Graph Spanners

15-859NN
5/5/21

Def 1) $G=(V,E)$ undirected, (unweighted) (today)

2) $H=(V,E_H)$, $E_H \subseteq E$

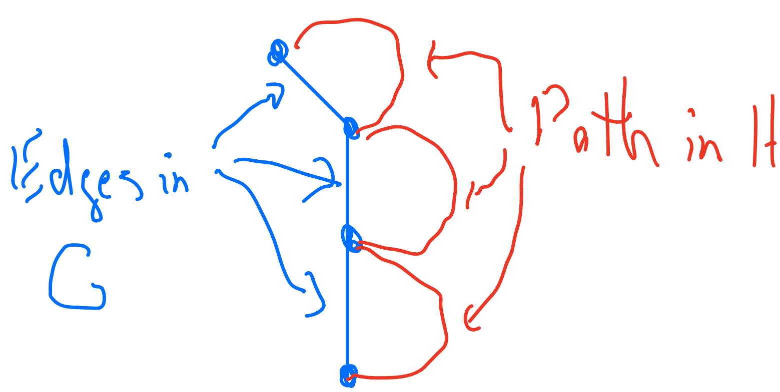
is a <u>k-spanner</u> of G if

$\forall x,y \in V \quad dist_H(x,y) \le k \cdot dist_G(x,y)$

Note: k is called the <u>stretch factor</u>.

Goal: $\text{Min } |E_H|$ for a given factor k.

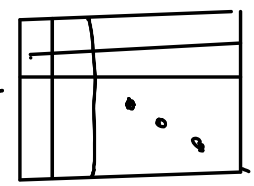
Note: Need only consider stretch of edges.



Known:

<u>Thm</u> $\exists (2k-1)$-spanner with $\frac{1}{2}(n^{1+1/k})$ edges.

Def: The <u>Girth</u> of G is min size cycle.

eg The mesh graph has girth 4. $M_n =$ 

Thus $H \subsetneq M_n$ the stretch ≥ 3.

The missing edge will have stretch 3.

Erdos Girth Conjecture

<u>Conjecture</u> (Erdos) $\exists G=(V,E)$

1) $|E| = \Omega(n^{1+1/k})$
2) $\mathrm{Girth}(G) \geq 2k+1$

Thus Thm is worst case optimal.

Today: $O(m)$ time algor. constructing $(4k+1)$-spanner with $O(n^{1+1/k})$ edges.

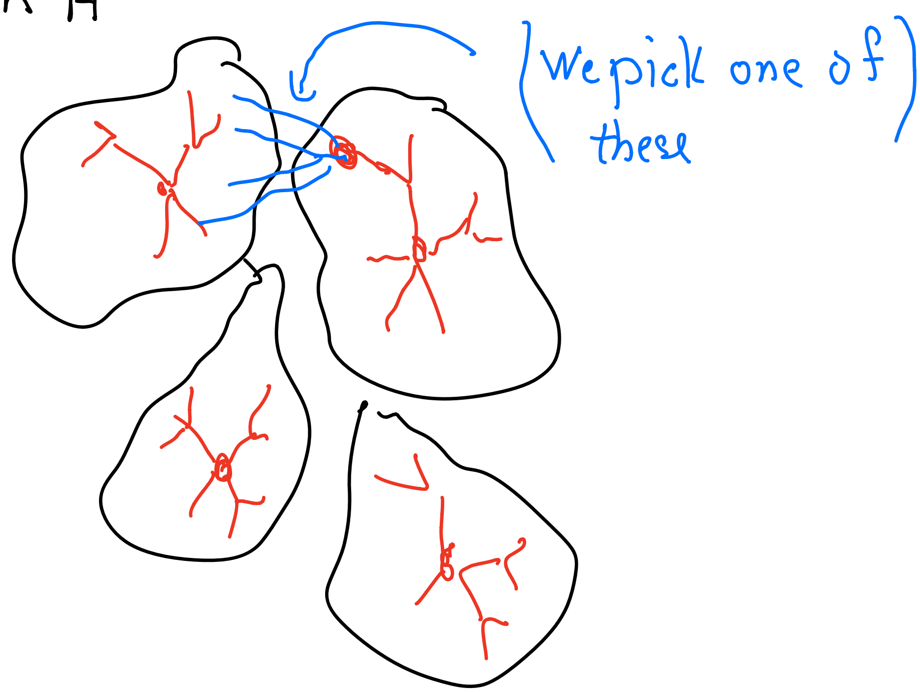
We settle for expected stretch & size.

