Lecture Notes for the Course on
"Distributed and Sequential Graph Algorithms"
Graph Spanners

Stefano Leucci

May 1, 2019

1 Spanners: definition

Let G be a graph with n vertices and m edges, and fix a *stretch* parameter $\alpha \geq 1$. A α-spanner of G is a spanning subgraph H of G (i.e., a graph H such that $V(H) = V(G)$) that satisfies:

$$d_H(s,t) \leq \alpha \cdot d_G(s,t) \; \forall s,t \in V(G)$$

It is possible to generalize the above definition by including an additive term. For $\alpha \geq 1$ and $\beta \geq 0$ and (α, β)-spanner of G is a spanning subgraph H of G such that:

$$d_H(s,t) \leq \alpha \cdot d_G(s,t) + \beta \; \forall s,t \in V(G), \tag{1}$$

That is a α-spanner is a $(\alpha, 0)$-spanner. Spanner for which $\beta = 0$ are called *purely multiplicative spanners*. Similarly, it is possible to have *purely additive spanners*, i.e., spanners for which $\alpha = 1$. The pair (α, β) is called *distortion*.

Spanners having distortion $(1 + \epsilon, 0)$, for any arbitrarily small constant $\epsilon > 0$, also exist and they are called *nearly-additive*.

In the following we will only focus on purely additive and purely multiplicative spanners with $\alpha, \beta = O(1)$. We will refer to the number of edges $|E(H)|$ of a spanner H as its *size*.

Our goal is to find spanners with *small size* and *low stretch/distortion*.

2 Multiplicative Spanners: The Greedy Spanner

A simple algorithm to construct a sparse α-spanner for $\alpha = 1, 3, 5, \dots$ was introduced in [Althöfer et al., 1993]. For the special case of unweighted graphs, the pseudocode of the algorithm in [Althöfer et al., 1993] is equivalent to the one shown in Algorithm 1, where we write $\alpha = 2k - 1$. The algorithm incrementally constructs the sought spanner H: initially H contains no edges at all; then, the edges of G are examined one by one. When an edge (u, v) is considered the algorithm *tests* whether there already exists an alternative path between u and v in H that uses at most $2k - 1$ edges. If this is the case, then the edge is discarded and will not be part of H. Otherwise, (u, v) is added to H and the next edge is considered. When all edges of G are exhausted, the current graph H is returned.

Notice that Algorithm 1 only explicitly check the distances between the pairs s, t for which $(s, t) \in E(G)$. One might then wonder whether this is sufficient to ensure that Equation (1) will be satisfied for all pairs $s, t \in V(G)$. The following lemma shows that this is indeed the case.

Lemma 1. *H is a $(2k-1)$-spanner of G.*

Proof. Fix any two vertices $s, t \in V(G)$ and consider an arbitrary shortest path $\pi = \langle s = u_0, u_1, \dots, u_\ell = t \rangle$ from s to t in G. We will construct a path π' in H by replacing each edge (u_{i-1}, u_i) of π that is not in H with a suitable *detour* π_i. More precisely, if $(u_{i-1}, u_i) \in E(H)$ we let π_i be the path consisting of the sole edge (u_{i-1}, u_i). Otherwise, if $(u_{i-1}, u_i) \notin E(H)$, then when the edge (u_{i-1}, u_i) was considered by the algorithm, the condition $d_H(u,v) > (2k-1)d_G(u,v)$ was not satisfied, i.e., H contained a path between u_{i-1} and u_i of length at most $2k-1$. We let π_i be such a path. See Figure 1 for a qualitative example.

