Sparse geometric graphs with small dilation*

Boris Aronov[1], Mark de Berg[2], Otfried Cheong[3], Joachim Gudmundsson[4], Herman Haverkort[2], Michiel Smid[5], and Antoine Vigneron[6]

[1] Department of Computer and Information Science, Polytechnic University, Brooklyn, New York, USA. http://cis.poly.edu/~aronov

[2] Department of Mathematics and Computing Science, TU Eindhoven, Eindhoven, the Netherlands. mdberg@win.tue.nl, cs.herman@haverkort.net

[3] Division of Computer Science, KAIST, Daejeon, South Korea. otfried@cs.kaist.ac.kr

[4] IMAGEN Program, National ICT Australia Ltd¶, Australia. joachim.gudmundsson@nicta.com.au

[5] School of Computer Science, Carleton University, Ottawa, Canada K1S 5B6. michiel@scs.carleton.ca

[6] Department of Computer Science, National University of Singapore, Singapore. antoine@comp.nus.edu.sg

Abstract. Given a set S of n points in $\mathbb{R}^D$, and an integer k such that $0 \leqslant k < n$, we show that a geometric graph with vertex set S, at most $n-1+k$ edges, and dilation $O(n/(k+1))$ can be computed in time $O(n \log n)$. We also construct n–point sets for which any geometric graph with $n-1+k$ edges has dilation $\Omega(n/(k+1))$; a slightly weaker statement holds if the points of S are required to be in convex position.

1 Preliminaries and introduction

A *geometric network* is an undirected graph whose vertices are points in $\mathbb{R}^D$. Geometric networks, especially geometric networks of points in the plane, arise in many applications. Road networks, railway networks, computer networks—any collection of objects that have some connections between them can be modeled as a geometric network. A natural and widely studied type of geometric network is the *Euclidean network*, where the weight of an edge is simply the Euclidean distance between its two endpoints. Such networks for points in $\mathbb{R}^D$ form the topic of study of our paper.

When designing a network for a given set S of points, several criteria have to be taken into account. In particular, in many applications it is important to ensure a short connection between every pair of points in S. For this it would be ideal to have a direct connection between every pair of points; the network would then be a complete graph. In most applications, however, this is unacceptable due to the high costs. Thus the question arises: is it possible to construct a network that guarantees a reasonably short connection between every pair of points while not using too many edges? This leads to the concept of *spanners*, which we define next.

Recall that the weight of an edge $e = (u, v)$ in a Euclidean network $G = (S, E)$ on a set S of n points is the Euclidean distance between u and v, which we denote by $d(u, v)$. The *graph distance* $d_G(u, v)$ between two vertices $u, v \in S$ is the length of a shortest path in G

* This work was supported by LG Electronics and NUS research grant R-252-000-166-112. Research by B.A. has been supported in part by NSF ITR Grant CCR-00-81964 and by a grant from US-Israel Binational Science Foundation. Part of the work was carried out while B.A. was visiting TU/e in February 2004 and in the summer of 2005. MdB was supported by the Netherlands' Organisation for Scientific Research (NWO) under project no. 639.023.301. MS was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

¶ National ICT Australia is funded through the Australian Government's Backing Australia's Ability initiative, in part through the Australian Research Council.

connecting u to v. The *dilation* (or: *stretch factor*) of G, denoted $\Delta(G)$, is the maximum factor by which graph distance d_G differs from the Euclidean distance d, namely

$$\Delta(G) := \max_{\substack{u,v \in S \\ u \neq v}} \frac{d_G(u,v)}{d(u,v)}.$$

The network G is a *t-spanner* for S if $\Delta(G) \leqslant t$.

Spanners find applications in robotics, network topology design, distributed systems, design of parallel machines, and many other areas and have been a subject of considerable research [3]. Recently spanners found interesting practical applications in metric space searching [15, 16] and broadcasting in communication networks [1, 9, 14]. The problem of constructing spanners has received considerable attention from a theoretical perspective—see the surveys [7, 18].

The complete graph has dilation 1, which is optimal, but we already noted that the complete graph is generally too costly. The main challenge is therefore to design *sparse* networks that have small dilation. There are several possible measures of sparseness, for example the total weight of the edges or the maximum degree of a vertex. The measure that we will focus on is the number of edges. Thus the main question we study is this: Given a set S of n points in $\mathbb{R}^D$, what is the best dilation one can achieve with a network on S that has few edges? Notice that the edges of the network are allowed to cross or overlap.