Todays Spanner Algorithm

Procedure: Spanner(G,k)

1) Set $\beta = \log n / 2k$ (thus $2k = \log n / \beta$)
2) $\{C_1, \dots, C_t\}$ = ExpDelay(G, β) (clusters)
3) For each C_i add BFS forest to H.
4) For each boundry vertex v
   add <u>one</u> edge to H for each adj cluster.

Return H



Ball Growing Using Exponential Delay

Procedure: ExpDelay(G, β)

1) Each vertex $v \in V$ draws $X_v \sim Exp(\beta)$
2) Each $v \in V$ computes $S_v = X_{max} - X_v$
3) Each $v \in V$ starts a BFS at time S_v
   a) If v is not owned at time S_v then v owns v.
   b) Each v is owned by first arrival vertex.

Def: $\bar{T}_i = X_{max} - T_i = X_i - dist(v_i, c)$

(early arrival) (Owner has max $\bar{T}_i$)

Since ExpDelay(G,β) is $O(m)$ time so is Spanner(G,k) $O(m)$ time.

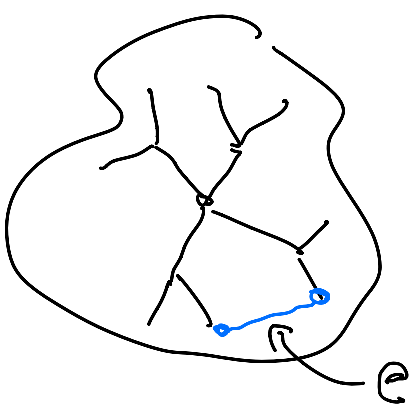
To Show:

1) Expected stretch is $4k+1$
2) Expected size of H is $O(n^{1+1/k})$.

We start with stretch

Note: Need only consider stretch of edges.

(Case 1) e is internal to a cluster.



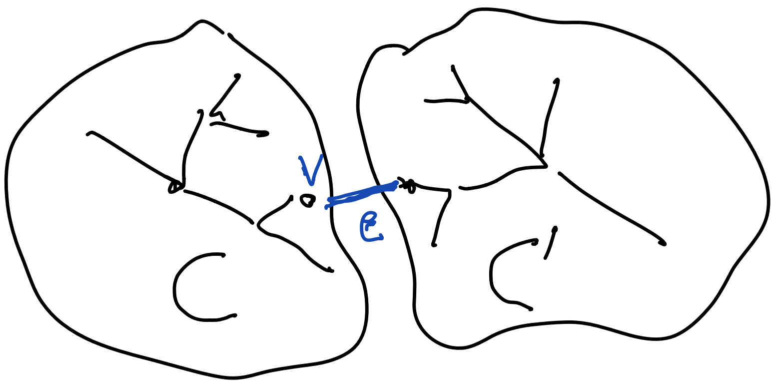
$\text{str}(e) \le 2\,\text{radius}(C)$

$$\mathbb{E}[\text{Maxradius}(C)] \le \frac{\ln n}{\beta} = 2k$$

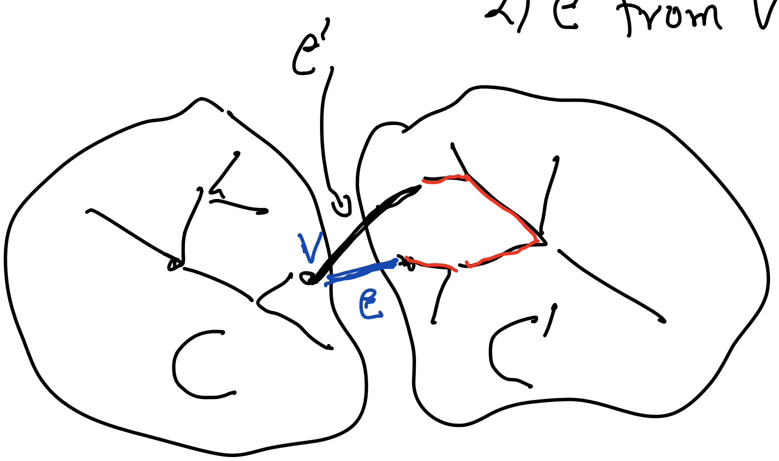
$$\mathbb{E}[\text{str}(e)] = 4k$$

(Case 2) Edge e is between C & C'
and e is added by bdry vertex v.

