On Pairwise Spanners

Marek Cygan, Fabrizio Grandoni, Telikepalli Kavitha

An (α, β)-spanner of $G = (V, E)$ is a subgraph $H = (V, E_H)$ such that $\forall u, v \in V$, $d_G(u, v) \leq d_H(u, v) \leq \alpha.d_G(u, v) + \beta$.

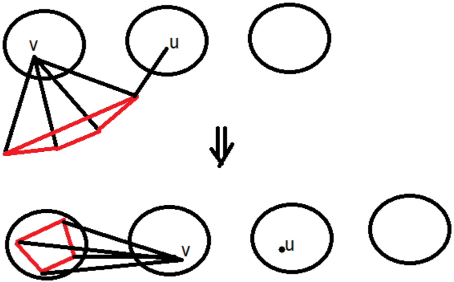
1. For any $S \subseteq V$, there is a polynomial time algorithm to compute a $(1, 2)$ $(S \times S)$-spanner of size $O(n\sqrt{|S|})$.
2. For any $S \subseteq V$ and integer $k \geq 1$, there is a polynomial time algorithm to compute a $(1, 2k)$ $(S \times V)$-spanner of size $O(n^{1+\frac{1}{2k+1}}(k|S|)^{\frac{k}{2k+1}})$.
3. For any $\epsilon > 0$ and any $P \subseteq V \times V$, there is a polynomial time algorithm to compute a $(1 + \epsilon, 4)$ P-spanner of size $O(n|P|^{\frac{1}{4}}\sqrt{\frac{\log n}{\epsilon}})$.
4. For any $P \subseteq V \times V$ and integer $k \geq 1$, there is a polynomial time algorithm to compute a $(1, 4k)$ P-spanner of size $O(n^{1+\frac{1}{2k+1}}(\sqrt{(4k+5)|P|})^{\frac{k}{2k+1}})$.

Our algorithms have two phases:

1. **Clustering phase:** In this phase, we partition a subset of V into clusters $C_1, ..., C_q$ and leave the remaining vertices *unclustered*. Our initial spanner $G_C = (V, E_C)$ will contain all the intra-cluster edges and a subset of the inter-cluster edges of G.
2. **Path-buying phase:** Here we add to the spanner some extra inter-cluster edges. To do this, we will assume there is a seller who has in his collection a sequence of paths. For the subsetwise spanner problem, the sequence of paths used is simply given by the shortest paths between the relevant pairs. There is a cost and a value assigned to every path in the seller's collection. We will buy a path from the seller and include it into our spanner if its cost is sufficiently low compared to its value.

Clustering: Given a graph $G = (V, E)$, We will compute a clustering of G with at most $n^{1-\beta}$ clusters and a subgraph G_C with $O(n^{1+\beta})$ edges.

Let U be the set of vertices not clustered yet (Initially $U := V$).
If there exists a vertex $v \in V$ with at least $\lceil n^{\beta} \rceil$ neighbours in U, then we create a new cluster C containing exactly $\lceil n^{\beta} \rceil$ arbitrary neighbours of v. Set $U := U \backslash C$ and add to G_c all the edges with both endpoints in $C \cup \{v\}$.
Otherwise we stop creating new clusters and declare U to the set of *unclustered* nodes.
Then add to G_C all edges incident to U.

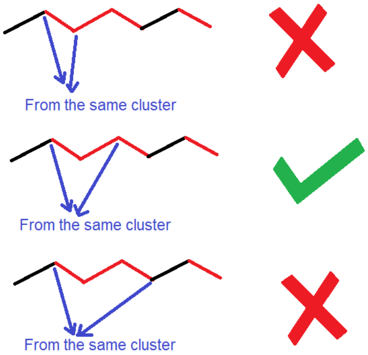


The clustering obtained has following two properties:

- **Missing-edge property:** If an edge $(u, v) \in E$ is absent in G_C, then u, v are from two different clusters.
- **Cluster-diameter property:** The distance in G_C between any two vertices of the same cluster is at most 2.

Lemma

If the shortest path P in G between any two vertices $u, v \in V$ contains t edges that are absent in G_C , then there are at least $\frac{t}{2}$ clusters intersecting with P.



