Lecture 16 (Part II: Graph Streaming)

$G = (V, E)$ or weighted $G = (V, E, W)$

Streaming models:

- insertion only: token inserts edges $e \in V \times V$
  - $|V| = n$ is known in advance
  - # edges is not known
  - promise: edges appear only once in stream
- insertion for weighted graphs:

  token is (e, w) $e \in E(G)$, w = weight of e.
- dynamic stream:

  token is $(e, b) \in E(G) \times \{+1, -1\}$

  $(e, +1) \to$ edge e is inserted

  $(e, -1)$ e is deleted
- turnstile $(e, m) \in E(G) \times \mathbb{Z}$

  $\uparrow$ multiplicity of edge.

Connectivity

- Maintain a spanning forest greedily (use union-find data structure to test if new edge makes a cycle with currently saved edges.

Space: $O(n \log n)$ $= o(n^2)$

$\uparrow$ bits

semistreaming.

Bipartiteness

- check if odd cycle. ✓

$O(n \log n)$ space.

Shortest path. : Answer query (x,y) $\forall x,y \in G$

- t-Spanner H for G is subgraph s.t

$$d_G(x,y) \le d(x,y) \le \underset{\substack{\uparrow \\ \text{stretch.}}}{t} \cdot d_G(x,y)$$

Alg:

$H \leftarrow \phi$

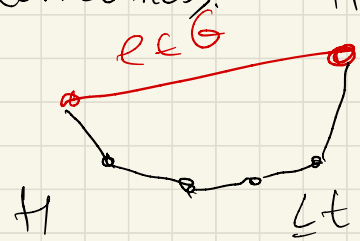
on edge (u,v)

if $d_H(u,v) \geq t+1$

$H \leftarrow H \cup \{(u,v)\}$

On query (x,y) output $d_H(x,y)$.

Correctness: H is spanner.



- each edge $e \in G$ is stretched by path of length $\le t$ in H

How large is the stored H?

Girth of G is length of smallest cycle

Obs: Girth of H above is $\geq t+2$

Alon Hoory Linial:

Any graph G on n vts, of girth $> k$ must have $\leq n + n^{O(1/k)}$ edges.

For t-approx: $O(n \log n)$ space.