DIAMETER

$G=(V,E)$ undirected

$$d(x,y) = |\text{shortest path from } x \text{ to } y| = d(y,x) \qquad x,y \in V$$

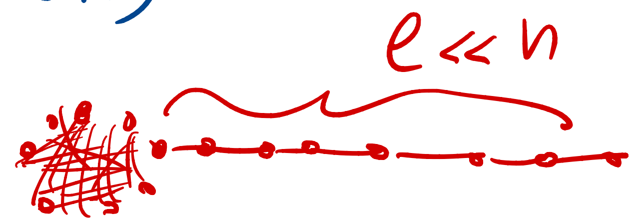
Milgram's experiment: avg distance
$$\sum_{\substack{x,y\in V\\ x\neq y}} \frac{d(x,y)}{\binom{n}{2}}$$

$n = |V|$

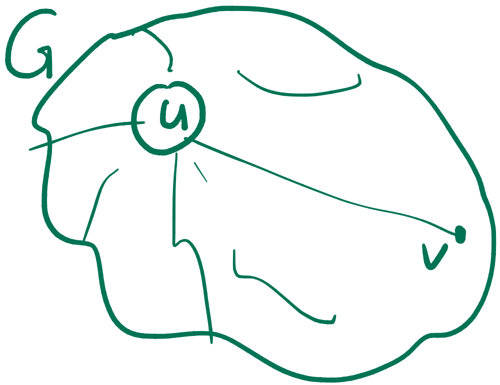
sketches can estimate average distance ⇝ sampling

$$D = \text{diameter of } G = \max_{\substack{x,y\in V\\ x\neq y}} d(x,y)$$

naive sampling does not help ⇝ CLIQUE



BASELINE $\quad BFS(u) \rightarrow ecc(u) = \max_{v \neq u} d(u,v)$



$$D = \max_{\substack{x,y \in V \\ x \neq y}} d(x,y) = \max_{u \in V} ecc(u)$$

Baseline requires $O(n)$ BFSes

$m = |E| \Rightarrow D$ can be computed in $O(nm)$ time

This is still the state of art in the worst case (more info → next class)

Sparse graphs: $m = O(n) \Rightarrow$ BASELINE is $O(n^2)$ time

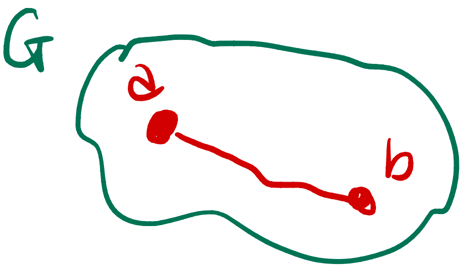
TODAY: Randomized algorithm taking $O(n\sqrt{n}\log n)$ time

GOAL compute an approximation $\tilde{D}$ of the diameter for SPARSE graphs

$$\frac{2}{3} D \leq \tilde{D} \leq D$$

in $O(n\sqrt{n}\log n)$ time w.h.p. $\left(\text{pr} \geq 1 - \frac{1}{n^{\text{const.}}}\right)$

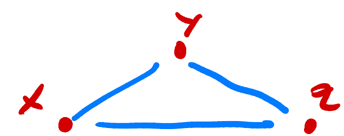
HIGH LEVEL IDEA:



$d(a,b) = D$ (other diametral nodes can exist)