Subsetwise spanners:

In the clustering phase, we obtain a cluster subgraph G_C with $O(n^{1+\beta})$ edges together with a set of at most $n^{1-\beta}$ clusters. Set $G_0 := G_C$.

In the path-buying phase, the seller has all the shortest paths between pairs of vertices in $S \times S$. For each such path P_i with endpoints u_i, v_i, we define its cost and value as follows:

- cost(P_i) is the number of edges on P_i that are missing in G_{i-1}.
- value(P_i) is the number of pairs (x, C) where $x \in \{u_i, v_i\}$ and C is a cluster intersecting P_i such that $dist_{G_{i-1}}(x, C) > dist_{P_i}(x, C)$.

If cost(P_i) ≤ 2 value(P_i), we buy the path P_i and set $G_i := G_{i-1} \cup P_i$. The final spanner is $H = G_{\binom{|S|}{2}}$.

Lemma

$\forall (u_i, v_i) \in S \times S$, $dist_H(u_i, v_i) \leq dist_G(u_i, v_i) + 2$.

Proof.

Trivially holds if the path P_i was bought.

Hence assume that $\text{cost}(P_i) > 2\ \text{value}(P_i)$. By lemma, there are at least $\frac{\text{cost}(P_i)}{2} >$ $\text{value}(P_i)$ clusters intersecting with P_i.

$\therefore$ There is one cluster C intersecting with P_i such that $dist_{G_{i-1}}(u_i, C) = dist_{P_i}(u_i, C)$ and $dist_{G_{i-1}}(v_i, C) = dist_{P_i}(v_i, C)$.

Then $dist_H(u_i, v_i) \leq dist_{G_{i-1}}(u_i, C) + dist_{G_{i-1}}(v_i, C) + 2$

$= dist_{P_i}(u_i, C) + dist_{P_i}(v_i, C) + 2$

$\leq dist_G(u_i, v_i) + 2$. ∎

After the clustering phase, G_C contains $O(n^{1+\beta})$ edges.
The number of edges added in the path-buying phase is $\sum_{\text{Path } P_i \text{ was bought}} \text{cost}(P_i)$
$\leq \sum_{\text{Path } P_i \text{ was bought}} 2.\text{value}(P_i)$.
Each pair (x, C) contributes to the above sum at most three times and number of such pairs is $|S|n^{1-\beta}$.
Therefore total number of edges in the spanner is $O(n^{1+\beta} + |S|n^{1-\beta}) = O(n\sqrt{|S|})$ when $\beta = \log_n \sqrt{|S|}$.

Intuition: If you ask me for a shortest path in my spanner H between a pair of vertices $(u, v) \in S \times S$, I want to return a path through some cluster C that intersects with the shortest path P between u, v in G. To be precise, I will return a concatenation of the following three paths: the shortest path from u to the closest node $x \in C$, the shortest path from v to the closest node $y \in C$ and the path of length at most most 2 between x and y.
Then $dist_H(u, v) \leq dist_H(u, C) + dist_H(v, C) + 2$.
Ideally we want $dist_H(u, C) = dist_P(u, C)$ and $dist_H(v, C) = dist_P(v, C)$ because in that case, $dist_H(u, v) \leq dist_P(u, C) + dist_P(v, C) + 2 \leq dist_G(u, v) + 2$.
Hence our path-buying strategy will simply be as follows: For every relevant pair of vertices u, v, if there exists a cluster C intersecting with the shortest path P between u, v such that $dist_H(u, C) = dist_P(u, C)$ and $dist_H(v, C) = dist_P(v, C)$, then don't buy P. Otherwise buy it.

Every time we buy a path P, the spanner's size increases by $\text{cost}(P)$.

There are $\geq \frac{\text{cost}(P)}{2}$ clusters intersecting with P.

$\implies$ There are $\geq \frac{\text{cost}(P)}{2}$ pairs (x, C) with $x \in \{u, v\}$ such that $dist_H(x, C) > dist_P(x, C)$.

$\implies$ At least $\frac{\text{cost}(P)}{2}$ vertex-cluster distances are decreased when we buy P.

Every vertex-cluster distance can be decreased at most thrice.

