CS 267 Lecture 7 Graph Spanners
Scribes: Po-Hsuan Wei, William Kuzmaul
Editor: Kevin Wu **Date:** October 18, 2016

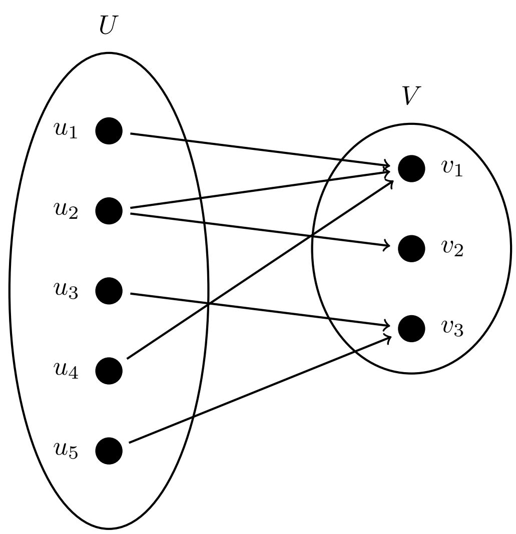
1 Graph Spanners

Our goal is to compress information about distances in a graph by looking at distances within a subgraph.

Definition 1.1. *An (α, β)-spanner of $G = (V, E)$ is a subgraph $H = (V, E_H)$, $E_H \in E$, such that $\forall u, v \in V$,*

$$d(u, v) \leq d_H(u, v) \leq \alpha d(u, v) + \beta.$$

If $\beta = 0$, then it's a multiplicative spanner (α-spanner); if $\alpha = 1$, then it's an additive spanner ($+\beta$-spanner). In general, directed graphs don't contain sparse spanners. An example is shown in the figure below, where G is a directed bipartite graph, with all its edges leave nodes in set U and incident on nodes in set V. In this case, any spanner with finite (α, β) must contain all edges in G. As a result, we will focus on undirected graphs in this class.



2 Multiplicative Spanners

Theorem 2.1. *Let $k \geq 1$ be an integer, then every n-node graph G with weighted edges contains a $(2k-1)$-spanner with $O\left(n^{1+\frac{1}{k}}\right)$ edges.*

Conjecture 2.1. (Erdős girth conjecture) *For integer $k \geq 2$ and sufficiently large n, there exist an n-node graph that contains no cycle of length $\leq 2k$ with $\Omega\left(n^{1+\frac{1}{k}}\right)$ edges.*

This conjecture is widely believed to be true. We know that the converse is true and it has been show for small k up to a certain number that this conjecture is true. If we believe this conjecture, then our bounds are tight.

Claim 2.1. *Under the Erdős girth conjecture, the bound in Theorem 2.1 is tight. Then $\forall n$, there exists a G on n nodes such that if H is a $2k-1$ spanner of G, then H has $\Omega\left(n^{1+\frac{1}{k}}\right)$ edges.*

Proof of Claim 2.1. Take $G = (V, E)$ from the Erdős girth conjecture; that is, let G be an unweighted graph on n edges with no cycle of length $\leq 2k$ and $\Omega\left(n^{1+\frac{1}{k}}\right)$ edges. We'll show that G has no non-trivial $2k-1$ spanners.
By contradiction. Assume there exists some subgraph $H \neq G$ that is a $2k-1$ spanner for G. Choose some edge $(u, v) \in E - E_H$. By the definition of a spanner, $d_H(u, v) \leq (2k-1)d(u, v) = 2k-1$. Therefore there exists some path P in E_H connecting u, v with length at most $2k-1$. However, adding (u, v) to P then completes a cycle in G of length at most $2k$, a contradiction since G has no cycles of length $\leq 2k$. This proves the claim. □

Proof of Theorem 2.1. We can generate a sufficient $2k-1$ spanner using the Create-Spanner algorithm and prove its correctness as follows:

Algorithm 1: Create-Spanner(G)

```
E_H ← ∅.
foreach (u, v) ∈ E in non-decreasing order do
    if d_H(u, v) > (2k − 1)w(u, v) then
        E_H ← E_H ∪ (u, v)
Return H.
```

Claim 2.2. *H is a $(2k-1)$-spanner. That is, $\forall u, v \in V, d_H(u, v) \leq (2k-1)d(u, v)$.*

Proof of Claim 2.2. Let u, v be vertices in V and P be their shortest path in G. For each edge $(x, y) \in P$, either:

- $(x, y) \in E_H$
- There is some path in H between x, y of length at most $(2k-1)w(x, y)$. If no such path exists, then (x, y) would have been added to E_H in Create-Spanner when it was considered.

