Graph Spanners

15-750
2/20/19

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

2) $H \subseteq G$ is a <u>k-spanner</u> of G

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

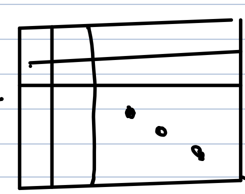
Note: k is called the stretch factor.

Goal: $\min |E_H|$ for a given factor k.

Known:

Thm $\exists (2k-1)$-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.

Erdos Girth Conjecture

__Conjecture__ (Erdos) $\exists G=(V,E)$

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

Thus Thm is worst case optimal.

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

we settle for expected stretch & size.

Procedure: Spanner(G,k)

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

Return H

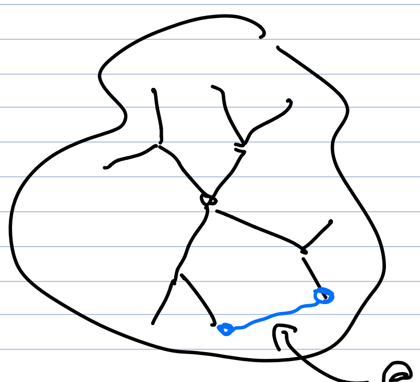
Since ExpDelay(G,β) is $O(m)$ time so is Spanner(G,k) $O(m)$ time.

To Show:

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

We start with stretch

(Case 1) e is internal to a cluster.



$str(e) \leq 2\,radius(C)$

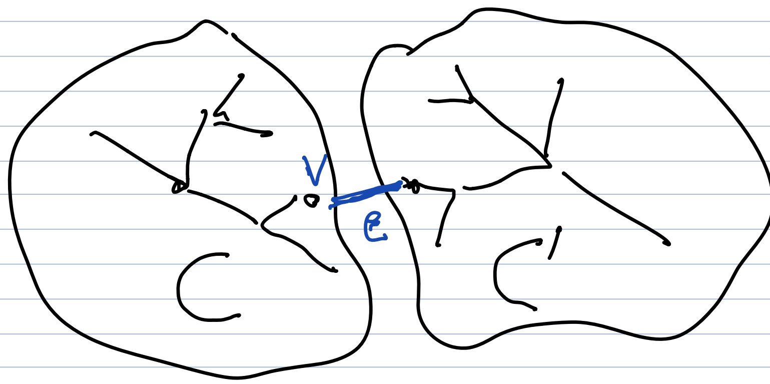
$Exp[rad(C)] = \frac{\ln n}{\beta} \approx 2k$

$Exp[str(e)] = 4k$

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

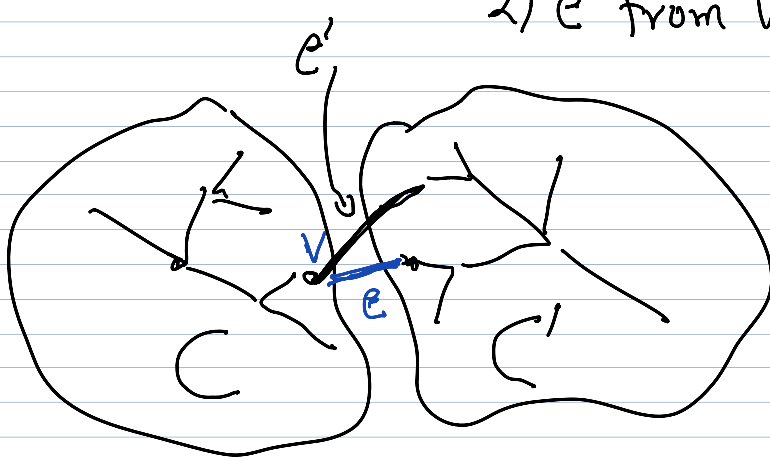
(Case a) e is only edge from v to C'.

In this case $e \in E_H$.



(Case b) $\exists e' \neq e$ 1) $e' \in E_H$

2) e' from v to C'



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

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

Thm $E[C_v] \leq e^{2\beta}$

Thus Expected number of intercluster

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

We need only prove Thm.

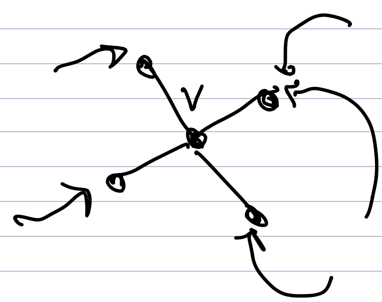
Question: How many clusters will a vertex see (share an edge with).

1) It will belong to one cluster.

2) How many edges to distinct clusters.

