1 Multiplicative Graph Spanners

1.1 Problem Statement

A fundamental challenge in storing and processing graphs is their size. A graph with n nodes can potentially have $\Omega(n^2)$ edges. Therefore, numerous methods have been developed to compress graphs into other graphs that have the same nodes as the original but significantly fewer edges, ideally only $O(n)$ edges. Typically, such compressed graphs only approximately preserve the same properties as the original, and there is often a trade-off between the size of the compressed graph and the quality of the approximation to the original.

In this section, we focus on compression methods that approximately preserve distances. The central definition for this is that of a spanner.

Definition 1.1. *A t-*spanner *(spanner with* stretch t*) of a graph $G = (V, E)$ is a subgraph $H = (V, F)$, for which*

$$\operatorname{dist}_H(u, v) \leq t \cdot \operatorname{dist}_G(u, v)$$

holds for all pairs of nodes $u, v \in V$.

The spanner H is a subgraph of G in the sense that the node set V of both graphs is the same, and the edge set F of H is a subset of the edge set E of G (i.e., $F \subseteq E$).

We first discuss two fundamental properties of this definition. First, the subgraph H never overestimates the distances in G, as all paths available in H are also available in G, including every shortest path in H.

Lemma 1.2. *For every spanner H of G, it holds that $\operatorname{dist}_H(u, v) \geq \operatorname{dist}_G(u, v)$ for every pair of nodes $u, v \in V$.*

Additionally, spanners can be characterized by the fact that the inequality from Definition 1.1 holds for all endpoints of edges.

Lemma 1.3. *A subgraph $H = (V, F)$ is a t-spanner of $G = (V, E)$ if and only if*

$$\operatorname{dist}_H(u, v) \leq t \cdot w_G(u, v)$$

holds for every edge $(u, v) \in E$.

The equivalence of the two formulations (all pairs of nodes vs. endpoints of edges) follows almost directly from considering all edges of a shortest path.

In the following, we will deal with spanners of undirected, unweighted graphs (where every edge has weight 1).

1.2 Greedy Spanner

1.2.1 3-Spanner

We first analyze the following (sequential) greedy algorithm [ADDJ+93] for computing a 3-spanner of a graph $G = (V, E)$.

```
F ← ∅
foreach (u, v) ∈ E do
    Let H = (V, F)
    if dist_H(u, v) > 3 then
        F ← F ∪ {(u, v)}
```

The algorithm iterates over the edges of G and checks for each edge whether a path of length at most 3 already exists between the endpoints of the edge in the spanner computed so far. If not, the edge is added to the spanner. The algorithm follows a *greedy* scheme because it decides for each edge individually whether it is necessary or not. The algorithm computes a spanner with stretch 3 *by design*, which we now formally prove.

Lemma 1.4. *The spanner $H = (V, F)$ computed by the greedy algorithm is a 3-spanner of G.*

Proof. Let $(u, v) \in E$ be an arbitrary edge of G. If (u, v) is included in the edge set F of H, then obviously $\mathrm{dist}_H(u, v) = 1 \leq 3$. Otherwise ($(u, v) \notin F$), consider the iteration in which $\mathrm{dist}_H(u, v)$ is checked (and the decision is made not to include (u, v) in F), and let $H' = (V, F')$ be the state of H during this iteration. Then, according to the algorithm, $\mathrm{dist}_{H'}(u, v) \leq 3$. Since H' is a subgraph of H, it also holds that $\mathrm{dist}_H(u, v) \leq \mathrm{dist}_{H'}(u, v)$, and thus $\mathrm{dist}_H(u, v) \leq 3$. Since (u, v) was an arbitrary edge, H is a 3-spanner of G by Lemma 1.3. □

The proof that the computed spanner H has only $O(n^{3/2})$ edges is somewhat more involved. The concept of the girth of a graph is useful here.

Definition 1.5. *The* girth *of a graph is the length of its shortest cycle.*

Lemma 1.6. *The girth of $H = (V, F)$ is greater than 4.*

Proof. Assume H has a cycle K of length at most 4 (i.e., with at most 4 edges). Let (u, v) be the edge of the cycle K that was added last to F by the algorithm. The edges in $K \setminus \{(u, v)\}$ form a path from u to v, which already exists in H during the iteration in which (u, v) is added to F. Before adding (u, v) to F, there is no path of length at most 3 from u to v. Therefore, the path $K \setminus \{(u, v)\}$ has length at least 4, and the cycle K thus has length at least 5. This contradicts the assumption that K has length at most 4. □

Lemma 1.7. *Every undirected graph with n nodes and girth greater than 4 has $O(n^{3/2})$ edges.*

We will need the following additional lemma for the proof of this lemma, which we will prove immediately afterward.

