i

Added $\sum_i d_{sp}(r, v_i)$

LAST Analysis

$\alpha > 1$

- (distance guarantee)

Holds by construction

- (weight guarantee)

$$d_{\underline{curr}}(r, v_i) > \alpha\, d_{sp}(r, v_i)$$

$$\sum_{i=1}^{k} \left[d_{sp}(r, v_{i-1}) + d_{walk}(v_{i-1}, v_i) > \alpha\, d_{sp}(r, v_i) \right]$$

$$2d(MST) \geq \sum_i d_{Walk}(v_{i-1}, v_i) > (\alpha - 1) \sum_{i=1}^{k-1} d_{sp}(r, v_i) + \alpha d_{sp}(r, v_k)$$

$$> (\alpha - 1) \sum_i d_{sp}(r, v_i)$$

$$\Rightarrow \quad \sum_i d_{sp}(r, v_i) < \frac{2}{\alpha - 1} d(MST)$$

Questions

Given an rooted undirected graph with edge lengths, is it NP-hard or easy to find

- A shortest path tree rooted at r of minimum total length
- A minimum length spanning tree that has the shortest maximum distance from the root to any vertex
- A minimum length spanning tree that has the shortest expansion of the distance from the root to any vertex

An application of LASTs – Cable Network Design

- Undirected graph with edge lengths l, root r, demands of flow to root f_v >0 for every vertex
- Must satisfy flows by building a cable network with cables of capacity u and cost c per unit length
- Lower bounds
  - Connectivity: All nodes must have a connection to the root
  - Routing: Every node's flow to the root must be on a path at least as long as the shortest path

Cable Network Design Approximation Algorithm

- Find an (a,b)-LAST T minimizing a+b
- Route every vertex's flow on T, comptuing the total flow f(e) on every tree edge e
- Install d f(e)/u e copies of the cable on e for all e

Analysis of approximation ratio

General Spanners

- Given an undirected graph with edge lengths, an (a,b)-spanner is a subgraph H such that
  - (distance guarantee)

    $$\forall x,y \quad d_H(x,y) \leq a \cdot d_G(x,y)$$

  - (weight guarantee)

    $$\sum_{e \in H} d_e \leq b \cdot d(MST)$$

- Theorem [CDNS '95]

$$\exists \left(\alpha, \left(3 + \frac{16\alpha}{\delta^2}\right) n^{\frac{2+\delta}{\alpha-1-\delta}}\right)$$

$$\forall \alpha - 1 > \delta > 0.$$

Eg: $\alpha = O(\lg n)$
$\delta = 2$
$\Rightarrow (O(\lg n), O(\lg n))$ sp
Spanner

$\alpha = O(\lg^2 n)$
$\delta = O(\lg n)$
$\Rightarrow (O(\lg^2 n), O(1))$ sp

Fix $\delta = 2$

- Background defn: The girth of an undirected graph is the length of shortest induced (chordless) cycle
- Lemma: For a graph with girth g,

$$\text{No. of edges of } G \cdot n \text{ d } n^{2/g-2} \text{ e}$$

Spanner construction

- Greedy Algorithm

- Observations
  - (distance guarantee) ✓
  - (MST inclusion)

Weight Analysis

- Partition non-MST edges of H by weight

$$H = E_{MST} \cup E_0 \cup E_1 \;.\; E_i \;\cup E_{lg n}$$

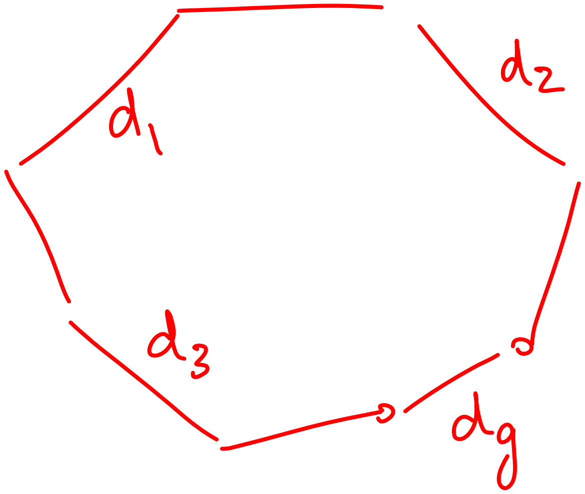
$$L = 2d(MST) \qquad l_e \in \left(\frac{2^{i-1}L}{n}, \frac{2^{i}L}{n}\right]$$

