Lecture 6: Distance-preserving trees.

G = edge length graph.

For a single root s, can find a shortest path tree from s. (Output of Dijkstra's algo.)

i.e. a spanning tree $T \subseteq G$ st $\forall x \in V$,

$$d_T(s,x) = d_G(s,x).$$

[Even for directed graphs].

Can we hope to have such a thing for APSP? No. Eg K_n: any tree you pick misses most edges. So most edges are distance ≥ 2, in original graph at distance 1. want $d_T(x,y) = d_G(x,y)$ $\forall x,y \in V$.

But factor 2 not so bad.

New relaxed condition $\quad d_T(x,y) \geq d_G(x,y)$ ← immediate if $T \subseteq G$

and $d_T(x,y) \leq \alpha \cdot d_G(x,y)$. for small α.

α (stretch)

Sadly: C_n: any spanning subtree T looks like a path, and the $(x,y) \in C_n \setminus T$ has $d_T(x,y) = n-1$, but $d_{C_n}(x,y) = 1$. $\alpha = n-1$. ☹

What can we do? Randomization to the rescue.

Randomized low stretch spanning tree:

Given a graph. $G = (V,E)$ find a distribution $\mathcal{D}_G$ over spanning trees of G

s.t ~~$d_T(x,y) \geq d_G(x$~~

$$\forall (x,y) \in E \quad \mathop{E}_{T \leftarrow \mathcal{D}_G}\left[d_T(x,y)\right] \leq \alpha\, d_G(x,y).$$

N.b. since T is a subtree of G, $d_T(x,y) \geq d_G(x,y)$.
each T in support of $\mathcal{D}$.

HW Claim: $\forall (x,y) \in V \times V$, $E[d_T(x,y)] \leq \alpha\, d_G(x,y)$.

Example: For C_n, pick a random edge and delete it. (at uniform)

Gives a uniform distribution over n trees.

$$E[d_T(u,v)] = \frac{n-1}{n} \cdot 1 + \frac{1}{n} \cdot n-1 = \frac{2(n-1)}{n} < 2. \quad ☺$$

neighbors

Why is this even interesting? Preserves distances "on average"

E.g. So if you pick $c \log n$ such trees, then $\{T_1, T_2 \dots T_{c\log n}\} \in \mathcal{T}$

$\forall x, y, \ \exists T_i \in \mathcal{T}$ st $d_{T_i}(x,y) \leq \alpha \, d_T(x,y)$.

More interesting: allows you to solve optimization problems on the graph by "reducing" to a tree

E.g: k-median on a graph. G, find $|C| = k$ st $\sum_{v \in V} d(v, C)$ min.

Solved this on a tree. But what if G is not a tree?

Algo: Pick a tree $T \leftarrow \mathcal{D}_G$

Solve it on T. Return solution C_T

Claim: $E\left[\sum_i d_G(v, C_T)\right] \leq \alpha . d_G(v, C^*)$ optimal for graph.

Pf: $. \ E\left[\sum_i d_T(v, C^*)\right] \leq \alpha \, d_G(v, C^*)$.

$. \ d_T(v, C_T) \leq d_T(v, C^*)$.

$. \ d_G(v, C_T) \leq d_T(v, C_T)$

$\boxtimes$

3

⇒ if G "embeds randomly into distribution of trees" with stretch α

$$\Rightarrow E[\text{cost of algorithm on } G] \leq \alpha \cdot OPT(G)$$

for k-median.

Works for any "linear" problem. · TSP

"any graph is close to being a tree, at least randomly".

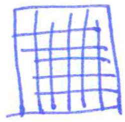
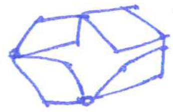
OK: So what can we show?

Thm: for any graph G, can ~~find a~~ sample from distribution $\mathcal{D}_{AN}$ in time $\tilde{O}(m)$

[Abraham Neiman] s.t. $\alpha_{AN} = O(\log n \log\log n)$.

Thm: $\exists$ graph G st. any distribution $\mathcal{D}$ must have $\alpha \geq \Omega(\log n)$.

[~~Bartal 96~~]

[Alon Karp Peleg West] ⇒ tight up to $O(\log\log n)$ factor.  or  fractal.

Today, prove a weaker bound. [due to Bartal '96].

(i) only for "metric graphs". and

(ii) weaker quantitative bounds.

