Advanced Topics in Theoretical Computer Science

University of BRISTOL

Lecture 10

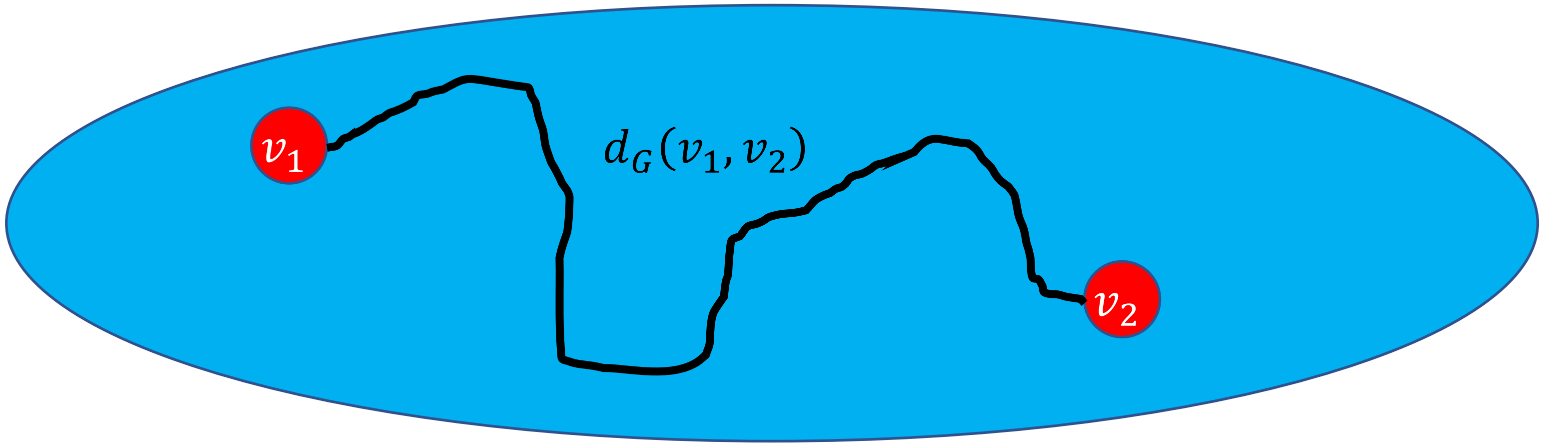
Spanners and Distance Estimation

Distance Estimation Problem

Task: Given $v_1, v_2 \in V$, what is the distance

$$d_G(v_1, v_2) = \min\{\mathrm{len}(\pi) : \pi \text{ is a path from } v_1 \text{ to } v_2\}$$

between v_1 and v_2 in G?



Can we compute $d_G(v_1, v_2)$ in a streaming fashion?

Distance Estimation Problem

Exact Distance Estimation in Data Streams:

Deciding whether two vertices are at distance exactly 3 (or at a larger distance) requires $\Omega(n^2)$ space, i.e., no sublinear space streaming algorithm exists.

Approximate Distance Estimation:

Given $v_1, v_2 \in V$, output estimated distance $\widehat{d(}v_1, v_2)$ such that

$$d_G(v_1, v_2) \leq \widehat{d(}v_1, v_2) \leq t \cdot d_G(v_1, v_2)$$

for some small value $t > 1$.

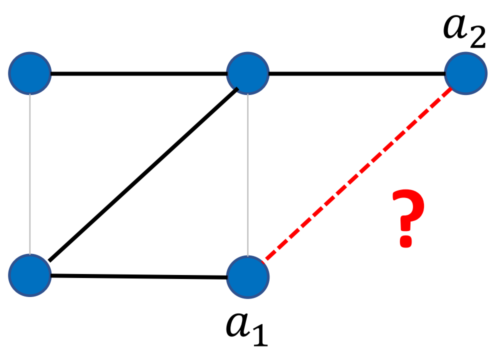
What is the trade-off between space and approximation guarantee t?