- Bound number of edges in each class using girth lemma

- Calculations for overall weight

Bounding the first class

Cycle $\in E_0$

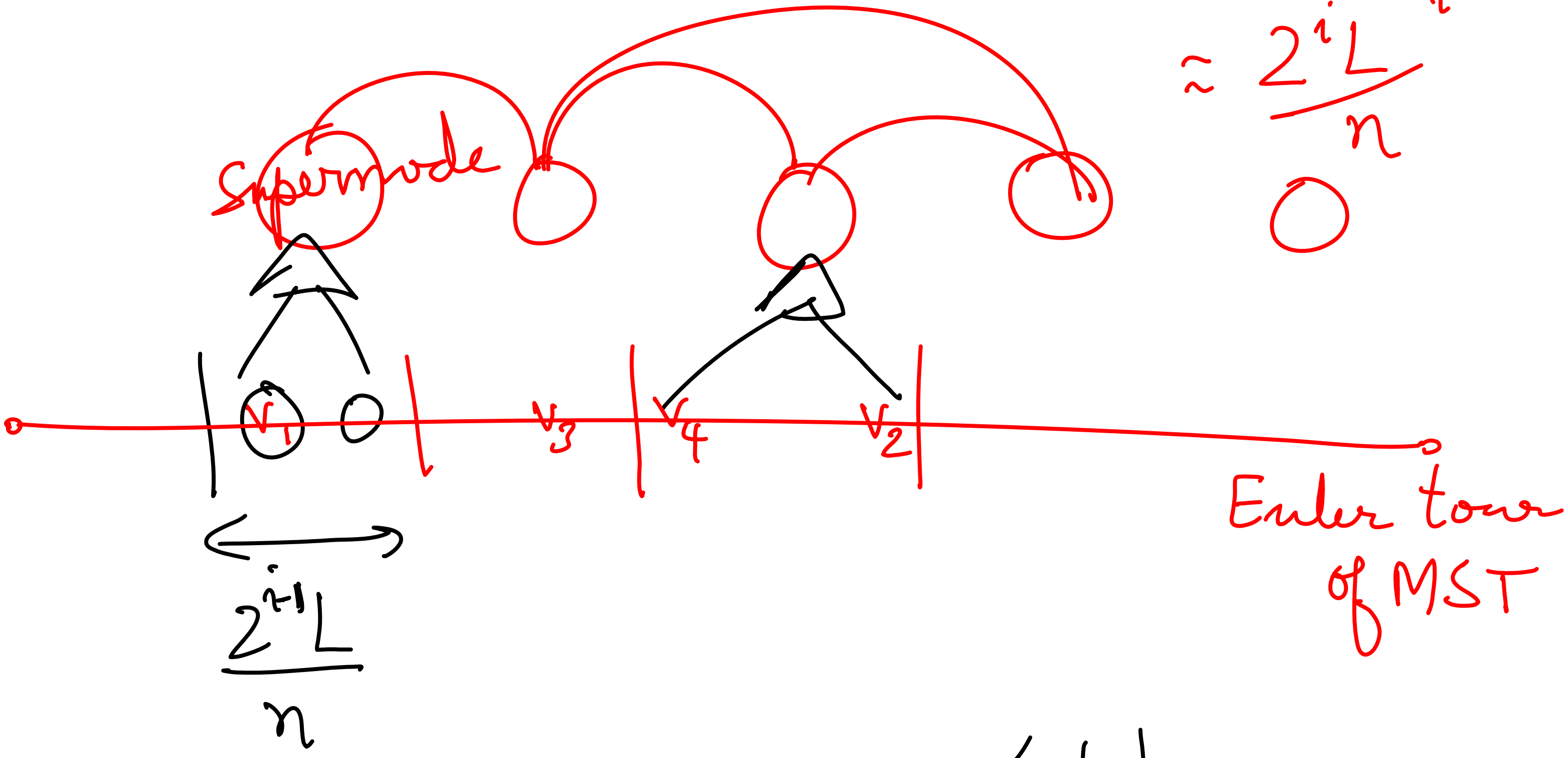


$d_1 \leq d_2 \ldots \leq d_g$

$$(g-1) \cdot d_g > \sum_{i=1}^{g-1} d_i > \alpha d_g \Rightarrow g > \alpha + 1$$

$$\Rightarrow |E_0| \leq n^{1+\frac{2}{g-2}} \leq n^{1+\frac{2}{\alpha-1}}$$

Bounding the size of a generic class E_i



$$\Rightarrow \quad g > \frac{\alpha+1}{2}$$

$$\Rightarrow |E_i| \le n_i^{1+\frac{2}{g-2}} \qquad n_i \longrightarrow \approx \frac{n}{2^i}$$

Bounding the length of a generic class

Bounding the total weight of all classes

Application of general spanners

- Cable network design application can be extended to arbitrary multi-commodity flow demand pairs
  - Setting appropriate values gives an O(log n) approximation [MP '96]

General questions

- Given an n-point metric, can you embed it into an l_1 metric isometrically? How many dimensions do you need?

General questions

- How much better can you do in the number of dimensions if the input metric was a tree metric (i.e., the metric value between any pair of points is the distance between them on a unique path in an underlying tree)? [Hint: Trees have a centroid decomposition, so aim for O(log n)]