$ecc(a) = ecc(b) = D$

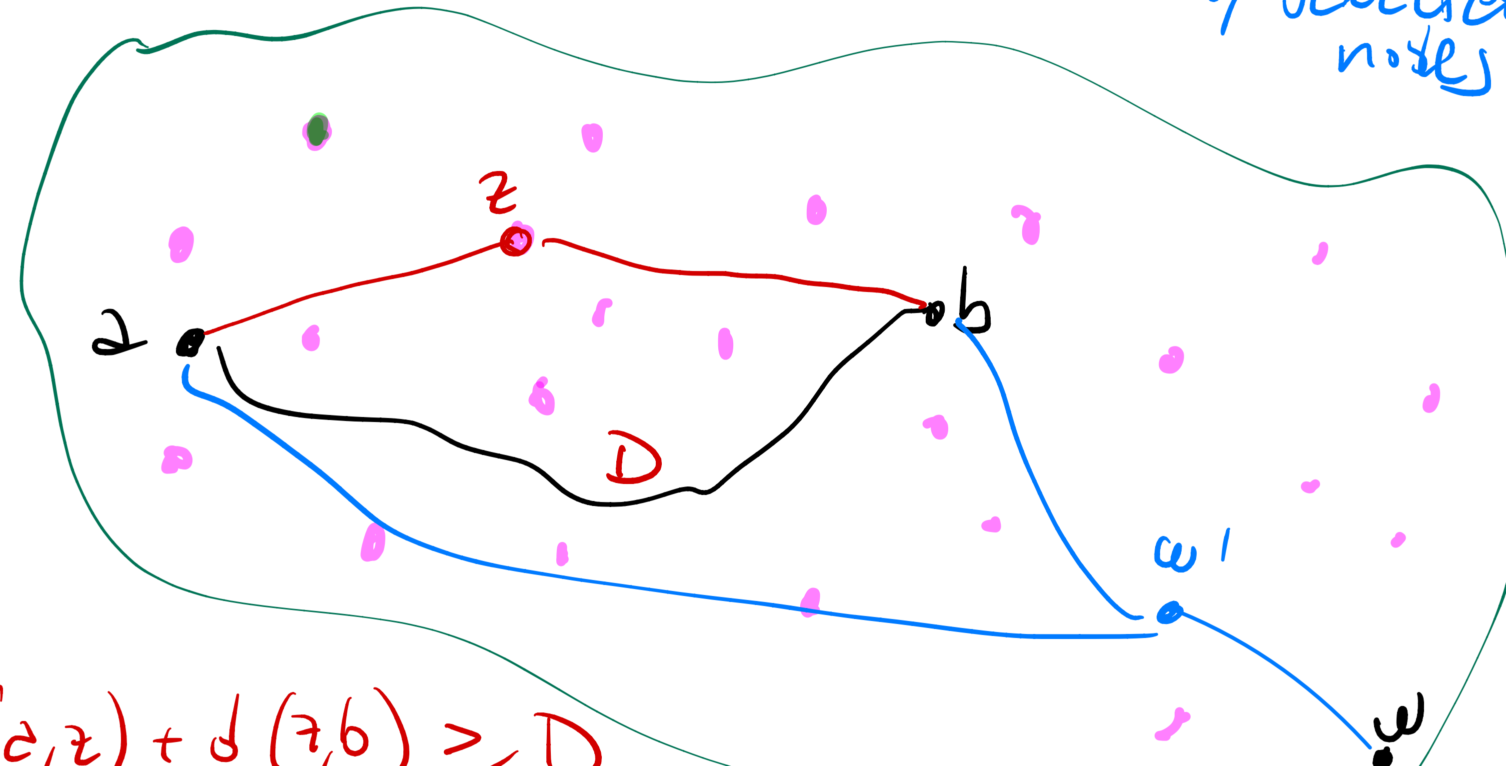
TI = triangle inequality



$d(x,y) + d(y,z) \geq d(x,z)$

$k = \sqrt{n}$ parameter

1. S = set of $O(\sqrt{n}\log n)$ randomly selected nodes



$d(a,z) + d(z,b) \geq D$

2. w = farthest from S

Approximation algorithm

$k = \sqrt{n}$

$G = (V, E)$ sparse
$n = |V|$, $m = |E|$

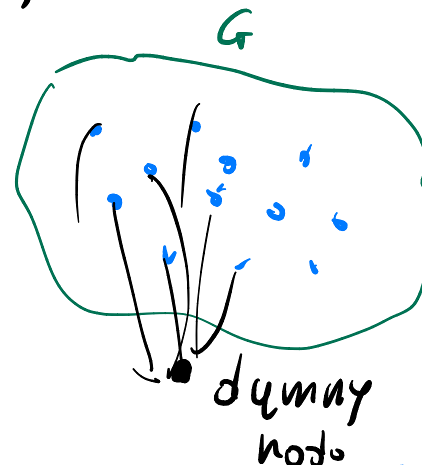
1. S = set of $\alpha \frac{n}{k} \ln n$ uniformly randomly selected nodes from V, $\alpha = \text{const} > 2$

