Graph Spanners

15-750
2/20/19

**Def:** 1) \( G = (V, E) \) undirected (unweighted)
2) \( H \subseteq G \) is a \( k \)-spanner of \( G \)
   \( \forall x, y \in V \); \( \text{dist}_H(x, y) \leq k \cdot \text{dist}_G(x, y) \)

**Note:** \( k \) is called the **stretch factor**.

**Goal:** \( \text{Min} |E_H| \) for a given factor \( k \).

**Known:**

**Thm:** \( \exists (2k-1) \)-spanner with
\( \frac{1}{2} (n^{1+1/k}) \) edges.

**Def:** The **Girth** of \( G \) is min size cycle.

eg The mesh graph \( M_n \) has girth 4.

Thus \( H \not\sim M_n \) the stretch \( \geq 3 \).
Erdos Girth Conjecture

Conjecture (Erdos) \exists G=(V,E)
1) |E| = \Omega(n^{1+\frac{1}{k}})
2) Girth(G) \geq 2k+1

Thus Thm is worst case optimal.

Today: \tilde{O}(m) algorithm constructing (4k+1)-spanner with \tilde{O}(n^{1+\frac{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, \ldots, C_{\beta}\} = \text{Exp Delay}(G, \beta) (clusters)
3) For each C_i add BFS forest to H.
4) For each boundary vertex v
   add one edge to H for each adj cluster.
Return H
Since $\text{Exp} \text{Delay} (G, \beta)$ 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 $O(N^{1+1/4})$.

We start with stretch

(Case 1) $e$ is internal to a cluster.

\[ \text{str}(e) \leq 2 \text{radius}(C) \]

\[ \text{Exp} [\text{rad}(C)] = \frac{\ln N}{\beta} > 2k \]

\[ \text{Exp} [\text{str}(e)] = 4k \]
(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 $C'$

$\text{str}(e) \leq \text{dia}(C') + 1 \Rightarrow \mathbb{E} \left[ \text{str}(e) \right] \leq 4k + 1$
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.

Let $V \in V$ consider random variable

$C_V = \# \text{ distinct clusters common to } v$

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

Thus Expected number of intercluster

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

We need only prove Thm.
Question: How many clusters will a 6 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

\[
C_B = \text{Cluster}(B) = \left| \{ \text{cluster} \mid \text{cluster} \cap B \neq \emptyset \} \right|
\]
\[ \text{Thm} \quad \text{Exp}[C_\beta] \leq e^{d\beta} \]

\[ A_\beta = \text{number of arrivals to } V \text{ within } d \text{ time of first to } V. \]

\[ \text{Note: } C_\beta \leq A_\beta \]

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

pf of claim.

Consider time of \( t \)th 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
Let's 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 $\leq \delta$. The prob $\approx (1 - e^{-d\beta})$

By independence we get $(1 - e^{-d\beta})^{t-1}$.
Proof 2: Consider the order statistics

\[ T_1 \leq \cdots \leq T_{n-1} \leq T_n \]

Consider random variables \( \text{GAP}_i = \bar{T}_{(n-i)} - \bar{T}_{(n)} \)

\[ \text{GAP}_i = \bar{T}_{(n)} - \bar{T}_{(n-i)} \]

In Probability-101 lecture we showed that \( \text{GAP}_i \sim \text{Exp}(\lambda) \)

Thus \( \text{Prob} \left[ A_{\beta} = t \right] = \text{Prob} \left[ \sum_{i=1}^{t-1} \text{GAP}_i \leq d \right] \)

Let's do case \( t=3 \)

\[ f(x) = \text{Prob} \left[ \text{GAP}_1 + \text{GAP}_2 = x \right] \]

\[ f(x) = \int_0^x \beta e^{-\beta y} \cdot 2\beta e^{-2\beta(x-y)} \, dy \quad (\text{Ind \ GAP}_3) \]

\[ = 2\beta e^{-\beta x} \int_0^x \beta e^{\beta y} \, dy \]

\[ = 2\beta e^{-\beta x} \left[ \frac{e^{\beta y}}{\beta} \right]_0^x \]

\[ = 2\beta e^{-\beta x} \left( e^{\beta x} - 1 \right) \]
\[
\begin{align*}
  f(x) &= 2\beta e^{-\beta x} - 2\beta e^{-2\beta x} \\
  F(y) &= \int_0^y f(x) = 2(1-e^{-\beta y}) - (1-e^{-2\beta y}) \\
  &= 1 - 2e^{-\beta y} + e^{-2\beta y} \\
  &= (1-e^{-\beta y})^2 \\
  &\quad \text{setting } y = d
\end{align*}
\]

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

\[
= \frac{1}{1-(1-e^{-\beta})} = e^{\beta}_{\text{d}\beta}
\]

\[
\text{QED} \quad \text{We use fact that } \sum_{i=0}^{\infty} \alpha_i \frac{1}{1-\alpha} = \frac{1}{1-\alpha}
\]