$\therefore$ Number of vertex-cluster distances is $|S|.n^{1-\beta} \geq \frac{1}{3} \sum_{\text{Path P was bought}} \frac{\text{cost}(P)}{2}$

$$\implies \sum_{\text{Path P was bought}} cost(P) = O(|S|.n^{1-\beta})$$

Sourcewise spanners:

New path-buying strategy: For each shortest path P, we buy it if it's sufficiently cheaper than its value. Otherwise we replace P with a slightly longer path P' between the same endpoints that is much cheaper and iterate the same process on P'. After a few iterations, the path becomes cheap enough and we include it into the spanner.

In the clustering phase, we obtain a cluster subgraph G_C with $O(n^{1+\beta})$ edges together with a set of at most $n^{1-\beta}$ clusters. Set $G_0 := G_C$.
In the path-buying phase, the seller has all the shortest paths between pairs of vertices in $S \times V$. Let P_i be a shortest path between $u_i \in S$ and $v_i \in V$. For each such path P_i, we will define a sequence of paths P_i^j for $0 \leq j \leq k$ maintaining the following invariants:

1. P_i^j is a path between u_i and v_i of length $\leq dist_G(u_i, v_i) + 2j$
2. Any cluster contains at most three points of P_i^j
3. $\text{cost}(P_i^j) \leq \dfrac{2n^{1-\beta}}{\gamma^j}$, where $\text{cost}(P_i^j)$ is the number of edges on P_i^j absent in G_{i-1} and $\gamma = (3n^{1-\beta})^{\frac{1}{k}}$.

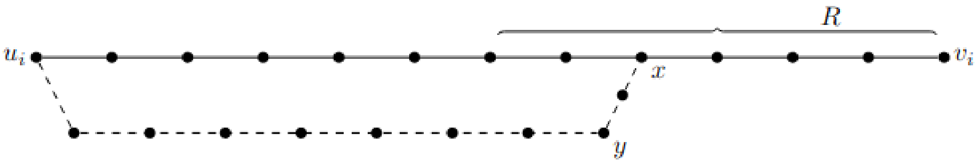
Our algorithm will buy exactly one path P_i^j for each i.

We set $P_i^0 := P_i$.
Now assume that we have constructed P_i^j. We define value(P_i^j) to be the number of clusters C intersecting with P_i^j such that $dist_{G_{i-1}}(u_i, C) > dist_{P_i^j}(u_i, C)$.

If cost(P_i^j) $\leq 3\gamma$ value(P_i^j), then we buy the path P_i^j and proceed with the next value of i.
Otherwise we construct P_i^{j+1} as follows: Let R be the longest suffix of P_i^j containing $\left\lfloor \frac{\text{cost}(P_i^j)}{\gamma} \right\rfloor$ edges that are absent in G_{i-1}. R will contain at least $\frac{\text{cost}(P_i^j)}{\gamma}$ clustered vertices and hence there are at least $\frac{\text{cost}(P_i^j)}{3\gamma}$ clusters intersecting with R.

As we did not buy P_i^j, there is a cluster C intersecting with R at a vertex x such that $dist_{G_{i-1}}(u_i, C) \leq dist_{P_i^j}(u_i, C)$.

We construct the path P_i^{j+1} by taking a shortest path in G_{i1} from u_i to the closest node $y \in C$, then we add a path of length at most two between y and x and finally add the suffix of R starting at x.



- We will definitely buy some P_i^j because cost(P_i^k) = 0.
- $\forall (u_i, v_i) \in S \times V$, $dist_H(u_i, v_i) \leq dist_G(u_i, v_i) + 2k$.
- After the clustering phase, G_C has $n^{1+\beta}$ edges. The number of edges added in the path-buying phase is $\sum_{P_i^j \text{ was bought}} \text{cost}(P_i^j) \leq 3\gamma \sum_{P_i^j \text{ was bought}} \text{value}(P_i^j)$ $\leq 3\gamma|S|(2k+3)n^{1-\beta}$.
  Hence number of edges in the spanner is $O(n^{1+\beta} + 3\gamma|S|(2k+3)n^{1-\beta}) = O(n^{1+\frac{1}{2k+1}}(k|S|)^{\frac{k}{2k+1}})$ for suitable value of β.