Lecture 23: Graph Compression

David Woodruff
Carnegie Mellon University

Outline

- Motivating Questions

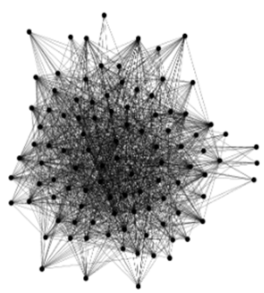
- Spanners
  - Multiplicative
  - Additive

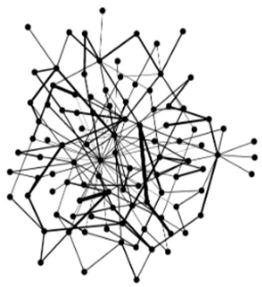
Motivating Questions

- You have an unweighted, undirected graph G = (V ,E) on n vertices
- Given vertices u and v, want to find a shortest path between u and v
  - Routing packets on a network
  - GPS: Fastest way to get from source to destination
- Problem: G may be a huge graph, and you can't afford to store it

Shortest Path Queries

- G = (V ,E) is an unweighted, undirected graph on n vertices
- |E| can be $\Theta(n^2)$, so want to "compress" G to fit in memory, but still want to answer shortest path queries
- Replace G with a subgraph H = (V, E')
  - Store H instead of G
  - Given query $d_G(u, v)$, respond with $d_H(u, v)$





- Suppose G = (V, E) is a clique
  - *If {u,v} not in H, what is $d_G(u, v)$ and what is $d_H(u, v)$?*
- Can we find a small subgraph H to approximate $d_G(u, v)$ for all u,v?

Outline

- Motivating Questions
- Spanners
  - Multiplicative
  - Additive

Spanners

- G = (V, E) is undirected, unweighted graph on n vertices
- $d_G(u, v)$ is shortest path distance from u to v
- A (k, b)-spanner of G is a subgraph H = (V, E') such that for all u,v in V
$$d_G(u, v) \leq d_H(u, v) \leq k \cdot d_G(u, v) + b$$
- If b = 0, H is a multiplicative spanner
- If k = 1, H is an additive spanner
- Do there exist (k,b)-spanners H with small |E'|?

Application of Spanners

- Shortest path query $d_G(u, v)$
  - Replace G with a (k, b)-spanner H with |E'| edges
  - Output $d_H(u, v)$
  - Approximation: $d_G(u, v) \leq d_H(u, v) \leq k \cdot d_G(u, v) + b$
  - Space: O(|E'| + n) instead of O(|E| + n)
  - Time: O(|E'|+ n) instead of O(|E|+ n)

- Faster if $|E'| \ll |E|$, but have to account for the time to create H

Outline

- Motivating Questions
- Spanners
  - Multiplicative
  - Additive

Multiplicative Spanners

- A (k, b)-spanner of G is a subgraph H = (V, E') such that for all u,v in V

$$d_G(u,v) \leq d_H(u,v) \leq k \cdot d_G(u,v) + b$$

- If b = 0, then $d_G(u,v) \leq d_H(u,v) \leq k \cdot d_G(u,v)$for all u,v in V
  - H = (V, E') is a k-multiplicative spanner
  - *How small can |E'| be?*

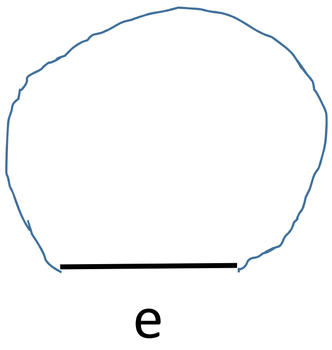
- If $d_G(u,v)$ = 1, then $d_H(u,v) \leq k$
- Conversely, if $d_H(u,v) \leq k$ for all *edges* {u,v} of G, then for any vertices u', v' ∈ V,

$$d_H(u',v') \leq k \cdot d_G(u',v')$$

- To construct H, just need for all edges {u,v} in E, $d_H(u,v) \leq k$

Greedy Algorithm for Multiplicative Spanners

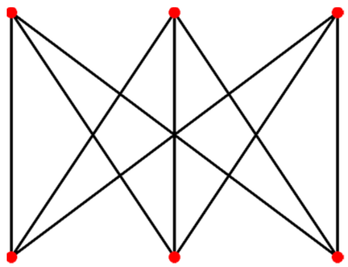
- Let's build H = (V, E') by walking through the edges of G



- Initialize H = (V, Ø)
  - For each edge e in G
  - If ______________________, then include e in H
- *That's the algorithm! What should ________________________ be?*
  - "If e doesn't form a cycle of length at most k+1 with the edges you've already included"
- *Why is this correct?*
  - For each edge not included, there's a path of length at most k between its endpoints
- How many edges does H have?