We define our path π' as the concatenation of all the paths π_i, i.e.,

$$\pi' = \pi_1 \circ \pi_2 \circ \cdots \circ \pi_\ell$$

By construction π' is a (non necessarily simple) path between u and v that is entirely contained in H. Moreover, we have:

$$d_H(s,t) \le |\pi'| \le \sum_{e \in E(\pi)} (2k-1) \le (2k-1) \cdot |\pi| = (2k-1) \cdot d_G(u,v). \qquad \square$$

But what about the *size* of H? After all we were looking for a *sparse* spanner. We can actually derive an upper bound to the number of edges in H by using tools from extremal graph theory [Bollobas, 2004]. We start with a definition:

Definition 1 (Girth)**.** *The* girth *of a graph is the length of its shortest cycle. If the graph is acyclic, its girth is defined to be $+\infty$.*

The key property that we are going to use is that if a graph has not small cycles then it cannot be too dense, or equivalently, if the graph's density is too high then there always exist a short cycle. Notice that, if we are aiming to prove a bound of $O(n^{1+x})$ to $|E(H)|$, then we can always iteratively remove all the vertices of degree smaller than n^x from H, as their contribution to $E(H)$ can be at most $n^{1+x} = O(n^{1+x})$. We can therefore focus on dense subgraphs of minimum degree $\kappa = n^x$. The following definition captures this notion.

Definition 2. *A κ-core of a graph G is a maximal connected subgraph in which each vertex has degree at least κ.*

Interestingly, the above observations provide with an algorithm to find all the κ-cores of a graph H: the κ-cores of H are all and only the connected components of the graph obtained by iteratively removing the vertices of degree at most κ from H.

Lemma 2. *Let H be a graph with n vertices and m edges. If H has girth at least $g = 2k+1$, then $m \le n^{1+\frac{1}{k}}$.*

Algorithm 1: Greedy-Spanner(G, k): returns a $(2k-1)$-spanner of G

$H \leftarrow (V(G), \emptyset)$;
foreach $(u,v) \in E(G)$ **do**
  if $d_H(u,v) > (2k-1)$ **then**
    $E(H) \leftarrow E(H) \cup \{(u,v)\}$;
return H

Proof. We start by computing a $(1+n^{\frac{1}{k}})$-core $\overline{H}$ of H. This can be done by using the greedy algorithm discussed above.

It follows that the edges in $E(H) \setminus E(\overline{H})$ are at most $n^{1+\frac{1}{k}}$, since each removed vertex had degree at most $n^{\frac{1}{k}}$ at the time of deletion. We now show that $\overline{H}$ must be empty.

Suppose towards a contradiction that there exists a vertex $r \in V(\overline{H})$ and let T be the graph obtained as the union of all the shortest paths from r that have length at most k. Since the girth of H (and hence of $\overline{H}$) is at least $2k+1$, T must be acyclic, i.e., T is a tree rooted in r (see Figure 3).

The root r of T has at least $1+n^{\frac{1}{k}}$ children, while every other vertex on levels $1,\dots,k-1$ of T has at least $n^{\frac{1}{k}}$ children. It follows that the number of leaves of T is at least:

$$\left(1+n^{\frac{1}{k}}\right)\cdot\left(n^{\frac{1}{k}}\right)^{k-1} = n^{\frac{k-1}{k}} + n > n,$$

which is clearly a contradiction since $|V(T)| \le |V(\overline{H})| \le |V(H)| \le n$. □

The above lemma reduces the problem of finding an upper bound $|E(H)|$ H to that of finding a lower bound to the girth of H. The test of Algorithm 1 ensures that no edge completing cycle of length smaller to $2k$ can ever be added to H, as the following lemma shows.

Lemma 3. *The girth of H is at least $2k+1$.*

Proof. Suppose towards a contradiction that H contains a cycle C of length at most $2k$. Let (u,v) be the last edge of C that was added to H. When (u,v) was considered all the other edges of C were already in H, implying $d_H(u,v) \le 2k-1$ (See Figure 4). It follows that the condition $d_H(u,v) > (2k-1)d_G(u,v) = 2k-1$ of is not satisfied, i.e., (u,v) cannot belong to H. □

Combining Lemma 1, Lemma 2, and Lemma 3, we immediately obtain:

Theorem 1. *H is a $(2k-1)$-spanner of G of size at most $n^{1+\frac{1}{k}}$.*

The above construction and analysis easily extends to non-negatively weighted graphs. The only modification needed in Algorithm 1 consists of replacing the test condition $d_H(u,v) > (2k-1)$ with $d_H(u,v) > (2k-1)w(u,v)$, where $w(u,v)$ denotes the weight of edge (u,v) in G.

