Approximating Average Distortion of Embeddings into Line

Kedar Dhamdhere

Carnegie Mellon University

Joint work with Anupam Gupta, R. Ravi

Finite metric spaces

(V, d) is a finite metric space if

- V is a finite set of points.
- The *distance* function $d : V \times V \rightarrow R^+$ that satisfies:

$$
\begin{aligned}
d(x,x) &= 0 \\
d(x,y) &= d(y,x) && \text{symmetry} \\
d(x,y) &\leq d(x,z) + d(y,z) && \triangle \text{ inequality}
\end{aligned}
$$

  for all $x, y, z \in V$.
- Synonymous with
  - Graph $G = (V, E)$ with edge lengths. Distance given by shortest paths.

Low Distortion Embeddings

(V, d) and (V', d'): Finite metric spaces.
A *non-contracting embedding* is a map $f : V \rightarrow V'$.

- For any $x, y \in V$,

$$d(x, y) \leq d'(f(x), f(y)) \leq c \cdot d(x, y)$$

Parameter c is called *distortion*.

Average Distortion

(V, d) and (V', d'): finite metric spaces.

- Average Distance:

$$\text{av}(d) = \frac{1}{n^2} \sum_{x,y \in V} d(x, y)$$

$$\text{av}(d') = \frac{1}{n^2} \sum_{x,y \in V} d'(f(x), f(y))$$

- Average Distortion [Rab03]:

$$\text{av}(d')/\text{av}(d)$$

Average Distortion into line

- Introduced by Rabinovich [Rab03].
- Related to sparsest cut.

- For a contracting embedding into line
  - $O(1)$ bound on average distortion of planar graphs.
  - $O(\log r)$ distortion for treewidth-r graphs.

Our model

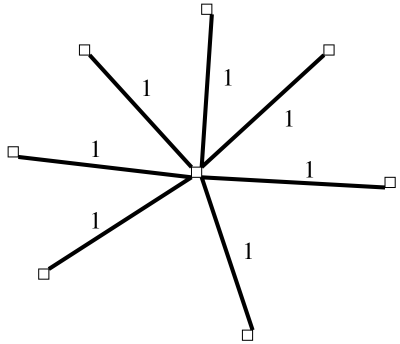
- Given finite metric (V, d).
- Host metric is the line (l_1^1, d').

Differences:

- *Non-contracting* $d'(x, y) \geq d(x, y)$ (for all $x, y \in V$).
- Find an embedding with *minimum average distortion*.

Simple lower bound on distortion

Consider a star on n nodes

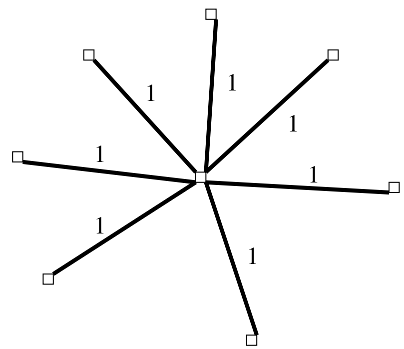


To embed into a line, in a non-contracting way

- Distortion = $\Omega(n)$, Average distortion = $\Omega(n)$
- Any embedding has distortion $O(n)$

Simple lower bound on distortion

Consider a star on n nodes



To embed into a line, in a non-contracting way

- Distortion = $\Omega(n)$, Average distortion = $\Omega(n)$
- Any embedding has distortion $O(n)$

Note:

- Average distortion can be as high as $\Omega(n)$.
- Yet $O(1)$-approximation for average distorion.

Absolute vs Relative Bounds

Absolute Bounds

- Best guarantee about “worst case” distortion.
- Guarantee on distortion is independent of input metric.

Absolute vs Relative Bounds

Absolute Bounds

- Best guarantee about "worst case" distortion.
- Guarantee on distortion is independent of input metric.

Relative Bound

- Given, as input, a finite metric, embed it into the host metric to (approximately) minimize distortion. [cf. Ravi's Talk]
- Comparing against the best possible distortion for the given input metric.

Note: Absolute bound $\rho \Rightarrow$ Relative bound ρ.