Lemma 1.8. *Every undirected graph with n nodes and minimum degree $\geq n^{1/2} + 1$ has girth at most 4.*

Proof of Lemma 1.7. Let G be a graph with girth greater than 4 and at least $3n^{3/2}$ edges. Repeatedly remove nodes with degree less than $n^{1/2} + 1$ (along with all incident edges) from G until every node has degree at least $n^{1/2} + 1$. Let $G' = (V', E')$ be the resulting graph. In total, at most $n \cdot (n^{1/2} + 1) \leq 2n^{3/2}$ edges are removed, so

$$|E'| \geq |E| - 2n^{3/2} \geq 3n^{3/2} - 2n^{3/2} = n^{3/2} > 0\,.$$

Thus, G' is not empty, and by construction has a minimum degree of at least $n^{1/2} + 1 \geq |V'|^{1/2} + 1$.

We can now apply Lemma 1.8 to G': G' has girth at most 4. Since G' is a subgraph of G, every cycle in G' also exists in G. It follows that G has girth at most 4, which contradicts our initial assumption. Thus, every graph with girth > 4 has fewer than $3n^{3/2} = O(n^{3/2})$ edges. □

Proof of Lemma 1.8. Assume that G has girth at least 5. We consider an arbitrary node v and the breadth-first search tree of depth 2 starting from v.

Every edge of a node at distance at most 1 from v to a node at distance at most 2 from v must be an edge of this node to its parent or to a child in the breadth-first search tree, as otherwise G would contain a cycle of length at most 4 (see Fig. 1.1). Every node at a distance less than 2 from v (i.e., the internal nodes in the breadth-first search tree of depth 2 from v) must therefore have at least $\sqrt{n}$ children in the breadth-first search tree, as at most one of its incident edges can go to a parent node.

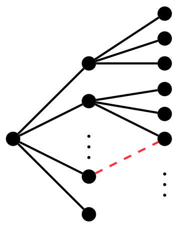

Figure 1.1: Breadth-first search tree of depth 2 from an arbitrary root node v. The red dashed edge cannot exist, as it would form a cycle of length 4.

The number of nodes at distance 2 from v is therefore at least $n^{1/2} \cdot n^{1/2} = (n^{1/2})^2 = n$. Together with v, we count at least $n + 1$ nodes for G. Due to this contradiction, our assumption was false, and we conclude that G has girth at most 4. □

Using the greedy algorithm, we have thus shown the following statement.

Theorem 1.9. *Every graph G with n nodes has a 3-spanner with $O(n^{3/2})$ edges.*

1.2.2 Generalization

The statement of Theorem 1.9 can be generalized to the following trade-off between stretch and the size of the spanner.

Theorem 1.10. *For every integer $k \geq 2$, every graph G with n nodes has a $(2k-1)$-spanner with $O(n^{1+1/k})$ edges.*

For $k = 2$, as we have already shown, this results in a 3-spanner of size $O(n^{3/2})$. For $k = \lceil \log n \rceil$, we have

$$n^{1/k} \leq n^{1/\log n} = \left(2^{\log n}\right)^{1/\log n} = 2^{(1/\log n) \cdot \log n} = 2$$

and thus $O(n^{1+1/k}) = O(n)$. This corresponds, up to constant factors, to the minimum required number of $\Omega(n)$ edges for connected graphs.

A $(2k-1)$-spanner with $O(n^{1+1/k})$ edges can also be computed using the greedy algorithm; we only need to generalize the tolerated path length in the spanner from 3 to $2k-1$.

$F \leftarrow \emptyset$
foreach $(u, v) \in E$ **do**
  Let $H = (V, F)$
  if $\text{dist}_H(u, v) > 2k - 1$ **then**
    $F \leftarrow F \cup \{(u, v)\}$

We now outline how the previous arguments need to be adapted to prove Theorem 1.10.

To prove that the stretch is $2k - 1$, the same proof as in Lemma 1.4 can be used, with the number 3 replaced by $2k - 1$. To prove that H has the desired size, we proceed similarly.