2. ω = farthest node from S

3. $Z = S \cup \underbrace{N_k(\omega)}_{\text{"}\{\text{the first } k \text{ nodes traversed by } BFS(\omega)\}\text{"}}$

"truncated BFS"

return $\tilde{D} = \max_{u \in Z} ecc(u)$



the dummy node is connected to all the node in S (and NOT to those in $V-S$)
+
BFS(dummy node)

α const. > 2
$k = \sqrt{n}$

Running time : $\underbrace{O(\alpha \frac{n}{k} \ln n)}_{\boxed{1}} + \underbrace{O(n)}_{\boxed{2} \text{ sparse } G} + \underbrace{O(|Z| \cdot n)}_{\boxed{3}} = O(n\sqrt{n} \lg n)$

BFS

a] S is a hitting set for each $N_k(u)$, $u \in V$ w.h.p.

$\hookrightarrow S \cap N_k(u) \neq \emptyset$

b] $D = 3h + z \qquad z \in \{0, 1, 2\}$

b.1 $2h+z \leq \tilde{D} \leq D$ when $z = 0, 1 \qquad (S)$

b.2 $2h+1 \leq \tilde{D} \leq D$ when $z = 2 \qquad (N_k(w))$

$\Rightarrow \frac{2}{3} D \leq \tilde{D} \leq D$

d) $Pr[S \text{ hitting set for } N_k(u), u \in V] =$

$1 - \underbrace{Pr[\exists u \in V : S \cap N_k(u) = \phi]}$ =

$p = Pr[S \cap N_k(u) = \phi] = \Big(1 - \underbrace{\frac{|N_k(u)|}{n}}\Big)^{|S|} = \left(1 - \frac{k}{n}\right)^{|S|}$

pr a given element of S is not in $N_k(u)$

$e^x \sim (1+x)$

$= \left(e^{-\frac{k}{n}}\right)^{|S|} = e^{-\cancel{\frac{k}{n}}(\alpha \cancel{\frac{n}{k}} \ln n)} = e^{-\alpha \ln n} = \frac{1}{n^\alpha}$

$|S| = \alpha \frac{n}{k} \ln n$

$= 1 - np = 1 - \frac{n}{n^\alpha} = 1 - \frac{1}{n^{\alpha-1}}, \quad \alpha > 2$

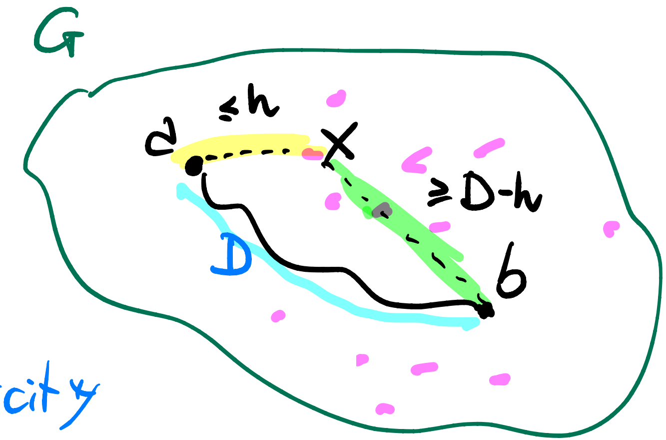
union bound

b) $D = 3h + z$ $\quad z \in \{0,1,2\}$ $\quad D = d(a,b)$

b1) $d(w, S) \leq h$

- Since w is the farthest node from S $\Rightarrow d(a,S) \leq h$
- Take $x \in S$ closest to a

$\Rightarrow ecc(x) \geq d(x,b)$ by definition of eccentricity



we T1 on x

By T1 $\Rightarrow d(x,a) + d(x,b) \geq d(a,b) = D \Rightarrow d(x,b) \geq D - \underbrace{d(x,a)}_{\leq h} \geq D - h$