Relative Bounds: Existing Work

- [LLR95] minimizing maximum distortion of embedding arbitrary finite metrics into l_2 via Semi-Definite Programming.
  $\Rightarrow 1$-approximation for maximum distortion problem.
- [WLB$^+$98] PTAS for minimum routing cost spanning tree.
  $\Rightarrow (1+\epsilon)$-approximation for average distortion of embedding arbitrary (graph) metrics into spanning tree metrics.

Open: Can one give an algorithm with $o(\log n)$ relative (average) distortion for embeddings into l_1?

Our results

Given a finite metric, embed it into a line in non-contracting fashion.

- $O(1)$-approximation for average distortion of embedding a general metric into line.
- Better bounds for when the input is a tree metric.
  - $(1+\epsilon)$-approximation in time $n^{O(\frac{\log n}{\epsilon})}$.
  - Polynomial-time *exact* algorithm for *tree-edge* average distortion.

Warm-up: Embedding into trees

Lower bound

- Let $star(x) = \sum_{y \in V} d(x, y)$

$$n^2 \cdot \text{av}(d) = \sum_x star(x)$$

- Let m be the point which has minimum $star(\cdot)$ value.

$$\text{av}(d) \geq \frac{1}{n} \cdot star(m)$$

Warm-up: Embedding into trees

Recall: $m = \text{argmin}_x\{star(x)\}$

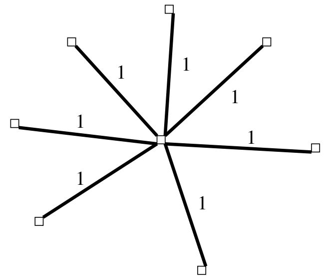
$$
\begin{aligned}
\text{av}(d) &= \frac{1}{n^2}\sum_{x,y} d(x,y) \\
&\leq \frac{1}{n^2}\sum_{x,y} d(x,m) + d(m,y) \\
&= \frac{2}{n}\cdot star(m)
\end{aligned}
$$

Theorem The shortest path tree rooted at m is a 2-approximation.

Getting a path

Remember: we wanted a line (path) metric, not <u>any</u> shortest path tree.

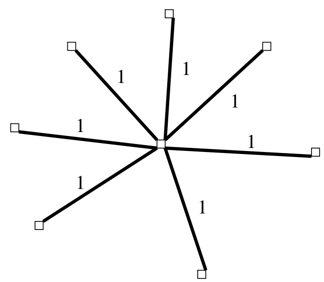
Tree could look like:



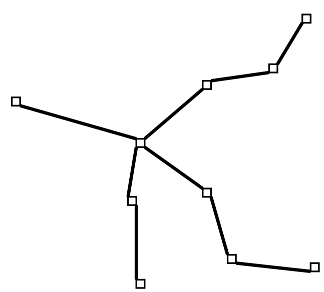
Getting a path

Remember: we wanted a line (path) metric, not any shortest path tree.

Tree could look like:



k*-spiders:* A tree with degree atmost two for all vertices except one, for which it could be upto k.

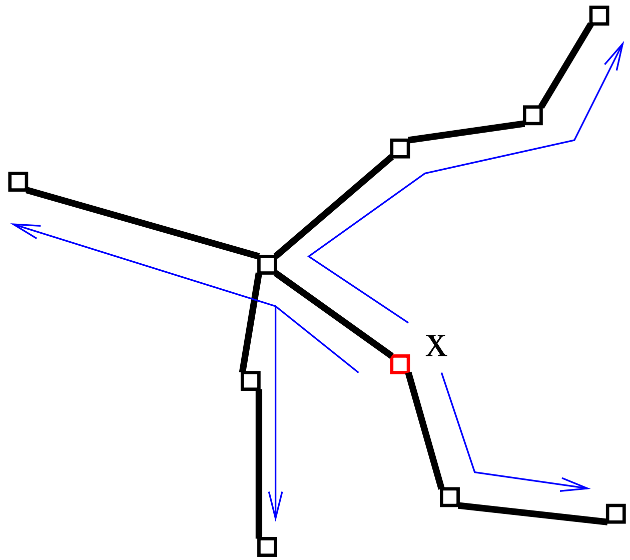


Embedding into k-spider

- k*-repairman tour:* Given k repairmen starting at a depot s. The k repairmen are to visit n customers in a metric space. The latency of a customer is her waiting time.
- k-repairman(x): the sum of latencies of all customers in a minimum k-repairman tour rooted at x.