(Case a) e is only edge from v to C'.
In this case $e \in E_H$.



(Case b) $\exists e' \neq e$ 1) $e' \in E_H$
2) e' from v to C'



$str(e) \leq dia(C') + 1$; $\mathbb{E}[str(e)] \leq 4k+1$

The Expected size of E_H

Two types of edges

1) Internal to a cluster (Forest Edges): at most $n-1$ such edges.

2) Intercluster edges:
   #boundry nodes $\leq n$
   #clusters common to a bdry node.

Let $v \in V$ consider random variable
C_v = # distinct clusters common to v

<u>Thm</u> $\mathbb{E}[C_v] \leq e^{2\beta}$

Thus Expected number of intercluster
$\leq n \cdot e^{2\beta} = n\, e^{\frac{\ln n}{k}} = n^{(1+1/k)}$. $\beta = \frac{\ln n}{2k}$

We need only prove Thm.

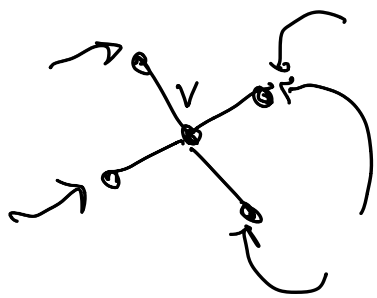
Question: How many clusters will a vertex see (share an edge with).

1) It will belong to one cluster.

2) How many edges to distinct clusters.

Back to horse racing.

Consider early arrivals to v.

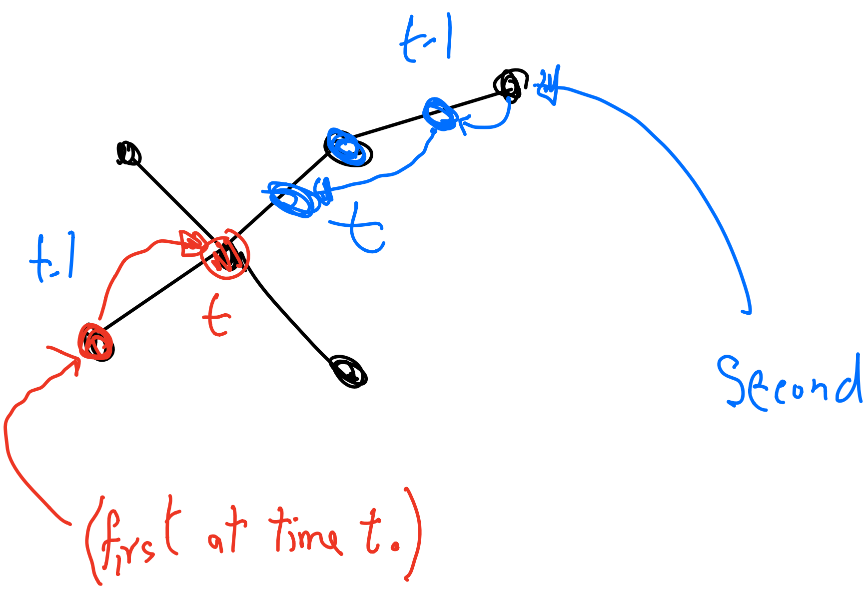


Note: A vertex must arrive within 2 units of first arrival to possibly own a neighbor of v.