This question has already received ample attention. For example, there are several algorithms [2, 13, 17, 19] that compute a $(1+\varepsilon)$-spanner for S, for any given constant $\varepsilon > 0$. The number of edges in these spanners is $O(n)$. Although the number of edges is linear in n, it can still be rather large due to the hidden constants in the O-notation that depend on ε and the dimension D. Das and Heffernan [4] showed how to compute in $O(n \log n)$ time, for any constant $\varepsilon' > 0$, a t–spanner with $(1+\varepsilon')n$ edges and degree three, where t only depends on ε' and D. We are interested in the case when the number of edges is close to n, not just linear in n. Any spanner must have at least $n-1$ edges, for otherwise the graph would not be connected, and the dilation would be infinite. This leads us to define the quantity $\Delta(S,k)$:

$$\Delta(S,k) := \min_{\substack{V(G)=S \\ |E(G)|=n-1+k}} \Delta(G).$$

Thus $\Delta(S,k)$ is the minimum dilation one can achieve with a network on S that has $n-1+k$ edges. The goal of our paper is not to give an algorithm for computing a minimum-dilation network with $n-1+k$ edges for the specific input graph. (Problems of this type have been studied by several authors [6, 8, 10]. In general they appear quite hard.) Rather, we will study the worst-case behavior of the function $\Delta(S,k)$: what is the best dilation one can guarantee for *any* set S of n points if one is allowed to use $n-1+k$ edges? In other words, we study the quantity

$$\delta(n,k) := \sup_{\substack{S \subset \mathbb{R}^D \\ |S|=n}} \Delta(S,k).$$

For the special case when the set S is in $\mathbb{R}^2$ and is required to be in convex position[1] we define the quantity $\delta_C(n,k)$ analogously.

[1] A set of points is *in convex position* if they all lie on the boundary of their convex hull.

The result of Das and Heffernan [4] mentioned above implies that, for any constant $\varepsilon' > 0$, $\delta(n, \varepsilon' n)$ is bounded by a constant. We are interested in what can be achieved for smaller values of k, in particular for $0 \leqslant k < n$.

In the above definitions we have placed a perhaps unnecessary restriction that the graph may use no other vertices besides the points of S. We will also consider networks whose vertex sets strictly contain S. In particular, we define a *Steiner tree* on S as a tree T with $S \subset V(T)$. The vertices in $V(T) \setminus S$ are called *Steiner points*. Note that T may have any number of vertices, the only restriction is on its topology.

Our results. We first show that any Steiner tree on a set S of n equally spaced points on a circle has dilation at least n/π. We prove in a similar way that $\delta(n, 0) \geqslant \frac{2}{\pi} n - 1$. We remark that Eppstein [7] gave a simpler proof of a similar bound for the case when no Steiner points are allowed. Our bound is tight in the sense that $\Delta(S, 0) = \frac{2}{\pi} n - 1 + o(1)$ when S is a set of n equally spaced points on a circle.

We then continue with the case $0 < k < n$. Here we give an example of a set S of n points for which any network with $n - 1 + k$ edges has dilation at least $\frac{2}{\pi} \lfloor n/(k+1) \rfloor - 1$, proving that $\delta(n, k) \geqslant \frac{2}{\pi} \lfloor n/(k+1) \rfloor - 1$. We also prove that for points in convex position the dilation can be almost as large as in the general case, i.e., $\delta_C(n, k) = \Omega(n/((k+1) \log n))$.

Next we study upper bounds. We describe an $O(n \log n)$ time algorithm that computes for a given set S and parameter $0 \leqslant k < n$ a network of dilation $O(n/(k+1))$. Combined with our lower bounds, this implies that $\delta(n, k) = \Theta(n/(k+1))$. In particular, our bounds apply to the case $k = o(n)$, which was left open by Das and Heffernan [4]. Notice that, for any constant $c \geqslant 1$, if $n \leqslant k \leqslant cn$, then we have $1 \leqslant \delta(n, k) \leqslant \delta(n, n-1)$, and thus $\delta(n, k) = \Theta(1)$. It means that our result $\delta(n, k) = \Theta(n/(k+1))$ generalizes to the case $0 \leqslant k \leqslant cn$ for any constant $c \geqslant 1$.

Our lower bounds use rather special point sets and it may be the case that more 'regular' point sets admit networks of smaller dilation. Therefore we also study the special case when S is a point set with so–called *bounded spread*. We show that such sets admits a network of dilation $O(\sqrt{n/(k+1)})$ and that this bound is asymptotically tight. We also obtain tight bounds for the case when S is a $\sqrt{n} \times \sqrt{n}$ grid.

Notation and terminology. Hereafter S will always denote a set of points in $\mathbb{R}^D$. Whenever it causes no confusion we do not distinguish an edge $e = (u, v)$ in the network under consideration and the line segment uv.

2 Lower bounds

In this section we prove lower bounds on the dilation that can be achieved with $n - 1 + k$ edges for $0 \leqslant k < n$. We prove this lower bound for point sets S in the plane $\mathbb{R}^2$.

2.1 Steiner trees

We first show a lower bound on the dilation of any Steiner tree for S. The lower bound for this case uses the set S of n points $p_1, p_2, \ldots, p_n$ spaced equally on the unit circle, as shown in Fig. 1(a).

Theorem 1. *For any $n > 1$, there is a set S of n points in convex position such that any Steiner tree on S has dilation at least $\frac{1}{\sin(\pi/n)} \geqslant \frac{n}{\pi}$.*

Proof. Consider the set S described above and illustrated in Fig. 1(a). Let o be the center of the circle, and let T be a Steiner tree on S. First, let us assume that o does not lie on an edge of the tree.

