Approximating metrics by simpler metrics

Suppose we are given a graph $G=(V,E)$ with edge weights $w_e > 0$. We can define "distance" of any two nodes u,v, $d_G(u,v)$ as the length of the shortest u-v path.

Note that $d(\cdot,\cdot)$ defines a metric on V:

$d(u,u)=0$ & $d(u,v)+d(v,w) \geq d(u,w)$

- For many problems that are difficult on graphs, they are easier on trees.
- Can we find trees as subgraphs that approximatly preserve distances of nodes in G?

low streth spanning trees:

Def: let T be a spanning tree of G and $\alpha \geq 1$. We say T is an α-Stretch tree of G if

$\forall\, u,v \in V$: $\qquad d_G(u,v) \leq d_T(u,v) \leq \alpha\, d_G(u,v)$

if we had such a tree we could solve the given problem on T and use that solution as an approximate solution for G.

For e.g. TSP is difficult even if G is metric (NP-hard)

We could try to solve TSP on T (which is trivial)

and this implies an α-approximate solution since:

if Π_G is the opt ordering for TSP on G then

$$\text{OPT} = \sum_{i \geq 1} d_G(\Pi_G(i), \Pi_G(i+1))$$

Now since T is an α-stretch tree then

$$\sum_{i \geq 1} d_T(\Pi_G(i), \Pi_G(i+1)) \leq \alpha \cdot \text{OPT}$$

Now if we find opt of T with ordering of Π_T then

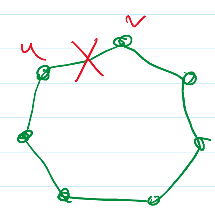
$$\sum_{i \geq 1} d_T(\Pi_T(i), \Pi_T(i+1)) \leq \sum_{i \geq 1} d_T(\Pi_G(i), \Pi_G(i+1))$$

and so we have an α-approx.

However, it turns out we cannot have a low stretch spanning tree in general.

Example: Consider C_n (cycle on n nodes)

for any spanning tree of C_n, two adj nodes in C_n are at distance $n-1$ of each other



Using randomness helps

What about having a family of trees s.t. we have the low stretch property on expectation if we draw a sample from this family?

Def: A rand. low-stretch spanning tree for a graph $G=(V,E)$ is a prob. distribution D over spanning trees of G s.t. $\forall\, u,v \in V$: $d_G(u,v) \leq d_T(u,v) \quad \forall\, T \in D$

$$\mathbb{E}_{T\sim D}[d_T(u,v)] \le \alpha\, d_G(u,v)$$

Is this good enough for approximation alg's? In most cases yes

For e.g. for TSP it implies the expected length of the solution is at most α.opt.

General paradigm if the objective is a linear func. of distances: take a rand. tree T from a low stretch dist. D; solve the problem on T and use that as an approximate solution for the problem on G.

First Attempt:

let Δ be the aspect ratio of metric $\frac{\max d(u,v)}{\min d(u,v)} = \Delta$

Theorem: Given any metric (V,d) with $|V|=n$ and aspect ratio Δ, there is an efficiently sampleable dist. D over spanning trees of V s.t. $\forall u,v \in V$

1) $\forall T \in \text{Support}(D)$, $d(u,v) \le d_T(u,v)$

2) $\mathbb{E}_{T\sim D}[d_T(u,v)] \le O(\log n \log \Delta)\, d(u,v)$

We use what is known as a "low diameter decomposition"

Suppose $r \in \mathbb{R}^+$ is given, we generate a partition of V into $V_1 \dots V_t$ s.t.

1) low diam: Each V_i has diameter at most r (think of a ball of radius $\frac{r}{2}$ around a node $c_i \in V_i$)

2) low cutting Prob: $\forall u,v \in V$, $\Pr[u,v \text{ lie in diff. } V_i] < \beta \cdot \frac{d(u,v)}{}$

2) low cutting prob: $\forall\, u,v \in V$, $\Pr[u,v$ lie in diff. $V_i] \leq \beta \cdot \frac{d(u,v)}{r}$

with $\beta = O(\log n)$.

For now suppose we know how to find such a decomp.