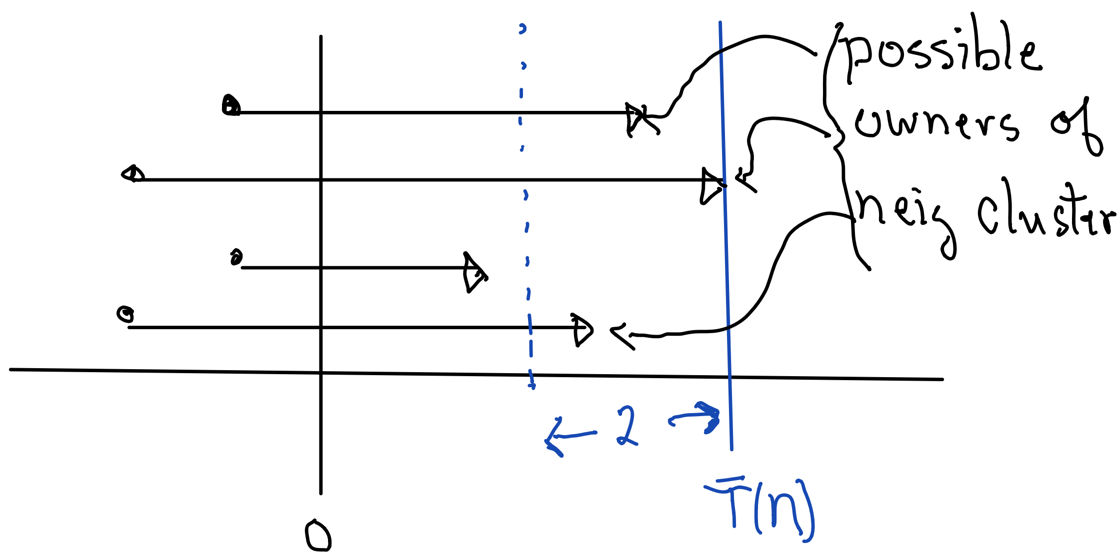
t = time of first arrival to x

Gap is 1½

Time t-1



Possible Neighboring Clusters to v.



We prove a more general thm:

Suppose B is a ball of G with
1) center v.
2) diameter d.

Consider random variable

$$C_B = \text{Cluster}(B) = |\{\text{cluster} \mid \text{cluster} \cap B \neq \emptyset\}|$$

<u>Thm</u> $\mathbb{E}[C_\beta] \leq e^{d\beta}$

$A_\beta \equiv$ number of arrivals to v within d time of first to v.

<u>Note:</u> $C_\beta \leq A_\beta$

<u>Claim:</u> $\text{Prob}\left[A_\beta \geq t\right] = \left(1 - e^{-d\beta}\right)^{t-1}$

<u>pf of claim</u>

Consider time of tth early arrival

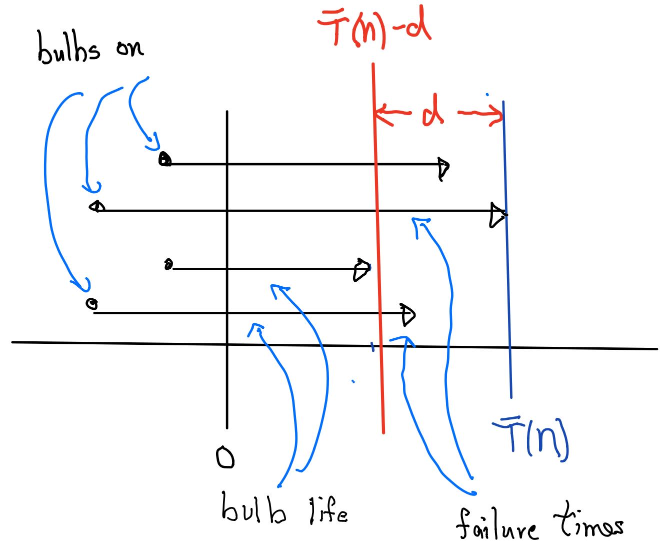
i.e. $T_{(n-t+1)} = T_t$

We give two proofs.