Lower bound for k-spider

- From a k-spider embedding, we can construct a k-repairman tour.



- k-repairman(x) $\leq \sum_y d_k(x, y)$

Lower bound for k-spider

Adding up . . .

$$\text{av}(d_k) \geq \frac{1}{n^2} \sum_x k\text{-repairman}(x)$$

Let m be the point with minimum k-repairman($\cdot$) value.

$$\text{av}(d_k) \geq (\frac{1}{n}) \cdot k\text{-repairman}(m)$$

k-spiders

Upper bound (same as before)

$$\text{av}(d) \leq \frac{2}{n} \cdot \text{k-repairman}(mx)$$

Theorem The best k-repairman tour rooted at m gives a 2-approximation for the average distortion of embeddings into k-spiders.
Theorem A ρ-approximation for k-repairman gives 2ρ-approximation for average distortion.

Currently $\rho = 6$ due to [CGRT 03].

Average distortion for line

Fact: A line is a 2-spider.
Theorem There is an $O(1)$-approximation algorithm for average distortion of embedding a finite metric into line.

Our results

Given a finite metric, embed it into a line in non-contracting fashion.

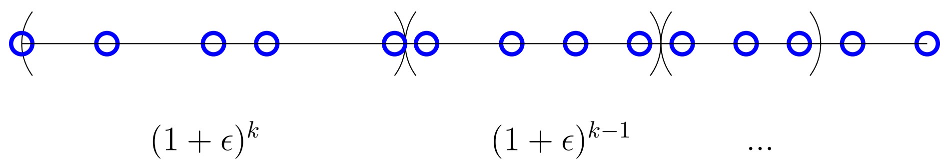
- O(1) approximation for average distortion of embedding a general metric into line.
- Better bounds for when the input is a tree metric.
  - $(1+\epsilon)$-approximation in time $n^{O(\frac{\log n}{\epsilon})}$.
  - Polynomial-time *exact* algorithm for *tree-edge* average distortion.

QPTAS for trees

- Idea based on QPTAS for minimum latency [AK99].
- **Fact** [AK99]: There exists a $(1+\epsilon)$-approximate minimum-latency tour that is a concatenation of $O(\frac{\log n}{\epsilon})$ TSP tours.
- We extend this idea for average distortion.

QPTAS for trees

- Divide OPT embedding into k ($\approx \log n/\epsilon$) segments.
- $(1+\epsilon)^{k-i}$ vertices assigned to segment i.
- Replace the embedding of each segment by an induced "TSP-like" path without increasing the distortion too much.



Proof Idea

- Divide the objective av(d) among the segments.
- Share of segment i can be written as

$$\alpha_i \cdot \text{Latency}(i) + \beta_i \cdot \text{TSP}(i)$$

where

Latency(i) $=$ Total Latency of segment i
TSP(i) $=$ Length of the embedding of segment i

Proof Idea

- Cost share of segment i can be written as

$$\alpha_i \cdot \text{Latency}(i) + \beta_i \cdot \text{TSP}(i)$$

- (Variant of [AK99]): Each segment itself can be modified to be a concatenation of $O(\frac{\log n}{\epsilon})$ TSP tours. This increases the distortion only by $1 + \epsilon$.

Theorem There is a near-optimal embedding that is a concatenation of $O(\frac{\log^2 n}{\epsilon^2})$ TSP tours.
Final Result:

- Can reduce the $O(\frac{\log^2 n}{\epsilon^2})$ to $O(\frac{\log n}{\epsilon^2})$ TSP tours.
- Solution computed by Dynamic Programming.

Our results

Given a finite metric, embed it into a line in non-contracting fashion.

- O(1) approximation for average distortion of embedding a general metric into line.
- Better bounds for when the input is a tree metric.
  - $(1+\epsilon)$-approximation in time $n^{O(\frac{\log n}{\epsilon})}$.
  - Polynomial-time *exact* algorithm for *tree-edge* average distortion.

Polynomial Time Algorithm

Tree-edge Distortion: av$_T(d) = \sum_{e \in T} d(e)$

Theorem There is a polynomial time algorithm that minimizes average tree-edge distortion.

