Distance Oracle

$x \rightarrow$, $y \rightarrow$ **Distance Oracle (DO)** $\rightarrow$ Exact / approx distance from x to y in G

Goal: Optimize
- query time (Time to report ans)
- size of oracle
- stretch

Eg. DO = Distance Matrix of n^2 size
(distance oracle)

query time $= O(1)$
(matrix can be stored as 2-d array)

size of DSO $= O(n^2)$

What if you have a graph with 10^{10} vertices, but your computer can store only 10^{15} byte information?

↳ It is not possible to store distance matrix

Result: For any n-vertex graph, and any k, we can have Distance Oracle (DO) s.t.

- size of DO $= O(k\, n^{1+1/k} \log n)$
- query time $= O(k)$
- dist stretch $= 2k-1$

decrease with k

increase with k

Implication:

k	stretch	query time	size
2	3	$O(2)$	$O(2 n^{1.5} \log n)$
3	5	$O(3)$	$O(3 n^{4/3} \log n)$
$\vdots$			
$k = \log n$	$2\log n - 1$	$O(\log n)$	$O(\log n \cdot 2 \cdot \log n)$

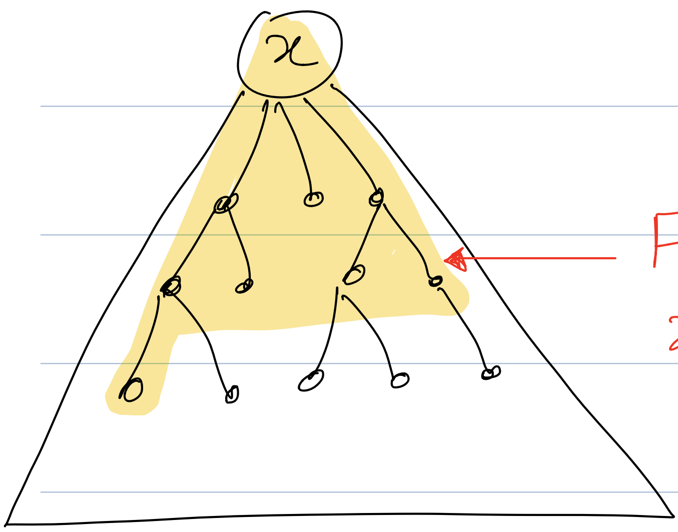
(or $n = 2^k$

(as $n^{1+1/k} = 2^k \cdot 2 = 2n$)

Tools used from Last Class (Hitting Set)

Lemma 1 : If $Z_1\ Z_2 \ldots Z_n$ are n sets each of size $\geq t$. Then for a random set R of size $\frac{4n\log n}{t}$ with high prob. $R \cap Z_i$ is non-empty for $1 \leq i \leq n$

⊛ Notation $C_t(x)$ = First t vertices in BFS(x) closest to x



First 9 vertices closest to x in BFS(x)

(Lemma 2)
Implication : If R is a random set of size $\frac{n\log n}{t}$ then with high probability $R \cap C_t(x)$ is non-empty for each $x \in V(G)$.

Construction of DO with stretch = 3 and size $O(n^{1.5}\log n)$

Some Ideas :

* For each v, you can store distance to $\sqrt{n}$ vertices.

* You can take random set A of $\sqrt{n}\log n$ vertices and store distance b/w all pairs in $A \times V$

Total size $= O(\sqrt{n}\log n)$

Lemma 3 : For any $x, y \in V(G)$, $a \in A$ if $\underline{d_G(x,a) \le d_G(x,y)}$ then $d_G(x,a) + d_G(a,y)$ is 3-approximation to $d_G(x,y)$

Pf : By Δ-lty, $d_G(x,y) \le d_G(x,a) + d_G(a,y)$

$\le d_G(x,a) + d_G(a,x) + d_G(x,y) \le 3 d_G(x,y)$

Oracle construction:

1. $A \leftarrow$ random set of $\frac{n \log n}{t}$ vertices

2. Store distance b/w all pairs in $A \times V$.

3. For each $x \in V(G)$,
   - store dist b/w $\{x\}$ and $C_t(x)$
   - Set a_x = vertex in A closest to x.

Size Analysis:

$$(n)|A| + nt$$

$$= n\left(\frac{n \log n}{t} + t\right)$$

$$= n\sqrt{n} \log n \quad \text{for } t = \sqrt{n}$$

Lemma 4: $a_x \in C_t(x)$, $\forall x$ with high probability.

Proof: Due to Lemma 2,

$A \cap C_t(x)$ is non-empty, $\forall x$

Query(x, y):

If $y \in C_t(x)$ report exact distance

Else output $d_G(x, a_x) + d_G(a_x, y)$

Lemma 5: Query(x, y) outputs 3-approximation to $d_G(x, y)$

Proof: If $y \in C_t(x)$, ans is exact.

If $y \notin C_t(x)$ then

$d_G(x, a_x) \leq d_G(x, y)$, so by Lemma 3 answer is 3-approx.

Ques: How to extend this to higher stretch?

Variants: DO for pairs in $V \times V_2$

Oracle construction:

1. $A_2 \leftarrow$ random set of $\frac{n \log n}{t_2}$ vertices
2. Store distance b/w all pairs in $A_2 \times V_2$.
3. For each $x \in V(G)$
   - store dist b/w $\{x\}$ and $C_t(x)$
   - Set $a_x =$ vertex in A_2 closest to x.