The first we consider time $T_{(n-t+1)}$ and we look forward in Time

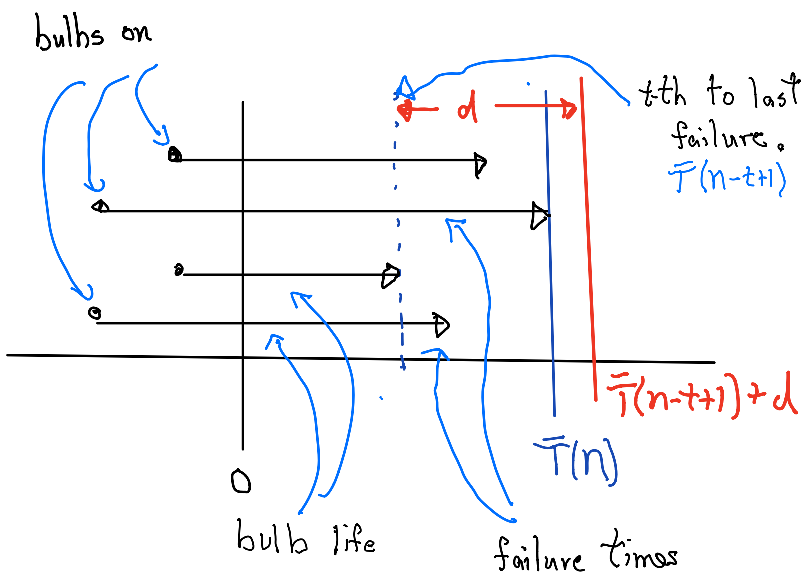
Lets use the light bulb analogy.

Formally: the event $A_\beta \geq t$ is true if the time-window $[\bar{T}(n)-d, \bar{T}(n)]$ has at least t failures.



<u>That is:</u> $\bar{T}(n) - \bar{T}(n-t+1) \leq d$ iff $A_\beta \geq t$

<u>View 2:</u> From time $\overline{T}(n-t+1)$ the remaining $t-1$ blubs must fail by time $\overline{T}(n-t+1)+d$.



At time $\overline{T}(n-t+1)$ there are $t-1$ memory less iid exponential random variables, one for each first $t-1$ early arrivals.

Each must take a value $\leq d$ which will happen with prob $(1-e^{-d\beta})$

By independence we get $(1-e^{-d\beta})^{t-1}$ is the prob they all are $\leq d$. $\square$

<u>Proof 2:</u> Consider the order statistics

$$\tilde{T}_{(1)} \leq \cdots \leq \tilde{T}_{(n-1)} \leq \tilde{T}_{(n)}$$

Consider random variables $GAP_i = \tilde{T}_{(n-i+1)} - \tilde{T}_{(n-i)}$

ie $GAP_1 = \tilde{T}_{(n)} - \tilde{T}_{(n-1)}$

In Probability-101 lecture we showed that

$$GAP_i \approx Exp(i\beta)$$

Thus $\text{Prob}[A_\beta \geq t] = \text{Prob}\left[\sum_{i=1}^{t-1} GAP_i \leq d\right]$

Lets do case $t=3$

$$f(x) = \text{Prob}[GAP_1 + GAP_2 = x]$$

Since Gaps are independent.

$$f(x) = \int_0^x \beta e^{-\beta y} \cdot 2\beta e^{-2\beta(x-y)} dy$$

$$= 2\beta e^{-2\beta x} \int_0^x \beta e^{\beta y} dy$$

$$= 2\beta e^{-2\beta x}\left[e^{\beta x} - 1\right]$$

$$f(x) = 2\beta e^{-\beta x} - 2\beta e^{-2\beta x}$$

$$F(y) = \int_0^y f(x) = 2\left(1 - e^{-\beta y}\right) - \left(1 - e^{-2\beta y}\right)$$

$$= 1 - 2e^{-\beta y} + e^{-2\beta y}$$

$$= \left(1 - e^{-\beta y}\right)^2$$

setting $y = d$ and $t = 3$ we get

$$\text{Prob}\left[A_\beta \geq 3\right] = \left(1 - e^{-\beta d}\right)^2$$

(this argument generalized to t)

$$\mathbb{E}[A_\beta] = \sum_{t=0}^{\infty} \mathrm{Prob}[A_\beta \geq t] = \sum_{t=1}^{\infty} (1-e^{-d\beta})^{t-1}$$

$$= \frac{1}{1-(1-e^{-d\beta})} = e^{d\beta}$$

<u>QED</u>

$$\boxed{\text{We use fact that } \sum_{i=0}^{\infty} \alpha^i = \frac{1}{1-\alpha}}$$