Embedding into Ditribution of Trees

Ashkan Mokarian

Max-Planck Institute of Informatics

Final Presentation for the course Random Discrete Structures, Summer 2014

Outline

Metric

Definition (metric)

A metric space is a pair (X,d), where X is a set of *points* and a function $d : X \times X \to \mathbb{R}^{\geq \nvdash}$ such that

1. $d(x, y) = 0 \Leftrightarrow x = y$
2. $d(x, y) = d(y, x)$ (symmetry)
3. $d(x, y) + d(y, z) \geq d(x, z)$ (triangle-inequality)

Embedding

Definition (embedding)

Given metric spaces (X, d) and (X', d') a map $f : X \to X'$ is called an embedding. An embedding is called distance-preserving or isometric if for all $x, y \in X$, $d(x, y) = d'(f(x), f(y))$.

Distortion

Definition (distortion)

Let f be an embedding from the finite metric space (X, d) into another finite metric (X', d'). We define

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

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

the *distortion* of an embedding f, $\text{distortion}(f)$, is defined as the product of $\text{expansion}(f)$ and $\text{contraction}(f)$. An embedding with $\text{distortion}(f) = 1$ is called *isometric*.

note: from here on, we'll show the distortion by D.

Distortion

notation: $(X,d) \overset{D}{\hookrightarrow} (X',d')$ means that there exist an embedding f from metric space (X,d) into (X',d') with distorion D and an $r > 0$ such that:

$$r \leq \frac{d'(f(x),f(y))}{d(x,y)} \leq Dr$$

for any $x, y \in X$.

Outline

Why trees?

- Many problems are simple on trees. Thus, embed a given metric into a tree.
- But Impossible with low distortion

Theorem

Every embedding of C_n into a tree T incurs distortion $\Omega(n)$.

- remedy: embed into *distribution* of trees.

Why trees?

- Many problems are simple on trees. Thus, embed a given metric into a tree.
- But Impossible with low distortion

Theorem

Every embedding of C_n into a tree T incurs distortion $\Omega(n)$.

- remedy: embed into *distribution* of trees.

Why trees?

- Many problems are simple on trees. Thus, embed a given metric into a tree.
- But Impossible with low distortion

Theorem

Every embedding of C_n into a tree T incurs distortion $\Omega(n)$.

- remedy: embed into *distribution* of trees.

Probabilistic definition of Distortion

Definition

Suppose $\mathcal{G}$ is a graph family. Then $(X, d) \overset{D}{\hookrightarrow} \mathrm{distrib}(\mathcal{G})$ means that there exist a distribution π on the graph family $\mathcal{G}$ and an $r > 0$ such that:

$$r \leq \frac{E_{g \leftarrow \pi}[d_g(x, y)]}{d(x, y)} \leq Dr$$

where $g \in G$ and $x, y \in X$.

from now on we use the notation $d_\pi(x, y) = E_{g \leftarrow \pi}[d_g(x, y)]$.

Outline

some definitions

Definition (diameter)

The diameter of a graph G (denoted by $\mathrm{diam}(G)$) is the least δ such that for all pairs of vertices $x, y \in V(G)$, we have $d_G(x, y) \leq \delta$. In other words $\mathrm{diam}(G) = \sup_{x,y \in V(G)} d_G(x, y)$.

Definition (weak-diameter)

Given a graph G and a subgraph $G' \subseteq G$, the weak diameter of G' with respect to G (denoted by $\mathrm{weak}_G\mathrm{diam}(G')$) is the least δ such that $d_G(x, y) \leq \delta$ for all $x, y \in V(G')$.

Bartal's theorem, 1996

Theorem

Given a metric (X, d) with diameter Δ, let $\mathcal{DT}$ be the set of all tree metrics that dominate d. Then

$$(X, d) \overset{\mathcal{O}(\log n \log \Delta)}{\hookrightarrow} \mathrm{distrib}(\mathcal{DT})$$

Probabilistic Decomposition (Cutting Schema)

Theorem

Given a graph $G = (V, E)$ with edge lengths, and a parameter δ, there exists a procedure that delets edges E' such that

1. *Each connected component C in $(V, E - E')$ has weak diameter smaller than δ.*
2. *Pr[edge e is cut] $\leq 4 \log n \times (d(e)/\delta)$.*

For now, assume such a randomized procedure exists and we will come back to it later.

