Lecture 12: Graph Spanners

Lecturer: *Huacheng Yu*

In this lecture, we will talk about *graph spanners*, which are *sparse* subgraphs that approximately preserves pairwise distances between vertices. This is useful for certain graph problems where the input graph may potentially be dense. If the graph problem only cares about the distances (e.g., all pairs shortest path), then one can hope to first compute a spanner, run the algorithm on this much sparser graph, and obtain an approximate answer.

Let G be a (maybe weighted) graph on n vertices. More formally, there are usually two types of approximations. The first is the multiplicative spanners.

Definition 1. *A t-multiplicative spanner (also called t-spanner) H is a subgraph of G such that $\forall x, y \in V$,*

$$d_G(x, y) \leq d_H(x, y) \leq t \cdot d_G(x, y),$$

where $d_G(\cdot, \cdot)$ is the distance function in graph G.

Since H is a subgraph, removing edges could only increase the distance, a t-spanner should not increase the distance by more than a factor of t.

The other notion is the additive spanners.

Definition 2. *A $+\beta$-multiplicative spanner (also called t-spanner) H is a subgraph of G such that $\forall x, y \in V$,*

$$d_G(x, y) \leq d_H(x, y) \leq d_G(x, y) + \beta.$$

We are only allowed to have an additive β error in the distance. Note that the additive spanner only makes sense the G is *unweighted* (otherwise the weights itself may be very large).

A more generalized notion of spanners is called (α, β)-spanner, which allows the error to be $\alpha d_G(x, y) + \beta$. In general, this error function of $d_G(x, y)$ that upper bounds $d_H(x, y)$ is called the *stretch function*. We will only focus on the multiplicative and additive spanners today.

Before we move on, we first note that the notion of spanner is only interesting when G is undirected. This is because for directed graph, it is possible that G is a dense bipartite graph with all the directed edges pointing from left side to the right. In this case, one cannot hope to remove any edge (u, v), since otherwise, u will not even be able to reach v.

Spanning trees can be viewed as a special case of spanners (additive $+(n-1)$ or n-multiplicative), which have the least number of edges with finite stretch.

1 Multiplicative spanners

We will prove the following theorem.

Theorem 1. *Let $k \geq 1$ be an integer. Every undirected (weighted) G has a $(2k-1)$-spanner H with $O(n^{1+1/k})$ edges.*

The spanner can be constructed using the following algorithm.

1. set $E_H \leftarrow \emptyset$
2. for edges $(u,v) \in E_G$ (in non-decreasing weight order)
3. if $d_H(u,v) > (2k-1)w(u,v)$
4. add (u,v) to E_H
5. return $H := (V, E_H)$

Correctness. To see that the algorithm does return a $(2k-1)$-spanner, note that an edge (u,v) either

- belongs to H, or
- have $d_H(u,v) \le (2k-1)w(u,v)$ when it is considered.

In either case, $d_H(u,v) \le (2k-1)w(u,v)$ holds for the final H. Therefore, let P be a shortest path from x to y in G, then we have

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

Sparsity. The sparsity of H is implied by the following two claims.

Claim 2. *H has no (simple) cycle with $\le 2k$ edges (i.e., the girth of H is at least $2k+1$).*

Proof. When we add an edge (u,v) to H, only edges with weights at most $w(u,v)$ exist in H, and $d_H(u,v) > (2k-1)w(u,v)$. Thus, this implies that before adding (u,v), there is no path between u and v that has at most $2k-1$ edges. In particular, this means that after adding (u,v), it does not create a new cycle with at most $2k$ edges.

Since the graph is empty initially, by induction, H does not any cycle with at most $2k$ edges. □

Claim 3. *Any graph H that does not have any cycle with $\le 2k$ edges must have at most $O(n^{1+1/k})$ edges.*

Proof. Suppose H has $n \cdot d$ edges. We first modify H by repeatedly removing vertices with degree $\le d/2$. Then we obtain a graph H' with $\le n$ vertices and at least $nd - nd/2 \ge nd/2$ edges. In particular, H' is not empty.

Consider a BFS tree from any vertex in H' for k steps. Since H' does not have any cycle with $\le 2k$ edges, the BFS tree does not have any cross edges in the first k steps. Moreover, since the minimum degree of H' is $d/2$, the first k levels must reach at least $(d/2-1)^k$ different vertices.