Def^n: a metric graph G is a complete graph $G = (V, \binom{V}{2})$ st. edge lengths satisfy the triangle inequality. $l_{uw} \leq l_{uv} + l_{wv}$ forall u, v, w.

Theorem [Bartal] Given a metric graph G, can sample from distribution $\mathcal{D}_{Bartal}$ st. $\alpha_{Bartal} = O(\log n \log \Delta)$

where $\Delta = \dfrac{\text{max distance in } G}{\text{min distance in } G}$.

HW: get rid of $\log\Delta$ replace by $O(\log n)$.

Thm [FRT]: $\alpha_{FRT} = O(\log n)$ and is tight even for metric graphs.

OK. let's prove it:

Crucial lemma: "Low diameter decompositions of graphs." LDD.

~~Thm~~ Lem: Given a metric graph G and parameter D, can find partition of vertices $V_1, V_2 \ldots V_t$

s.t. (1) $d_G(x,y) \le D \quad \forall x,y \in V_i \; \forall i$

(2) $\Pr[(x,y) \text{ not in same cluster } V_i] \le \frac{d_G(x,y)}{D} \underset{\beta}{\underset{\parallel}{O(\log n)}}$.

Intuition: for a path, can set $\alpha = 1$.

for a grid, $\alpha = 2$. Natural to imagine hypercube $\alpha = \log n$.

In fact. for all graphs $\alpha = \log n$.

Assuming ~~proof~~ lemma lets prove the theorem.

Here's a construction. Say mindist $= 1$. $\Rightarrow$ max distance $= \Delta$.

Bartal$(H, 2^i)$ { // precondition: $\max_{x,y} d_H(x,y) \le 2^i$

. Use LDD_β with $D = 2^{i-1}$ to get $V_1, V_2 \ldots V_t$ partition of $H(V)$.

. ~~Build tree~~ Bartal$(H[V_i]) \longrightarrow$ tree T_i with root r_i

. attach r_1 to all $r_2 . r_3 \ldots r_t . \longrightarrow T$

Return T.

Claim: $E[d_T(x,y)] \le \cancel{4}\,8\beta . \cancel{\log \Delta} . i . \; d_H(x,y)$.

$\Rightarrow$ if we call original graph G with Bartal$(G, 2^{\lceil \log_2 \Delta \rceil})$

then $E[d_T(x,y)] \le O(\log \Delta) 4\beta \cdot d_G(x,y)$.

Pf of claim : look at $x, y \in H$.

$$\Pr[(x,y) \text{ sep in } T_i \& T_j] \times [d(r_i, r_1) + d(r_1, r_j) + d(x, r_i) + d(y, r_j)] + \Pr[(x,y) \text{ not sep}] . \, d_{T_i}(x,y).$$

$$\leq \frac{d(x,y)}{2^{i-1}} . \beta . \times [2^i + 2^i] + 1 . 4\beta . (i-1) . d(x,y)$$

by induction

Look at class notes, final + . . .

$$= 4\beta \, d(x,y) + 4\beta (i-1) \, d(x,y) = 4\beta i \, d(x,y)$$

Metric Graph? We added edges r_1, r_i. Why do we know such edges exist.

B/c complete graph.

Finally: Procedure for $LDD_\beta(G, D)$?

meaningful if $2 \lg n \leq D$.

Pick a vertex v, pick a random radius R_v from $\text{Geom}\left(\min\left(1, \frac{4 \log n}{D}\right)\right)$

$$E[R_v] \simeq \frac{D}{4 \lg n}. \quad \Pr[R_v > D] \leq \frac{1}{n^4}.$$

~~Define $V_1 \leftarrow \text{Ball}(v, \min(R_v, D))$. Repeat in rest of G.~~

if Any of the R_vs reach more than D, repeat entire experiment.
(will do this twice in expectation).

. Clearly diameter $\leq D$.

. What is prob (x,y) separated? View this as coin tosses.

Memoryless property of exponential ⇒ if the R_v does not reach x or y does not matter.

So sps it reaches x. Then Prob that it cuts y Af is

~~(a) it reaches $2\lg n$ ⇒ $\Pr[A \text{ that}] \leq \frac{1}{n^3}$~~

it gets heads in the $d(x,y)$ tosses

$$\text{w.p.} \leq \frac{d(x,y)}{p} = \frac{d(x,y)(4 \lg n)}{D}.$$

& twice because of $\Pr[\text{reach} \geq D]$.