$\Rightarrow ecc(x) \geq D - h = (3h+z) - h = 2h + z$

$\boxed{T1: a, b, x}$

DONE

b.2) $d(w,S) > h$

- $ecc(w) \geq 2h+z$, we are happy!
- hence, suppose $ecc(w) < 2h+z$

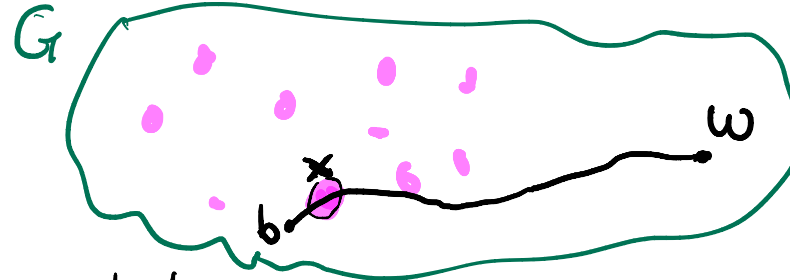
  $\Rightarrow$ consider the shortest path from w to b

  since $ecc(w) < 2h+z \Rightarrow d(w,b) < 2h+z$ (otherwise, larger $ecc(w)$)

  $\Rightarrow$ Since S is a hitting set for $N_h(w)$ w.h.p.

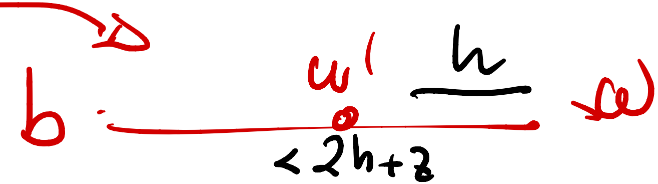
  $\Rightarrow \exists x \in S \cap N_h(w)$

  $\Rightarrow$ since we assume $d(w,S) > h \Rightarrow d(w,x) > h$



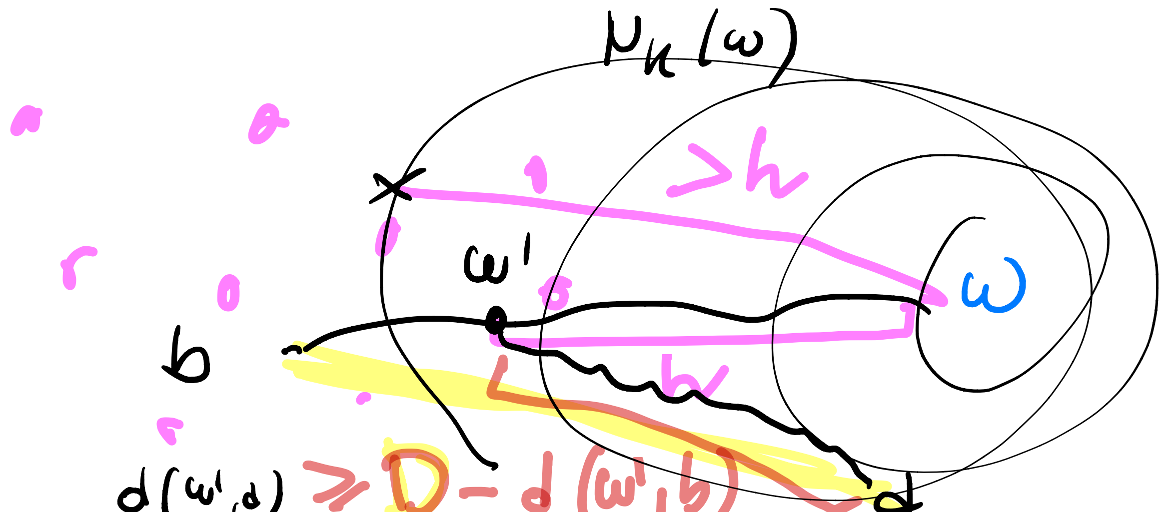
$\exists\, w'$ along the path s.t. $d(w', w) = h$

$\Rightarrow \underbrace{d(w', b) < h+2}$



$\overset{T1}{\Rightarrow} d(w', a) \geq D - d(w', b) \geq D - (h+2-1) = (3h+2) - (h+2-1) = 2h+1$

DONE



$\boxed{T1: a, b, w'}$

$d(w', a) \geq D - d(w', b)$