

Lecture 13: Graph spanners via low diameter decomposition

Lecturer: Gary Miller

Scribes: Jueheng Zhu & Tianyi Yang

1 Graph Spanners

Definition 1.1. Let $G = (V, E)$ be an undirected, unweighted graph. Then $H \subseteq G$ is a k -**spanner** of G if

$$\forall x, y \in V, \quad \text{dist}_H(x, y) \leq k \cdot \text{dist}_G(x, y)$$

where $\text{dist}_G(x, y)$ denotes the length of the shortest path between x and y on G . Here k is called the **stretch factor**.

We are interested in finding the k -spanner with the least number of edges for a given stretch factor k . We next state a known theorem on the stretch and size of a spanner..

Theorem 1.2. $\exists (2k - 1)$ -spanner with $1/2(n^{1+1/k})$ edges.

Definition 1.3. The **girth** of a graph G is size of its smallest cycle.

Example 1.4. The mesh graph has girth 4. Thus for any $H \subsetneq M_n$, the stretch ≥ 3

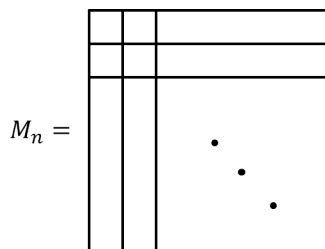


Figure 1: Mesh graph of size n

1.1 Erdos Girth Conjecture

Conjecture 1.5. There exists $G = (V, E)$ such that

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

Note that if the above conjecture is true, Theorem 1.2 is worst case tight.

Here we informally prove a weaker version of Theorem 1.2, which is stated as the following Lemma.

Lemma 1.6. There exists an $O(m)$ algorithm constructing $(4k + 1)$ -spanner with $O(n^{1+1/k})$ edges.

We settle for expected stretch & size. In the homework we will remove the expectation and give an efficient algorithm for finding a spanner.

Algorithm To construct $\text{Spanner}(G, k)$

1. Set $\beta = \log \left(\frac{n}{2k} \right)$
2. Let $\{C_1, \dots, C_t\} = \text{ExpDelay}(G, \beta)$ (The clusters generated)
3. For each C_i , add its BFS forest to H
4. For each boundary vertex v , add one edge from v to each adjacent cluster.
5. Return H

Proof of Lemma 1.6

Proof. First, since $\text{ExpDelay}(G, \beta)$ is $O(m)$, so is $\text{spanner}(G, k)$. It remains to show that the expected stretch is $4k + 1$ and the expected size of H is $O(n^{1+1/k})$ (Recall here we are only concerned with expectation). We start with stretch. For an edge e , we define $\text{str}(e)$ to be the stretch for an single edge e . It then suffices to show that the expected stretch is $\text{str}(e) \leq 4k + 1$ for all e in the edge set of $\text{Spanner}(G, k)$.

- (Case 1) e is internal to a cluster

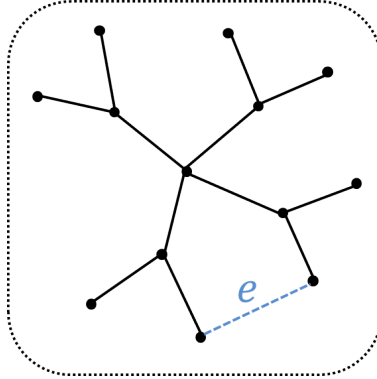


Figure 2: e is internal to a cluster

Then $\text{str}(e) \leq 2\text{radius}(C)$. Recall $\mathbb{E}[\text{radius}(C)] = \frac{\ln n}{\beta} = 2k$. Therefore

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

- (Case 2a) e is between C and C' and e is added to H by boundary vertex v .

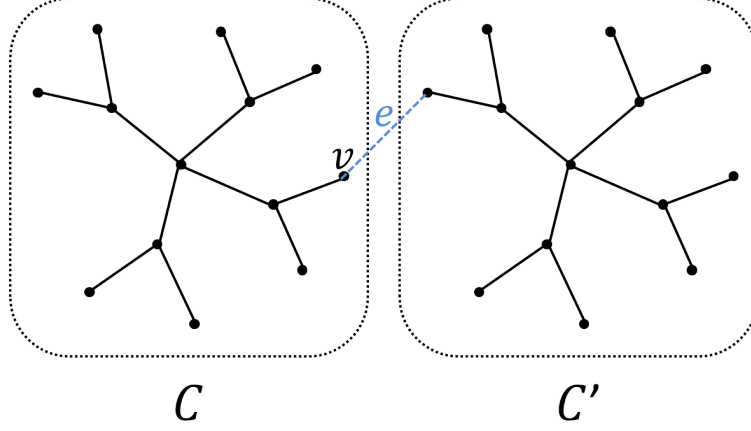


Figure 3: e connects v and C' and $e \in E_H$

In this case $e \in E_H$ and $str(e) = 1$.