Bounding the Number of Edges in H

- H doesn't have a cycle of length at most k+1. *Why?*
- Minimum cycle length is called the girth
- What's the maximum number of edges in a graph with girth at least k+2?
  - *What if k = 2?*
  - A complete bipartite graph has $\Omega(n^2)$ edges, and girth 4
  - *What if k = 3?*
  - At most $O(n^{\frac{3}{2}})$ edges!
  - For k=2t or k=2t-1 for an integer t, at most $O(n^{1+\frac{1}{t}})$ edges, so H is tiny!



Bounding the Number of Edges in H

- Theorem: for k=2t or k=2t-1, a graph with girth at least k+2 has $O(n^{1+\frac{1}{t}})$ edges
- Lemma: let $\bar{d}$ = 2m/n be the average degree in a graph G with m edges and n nodes. There is a non-empty subgraph G' of G with minimum degree $\bar{d}/2$
- Proof: Initialize $V_0 = V$ and $E_0 = E$
  - i = 0
  - While there is a vertex v of degree at most $|E_i|/|V_i|$,
    - $i \leftarrow i + 1$
    - $V_i \leftarrow V_{i-1} \setminus \{v\}$
    - $E_i \leftarrow E_{i-1} \setminus \{\{v, w\} \text{ for all neighbors w of v}\}$
  - Output G' = (V_i, E_i)
  - G' is non-empty because $\frac{|E_i|}{|V_i|} \geq \frac{|E_{i-1}|}{|V_{i-1}|} \geq \cdots \geq \frac{|E|}{|V|} = \frac{m}{n} > 0$

For $t \leq \frac{x}{y}$ we have:

$$\frac{x-t}{y-1} \geq \frac{x-\frac{x}{y}}{y-1} = \frac{x\left(1-\frac{1}{y}\right)}{y\left(1-\frac{1}{y}\right)} = \frac{x}{y}$$

Bounding the Number of Edges in H

- Theorem: for k = 2t or k=2t-1, a graph with girth at least k+2 has $O(n^{1+\frac{1}{t}})$ edges

- Proof:
  - By lemma, a graph G has a non-empty subgraph G' with min degree $\overline{d}/2$
  - Grow a breadth-first-search (BFS) tree from a node v ∈ G'
  - G' has girth k+2
  - At level t in the BFS tree, there are at least $\left(\frac{\overline{d}}{2} - 1\right)^t$ *distinct* nodes
  - $\left(\frac{\overline{d}}{2} - 1\right)^t \leq n$, so $\left(\frac{m}{n} - 1\right)^t \leq n$, and solving gives $m \leq n + n^{1+\frac{1}{t}}$

Can we do Better?

- Girth conjecture: for k = 2t or k=2t-1, there are graphs with girth k+2 and $\Omega(n^{1+\frac{1}{t}})$ edges
- Implies any k-multiplicative spanner has $\Omega(n^{1+\frac{1}{t}})$ edges. *Why?*
- If we delete any edge {u,v} in G, the distance from u to v increases from 1 to k+1
- Only k-spanner of G is G itself
- Girth conjecture true for k = 1, 2, 3, 5

Where are We?