Then we can complete proof of Theorem:

Proof: We apply the following recursively: take as input (U, i) where $U \subseteq V$ and U has diameter $\leq 2^i$ and run Tree-Embed that returns a rooted tree T, root r.

TreeEmbed (U, i):

- Find low diam. decomp. $U_1 \dots U_t$ with $r = 2^{i-1}$
- Recurse on $U_1 \dots U_t$ to obtain trees $T_1 \dots T_t$ rooted at $r_1 \dots r_t$; if U_i is a single point just return it
- Add an edge between r and r_i with length 2^i
- return (T, r)

- Assuming that low-diam decomp. is randomized this returns a dist. over trees.
- To build tree T for V, we first scale distances s.t. $d(u,v) \geq 1$ and $d(u,v) \leq \Delta = 2^\delta$; then call Tree Embed (V, δ).

Lemma 1: $\forall u,v \in V$: $d_T(u,v) \geq d(u,v)$

Proof: Suppose $d(u,v) \in (2^{i-1}, 2^i]$ and consider the iteration Tree_Embed(U, i) where $u \in U$. if $v \in U$ then since $d(u,v) > 2^{i-1}$, then u & v will fall into two diff. parts in the low-diam. decomposition. Then $d_T(u,v) \geq 2^i$ as there is an edge of length 2^i between roots $r_1 .. r_t$ and r.

if $v \notin U$ then u,v are separated at a higher level i' and $d_T(u,v) \geq 2^{i'} > 2^i$ ∎

Lemma 2: For all $u,v \in V$: $E_{T \sim D}[d_T(u,v)] \leq O(\log n \log \Delta)\, d(u,v)$.

Proof: Suppose (T, r) is the result of call to Tree_Embed(U, i). We prove by induction that $d_T(r, x) \leq 2^{i+1}$ for all $x \in U$. Note that $x \in U_j$ for some part U_j in the decomposition with diam 2^{i-1}

$\longrightarrow$ distance x and root $r_j \leq 2^{i-1+1} = 2^i$

Since r_j (root of U_j) is connected to r with an edge of 2^i we have $d_T(r, x) \leq 2^i \times 2 = 2^{i+1}$.

This implies $\forall x,y \in U$: $d_T(x,y) \leq 2^{i+2}$.

Thus:

$$d_T(x,y) \le \sum_{i=\delta}^{0} \Pr[x,y \text{ first separated at level } i]\cdot 2^{i+2}$$

$$\le \sum_{i=\delta}^{0} \beta\cdot \frac{d(x,y)}{2^{i-1}}\cdot 2^{i+2}$$

$$= 8\beta(\delta+1)\,d(x,y)$$

with $\beta = O(\log n)$ and $\delta = O(\log \Delta)$ the theorem follows ∎

The first result on this was given by Bartal'96 & was subsequently improved to $\alpha = O(\log n \log\log n)$

The question is how to obtain the low-diam. decomp.

We describe a method that actually gives a stronger bound, however, the tree T generated might be over a set $V' \supseteq V$; so it may not be a spanning tree of G but instead the leaves of T will be V.

We prove the following stronger bound.

Theorem: [FRT '04] Given a metric (V,d) with $d(u,v) \ge 1$ there is a rand. polytime alg. that produces a weighted tree $T=(V',E)$, $V \subseteq V'$ s.t.

1) $\forall u,v$: $d(u,v) \le d_T(u,v)$

2) $\mathbb{E}_T[d_T(u,v)] \le O(\log n)\cdot d(u,v)$

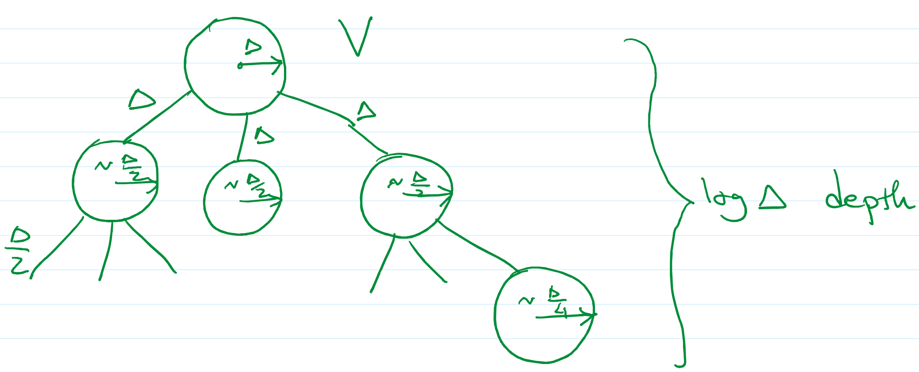
Hierarchical decomposition

As before assume $\Delta = \max_{u,v} d(u,v) = 2^\delta$

We describe a hierarchical decomposition where each level is a low-diameter decomposition of the prev. level.