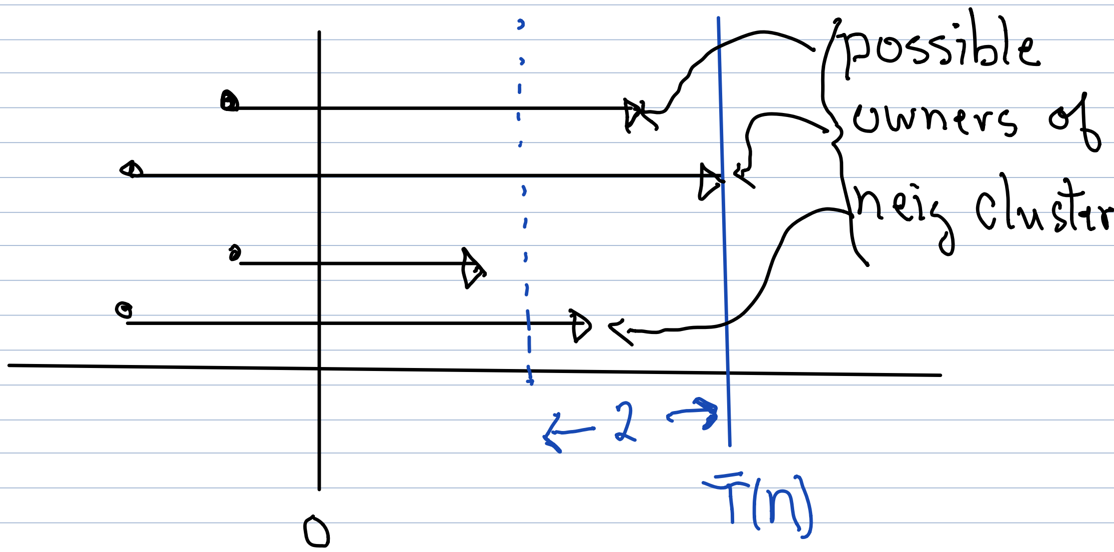
Back to horse racing.

Consider early arrivals to v.



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

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> $\text{Exp}[C_\beta] \leq e^{d\beta}$

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

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

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

<u>pf of claim</u>

Consider time of tth early arrival

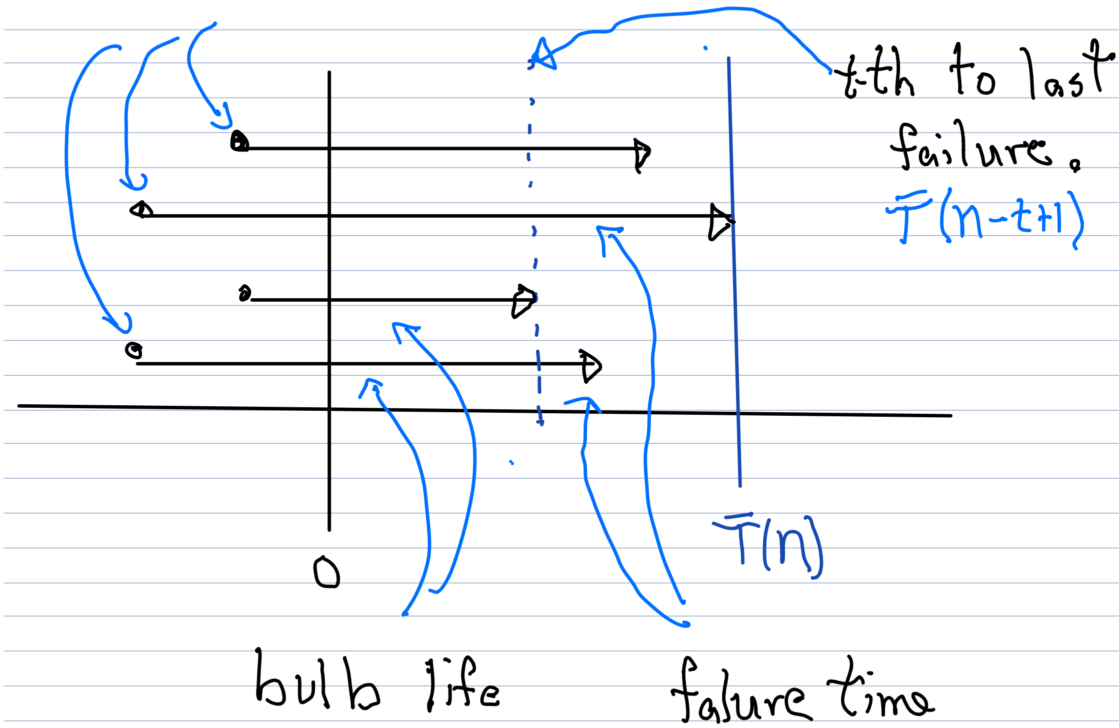
i.e. $T_{(n-t+1)} = T_t$

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.

When we turn on the bulb.



At time T_t there are $t-1$ memory less iid exponential random variables, one for each first $t-1$ early arrivals.

Each must take a value $\leq d$. The prob $= (1-e^{-d\beta})$

By independence we get $(1-e^{-d\beta})^{t-1}$.

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

Lets do case $t=3$

$$f(x) = Prob[GAP_1 + GAP_2 = x]$$

$$f(x) = \int_0^x \beta e^{-\beta y} \cdot 2\beta e^{-2\beta(x-y)} dy \quad (\text{Ind } GAP_3)$$

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

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

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

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

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

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

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

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

<u>QED</u>

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