2.1 Can we do better?

A natural question is whether we can improve over the trade-off between size and stretch achieved by the above simple algorithm. Perhaps surprisingly, it is possible to show that the greedy spanner is asymptotically optimal unless the following conjecture fails.

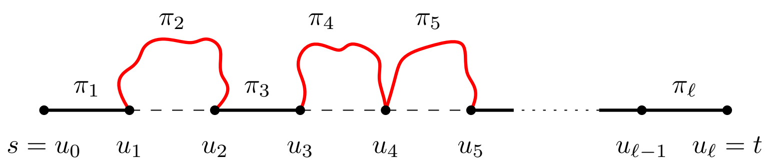

Figure 1: A qualitative representation of the path π' between u and v in H constructed in the proof of Lemma 1. Solid lines represent edges and paths in H. Dashed edges belong to G but not to H. For each dashed edge (u_{i-1},u_i) the corresponding detour π_i is shown in red. The path π' consists of all the bold edges.

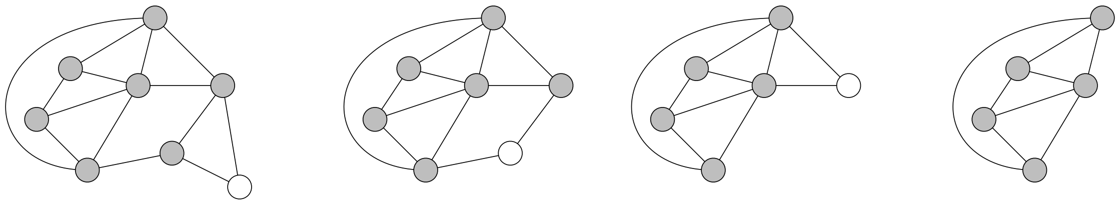

Figure 2: Example of the greedy algorithm for computing a 3-core of graph. Vertices with degree smaller than 3 are shown in white and are iteratively removed (from left to right). The rightmost graph is the 3 core.

Conjecture 1 (Erdős Girth Conjecture [Erdős, 1964])**.** *For every* $g \in [2k+1, 2k+2]$*,* $k \geq 1$*, there exists a graph with girth* g *and* $\Omega(n^{1+\frac{1}{k}})$ *edges.*

Example 1. *The complete graph* K_n *has girth* 3 *(*$k = 1$*) and* $\frac{n(n-1)}{2} = \Theta(n^2)$ *edges.*

Example 2. *The complete bipartite graph* $K_{n/2,n/2}$ *has girth* 4 *(*$k = 2$*) and* $\frac{n^2}{4} = \Theta(n^2)$ *edges.*

So far, the Erdős Girth Conjecture has been proven for for $k = 1, 2, 3, 5$ [Wenger, 1991]. An interesting consequence of the existence of such graph is that they cannot be further sparsified:

Observation 1. *Let* G *be a graph as in Conjecture 1. The only* $(2k-1)$*-spanner of* G *is* G *itself.*

Proof. Let H be a $(2k-1)$-spanner of G and suppose that $H \neq G$. Then, there exist an edge $(u, v) \in E(G) \setminus E(H)$. Since the girth of G (and hence of H) is at least $2k + 1$, we have that $d_H(u, v) \geq 2k$, yielding the contradiction: $2k \leq d_H(u,) \leq (2k-1)d_G(u, v) = 2k - 1$. □

Corollary 1. *The greedy spanner is optimal unless the Erdős Girth Conjecture fails.*