Lemma 1.11. *The girth of $H = (V, F)$ is greater than $2k$.*

Proof. Assume H has a cycle K of length at most $2k$ (i.e., with at most $2k$ edges). Let (u, v) be the edge of the cycle K that was added last to F by the algorithm. The edges in $K \setminus \{(u, v)\}$ form a path from u to v, which already exists in H during the iteration in which (u, v) is added to F. Before adding (u, v) to F, there is no path of length at most $2k - 1$ from u to v. Therefore, the path $K \setminus \{(u, v)\}$ has length at least $2k$, and the cycle K thus has length at least $2k + 1$. This contradicts the assumption that K has length at most $2k$. □

Lemma 1.12. *Every undirected graph with n nodes and girth greater than $2k$ has $O(n^{1+1/k})$ edges.*

To prove this lemma, the proof of Lemma 1.7 can be adapted with different numbers. Similarly, the proof of the auxiliary lemma must be slightly adjusted.

Lemma 1.13. *Every undirected graph with n nodes and minimum degree $\geq n^{1/k} + 1$ has girth at most $2k$.*

Proof. Assume that G has girth at least $2k + 1$. We consider an arbitrary node v and the breadth-first search tree of depth k starting from v.

Every edge of a node at distance at most $k - 1$ from v to a node at distance at most k from v must be an edge of this node to its parent or to a child in the breadth-first search tree, as otherwise G would contain a cycle of length at most $2k$. Every node at a distance less than k from v (i.e., the internal nodes in the breadth-first search tree of depth k from v) must therefore have at least $n^{1/k}$ children in the breadth-first search tree, as at most one of its incident edges can go to a parent node.

The number of nodes at distance k from v is therefore at least $(n^{1/k})^k = n$. Together with v, we count at least $n + 1$ nodes for G. Due to this contradiction, our assumption was false, and we conclude that G has girth at most $2k$. □

1.2.3 Optimality Conjecture

The trade-off between stretch and the size of the spanner computed by the greedy algorithm is based on graph-theoretic considerations regarding the size of graphs and their girth. The *Girth Conjecture* by Paul Erdős [Erd63] states that for every $k \geq 2$ and sufficiently large n, there exists a graph with n nodes and $\Omega(n^{1+1/k})$ edges that has no cycle of length at most $2k$. This conjecture has only been proven for small values of k, and it remains an open problem in combinatorics whether the generalization holds for arbitrary k. If the Girth Conjecture is true, then the $2k - 1/O(n^{1+1/k})$ trade-off is – up to constants in size – optimal.

Lemma 1.14. *If the Girth Conjecture holds, then for all sufficiently large n, there exists a graph G with n nodes such that every $(2k-1)$-spanner of G has at least $\Omega(n^{1+1/k})$ edges.*

Proof. Let G be a graph as described in the Girth Conjecture, i.e., G has $\Omega(n^{1+1/k})$ edges and girth greater than $2k$. Assume there exists a non-trivial $(2k-1)$-spanner H of G. Non-trivial means there is at least one edge (u, v) in G that does not appear in H. Since H is a $(2k-1)$-spanner of G, there exists a path P of length at most $2k-1$ from u to v in H. However, $P \cup \{(u, v)\}$ forms a cycle of length at most $2k$ in G, which contradicts the assumption that the girth of G is greater than $2k$. Thus, G is the only $(2k-1)$-spanner of G. It follows that every $(2k-1)$-spanner of G has as many edges as G, namely $\Omega(n^{1+1/k})$ edges. □

Finding a better trade-off between stretch and the asymptotic size of spanners would therefore require disproving the Girth Conjecture, which would itself be a breakthrough in combinatorics.

1.3 References

[ABSH+20] Abu Reyan Ahmed, Greg Bodwin, Faryad Darabi Sahneh, Keaton Hamm, Mohammad Javad Latifi Jebelli, Stephen G. Kobourov, and Richard Spence. "Graph spanners: A tutorial review". In: *Computer Science Review* 37 (2020), p. 100253. DOI: 10.1016/j.cosrev.2020.100253.

[ADDJ+93] Ingo Althöfer, Gautam Das, David P. Dobkin, Deborah Joseph, and José Soares. "On Sparse Spanners of Weighted Graphs". In: *Discret. Comput. Geom.* 9 (1993), pp. 81–100. DOI: 10.1007/BF02189308 (cit. on p. 1).

[BS07] Surender Baswana and Sandeep Sen. "A simple and linear time randomized algorithm for computing sparse spanners in weighted graphs". In: *Random Structures and Algorithms* 30.4 (2007), pp. 532–563. DOI: 10.1002/rsa.20130.

[Erd63] Paul Erdős. "Extremal problems in graph theory". In: *Proceedings of the Symposium on Theory of Graphs and its Applications*. 1963, p. 2936 (cit. on p. 4).

These lecture notes have been translated from a German version by an AI chatbot without revising its output. The content is partially based on a lecture by Virginia Vassilevska Williams.