- (case 2b) e is between C and C' and e is not added to H by boundary vertex v .

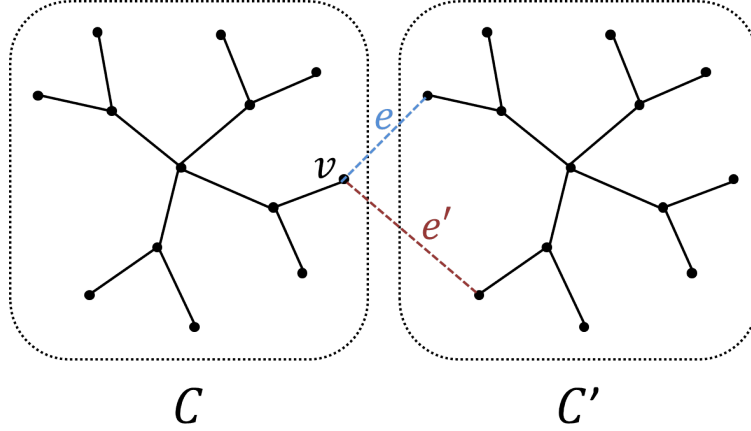


Figure 4: e connects v and C' and $e \in E_H$

Then by the procedure, there must exist e' from v to C' . Hence $str(e) \leq dia(C') + 1$. Thus $E[str(e)] \leq 4k + 1$

Therefore expected stretch is no more than $4k + 1$.

We now analyze the expected size of E_H . There are two types of edges in E_H :

1. edges internal to a cluster. There will be at most $n - 1$ of these since the union of all clusters is a forest.
2. Inter-cluster edges. The expected amount of these depends on the number of boundary nodes and the number of distinct clusters common to each boundary nodes. The former is bounded by n and we claim that the latter in expectation is bounded by $e^{2\beta}$. As a result

$$\mathbb{E}[\text{Number of inter-cluster edges}] \leq ne^{2\beta} = ne^{\frac{\ln n}{k}} = n^{1+1/k}$$

It remains to prove the claim, which we defer to the following section. □

Let $v \in V$. Consider the random variable

$$C_v = \text{Number of distinct clusters common to } v$$

Then our claim can be expressed as the following theorem:

Theorem 1.7. $\mathbb{E}[C_v] \leq e^{2\beta}$

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 .

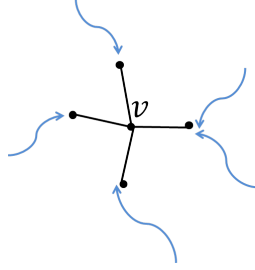


Figure 5: Arrivals to vertex v

An early arrival must arrive within 2 units to possibly own a neighbor of v

Possible Neighboring clusters to v :

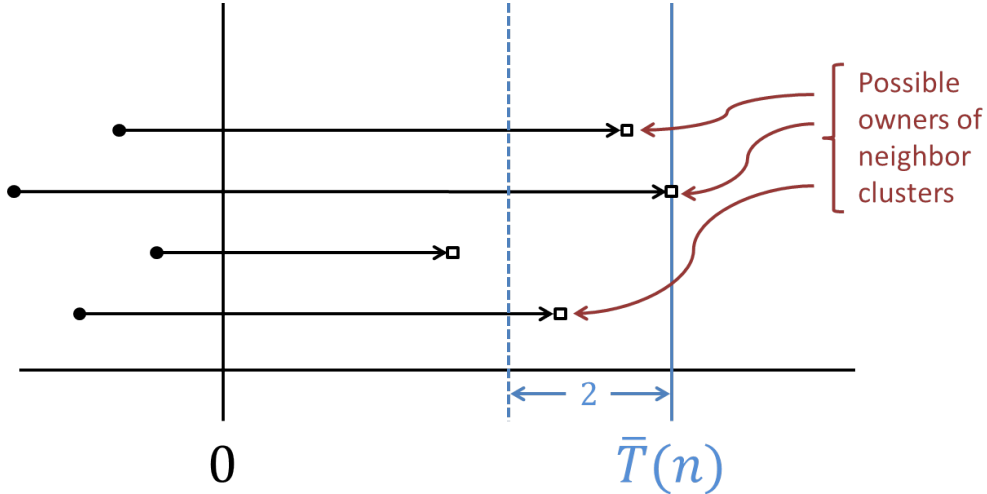


Figure 6: neighboring clusters of v according to arrival times

We prove a more general theorem: Suppose B is a ball of G with center v , diameter d . Consider random variable $C_B = \text{Cluster}(B) = |\{\text{cluster} | \text{cluster} \cap B \neq \emptyset\}|$

Theorem 1.8. $\mathbb{E}[C_B] \leq e^{d\beta}$

Let A_B = number of arrivals within d time of first. Note $C_B \leq A_B$. This is because for each cluster not disjoint with B , its center must arrive at V within d time from the first.