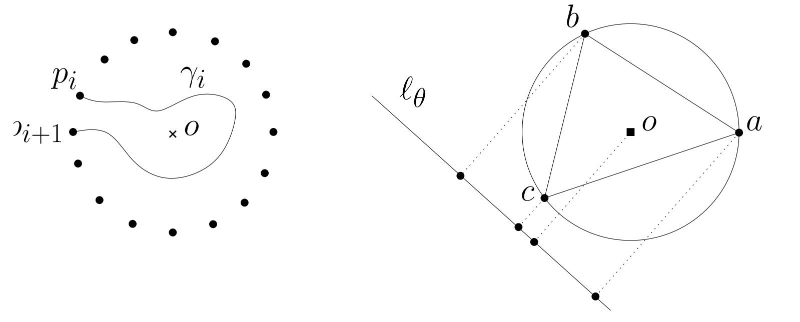

Fig. 1. (a) The homotopy class of the path γ_i. (b) Illustrating the proof of Lemma 1.

Let x and y be two points and let γ and γ' be two paths from x to y avoiding o. We call γ and γ' *(homotopy) equivalent* if γ can be deformed continuously into γ' without ever passing through the point o, that is, if γ and γ' belong to the same homotopy class in the punctured plane $\mathbb{R}^2 \setminus \{o\}$.

Let γ_i be the unique path in T from p_i to p_{i+1} (where $p_{n+1} := p_1$). We argue that there must be at least one index i for which γ_i is not equivalent to the straight segment p_ip_{i+1}, as illustrated in Fig. 1(a).

We argue by contradiction. Let Γ be the closed loop formed as the concatenation of $\gamma_1, \ldots, \gamma_n$, and let Γ' be the closed loop formed as the concatenation of the straight segments p_ip_{i+1}, for $i = 1 \ldots n$. If γ_i is equivalent to p_ip_{i+1}, for all i, then Γ and Γ' are equivalent. We now observe that, since Γ' is a simple closed loop surrounding o, it cannot be contracted to a point in the punctured plane (formally, it has winding number 1 around o). On the other hand, Γ is contained in the tree $T \not\ni o$ (viewed as a formal union[2] of its edges) and hence must be contractible in $\mathbb{R}^2 \setminus \{o\}$. Hence Γ and Γ' cannot be equivalent, a contradiction.

Consider now a path γ_i not equivalent to the segment p_ip_{i+1}. Then γ_i must "go around" o, and so its length is at least 2. The distance between p_i and p_{i+1}, on the other hand, is $2\sin(\pi/n)$, implying the theorem.

Now consider the case where o lies on an edge of T. Assume for a contradiction that there is a spanning tree T that has dilation $1/\sin(\pi/n) - \varepsilon$ for some $\varepsilon > 0$. Let o' be a point not on T at distance $\varepsilon/100$ from o. Then we can use the argument above to show that there are two consecutive points whose path in T must go around o'. By the choice of o' such a path must have dilation larger than $1/\sin(\pi/n) - \varepsilon$, a contradiction. ☐

[2] Notice that T may not be properly embedded in the plane, i.e., the edges of T may cross or overlap. However, viewed not as a subset of the plane, but rather as an abstract simplicial complex, T is certainly simply connected and Γ is a closed curve contained in it and thus contractible, *within* T, to a point. Therefore it is also contractible in the punctured plane, as claimed.

2.2 The case $k = 0$

If we require the tree to be a spanning tree without Steiner points, then the path γ_i in the above proof must not only "go around" o, but must do so using points p_j on the circle only. We can use this to improve the constant in Theorem 1, as follows. Let p_i and p_{i+1} be a pair of consecutive points such that the path γ_i is not equivalent to the segment p_ip_{i+1}. Consider the loop formed by γ_i and $p_{i+1}p_i$. It consists of straight segments visiting some of the points of S. Let C be the convex hull of this loop. We can deal with the case where the center o lies on the boundary of C by moving it slightly, as we did in the proof of Theorem 1. Therefore we can assume that o lies in the interior of C (otherwise, the loop is contractible in the punctured plane) and there exist three vertices v_1, v_2, and v_3 of C such that $o \in \triangle v_1v_2v_3$. Since the loop visits each of these three vertices once, its length is at least the perimeter of $\triangle v_1v_2v_3$, which is at least 4 by Lemma 1 below. Therefore, we have proven

Corollary 1. *For any $n > 1$,*

$$\delta_C(n, 0) \geqslant \frac{4 - 2\sin(\pi/n)}{2\sin(\pi/n)} \geqslant \frac{2n}{\pi} - 1.$$

Lemma 1. *Any triangle inscribed in a unit circle and containing the circle center has perimeter at least 4.*

Proof. Let o be the circle center, and let a, b, and c be three points at distance one from o such that o is contained in the triangle $\triangle abc$. We will prove that the perimeter $p(\triangle abc)$ is at least 4.

We need the following definition: For a compact convex set C in the plane and $0 \leqslant \theta < \pi$, let $w(C, \theta)$ denote the *width of C in direction θ*. More precisely, if ℓ_θ is the line through the origin with normal vector $(\cos\theta, \sin\theta)$, then $w(C, \theta)$ is the length of the orthogonal projection of C to ℓ_θ.

The Cauchy-Crofton formula [5] allows us to express the perimeter $p(C)$ of a compact convex set C in the plane as $p(C) = \int_0^\pi w(C, \theta)d\theta$.