- The root corresponds to V
- each leaf node corresponds to a single point in V
- A node of the tree in level i corresponds to a set $S \subseteq V$ and children of S correspond to a partition of S; Vertices of S are at dist $[2^{i-1}, 2^i)$ from a centre node, so it is a ball with radius in $[2^{i-1}, 2^i)$.

The length of edges joining a node at level $i-1$ to its parent is 2^i



Rand. Heirarchical Decomposition

1) Pick a rand. permutation π of V

2) Set Δ to the smallest power of 2 $\geq 2\max_{u,v} d(u,v)$

3) Pick $r_0 \in [\frac{1}{2}, 1)$ and set $r_i = r_0 \cdot 2^i$ $\quad 1 \leq i \leq \delta = \log \Delta$

4) $C_\delta = V$ be the root; create a node for i

5) for $i \leftarrow \delta$ down to 1 do

  $C_{i-1} \leftarrow \phi$

  for all $C \in C_i$ do

    $S \leftarrow C$

    for $j \leftarrow 1$ to n do

      if $B(\pi(j), r_{i-1}) \cap S \neq \phi$ then

        Add $B(\pi(j), r_{i-1}) \cap S$ to C_{i-1}

        and remove it from S

  Create tree nodes for sets in C_{i-1}

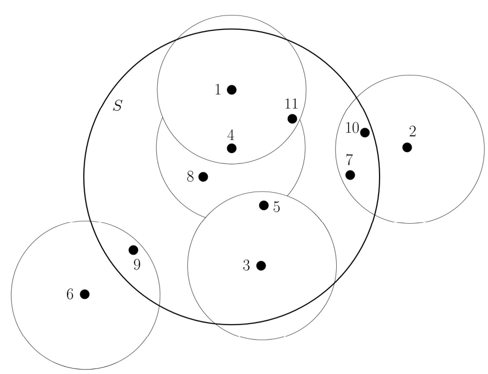


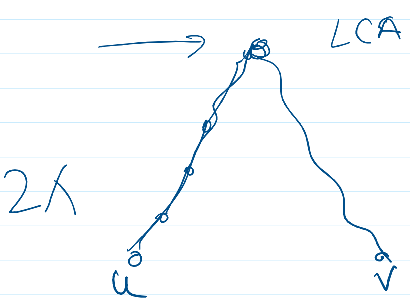
Figure 8.8: An example of the partitioning of set $S = \{1, 3, 4, 5, 7, 8, 9, 10, 11\}$ in the hierarchical cut decomposition. Suppose that the random permutation is the identity.

Lemma 1: $d(u,v) \leq d_T(u,v)$ $\forall u,v$ and if the least common ancestor of u,v is at level i then $d_T(u,v) \leq 2^{i+2}$

[level 0 : leaf nodes, level δ : root]

proof: For each set S at level i, diameter is $\leq 2^{i+1}$

For any pair u,v, they cannot belong to a set at a level $\lfloor \log d(u,v) \rfloor - 1$ or lower. $\longrightarrow$ $d_T(u,v) \geq 2\sum_{j=1}^{\lfloor \log d(u,v) \rfloor} 2^j$



if LCA is at level i

$$d_T(u,v) = 2\sum_{j=1}^{i} 2^j = 2(2^{i+1} - 2) \leq 2^{i+2}$$

Our next goal is to show that $E[d_T(u,v)] \leq O(\log n)\, d(u,v)$

Def: We say w settles u,v at level i if w is the first vertex in π s.t. at least one of $u,v \in B(w, r_i)$