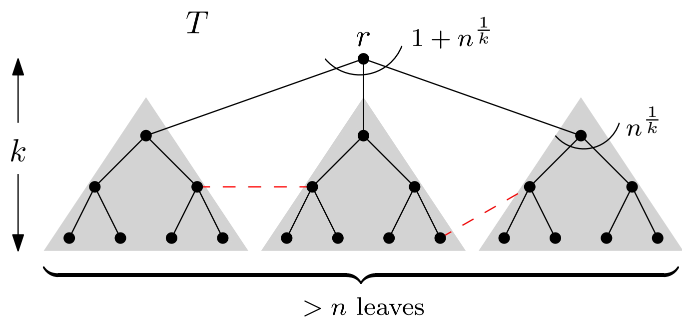

Figure 3: The tree T used in the proof of Lemma 2. The existence of any additional edge incident to a non-leaf vertex (e.g., the dashed red edges) would create a cycle of length at most $2k$, contradiction $g \geq 2k+1$. Each of the $1+n^{\frac{1}{k}}$ gray subtrees has at least $n^{\frac{k-1}{k}}$ leaves, implying that T has more than n leaves.

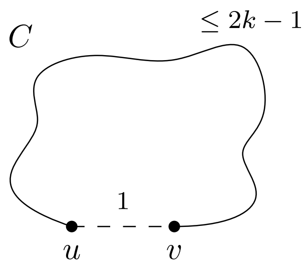

Figure 4: No cycle C of length at most $2k$ can exist in H, since the last edge (u, v) to be added to H would not satisfy $d_H(u, v) > (2k-1)d_G(u, v) = 2k-1$.

3 Additive Spanners

A simple (1, 2)-spanner

In this section we will first describe a simple *randomized* algorithm that allows us to compute a spanner H with $\tilde{O}(n^{\frac{3}{2}})$ edges and that preserves the distance in G up to an additive error of 2. We will then derive a *deterministic* version of the algorithm that also improves the size to $O(n^{\frac{3}{2}})$, which will turn out to be asymptotically tight. The presented construction is based on the work of [Aingworth et al., 1999].

A randomized algorithm

We will say that a vertex v of G is *heavy* if its degree $\delta(v)$ is at least $\sqrt{n}$. Otherwise, we say that v is *light*. Similarly, we say that an edge is *heavy* if *both* its endpoints are heavy, and *light* otherwise. The idea behind the algorithm is the following: even if all the light edges are added to the spanner H, their combined contribution to $|E(H)|$ will be at most $n^{\frac{3}{2}}$. We therefore only need to approximate the distances between pair of vertices that are connected by shortest paths traversing one or more heavy vertices. To this aim we will add all the shortest paths emanating from a certain set random S of *source* vertices. If the neighborhood $N(v)$ of an heavy vertex v happens to contain a source in $x \in S$, then all the shortest paths emanating from v can be approximated by the first going from v to x and then following the shortest path from x. The pseudocode is shown in Algorithm 2. See also Figure 5 for a qualitative representation of the resulting spanner.

Algorithm 2: randomized $(1, 2)$-spanner(G)

$H \leftarrow (V(G), \emptyset)$;
Create $S \subseteq V(G)$ by independently adding each vertex in $V(G)$ to S with probability $p = \frac{3 \ln n}{\sqrt{n}}$;
foreach $v \in S$ **do**
  Add a BFS tree of G rooted in v to H;
$E(H) \leftarrow E(H) \cup \{e \in E(G) \,:\, e \text{ is light}\}$
return H

Lemma 4. $|E(H)| = O(n^{\frac{3}{2}} \ln n)$ *with probability at least* $1 - n^{-2}$.

Proof. We start by proving an upper bound to $|S|$. Notice that $|S|$ is a binomial random variable with parameters n and $p = \frac{3 \ln n}{\sqrt{n}}$. Therefore, the expected number of vertices in S is $\mathbb{E}[|S|] = n \cdot p = 3\sqrt{n} \ln n$ and by using a Chernoff bound with $\epsilon = 1$, we obtain:

$$\Pr(|S| \geq 6\sqrt{n} \ln n) = \Pr(|S| \geq (1+\epsilon)\mathbb{E}[|S|]) \leq e^{-\frac{1}{3}\epsilon \mathbb{E}[|S|]} = e^{-\frac{1}{3} 3\sqrt{n} \ln n} \leq n^{-\sqrt{n}} \leq n^{-2}.$$

That is, with probability at least $1 - n^{-2}$, $|S| < 6\sqrt{n} \ln n$. Since each vertex in S causes the addition of at most $n - 1$ edges, their total contribution to $|E(H)|$ is at most $6n^{1+\frac{3}{2}} \ln n$.