We apply this formula to $\triangle abc$, and consider its projection on the line ℓ_θ (for some θ). The endpoints of the projection of $\triangle abc$ are projections of two of the points a, b, and c. Without loss of generality, for a given θ, let these be a and b, and assume that c projects onto the projection of the segment ob. Then we clearly have $w(oc, \theta) \leqslant w(ob, \theta)$. Since o lies in C, we also have $w(oc, \theta) \leqslant w(oa, \theta)$ (consider the angles the segments oa, ob and oc make with the line ℓ_θ). This implies $3w(oc, \theta) \leqslant w(oa, \theta) + w(ob, \theta) + w(oc, \theta)$, and therefore

$$\begin{aligned} w(\triangle abc, \theta) &= w(oa, \theta) + w(ob, \theta) \\ &= w(oa, \theta) + w(ob, \theta) + w(oc, \theta) - w(oc, \theta) \\ &\geqslant \tfrac{2}{3}(w(oa, \theta) + w(ob, \theta) + w(oc, \theta)). \end{aligned}$$

Integrating θ from 0 to π and applying the Cauchy-Crofton formula gives $p(\triangle abc) \geqslant \frac{2}{3}(p(oa) + p(ob) + p(oc))$. Since oa, ob, and oc are segments of length 1, each has perimeter 2, and so we have $p(\triangle abc) \geqslant \frac{2}{3} \cdot 6 = 4$. □

2.3 The general case

We now turn to the general case, and we consider graphs with $n - 1 + k$ edges for $0 < k < n$.

Theorem 2. *For any n and any k with $0 < k < n$,*

$$\delta(n,k) \geqslant \frac{2}{\pi} \cdot \left\lfloor \frac{n}{k+1} \right\rfloor - 1.$$

Proof. Our example S consists of $k+1$ copies of the set used in Theorem 1. More precisely, we choose sets S_i, for $1 \leqslant i \leqslant k+1$, each consisting of at least $\lfloor n/(k+1) \rfloor$ points. We place the points in S_i equally spaced on a unit-radius circle with center at $(2ni, 0)$, as in Fig. 2. The set S is the union of $S_1, \ldots, S_{k+1}$; we choose the sizes of the S_i such that S contains n points.

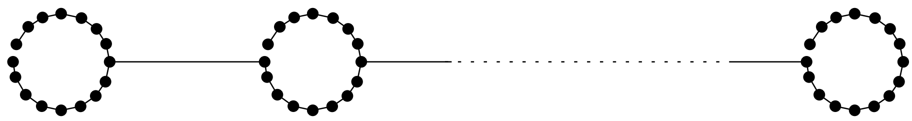

Fig. 2. Illustrating the point set S constructed in the proof of Theorem 2.

Let G be a graph with vertex set S and $n-1+k$ edges. We call an edge of G *short* if its endpoints lie in the same set S_i, and *long* otherwise. Since G is connected, there are at least k long edges, and therefore at most $n-1$ short edges. Since $\sum |S_i| = n$, this implies that there is a set S_i such that the number of short edges with endpoints in S_i is at most $|S_i| - 1$. Let G' be the induced subgraph of S_i. By Corollary 1, its dilation is at least

$$\frac{2 - \sin(\pi/\lfloor n/(k+1) \rfloor)}{\sin(\pi/\lfloor n/(k+1) \rfloor)} \geqslant \frac{2}{\pi} \cdot \left\lfloor \frac{n}{k+1} \right\rfloor - 1.$$

Since any path connecting two points in S_i using a long edge has dilation at least n, this implies the claimed lower bound on the dilation of G. □

2.4 Points in convex position

The point set of Theorem 1 is in convex position, but works as a lower bound only for $k = 0$. In fact, by adding a single edge (the case $k = 1$) one can reduce the dilation to a constant. Now consider n points that lie on the boundary of a planar convex figure with aspect ratio at most ρ, that is, with the ratio of diameter to width at most ρ. It is not difficult to see that connecting the points along the boundary—hence, using n edges—leads to a graph with dilation $\Theta(\rho)$. However, the following theorem shows that for large aspect ratio, one cannot do much better than in the general case.

Theorem 3. *For any n and any k with $0 \leqslant k < n$,*

$$\delta_{\mathrm{C}}(n,k) = \Omega\Big(\frac{n}{(k+1)\log n}\Big).$$

Proof. The proof will be shown for $k = 0$, but the construction can be generalized to hold for any $k > 1$ by using the same idea as in the proof of Theorem 2; placing $k+1$ copies of the construction along a horizontal line, with sufficient space between consecutive copies.

For the case when $k = 0$, set $n = 4m + 2$ and let $o := (0, 0)$. Consider the function $f(i) = (1 + \frac{\ln n}{n})^i - 1$. Our construction consist of the n points S with coordinates $(f(i), 1)$, $(f(i), -1)$, $(-f(i), 1)$, and $(-f(i), -1)$, where $0 \leqslant i \leqslant m$, as shown in Fig. 3. (This set is not in convex position in the sense that all points are extreme and the reader should verify that a slight perturbation can be applied to turn S into a set in strictly convex position.) Consider a minimum dilation tree T of S. As in the proof of Lemma 1, we can now argue that there must exist two "consecutive points" p and q in S such that the path in T connecting them is not (homotopy) equivalent to the straight-line segment between them in the punctured plane $\mathbb{R}^2 \backslash \{o\}$.