We say w cuts u,v at level i if exactly one of $u,v \in B(w, r_i)$

$X_{i,w}$: event that w cuts u,v at level i

$S_{i,w}$: " " " settles " " " "

$$d_T(u,v) \leq \max_{i=0 \dots \log \Delta} 2^{i+2} \quad \text{where } \exists\, w \in V : X_{iw} \wedge S_{iw}$$

$$E[d_T(u,v)] \leq \sum_{w \in V} \sum_{i=0}^{\log \Delta} \Pr[X_{iw} \wedge S_{iw}] \cdot 2^{i+2}$$

$\log \Delta$

Lemma 2: $\sum_{i=0}^{\lg \Delta} \Pr[X_{iw}] \cdot 2^{i+2} \le 8 d_{uv}$

lemma 3: $\Pr\left[S_{iw} \mid X_{iw}\right] \le \frac{1}{j}$ where w is the j'th closest node to u,v (in π)

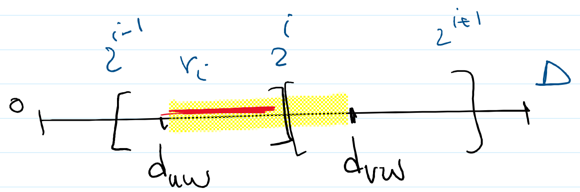
$$E[d_T(u,v)] \le \sum_{w \in V} \sum_{i=0}^{\lg \Delta} \Pr[S_{iw} \mid X_{iw}] \cdot \Pr[X_{iw}]\, 2^{i+2}$$

$$\le \sum_{w \in V} \frac{1}{\mathrm{rank}(w)} \sum_{i} \Pr[X_{iw}] \cdot 2^{i+2}$$

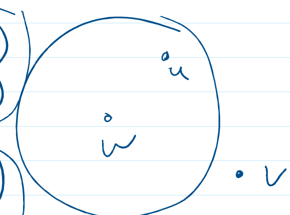
$$\le 8 d_{uv} \cdot O(\ln n)$$

Proof of lemma 2: bound $\Pr[X_{iw}]$ WLOG $d_{uw} \le d_{vw}$

Prob w cuts u,v on level i = $\Pr\left[u \in B(w, r_i),\ v \notin B(w, r_i) \right]$



$r_i \in [2^{i-1}, 2^i)$



$$\Pr[X_{iw}] = \frac{\left|[2^{i-1}, 2^i) \cap [d_{uw}, d_{vw})\right|}{\left|[2^{i-1}, 2^i)\right|}$$

$\left|[2^{i-1}, 2^i)\right| = 2^{i-1}$

for $i = 0 \ldots \lg \Delta - 1$ partitions $[\frac{1}{2}, \frac{\Delta}{2})$

so $\Pr[X_{iw}] \cdot 2^{i+2} = \frac{2^{i+2}}{2^{i-1}} \left| [2^{i-1}, 2^i) \cap [d_{uw}, d_{vw}) \right|$

$$\sum \Pr[X_{iw}] \cdot 2^{i+2} = 8 \left| [d_{uw}, d_{vw}) \right| \le 8 (d_{vw} - d_{uw})$$

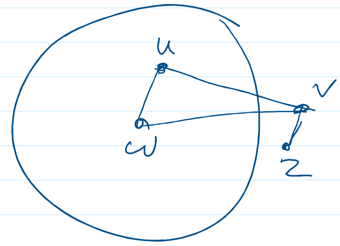
$$\sum_{i \geq 0} \Pr[X_{iw}] \cdot 2 = 8 \left| [d_{uw}, d_{vw}) \right| \leq 8 (d_{vw} - d_{uw}) \leq 8 d_{uv}$$

Proof of lemma 3:

$\Pr[S_{iw} \mid X_{iw}] \leq \frac{1}{j}$ → rank of w

Consider nodes of V in the order of their dist to u,v

if X_{iw} happens then one of $u,v \in B(w, r_i)$ ⟶ any vertex z closer than w to u,v will have $\geq$ one of u,v in $B(z, r_i)$.



and so S_{iw} cannot happen if z comes before w in Π.

Therefore, among the j closest nodes to u,v, w must be the first appearing in Π. $= \frac{1}{j}$

$w \quad u \quad z \quad v$