Therefore

$$d_H(u, v) \leq \sum_{(x,y)\in P} d_H(x, y) \leq \sum_{(x,y)\in P} (2k-1)w(x, y) = (2k-1)w(P) = (2k-1)d(u, v).$$

□

Claim 2.3. *H has girth greater than $2k$.*

Proof of Claim 2.3. Assume H has a cycle C of length $\leq 2k$. Let (u, v) be the highest weighted edge of C. Then (u, v) is the last edge in C added to E_H, satisfying

$$\sum_{\substack{(x,y)\in C,\\ (x,y)\neq(u,v)}} w(x, y) > (2k-1)w(u, v)$$

On the other hand, the path $C \backslash (u, v)$ has length at most $(2k-1)w(u, v)$, giving

$$\sum_{\substack{(x,y)\in C,\\ (x,y)\neq(u,v)}} w(x, y) \leq (2k-1)w(u, v)$$

This gives a contradiction. □

Claim 2.4. *Any n-node graph with girth greater than $2k$ has $O\left(n^{1+\frac{1}{k}}\right)$ edges.*

Proof of Claim 2.4. Let H be any graph with girth greater than $2k$ and at least $10n^{1+\frac{1}{k}}$ edges. Modify the graph by repeatedly removing any nodes of degree $\leq \lceil n^{\frac{1}{k}} \rceil$, and any edges incident to that node, until no such nodes exist. The total number of edges removed in this way at most $2n \cdot \lceil n^{\frac{1}{k}} \rceil$, which means that at least $8n^{1+\frac{1}{k}}$ edges remain (and so the graph is not empty).

The minimum degree of the resulting subgraph is greater than $\lceil n^{\frac{1}{k}} \rceil$. However, by homework problem **1.3b**, this means that the subgraph has girth at most $2k$, and therefore the original graph does as well. This is a contradiction. □

The subgraph returned by Create-Spanner is a $(2k-1)$-spanner by Claim 2.2, and has $O\left(n^{1+\frac{1}{k}}\right)$ edges by Claim 2.3, Claim 2.4. This completes the proof of the theorem. □

3 Additive Spanners

We restrict ourselves to unweighted graphs, in which an additive spanner makes sense.

Theorem 3.1. *Any n-node graph G has a $+2$-spanner with $O\left(n^{\frac{3}{2}} \log n\right)$ edges.*

Theorem 3.2. *Any n-node graph G has a $+4$-spanner with $\widetilde{O}\left(n^{\frac{7}{5}}\right)$ edges.*

Theorem 3.3. *Any n-node graph G has a $+6$-spanner with $\widetilde{O}\left(n^{\frac{4}{3}}\right)$ edges.*

Theorem 3.4. (Abboud and Bodwin '15) *There is no additive spanner on $O\left(n^{\frac{4}{3}-\epsilon}\right)$ edges for n-node graphs where $\epsilon > 0$.*

We know that Theorem 3.1 is tight, but we are unsure whether Theorem 3.2 or Theorem 3.3 is tight.

We first prove Theorem 3.1

Proof of Theorem 3.1.

The proof of this theorem is very similar to the +2-approximation to the APSP problem. Let S be a hitting set for $\{N(v) \mid deg(v) \geq \sqrt{n}\}$. From previous lecture on hitting sets, we know that we can choose S with $|S| = O(\sqrt{n} \log n)$. Do a BFS search from each $s \in S$, and add the resulting n edges to E_H. For every $u \in V$ with $deg(u) < \sqrt{n}$, add all edges incident to u to E_H. By construction, $E_H = O(|S| \cdot n) + O(n\sqrt{n}) = O\left(n^{\frac{3}{2}} \log n\right)$. Consider any pair of edges $(u, v) \in V$ with shortest path P in G. We have two cases:

- P contains only low-degree nodes. Then P is entirely contained in E_H, so $d_H(u, v) = d(u, v)$.
- P contains a high-degree node x. Let $s \in S$ be a node adjacent to x. Then we can approximate the distance from u to v by appending the paths $(u, s), (v, s)$, since E_H contains shortest paths from s to every other element. Thus

$$d_H(u, v) \leq d_H(u, s) + d_H(v, s) = d(u, s) + d(v, s) \leq (d(u, x) + 1) + (d(v, x) + 1) = d(u, v) + 2.$$

Therefore the H constructed by this algorithm is a +2-spanner, as desired. □

We can improve on the size of this spanner slightly with worse constant errors.