To bound the number of light edges added to H, notice that each of those edges as at least one light endpoint. Since each light vertex has degree less than $\sqrt{n}$, the total number of light edges can be at most $n^{\frac{3}{2}}$. □

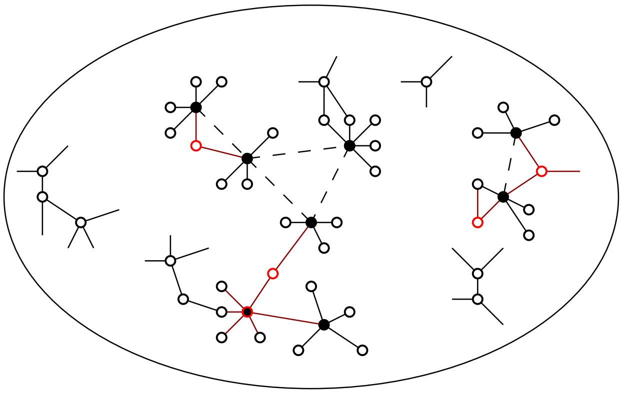

Figure 5: Qualitative (and partial) example of the spanner returned by Algorithm 2. Black (resp. white) vertices are heavy (resp. light). Vertices whose outline is highlighted in red belong to the randomly selected set of sources S. Solid edges belong to both G and H, while dashed edges belong to $E(G) \setminus E(H)$. Edges highlighted in red are guaranteed to belong to a BFS rooted in some vertex in S. For the sake of readability, the BFS trees themselves are not depicted.

Lemma 5. *With probability at least $1 - n^{-2}$, each heavy vertex has at least one neighbor in S.*

Proof. Let v be a heavy vertex. Since $\delta_G(v) \geq \sqrt{n}$, the probability that $N_G(v)$ contains no vertex in S is at most:

$$\Pr(N_G(v) \cap S = \emptyset) \leq (1-p)^{\delta_G(v)} \leq \left(1 - \frac{3 \ln n}{\sqrt{n}}\right)^{\sqrt{n}} = \left(1 - \frac{1}{x}\right)^{x \cdot 3 \ln n} \leq e^{-3 \ln n} = n^{-3},$$

where we used the substitution $x = \frac{1}{p} = \frac{\sqrt{n}}{3 \ln n}$ and the inequality $(1 - \frac{1}{x})^x \leq \frac{1}{x}$ for $x \geq 1$.

By using the union bound on the (at most n) heavy vertices, the probability that at least one heavy vertex has no neighbor in S is at most:

$$\sum_{v\,:\,v \text{ is heavy}} \Pr(N_G(v) \cap S = \emptyset) \leq n \cdot n^{-3} = n^{-2}. \qquad \square$$

Lemma 6. *With probability at least $1 - n^{-2}$, H is a $(1, 2)$-spanner of G.*

Proof. Suppose that the condition of Lemma 5 holds (this happens with probability at least $1 - n^{-2}$).

Let $s, t \in V(G)$ and consider a shortest path π between s and t in G.

If π contains no heavy vertex, then all the edges in π belong to H and we are done. Otherwise, let $u \in V(G)$ be the *first* heavy vertex encountered when π is traversed from s to t.

It follows that the subpath of π from s to u is entirely contained in H (as all of its vertices except for u are light), therefore $d_H(s, u) = d_G(s, u)$. Since u is a heavy vertex, it has at least one neighbor $v \in S$. Using the fact that H contains a BFS tree from v, we know that $(u, v) \in E(H)$ and that $d_H(v, t) = d_G(v, t) \leq 1 + d_G(u, v)$, where we used the triangle inequality (see also Figure 6). We can then write:

$$\begin{aligned} d_H(s,t) &= d_H(s,u) + d_H(u,t) \\ &\leq d_G(s,u) + 1 + d_H(v,t) \\ &\leq d_G(s,u) + 1 + d_G(u,t) + 1 \\ &= d_G(s,t) + 2. \end{aligned} \qquad \square$$

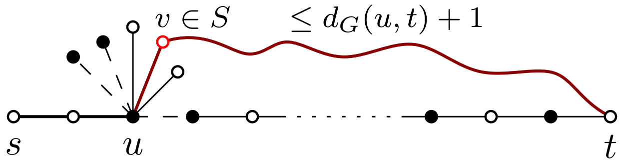

Figure 6: Graphical representation of the proof of Lemma 6. The same graphical conventions of Figure 5 are used. In addition, a path in H using at most $d_G(s,t)+2$ edges is shown in bold.