Size Analysis:

$$|V_2| \, (\cancel{n}) \, |A_2| + n \, t_2 = \frac{|V_2| \, n \log n}{t_2} + n \cdot t_2$$

This for $t_2 = \sqrt{|V_2|} \log n$ is

$$O(n \sqrt{|V_2|} \log n)$$

Query algorithm & correctness is same as before

—— x —— x ——

(*) What will be stretch & size if you plugin this in your previous construction?

H.W. Resultant stretch $= 1 + 2(3)$

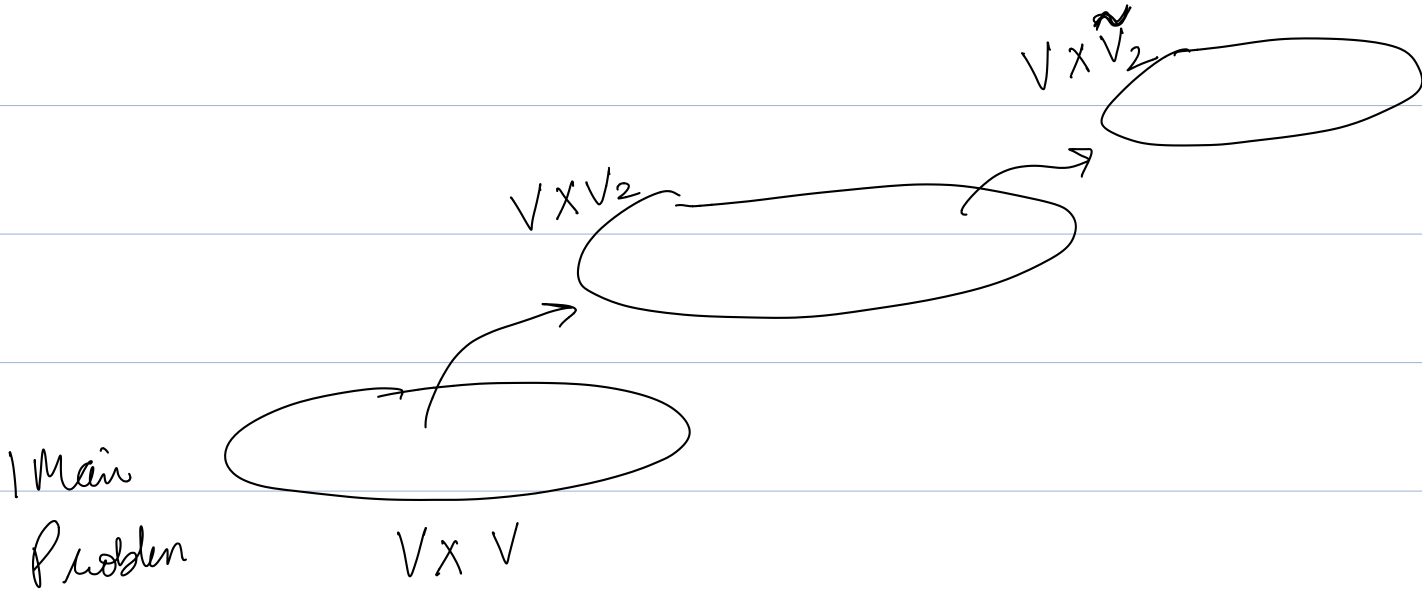
$$\text{old size} = n\,\overset{\frac{n\log n}{t}}{A} + nt$$

new size =

$$O\left(n\sqrt{\frac{n\log n}{t}}\log n + nt\right)$$

$$= O(n^{4/3}\log n) \text{ for } t = n^{1/3}\log n$$

$\boxed{\text{H.W.}}$ What will be the solⁿ with 3- level hierarchy.



To get $O(n^{1+\frac{1}{k}})$ solⁿ, you will have k- level hierarchy

& by more careful analysis you can get stretch $= 2k-1$.

(One of the assignment problem).

———— x ———— x ——

(*) DO of stretch 5, size $= O(n^{4/3} \log n)$

Oracle construction:

1. $A \leftarrow$ random set of $\frac{n \log n}{t}$ vertices
2. Store distance b/w all pairs in $A \times A$.
3. For each $x \in V(G)$,
   - store dist b/w $\{x\}$ and $C_t(x)$
   - Set $a_x =$ vertex in A closest to x.

Size Analysis:

$$|A| \cdot |A| + nt$$

$$= \left(\frac{n \log n}{t}\right)^2 + nt$$

$$= n^{4/3} \log n \quad \text{for } t = n^{1/3} \log n$$

Query(x, y):

- If $y \in C_t(x)$ report exact distance

- If $x \in C_t(y)$ report exact distance

Else output $d_G(x, a_x) + d_G(a_y, y)$
$+ d_G(a_x, a_y)$

H.W.

Lemma : Query(x,y) outputs 5-approximation to $d_G(x,y)$

Proof: If $y \in C_t(x)$, ans is exact.

If $x \in C_t(y)$, ans is exact.

If $y \notin C_t(x)$ & $x \notin C_t(y)$, then

- $d_G(x, a_x) \leq d_G(x,y)$
- $d_G(y, a_y) \leq d_G(x,y)$
- $d_G(a_x, a_y) \leq d_G(a_x, x) + d_G(x,y) + d_G(y, a_y)$
$\leq 3 d_G(x,y)$

So, the stretch $= 1+1+3 = 5.$

$\square$