- Can find a (2t-1)-spanner with $O(n^{1+\frac{1}{t}})$ edges
- Can approximate $d_G(u, v)$ for any u,v up to a multiplicative factor 2t-1
- Don't store G, just store H. Only O(|E'|) = $O(n^{1+\frac{1}{t}})$ instead of $O(n^2)$ edges
- Time to compute $d_H(u, v)$, given H, is O(|E'|) = $O(n^{1+\frac{1}{t}})$
  - Faster than the $O(n^2)$ time to query a dense graph G
  - Greedy algorithm to find H is slow, but can find H in O(|E|+n) time

Outline

- Motivating Questions
- Spanners
  - Multiplicative
  - Additive

Additive Spanners

- A (k, b)-spanner of G is a subgraph H = (V, E') such that for all u,v in V
$$d_G(u,v) \le d_H(u,v) \le k \cdot d_G(u,v) + b$$

- If k = 1, then $d_G(u,v) \le d_H(u,v) \le d_G(u,v) + b$ for all u,v in V
  - H = (V, E') is a b-additive spanner
  - *How small can |E'| be?*

- For multiplicative spanners, sufficient to show for all edges {u,v} in G, $d_H(u,v) \le k$
  - Insufficient for additive spanners to show $d_H(u,v) \le b + 1$ for all edges {u,v} in G

- Would you believe: there is a 2-additive spanner with $O(n^{3/2} \log n)$ edges?

Additive Spanner Algorithm

- The algorithm has two parts

(1) Include in H all edges incident to vertices of degree at most $\sqrt{n}$

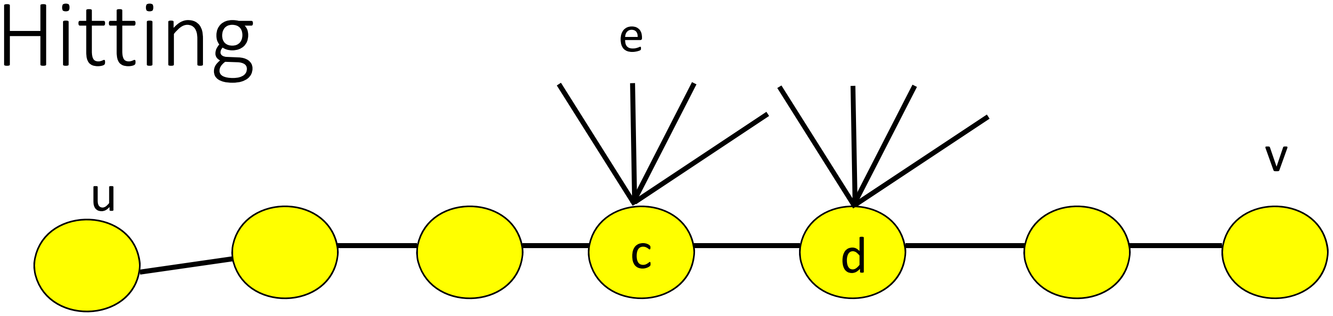
- at most $n^{3/2}$ edges (why?)

(2) Randomly sample a set S of $2\sqrt{n} \cdot \ln n$ vertices and include a BFS tree rooted at each vertex in S, in H

- at most $2n^{3/2}$ln n edges (why?)

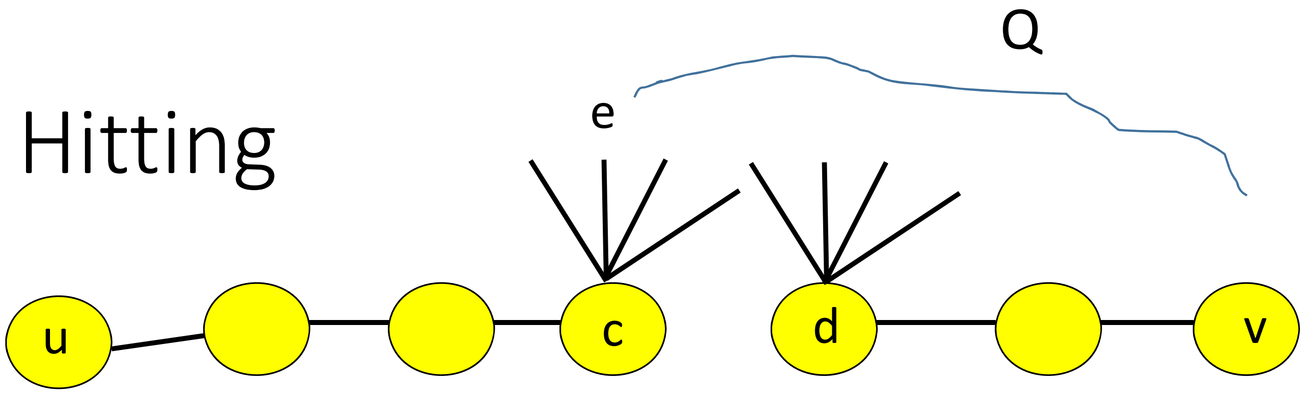
H has $O(n^{\frac{3}{2}} \log n)$ edges. Why is it a 2-additive spanner?