Claim 1.9. $\text{Prob}[A_B \geq t] = (1 - e^{-d\beta})^{t-1}$

Proof of claim 1.9

Proof. Let's go back to the light bulb analogy. Recall in this analogy the last failure corresponds

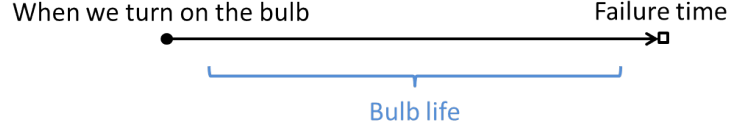


Figure 7: Light bulb analogy: legend

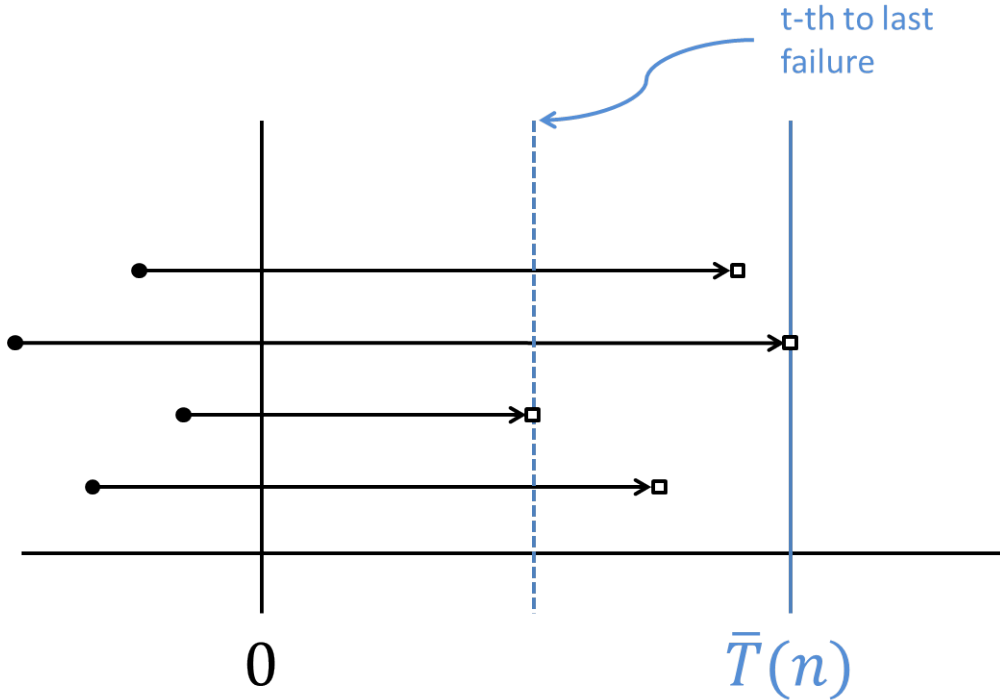


Figure 8: Light bulb analogy: graph

to the first arrival. Let $\bar{T}(k)$ be the random variable denoting the time at which the k bulbs have failed.

Note the following equivalences

- $A_B \geq t$
- $\Leftrightarrow$ There are at least t failures between $\bar{T}(n) - d$ and $\bar{T}(n)$, excluding the last failure.
- $\Leftrightarrow$ The t -th to last failure occurs after $\bar{T}(n) - d$. That is, $\bar{T}(n - t + 1) \geq \bar{T}(n) - d$.
- $\Leftrightarrow \bar{T}(n - t + 1) + d \geq \bar{T}(n)$.

By the memoryless property of the $t - 1$ light bulbs that have not yet failed that are i.i.d exponential random variables, we have

$$Prob [\bar{T}(n - t + 1) + d \geq \bar{T}(n)] = (1 - e^{-d\beta})^{t-1}$$

Effectively we are treating the t -th to last failure as the new starting time and considering only the remaining $t - 1$ light bulbs. The fact that the last failures among these $t - 1$ light bulbs occur before d implies all $t - 1$ light bulbs failure before time d . Since the failure times are i.i.d and exponential we have

$$P [d \geq \bar{T}(t)] = (1 - e^{-d\beta})^{t-1}$$

□

Theorem 1.8 then follows Claim 1.9 because

$$\begin{aligned} \mathbb{E}[C_B] &\leq E[A_B] \\ &= \sum_{t=1}^{\infty} Prob[A_B \geq t] \\ &= \sum_{t=1}^{\infty} (1 - e^{-d\beta})^{t-1} \\ &= \frac{1}{1 - (1 - e^{-d\beta})} \\ &= e^{d\beta} \end{aligned}$$