Algorithm

Distance Estimation Algorithm:

Input: Vertices v_1, v_2, parameter t

1. $H \leftarrow \emptyset$
2. While(stream not yet empty)
   a) Let $e = a_1 a_2$ be next edge in stream
   b) **if** ($d_H(a_1, a_2) \geq t + 1$) **then** $H \leftarrow H \cup \{e\}$
3. **return** $d_H(v_1, v_2)$



Example: $t = 2$
H: black edges

$d_H(a_1, a_2) := d_{(V,H)}(a_1, a_2)$: Shortest path between a_1 and a_2 using only edges from H

Analysis:

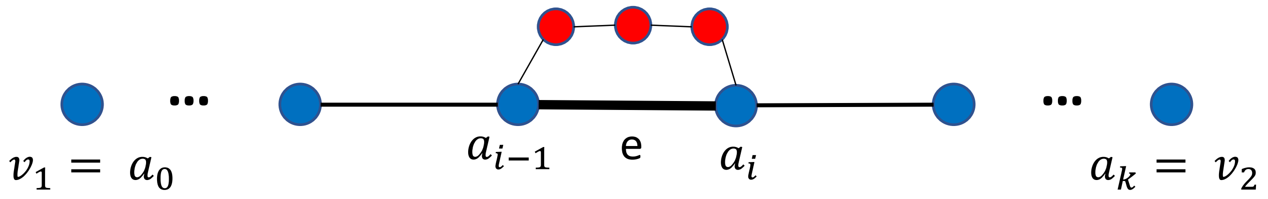
1. Approximation Guarantee?
2. Space?

Algorithm: Approximation Guarantee

Theorem. $d_G(v_1, v_2) \le d_H(v_1, v_2) \le t \cdot d_G(v_1, v_2)$, for any pair of vertices v_1, v_2

Proof.

- Let π be a shortest v_1, v_2 path in G and let $v_1 = a_0, a_1, a_2, \dots, a_k = v_2$ be the vertices in π in order. Then, $d_G(v_1, v_2) = k$.



- Consider any index i and let $e = a_{i-1}a_i$. Then:
  If $e \in H$ then $d_H(a_{i-1}, a_i) = 1$,
  If $e \notin H$ then at time when e appeared in the stream, we had $d_{H'}(a_{i-1}, a_i) \le t$, where H' was the value of H at that time. Thus: $d_H(a_{i-1}, a_i) \le t$.
- We therefore obtain:

$$d_H(v_1, v_2) \le \sum_{i=1}^{k} d_H(a_{i-1}, a_i) \le k \cdot t = t \cdot d_G(v_1, v_2) .$$

$\square$

Spanners

Observe:

- Theorem on last slide holds for any pair of vertices v_1, v_2
- In particular, the construction of set H is independent from v_1, v_2
- We can therefore estimate the distance of any pair of vertices using edges H!

Spanner:

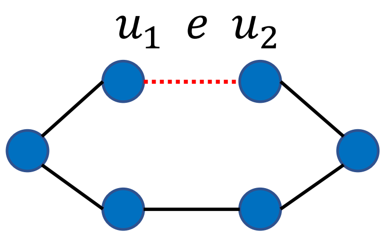
- Set H can be regarded as a data structure that allows us to query approximate distances
- Such a data structure is known as a *t-Spanner*, where t indicates the approximation factor

Algorithm: Space Complexity

Space Complexity:

Clearly $\mathrm{O}(|H|\ \log n)$, but how large can H get?

→ Reduce to a question in extremal combinatorics!



Lemma. (V, H) does not contain any cycle of length at most $t + 1$.

Proof. *(by contradiction)*

- Suppose that (V, H) contained a cycle C of length at most $t + 1$
- Let $e = u_1 u_2$ be the last edge inserted into H that completed the cycle C
- Then, since C is of length at most $t + 1$, $d_{H'}(u_1, u_2) \le t$, where H' was the value of H when e was inserted
- Contradiction to the fact that algorithm would then not have inserted e into H.

□

Algorithm: Space Complexity

Definition (girth): The girth $\gamma(G)$ of a graph G is the length of its shortest cycle.

Observe: (V, H) has girth $t + 2$

Extremal Combinatorics:

What is the maximum number of edges that a graph of girth k can have?