Proof of Theorem 3.2. In this section, a node in an n-node graph is *high degree* if it has degree greater than $n^{2/5}$ and is *low degree* otherwise. The algorithm for constructing a $+4$-spanner H from G is as follows:

1. For all low degree $v \in V$, add v's $\leq n^{2/5}$ edges to H. This deals with the case where the shortest path consists of only low degree nodes. We add at most $O(nD)$ edges to H, which contributes at most $n \cdot n^{2/5} = n^{7/5}$ edges.

2. Let S be a random subset of V on $O(n^{2/5} \log n)$ nodes[1]. For each $s \in S$, add the entire breadth-first-search tree rooted at s into H. Here, we add $O(n^{2/5} \log n \cdot n) \in \tilde{O}(n^{7/5})$ edges because each BFS-tree has at most n edges.

3. Let T be a random sample on $O(n^{3/5} \log n)$ nodes. With high probability, T hits the neighborhood of each high degree node. For each high degree node v, we pick one of its neighbors in T, $t(v)$ and we add $(v, t(v))$ to H (adding $O(n)$ edges; $t(v)$ exists for each high degree v since T is a hittings set).

4. For each $t \in T$, we define t's *ball* as the set of high degree vertices assigned to it:

$$B(t) = \{v \in V | v \text{ high degree }, t = t(v)\}.$$

   We continue adding edges to H with Algorithm 2.

Algorithm 2: Adding final edges to H

```
foreach distinct t, t' ∈ T do
    P_{t,t'} = {};
    foreach u ∈ B(t), v ∈ B(t') do
        if shortest path P from u → v has < n^{1/5} high degree nodes then
            P_{t,t'}.insert(P);
    if P = {} then
        break;
    p = shortest path in P_{t,t'};
    foreach edge e ∈ p do
        H.insert(e);
```

The path added to H corresponding to t and t' will be referred to as the (t, t')-*linking path*.

How many edges are added to H? There are $\tilde{O}(n^{6/5})$ pairs of t, t'. For each pair, we may add the edges from some path P connecting u and v in $B(t)$ and $B(t')$ respectively. The only edges in P not already in H are those between two high degree nodes (by Step (1)), of which there are $\leq n^{1/5}$ in P. Thus each pair t, t' adds $O(n^{1/5})$ edges to H. Summing over the $\tilde{O}(n^{6/5})$ pairs, $\tilde{O}(n^{7/5})$ edges are added.

We have constructed H to be of size $\tilde{O}(n^{7/5})$. It remains to show that H is an additive 4-spanner of G.

If a pair of vertices u and v in G have a shortest path using no high degree nodes, then that path is in H due to Step (1).

The following claim helps resolve the case where u and v have a shortest path using more than $n^{1/5}$ high degree nodes.

Claim 3.1. *Let P be a uv-shortest path for some $u, v \in V$. If P has $> n^{1/5}$ high degree nodes, then P has $\geq \Omega(n^{3/5})$ distinct neighbors.*

[1] The constant behind the big O needs to be picked as sufficiently large for a hitting set argument appearing later to work.

Proof. Each node x in G can have at most 3 neighbors in P. Otherwise, P could be shortened by taking a path through x, a contradiction since P is a uv-shortest path for some $u, v \in V$. Since P has $> n^{1/5}$ high-degree nodes, each of which has at least $n^{2/5}$ neighbors, it follows that P has $> n^{3/5}$ edges incident on it. Each neighbor of P can account for at most 3 of these edges, and hence P must have $\geq n^{3/5}/3$ distinct neighbors. □

It follows from Claim 3.1 that for every $u, v \in V$ with a shortest path hitting $> n^{1/5}$ high degree nodes, S hits a neighbor node of a shortest path from u to v. In other words, such paths from u to v each have enough distinct neighbor nodes ($\Omega(n^{3/5})$) that S forms a hitting set on them. Thus the edges added in Step (2) include a +2-approximation for a shortest path between such u and v.