We first assume that $p = (f(i), z)$ and $q = (f(i + 1), z)$, where z is either 1 or -1 and $0 \leqslant i < m$. The Euclidean distance between them is $f(i + 1) - f(i) = \frac{\ln n}{n}(1 + f(i))$, and the length of the shortest path in T is at least $\sqrt{2}(1 + f(i))$. The two bounds imply that the dilation of T is at least $\frac{\sqrt{2}n}{\ln n}$ in this case.

Up to symmetry, the only remaining case is that the two points are $p = (-f(n), 1)$ and $q = (-f(n), -1)$, or $p = (f(n), 1)$ and $q = (f(n), -1)$. In both cases the Euclidean distance between them is 2, and the length of the shortest path in T between them is at least $2 \cdot f(n)$, thus T has dilation $f(n)$ in this case. It remains to bound $f(n)$; which can be done by using the inequality $(1 + t/n)^n \geqslant e^t(1 - t^2/n)$. We obtain

$$f(n) = (1 + \frac{\ln n}{n})^n - 1 \geqslant n(1 - \frac{\ln^2 n}{n}) - 1 = \Omega(n),$$

which concludes the proof of the theorem. □

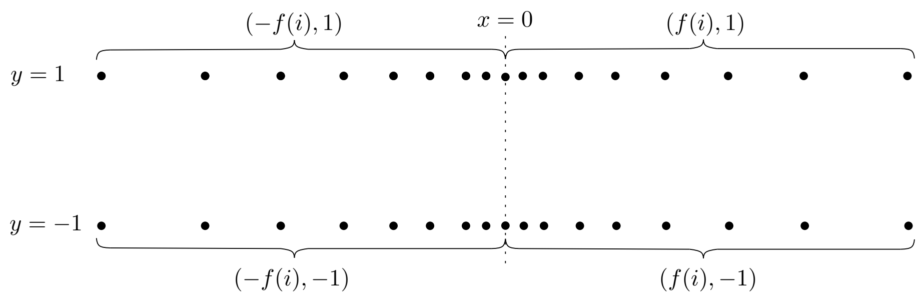

Fig. 3. Illustrating the lower bound construction used in the proof of Theorem 3 with $n = 34$.

3 A constructive upper bound

In this section we show an upper bound on the dilation achievable with k extra edges. We make use of the fact (observed by Eppstein [7]) that a minimum spanning tree has linear dilation. More precisely, we will use the following lemma.

Lemma 2. *Let S be a set of n points in $\mathbb{R}^D$, and let T be a minimum spanning tree of S. Then T has dilation at most $n - 1$. In other words, $\Delta(S, 0) \leqslant \Delta(T) \leqslant n - 1$ and, as it holds for all S with $|S| = n$, we have $\delta(n, 0) \leqslant n - 1$.*

Proof. Let $p, q \in S$ and consider the path γ connecting p and q in T. Since T is a minimum spanning tree, any edge in γ has length at least $d(p, q)$. Since γ consists of at most $n-1$ edges, the dilation of γ is at most $n-1$. □

3.1 The planar case, $D = 2$

The following algorithm builds a spanner with at most $n - 1 + k$ edges:

Algorithm 1 SPARSESPANNER(S, k)

Input A set S of n points in the plane and an integer $k \geqslant 0$.
Output A graph G=(S,E).

1: Compute a Delaunay triangulation of S.
2: Compute a minimum spanning tree T of S.
3: **if** $k = 0$ **then**
4:    **return** T.
5: Let $m \leftarrow \lfloor (k+5)/2 \rfloor$.
6: Compute m disjoint subtrees of T, each containing $O(n/m)$ points, by removing $m-1$ edges.
7: $E \leftarrow \emptyset$.
8: **for** each subtree T' **do**
9:    add the edges of T' to E.
10: **for** each pair of subtrees T' and T'' **do**
11:    **if** there is a Delaunay edge (p, q) with $p \in T'$, $q \in T''$ **then**
12:       add the shortest such edge (p, q) to E.
13: **return** $G = (S, E)$.

We first prove the correctness of the algorithm.

Lemma 3. *Algorithm* SPARSESPANNER *returns a graph G with at most $n - 1 + k$ edges and dilation bounded by $O(n/(k + 1))$.*

Proof. Lemma 2 shows that our algorithm is correct if $k = 0$, so from now on we assume that $k \geqslant 1$, and thus $m \geqslant 3$. Consider the graph G' obtained from the output graph G by contracting each subtree T' created in step 6 to a single node. G' is a planar graph with $m \geqslant 3$ vertices, without loops or multiple edges, and so it has at most $3m - 6$ edges. The total number of edges in the output graph is therefore at most

$$n - 1 - (m - 1) + 3m - 6 = n + 2m - 6 \leqslant n + k - 1.$$

Next we prove the dilation of G. Consider two points $x, y \in S$. Let $x = x_0, x_1, \ldots, x_j = y$ be a shortest path from x to y in the Delaunay graph. Its dilation is bounded by $2\pi/(3\cos(\pi/6)) = O(1)$, as shown in [12]. We claim that each edge $x_i x_{i+1}$ in the Delaunay graph has a path in G with dilation $O(n/k)$. The concatenation of these paths yields a path from x to y with dilation $O(n/k)$, proving the lemma.