Path Hitting



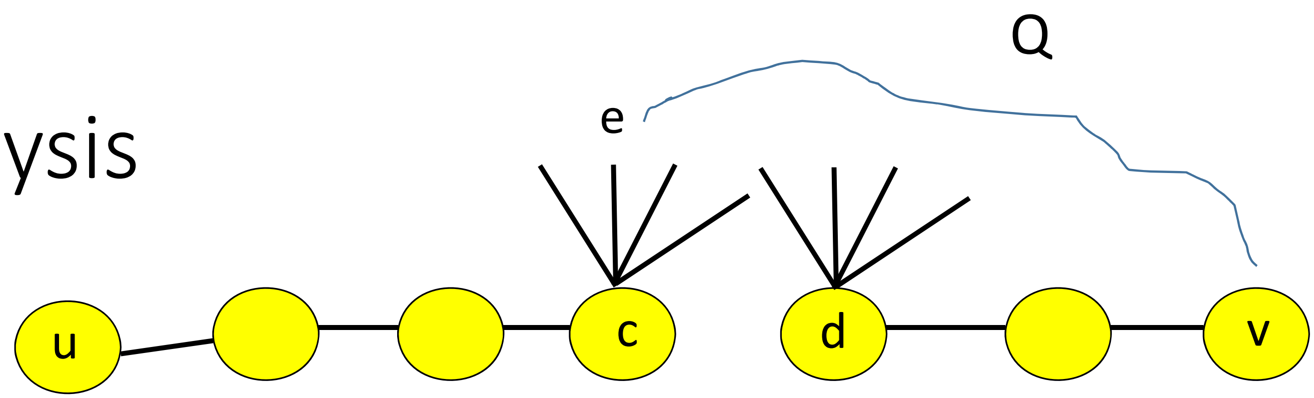
- Consider a shortest path P from u to v in G
- If all nodes on P have degree $\leq \sqrt{n}$, then all edges in P are included in the spanner H
- Otherwise consider the first edge {c,d} in P, but not in H
  - c and d have degree at least $\sqrt{n}$
- Since we randomly sample a set S of size $2\sqrt{n} \cdot \ln n$, with high probability, we sample a neighbor e of c (probability we don't sample a neighbor of c at most $\left(1 - \frac{\sqrt{n}}{n}\right)^{2\sqrt{n}\ln n} \leq \frac{1}{n^2}$)

Path Hitting



- For each of our sampled vertices in S, we grew a BFS tree
- Let $\mathrm{T_e}$ be the BFS tree rooted at e included in H
- Let Q be the path from e to v in $\mathrm{T_e}$
- Consider the path P' in H which follows P from u to c, then traverses edge {c, e}, then follows Q to v. *How long is P'?*

Analysis



- Consider the path P’ in H which follows P from u to c, then traverses edge {c, e}, then follows Q to v. How long is P’?

- Q is a shortest path from e to v in G!

- $d_Q(e, v) \leq 1 + d_P(c, v)$

- $d_{P'}(u, v) = d_P(u, c) + 1 + d_Q(e, v) \leq d_P(u, c) + d_P(c, v) + 2 \; = d_P(u, v) + 2$

Additive Spanner Notes

- Can find a 2-additive spanner with $O(n^{3/2} \log n)$ edges
  - Can get $O(n^{3/2})$ edges
- Can find a 4-additive spanner with $n^{7/5}$poly(log n) edges
- Can find a 6-additive spanner with $O(n^{4/3})$ edges
- For any constant C > 0, any C-additive spanner requires $\Omega(n^{4/3})$ edges