Bartal's tree construction

Theorem

Given a metric (X, d) with diameter Δ, let $\mathcal{DT}$ be the set of all tree metrics that dominate d. Then

$$(X, d) \overset{\mathcal{O}(\log n \log \Delta)}{\hookrightarrow} \mathrm{distrib}(\mathcal{DT})$$

some lemmas about the tree constructed

Lemma

consider a level j component H, and a level $j+1$ component H' formed by partitioning H. Let T' be the tree corresponding to H' (with root r'), and let T be the tree corresponding to H (with root r).

- *length of the edge $\{r, r'\}$ in T is $\Delta/2^j$.*
- *The distance of any leaf in T from the root r is at most $2\Delta/2^j$.*
- *The diameter of T is at most $4\Delta/2^j$.*

Lemma

Any tree constructed as above dominates d, i.e., $d_T(x,y) \geq d(x,y)$.

some lemmas about the tree constructed

Lemma

consider a level j component H, and a level $j+1$ component H' formed by partitioning H. Let T' be the tree corresponding to H' (with root r'), and let T be the tree corresponding to H (with root r).

- *length of the edge $\{r, r'\}$ in T is $\Delta/2^j$.*
- *The distance of any leaf in T from the root r is at most $2\Delta/2^j$.*
- *The diameter of T is at most $4\Delta/2^j$.*

Lemma

Any tree constructed as above dominates d, i.e., $d_T(x, y) \geq d(x, y)$.

Bartal's thm - proof

Theorem

Given a metric (X, d) with diameter Δ, there exist an embedding into a distribution of trees such that

1. *all trees constructed dominate d.*
2. *The embedding has distortion $\in \mathcal{O}(\log n \log \Delta)$*

Recap:
The diameter of T is at most $4\Delta/2^j$ (one of the properties of the tree constructed)
Pr[edge e is cut] $\leq 4 \log n \times (d(e)/\delta)$ (by the randomized cutting procedure).

Outline

1. Basic Definitions
2. Embedding into random trees
   - Motivation
   - Bartal's thm, $\mathcal{O}(\log n \log \Delta)$
   - [CKR01]-Cutting Scheme
   - [FRT03], $\mathcal{O}(\log n)$

randomized cutting procedure

Theorem

Given a graph $G = (V, E)$ with edge lengths, and a parameter δ, there exists a procedure that delets edges E' such that

1. *Each connected component C in $(V, E - E')$ has weak diameter smaller than δ.*
2. *Pr[edge e is cut]* $\leq 4 \log n \times (d(e)/\delta)$.

The procedure that is going to be explained, was first introduced in a paper of Calinescu, Karloff and Rabani(2001), and was later used by [FRT03] to achive the better bound on distortion.

Bartal uses some other method in his original paper.

CKR-cutting procedure

Algorithm

INPUT: G,δ

1. Pick a radius R uniformly at random from $[\delta/4, \delta/2]$
2. Pick a random permutation $\sigma \in S_n$, which defines an order $<_\sigma$ on the vertices
3. For every vertex v_i, define a ball $B_i = B(v_i, R)$
4. Assign each vertex to the first ball it lies in. define $\hat{B}_i = B_i - \cup_{j<_\sigma i} B_j$
5. Delete all edges in the cut $(\hat{B}_i, V \setminus \hat{B}_i)$

proof

Theorem

Given a graph $G = (V, E)$ with edge lengths, and a parameter δ, there exists a procedure that deletes edges E' such that

1. *Each connected component C in $(V, E - E')$ has weak diameter smaller than δ.*
2. *Pr[edge e is cut]* $\leq 4 \log n \times (d(e)/\delta)$.

can the bound be improved?

proof

Theorem

Given a graph $G = (V, E)$ with edge lengths, and a parameter δ, there exists a procedure that deletes edges E' such that

1. *Each connected component C in $(V, E - E')$ has weak diameter smaller than δ.*
2. *Pr[edge e is cut]* $\leq 4 \log n \times (d(e)/\delta)$.

can the bound be improved?

Outline

better bound for distortion

Theorem

Every metric space with n points can be embedded by a distribution of trees with distortion $\mathcal{O}(\log n)$.

Summary

- any metric space can be embedded into a distribution of trees with low distortion
- a randomized algorithm for getting decomposition of arbitrary size component upper limit has been explained

Thanks I

- Yair Bartal.
  Probabilistic approximations of metric spaces and its algorithmic applications.
  In Proceedings of the 37th IEEE Symposium on Foundations of Computer Science (FOCS), pages 184-193, 1996

- Fakcharoenphol, Rao, and Talwar.
  A tight bound on approximating arbitrary metrics by tree metrics.
  ACM Symposium on Theory of Computing(STOC), 2003.