If x_i and x_{i+1} fall into the same subtree T', then Lemma 2 implies a dilation of $O(n/m) = O(n/k)$.

It remains to consider the case $x_i \in T'$, $x_{i+1} \in T''$, where T' and T'' are distinct subtrees of T. Let (p, q) be the edge with $p \in T'$, $q \in T''$ inserted in step 12—such an edge exists, since at least one Delaunay edge between T' and T'' has been considered, namely (x_i, x_{i+1}). Let δ' be a path from x_i to p in T', and let δ'' be a path from q to x_{i+1} in T''. Both paths have

$O(n/k)$ edges. By construction, we have $d(p,q) \leqslant d(x_i, x_{i+1})$. We claim that every edge e on δ' and δ'' has length at most $d(x_i, x_{i+1})$. Indeed, if e had length larger than $d(x_i, x_{i+1})$, then we could remove it from T and insert either (x_i, x_{i+1}) or (p, q) to obtain a better spanning tree. Therefore the concatenation of δ', the edge (p, q), and δ'' is a path of $O(n/k)$ edges, each of length at most $d(x_i, x_{i+1})$, and so the dilation of this path is $O(n/k)$. □

Theorem 4. *Given a set S of n points in the plane and an integer $k \geqslant 0$, a graph G with vertex set S, $n-1+k$ edges, and dilation $O(n/(k+1))$ can be constructed in $O(n \log n)$ time.*

Proof. We use algorithm SparseSpanner. Its correctness has been proven in Lemma 3. Steps 1 and 2 can be implemented in $O(n \log n)$ time [7]. Step 6 can be implemented in linear time as follows. Orient T by arbitrarily choosing a root node. Traverse T in postorder, keeping track of the size $|T_v|$ of the (remaining) subtree T_v rooted at the current node v. When $|T_v|$ reaches n/m, cut T_v off the main tree by removing the edge connecting v to its parent. Each of the trees rooted at the children of v has size smaller than n/m, and the maximum degree of any node of T is at most six [7, 20]. Therefore, at the time T_v is cut off from the main tree, we have $n/m \leqslant |T_v| < 1 + 5(n/m)$. The argument does not apply when we reach the root, so we are left with one tree that can be arbitrarily small, but has fewer than $6n/m$ vertices. To implement steps 10–12, we scan the edges of the Delaunay triangulation, and keep the shortest edge connecting each pair of subtrees. It can be done in $O(n \log n)$ time. □

3.2 The higher-dimensional case, $D \geqslant 2$

The algorithm of Section 3.1 uses the following three facts. First, the Delaunay triangulation and the minimum spanning tree can be computed in $O(n \log n)$ time. Second, the dilation of the Delaunay triangulation is bounded by a constant. Third, the Delaunay triangulation is a planar graph. Observe that in dimension $D \geqslant 3$, it is unlikely that the minimum spanning tree can be computed in $O(n \log n)$ time. Also, for $D \geqslant 3$, no non-trivial upper bound on the dilation of the Delaunay triangulation is known. Finally, again for $D \geqslant 3$, the Delaunay triangulation is not a planar graph; in particular, it may have $\Theta(n^2)$ edges.

In this section, we show that, by using a minimum spanning tree *of a bounded degree spanner* for S (as opposed to a minimum spanning tree of the point set itself), the results of Section 3.1 are in fact valid for any constant dimension $D \geqslant 2$. Moreover, we show that this result can be obtained by a graph having degree five.

Properties of the minimum spanning tree of a spanner: Let $t \geqslant 1$ be a real number, let G be an arbitrary t-spanner for S, and let T be a minimum spanning tree of G. In the following three lemmas, we prove that T has "approximately" the same properties as an exact minimum spanning tree of the point set S.

Lemma 4. *Let p and q be two points of S. Then every edge on the path in T between p and q has length at most $t \cdot d(p, q)$.*

Proof. Let P be the path in T between p and q, and let (x, y) be an arbitrary edge on P. We will prove by contradiction that $d(x, y) \leqslant t \cdot d(p, q)$. Hence, we assume that $d(x, y) > t \cdot d(p, q)$.

Let Q be a t-spanner path in G between p and q. Since the length of Q is at most $t \cdot d(p, q)$, every edge of Q has length at most $t \cdot d(p, q)$. In particular, (x, y) is not an edge of Q. We may assume without loss of generality that x is between p and y on the path P. Starting at

x, follow the path P towards p, and let x' be the first vertex that is on Q. Similarly, starting at y, follow the path P towards q, and let y' be the first vertex that is on Q. Let P' be the subpath of P between the vertices x' and y', and let Q' be the subpath of Q between the vertices x' and y'. Then, P' and Q' do not have any edge in common, and these two subpaths form a simple cycle in G that contains the edge (x, y).