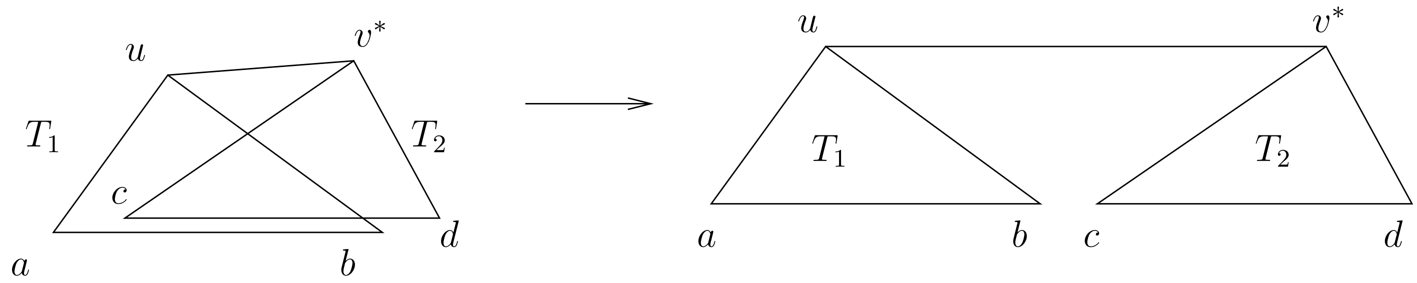
Main Idea:

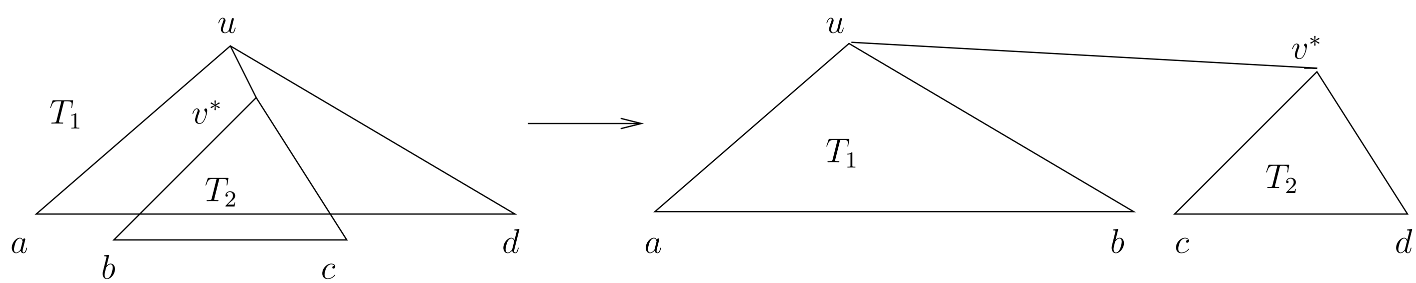
- The best embedding is an Eulerian tour truncated at an appropriately defined centroid.
- This Eulerian tour can be found efficiently.

Similar to Minimum Linear Arrangement of trees ([Shi79, Chu84]).

Main Idea

Local interchanges reduce average tree-edge distortion.

1. 

2. 

v^* is the *centroid* of the tree.

Open Questions

- PTAS for average distortion for embedding arbitray metrics into line? Tree metriics into line?
- Approximating the maximum distortion of embedding a (simple) metric space (e.g. trees) into line or l_1?

References

(AK99) Sanjeev Arora and George Karakostas. Approximation schemes for minimum latency problems. In *Proceedings of the 31st Annual ACM Symposium on the Theory of Computing (STOC)*, pages 688–693, 1999.

(Chu84) F. R. K. Chung. On optimal linear arrangements of trees. *Computers and Mathematics with Applications*, 10, 1984.

(LLR95) Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15:215–245, 1995.

(Rab03) Y. Rabinovich. On average distortion of embedding metrics into l_1 and into the line. In *35th Annual ACM Symposium on the Theory of Computing (STOC)*, 2003.

(Shi79) Yossi Shiloach. A minimum linear arrangement algorithm for undirected trees. *SIAM J.* COMPUT., 8(1), February 1979.

(WLB+98) Bang Ye Wu, Giuseppe Lancia, Vineet Bafna, Kun-Mao Chao, R. Ravi, and Chuan Yi Tan. A polynomial time approximation scheme for minimum routing cost spanning trees. In *Symposium on Discrete Algorithms*, pages 21–32, 1998.