The following theorem follows directly from the combination of Lemma 4 and Lemma 6.

Theorem 2. *H is a $(1,2)$-spanner of G of size $O(n^{\frac{3}{2}} \ln n)$.*

A deterministic algorithm

We now show how the above randomized algorithm can be be transformed in a deterministic one. To this aim we will drop the distinction between *heavy* and *light* nodes and edges and instead distinguish between *marked* and *unmarked* nodes. A node will be marked if it has at least one neighbor in the set S and unmarked otherwise. The algorithm starts with an empty set S and iteratively adds to S one vertex v that is adjacent to at least $\sqrt{n}$ unmarked vertices. Then, it proceeds similarly to Algorithm 2. The pseudocode is shown in Algorithm 3, where $\bar{\delta}(v)$ denotes the number of neighbors of v that are unmarked.

Algorithm 3: deterministic $(1,2)$-spanner(G)

$H \leftarrow (V(G), \emptyset)$;
$S \leftarrow \emptyset$;
while $\exists v \in V(G) \, : \, \bar{\delta}(v) > \sqrt{n}$ **do**
  $S \leftarrow S \cup \{v\}$;
  Mark all neighbors of v;
foreach $v \in S$ **do**
  Add a BFS tree of G rooted in v to H;
$E(H) \leftarrow E(H) \cup \{e \in E(G) \, : \, e \text{ is unmarked}\}$
return H

Lemma 7. *$|S| \leq \sqrt{n}$.*

Proof. Each time a vertex v is added to S, at least $\bar{\delta}(v) \geq \sqrt{n}$ previously unmarked vertices are marked. Since initially all vertices are unmarked, we must have $|S|\sqrt{n} \leq n$, i.e., $|S| \leq \sqrt{n}$. □

Lemma 8. *$|E(H)| = O(n^{\frac{3}{2}})$.*

Proof. Each vertex in S causes the addition of at most $n-1$ edges. Using Lemma 7, we have: $|S| \cdot (n-1) \leq n\sqrt{n}$.

Each unmarked edge (u,v) is incident to at least one unmarked node, say v w.l.o.g. Then (u,v) contributes 1 to $\bar{\delta}(v)$. This implies total number of unmarked edges is at most $\sum_{v \in V(G)} \bar{\delta}(v) \leq n\sqrt{n}$. □

Lemma 9. *H is a $(1,2)$-spanner of G.*

Proof. Let $s, t \in V(G)$ and consider a shortest path π between s and t in G.

If π contains no marked vertex, then all the edges in π are also unmarked and belong to H. Otherwise, let $u \in V(G)$ be the *first* marked vertex encountered when π is traversed from s to t.

Since u is marked, there must be an edge $(u, v) \in E(G)$ with $v \in S$. Moreover, since H contains a BFS tree from each vertex in S, the edge (u, v) is also in $E(H)$. The rest of the proof is now identical to that of Lemma 6. □

Theorem 3. *H is a $(1,2)$-spanner of G of size $O(n^{\frac{3}{2}})$.*

Proof. The claim follows directly from the combination of Lemma 7 and Lemma 9. □

Observation 2. *Assuming Erdős Girth Conjecture, a $(1,2k)$-spanner or $(1,2k+1)$-spanner must contain at least $\Omega(n^{1+\frac{1}{k+1}})$ edges in the worst case.*

Actually, for $k = O(1)$, [Woodruff, 2006] showed that the above lower bound holds *unconditionally*, i.e., regardless of the Erdős Girth Conjecture. By substituting $k = 1$, we then obtain an *unconditional* lower bound of $\Omega(n^{\frac{3}{2}})$ on the worst-case size of $\beta = 2k = 2$-additive spanners.