Let G' be the graph obtained from T, by adding all edges of Q' (that are not in T yet), and deleting the edge (x, y). Then G' is a connected subgraph of G on the point set S, and, since the weight of Q' is less than the weight of (x, y), the weight of G' is less than the weight of T. This is a contradiction and, thus, we have shown that $d(x, y) \leqslant t \cdot d(p, q)$. □

Lemma 5. *The minimum spanning tree T of the t-spanner G is a $(t(n-1))$-spanner for S.*

Proof. Let p and q be two distinct points of S, and let P be the path in T between p and q. By Lemma 4, each edge of P has length at most $t \cdot d(p, q)$. Since P contains at most $n - 1$ edges, it follows that the length of P is at most $t(n-1) \cdot d(p, q)$. □

Lemma 6. *Let m be an integer with $1 \leqslant m \leqslant n-1$, and let T' and T'' be two vertex-disjoint subtrees of T, each consisting of at most m vertices. Let p be a vertex of T', let q be a vertex of T'', and let P be the path in T between p and q. If x is a vertex of T' that is on the subpath of P within T', and y is a vertex of T'' that is on the subpath of P within T'', then*

$$d(x, y) \leqslant (2t(m-1) + 1)d(p, q).$$

Proof. Let P' be the subpath of P between p and x. By Lemma 4, each edge of P' has length at most $t \cdot d(p, q)$. Since P' contains at most $m - 1$ edges, it follows that this path has length at most $t(m-1) \cdot d(p, q)$. On the other hand, since P' is a path between p and x, its length is at least $d(p, x)$. Thus, we have $d(p, x) \leqslant t(m-1) \cdot d(p, q)$. A symmetric argument can be used to show that $d(q, y) \leqslant t(m-1) \cdot d(p, q)$. Therefore, we have

$$\begin{aligned} d(x, y) &\leqslant d(x, p) + d(p, q) + d(q, y) \\ &\leqslant t(m-1) \cdot d(p, q) + d(p, q) + t(m-1) \cdot d(p, q), \end{aligned}$$

completing the proof of the lemma. □

A graph with $n + O(k)$ edges and dilation $O(n/k)$: Let let k be an integer with $1 \leqslant k \leqslant n$. Fix a constant $t > 1$, and let G be a t-spanner for S whose degree is bounded by a constant that only depends on the dimension D. Clearly, the minimum spanning tree T of G has bounded degree as well. Thus, T contains a *centroid edge*, i.e., an edge whose removal from T yields two subtrees, each consisting of at most αn vertices, for some constant $\alpha < 1$ that depends on the degree of T. In fact, a centroid edge can be computed in $O(n)$ time. By repeatedly choosing a centroid edge in the currently largest subtree, we can remove $\ell = O(k)$ edges from T, and obtain vertex-disjoint subtrees $T_0, T_1, \ldots, T_\ell$, each containing $O(n/k)$ vertices. Observe that the vertex sets of these subtrees form a partition of S. Let X be the set of endpoints of the ℓ edges that are removed from T. Then, the size of X is at most 2ℓ, which is $O(k)$.

We define G' to be the graph with vertex set S that is the union of

1. the trees $T_0, T_1, \ldots, T_\ell$, and
2. a t-spanner G'' for the set X, consisting of $O(k)$ edges.

We first observe that the number of edges of G' is bounded from above by $n - 1 + O(k)$.

Lemma 7. *The graph G' has dilation $O(n/k)$.*

Proof. Let p and q be two distinct points of S. Let i and j be the indices such that p is a vertex of the subtree T_i and q is a vertex of the subtree T_j.

First assume that $i = j$. Let P be the path in T_i between p and q. Then, P is a path in G'. By Lemma 4, each edge on P has length at most $t \cdot d(p, q)$. Since T_i contains $O(n/k)$ vertices, the number of edges on P is $O(n/k)$. Therefore, since t is a constant, the length of P is $O(n/k) \cdot d(p, q)$.

Now assume that $i \neq j$. Let P be the path in T between p and q. Let (x, x') be the edge of P for which x is a vertex of T_i, but x' is not a vertex of T_i. Similarly, let (y, y') be the edge of P for which y is a vertex of T_j, but y' is not a vertex of T_j. Then, both (x, x') and (y, y') are edges of T that have been removed when the subtrees were constructed. Hence, x and y are both contained in X and, therefore, are vertices of G''. Let P_i be the path in T_i between p and x, let P_{xy} be a t-spanner path in G'' between x and y, and let P_j be the path in T_j between y and q. The concatenation Q of P_i, P_{xy}, and P_j is a path in G' between p and q.

Since both P_i and P_j are subpaths of P, it follows from Lemma 4 that each edge on P_i and P_j has length at most $t \cdot d(p, q)$. Since T_i and T_j contain $O(n/k)$ vertices, it follows that the sum of the lengths of P_i and P_j is $O(n/k) \cdot d(p, q)$. The length of P_{xy} is at most $t \cdot d(x, y)$ which, by Lemma 6, is also $O(n/k) \cdot d(p, q)$. Thus, the length of Q is $O(n/k) \cdot d(p, q)$. □

Until now, we took for G an arbitrary spanner of bounded degree. Now let G be the t-spanner of Das and Heffernan [4]. This spanner can be computed in $O(n \log n)$ time, and each vertex has degree at most three. Given G, its minimum spanning tree T can be computed in $O(n \log n)$ time. Since a centroid edge can be computed in $O(n)$ time, the subtrees $T_0, T_1, \ldots, T_\ell$ can be computed in $O(n \log n)$ time. Finally, we take for G'' the t-spanner of Das and Heffernan. This spanner G'' can be computed in $O(k \log k) = O(n \log n)$ time, and each vertex has degree at most three.