The only remaining case is uv-shortest paths hitting between 1 and $n^{1/5}$ high degree nodes.

Claim 3.2. *If a uv-shortest path hits between* 1 *and* $n^{1/5}$ *high degree nodes, then after Step (4),* $d_H(u, v) \leq d(u, v) + 4$.

Proof. Let P be a shortest path from u to v hitting between 1 and $n^{1/5}$ high degree nodes. Let h_1 be the first and h_2 be the last high degree nodes in P (possibly not distinct). Let $t_1 = t(h_1)$ and $t_2 = t(h_2)$. Let x and y be the neighbors of t_1 and t_2 respectively connected by the (t_1, t_2)-linking path. Note that the (t_1, t_2)-linking path exists because there is at least one pair of elements of $B(t_1)$ and $B(t_2)$ connected by a shortest path using $\leq n^{1/5}$ nodes, namely h_1 and h_2.

Since the subset of P from u to h_1 and from h_2 to v is in H (Step (1)), we have

$$d_H(u, v) \leq d(u, h_1) + d_H(h_1, h_2) + d(h_2, v).$$

Thus we need only show that $d_H(h_1, h_2) \leq d(h_1, h_2) + 4$. Observe that

$$d_H(h_1, h_2) \leq d(h_1, x) + d(x, y) + d(y, h_2) = d(x, y) + 4.$$

Recall that x and y have the shortest path between them in G of any pair of elements in $B(t_1)$ and $B(t_2)$, excluding paths with greater than $n^{1/5}$ high degree nodes. Namely, since $h_1 \in B(t_1)$ and $h_2 \in B(t_2)$, we have that $d(x, y) \leq d(h_1, h_2)$. Plugging this in yields

$$d_H(h_1, h_2) \leq d(h_1, h_2) + 4,$$

completing the proof.

□

□

We outline the proof for constructing a +6-spanner H for G as follows:

Proof of Theorem 3.3. [Sketch] We do the following steps. Pick parameters D_1, D_2 where $D_1 \leq D_2 = \frac{n}{D_1}$. Nodes of $deg \leq D_1$ are "low" degree, and those of $deg \geq D_2$ are high degree. Otherwise, the nodes are 'medium' degree nodes.

1. Add all edges to H that are incident to nodes of degree $\leq D_1$. This is roughly nD_1 or roughly $n^{\frac{4}{3}}$ edges.

2. Add BFS trees to H from each node in random sample S on $\frac{n}{D_2} \log n$ nodes. This is $nD_1 \log n \approx n^{\frac{4}{3}} \log n$ edges.

3. $\forall i \in 0, ..., \log n$, pick random samples T_i of size $\approx \frac{n}{D_1 \cdot 2^i} \log n$. $\forall t_i \in T_i, \forall t \in T_0$, let P_i be the shortest path between t and t_i among those missing $\leq 2^{i+1}$ edges and add all its edges to H.

   This is $\sum_i \frac{n}{D_1 2^i} \cdot \frac{n}{D_1} 2^{i+1} \leq \sum_i \frac{2n^2}{D_1^2} \leq \tilde{O}(\frac{n^2}{D_1^2}) \leq \tilde{O}(n^{\frac{4}{3}})$

For T_0, this set contains nodes that are neighbors of both the first medium degree node on P as well as the last medium degree node on P. Consider some node x on P that has a neighbor contained in set T_i. Then the path from s to some node in T_0 is 1 and the path from some $t' \in T_0$ where t' is the neighbor of the first medium degree node on P and some $t_i \in T_i$ is $\leq 2 + d(s, x)$ and similarly this holds for some $t'' \in T_0$ where t'' is the neighbor of the last medium degree node on P. This gives us an additive 6 approximation as there is at most $1 + 2 + 2 + 1$ additive-error. □

References

[1] Amir Abboud and Greg Bodwin, *The 4/3 Additive Spanner Exponent is Tight*, CoRR, 2015.