Theorem. Let $G = (V, E)$ be an n-vertex graph with m edges and girth $\gamma(G) \geq k$. Then:

$$m \leq n + n^{1+\frac{1}{\left\lfloor\frac{k-1}{2}\right\rfloor}}.$$

Corollary. The space complexity of our algorithm is $O\left(n^{1+\frac{2}{t}} \log n\right)$.

$$|H| \leq n + n^{1+\frac{1}{\left\lfloor\frac{t+2-1}{2}\right\rfloor}} = n + n^{1+\frac{1}{\left\lfloor\frac{t+1}{2}\right\rfloor}} \leq n + n^{1+\frac{1}{\frac{t}{2}}} = n + n^{1+\frac{2}{t}} = O\left(n^{1+\frac{2}{t}}\right).$$

(using $\left\lfloor\frac{t+1}{2}\right\rfloor \geq \frac{t}{2}$ which holds for integral t)

Algorithm: Space Complexity

Lemma. Let G be an n-vertex graph with average degree $d = \frac{2m}{n}$. Let F be the graph obtained from G by repeatedly deleting nodes of degree at most $\frac{d}{2}$. Then:

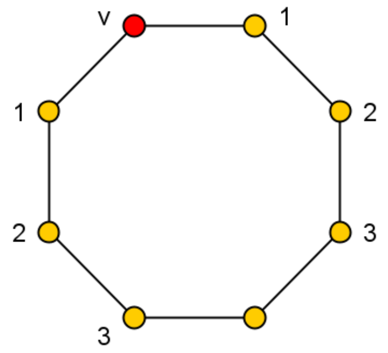
- The minimum degree in F is at least $\frac{d}{2}$.
- F is non-empty.

Proof.

By contradiction, suppose that F is empty. Then, when we removed the last vertex v, we removed no incident edge to v. We therefore removed at most $\frac{(n-1)d}{2} = m\ \frac{n-1}{n}$ edges and thus there are $\frac{m}{n}$ edges left in F, a contradiction to F being empty.

□

Algorithm: Space Complexity



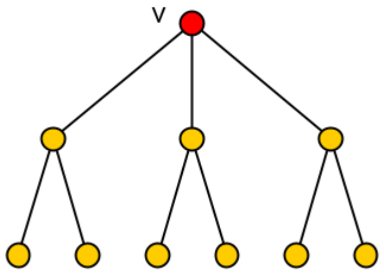
Theorem. Let $G = (V, E)$ be an n-vertex graph with m edges and girth $\gamma(G) \geq k$. Then:

$$m \leq n + n^{1+\frac{1}{\left\lfloor\frac{k-1}{2}\right\rfloor}}.$$

Proof:

- Let $d = \frac{2m}{n}$ be the average degree.
- If $d \leq 4$ then $m \leq 2n$ which implies the theorem. Suppose hence that $d > 4$.
- Let F be graph obtained by repeatedly removing from G all vertices of degree at most $\frac{d}{2}$. Then, F has minimum degree $\frac{d}{2}$ and is non-empty.
- Let $l = \left\lfloor\frac{k-1}{2}\right\rfloor$ and observe that $\gamma(F) \geq \gamma(G) \geq k$.
- Consider any vertex v. The distance-l neighborhood of v is acyclic and therefore a tree (if not then there was a cycle of length less than k). The branching factor of this tree is at least $\frac{d}{2} - 1 > 1$ (using $d > 4$). The tree therefore has at least $(\frac{d}{2} - 1)^l$ vertices. Hence,

$$n \geq \left(\frac{d}{2} - 1\right)^l = \left(\frac{m}{n} - 1\right)^l \Rightarrow m \leq n + n^{1+\frac{1}{l}}$$



□

Summary

Summary:

- There is a streaming algorithm with space $O\left(n^{1+\frac{2}{t}}\log n\right)$ for computing a t-spanner.
- For $t \geq 3$, the space bound is $\mathrm{o}(n^2)$!
- Analysis: Space complexity inherent in algorithm, can be bounded by a result from extremal combinatorics
- Slightly improved bounds are possible (see references)

References:

- Surender Baswana: “Streaming algorithm for graph spanners - single pass and constant processing time per edge”. Inf. Process. Lett. 106(3): 110-114 (2008)
- Joan Feigenbaum, Sampath Kannan, Andrew McGregor, Siddharth Suri, Jian Zhang: “Graph Distances in the Data-Stream Model”. SIAM J. Comput. 38(5): 1709-1727 (2008)