Corollary 2. *The size of H is asymptotically optimal.*

4 Other Additive Spanners

In addition to the $(1,2)$-spanner of [Aingworth et al., 1999] discussed above, two other additive spanners are currently known, namely the $(1,4)$-spanner of [Chechik, 2013] which has a size of $\tilde{O}(n^{\frac{7}{5}})$, and the $(1,6)$-spanner of [Baswana et al., 2010] which has a size of $O(n^{\frac{4}{3}})$.

One might expect that sparser additive spanners could be obtained by worsening their additive stretch. Quite surprisingly, that turns out not to be the case: Abboud and Bodwin showed that there exists no $(1,\beta)$-spanner H with $\beta = n^{o(1)}$ such that $|E(H)| = O(n^{\frac{4}{3}-o(1)})$ [Abboud and Bodwin, 2017].

The following table summarizes the spanners discussed in these notes.

Stretch	Upper Bound	Lower Bound
$2k-1$	$O(n^{1+\frac{1}{k}})$ [Althöfer et al., 1993]	$\Omega(n^{1+\frac{1}{k}})$ (conditional)
$2k$	(dominated)	$\Omega(n^{1+\frac{1}{k}})$ (conditional)
$(1,2)$	$O(n^{\frac{3}{2}})$ [Aingworth et al., 1999]	$\Omega(n^{\frac{3}{2}})$
$(1,3)$	(dominated)	$\Omega(n^{\frac{3}{2}})$
$(1,4)$	$\tilde{O}(n^{\frac{7}{5}})$ [Chechik, 2013]	$\Omega(n^{\frac{4}{3}})$
$(1,5)$	???	$\Omega(n^{\frac{4}{3}})$
$(1,6)$	$O(n^{\frac{4}{3}})$ [Baswana et al., 2010]	$\Omega(n^{\frac{4}{3}-\epsilon})$, for any constant $\epsilon > 0$
$(1,k)$, $6 < k = n^{o(1)}$	(dominated)	$\Omega(n^{\frac{4}{3}-\epsilon})$ [Abboud and Bodwin, 2017]

References

[Abboud and Bodwin, 2017] Abboud, A. and Bodwin, G. (2017). The 4/3 additive spanner exponent is tight. *J. ACM*, 64(4):28:1–28:20.

[Aingworth et al., 1999] Aingworth, D., Chekuri, C., Indyk, P., and Motwani, R. (1999). Fast estimation of diameter and shortest paths (without matrix multiplication). *SIAM J. Comput.*, 28(4):1167–1181.

[Althöfer et al., 1993] Althöfer, I., Das, G., Dobkin, D. P., Joseph, D., and Soares, J. (1993). On sparse spanners of weighted graphs. *Discrete & Computational Geometry*, 9:81–100.

[Baswana et al., 2010] Baswana, S., Kavitha, T., Mehlhorn, K., and Pettie, S. (2010). Additive spanners and (alpha, beta)-spanners. *ACM Trans. Algorithms*, 7(1):5:1–5:26.

[Bollobas, 2004] Bollobas, B. (2004). *Extremal Graph Theory*. Dover Publications, Inc., New York, NY, USA.

[Chechik, 2013] Chechik, S. (2013). New additive spanners. In *SODA*, pages 498–512. SIAM.

[Erdős, 1964] Erdős, P. (1964). Extremal problems in graph theory. In *IN "THEORY OF GRAPHS AND ITS APPLICATIONS," PROC. SYMPOS. SMOLENICE*, pages 29–36.

[Wenger, 1991] Wenger, R. (1991). Extremal graphs with no C^4's, C^6's, or C^{10}'s. *J. Comb. Theory, Ser. B*, 52(1):113–116.

[Woodruff, 2006] Woodruff, D. P. (2006). Lower bounds for additive spanners, emulators, and more. In *FOCS*, pages 389–398. IEEE Computer Society.