Therefore, we must have $(d/2-1)^k \le n$, implying that $d \le O(n^{1/k})$. This proves the claim. □

Combining the two claims, we prove that H has at most $O(n^{1+1/k})$ edges.

2 Additive spanner

Recall that we work with unweighted G for additive spanners. It is known that

- G has a +2-spanner with $\tilde{O}(n^{3/2})$ edges;
- G has a +4-spanner with $\tilde{O}(n^{7/5})$ edges;
- G has a +6-spanner with $\tilde{O}(n^{4/3})$ edges,

where $\tilde{O}(f)$ denotes $O(f \operatorname{poly}\log n)$.

However, one cannot go below 4/3 in the exponent.

Theorem 4 ([1])**.** *There exists graphs with n vertices s.t. any spanner with $O(n^{4/3-\varepsilon})$ edges, for some constant $\varepsilon > 0$, must have at least $+n^{\delta}$ error for some $\delta > 0$.*

Today, we will focus on the +2-spanner. We will first partition the vertex set into two parts based on the degree in G: V_{HD} which consists of vertices with degree $\geq \sqrt{n}$, and V_{LD} consisting of vertices with degree $< \sqrt{n}$. Since low degree vertices do not have too many edges incident to them, we can include all these edges in H. In particular, if a shortest path only has low degree vertices, then the entire path is in H. Thus, we only need to focus on paths with some high degree vertices on it.

The first step is to find a set S that intersects the neighborhoods of all high degree vertices. Then we will include the BFS tree from all $s \in S$ in H. If a path has a high degree vertex, then one can take a detour through s, and it turns out that this detour has at most +2 error.

Denote by $N(x)$ the set of neighbors of vertex x.

Claim 5. *$\exists S \subseteq V$ of size $O(\sqrt{n}\log n)$ such that $\forall x \in V_{\text{HD}}$, we have $N(x) \cap S \neq \emptyset$.*

Proof. The claim can be proved using a probabilistic argument. We sample a random S of size $2\sqrt{n}\log n$. Fix some $x \in V_{\text{HD}}$, we have

$$\begin{aligned}
\Pr_S[N(x) \cap S = \emptyset] &= \frac{\binom{n-|N(x)|}{|S|}}{\binom{n}{|S|}} \\
&= \frac{(n-|N(x)|)\cdots(n-|N(x)|-|S|+1)}{n(n-1)\cdots(n-|S|+1)} \\
&\leq \left(1 - \frac{|N(x)|}{n}\right)^{|S|} \\
&\leq e^{-|N(x)|\cdot|S|/n} \\
&\leq \frac{1}{n^2}.
\end{aligned}$$

Thus, by union bound, $\exists x \in V_{\text{HD}}$ such that $N(x) \cap S = \emptyset$ with probability at most $1/n$. □

For every $s \in S$, we fix a BFS tree T_s from s. E_H consists of the following edges:

1. all edges incident to some $x \in V_{\text{LD}}$,
2. all edges in T_s for some $s \in S$.

Sparsity. H has at most $\sqrt{n} \cdot |V_{\text{LD}}| + (n-1) \cdot |S| = \tilde{O}(n^{3/2})$ edges.

Correctness. Let P be a shortest path from x to y in G. If P is disjoint from V_{HD}, then $P \subset H$. In particular, $d_H(x, y) = d_G(x, y)$.

Otherwise, let u be a high degree vertex in P, and let $s_u \in S \cap N(u)$. Since H includes T_{s_u}, we have that $d_H(x, s_u) = d_G(x, s_u)$ and $d_H(s_u, y) = d_G(s_u, y)$. Therefore,

$$\begin{aligned} d_H(x, y) &\leq d_H(x, s_u) + d_H(s_u, y) = d_G(x, s_u) + d_G(s_u, y) \\ &\leq d_G(x, u) + 1 + d_G(u, y) + 1 = d_G(x, y) + 2. \end{aligned}$$

Remark 1. *There is no non-trivial* $+1$*-spanner (consider a complete bipartite graph). A more sophisticated construction shows that there is a bipartite graph with* $\Omega(n^{3/2})$ *edges that does not have a* 4*-cycle. In particular, this graph does not have any nontrivial* $+3$*-spanner.*

References

[1] Amir Abboud, and Greg Bodwin. "The 4/3 additive spanner exponent is tight." Journal of the ACM (JACM) 64.4 (2017): 1-20.