For these choices of G and G'', the graph G' has dilation $O(n/k)$, it contains $n - 1 + O(k)$ edges, and it can be computed in $O(n \log n)$ time. We analyze the degree of G': Consider any vertex p of G'. If $p \notin X$, then the degree of p in G' is equal to the degree of p in T, which is at most three. Assume that $p \in X$. The graph G'' contains at most three edges that are incident to p. Similarly, the tree T contains at most three edges that are incident to p, but, since $p \in X$, at least one of these three edges is not an edge of G'. Therefore, the degree of p in G' is at most five. Thus, each vertex of G' has degree at most five.

As we have seen above, the graph G' contains at most $n - 1 + O(k)$ edges. Let c be a constant such that it contains at most $n - 1 + ck$ edges.

The final construction: We are now ready to prove the main result of this section. Let k be an integer with $0 \leqslant k \leqslant n$. Consider the constant c that was introduced above.

First assume that $k < c$. Let G be a t-spanner for S, for some constant t, in which each vertex has degree at most three, and let G' be a minimum spanning tree of G. Then, G' has $n - 1 \leqslant n - 1 + k$ edges, degree at most three and, by Lemma 5, the stretch factor of G' is at most $t(n - 1)$, which is $O(n/(k + 1))$.

If $c \leqslant k \leqslant n$, then we apply the previous results with k replaced by k/c. This gives a graph G' with at most $n - 1 + k$ edges, degree at most five, and dilation $O(n/(k + 1))$. Thus, we have proved the following result:

Theorem 5. *Given a set S of n points in $\mathbb{R}^D$ and an integer k with $0 \leqslant k \leqslant n$, a graph G with vertex set S, $n-1+k$ edges, degree at most five, and dilation $O(n/(k+1))$ can be constructed in $O(n \log n)$ time.*

4 Bounded Spread

In this section we consider the case when the set of input points have bounded spread. The *spread* of a set of points S, denoted $s(S)$, is the ratio between the longest and shortest pairwise distances in S. In the Euclidean plane we have $s(S) = \Omega(\sqrt{n})$.

We define the function

$$\delta(n, s, k) = \sup\{\Delta(S, k) \mid S \subset \mathbb{R}^2, |S| = n, s(S) = s\}.$$

Theorem 6. *For any n and any k with $0 \leqslant k < n$,*

$$\delta(n, s, k) = O(s/\sqrt{k+1}),$$

where $s = O(n/\sqrt{k})$.

Proof. Assume without loss of generality that the smallest interpoint distance in S is 1. Let B be a bounding square of S with side length s.

In the case when $k = 0$ partition B into $2s$ vertical strips of width $1/2$. Connect the points in each strip by a path visiting the points from top-to-bottom. Since the smallest distance between two consecutive points along the path is 1 the dilation of a path is at most $2/\sqrt{3}$. Let u denote the topmost point in the leftmost non-empty strip. In the final step of the algorithm the topmost point of each path is connected to u by an edge, see Fig. 4a. The resulting graph has dilation at most $2(s + \frac{2}{\sqrt{3}}s)$, hence, the theorem holds for $k = 0$.

For $k > 0$, let m be a positive integer to be defined below. Partition B into m^2 smaller squares with side length s/m. Consider the points in each subsquare and note that the spread of the point set in each square is bounded by $\sqrt{2}s/m$. Using the approach described in the previous paragraphs we can construct a tree for the points in each square with dilation and diameter at most $2(\sqrt{2}+\frac{2}{\sqrt{3}}) \cdot \frac{s}{m}$. The union of the trees is denoted T. Select an arbitrary point from each square and construct the Delaunay triangulation, denoted D, of the representative points. Keil and Gutwin [11, 12] proved that the Delaunay triangulation of a point set in the Euclidean plane is at most $\frac{2\pi}{3\cos(\pi/6)} \approx 2.42$. The final graph G is the union of T and D.

It remains to prove that the dilation of G is $O(s/m)$ for every pair of points $p, q \in S$. If p and q lie in the same subsquare then the bound follows from the construction of the tree. Otherwise, if p and q lie in different squares let p' and q' be the representative points in the subsquares containing p and q respectively, as shown in Fig. 4b. The length of the path is then:

$$\begin{aligned} d_G(p, q) &\leqslant d_G(p, p') + d_G(p', q') + d_G(q', q) \\ &\leqslant 4(\sqrt{2} + \frac{2}{\sqrt{3}}) \cdot \frac{s}{m} + d_G(p', q') \\ &\leqslant 4(\sqrt{2} + \frac{2}{\sqrt{3}}) \cdot \frac{s}{m} + \frac{2\pi}{3\cos(\pi/6)} \cdot d(p', q') \\ &< 4(\frac{3}{\sqrt{2}} + \frac{2}{\sqrt{3}}) \cdot \frac{s}{m} + \frac{2\pi}{3\cos(\pi/6)} \cdot d(p, q) \end{aligned}$$

Setting $m := \sqrt{k+1}$ completes the proof of the theorem. □

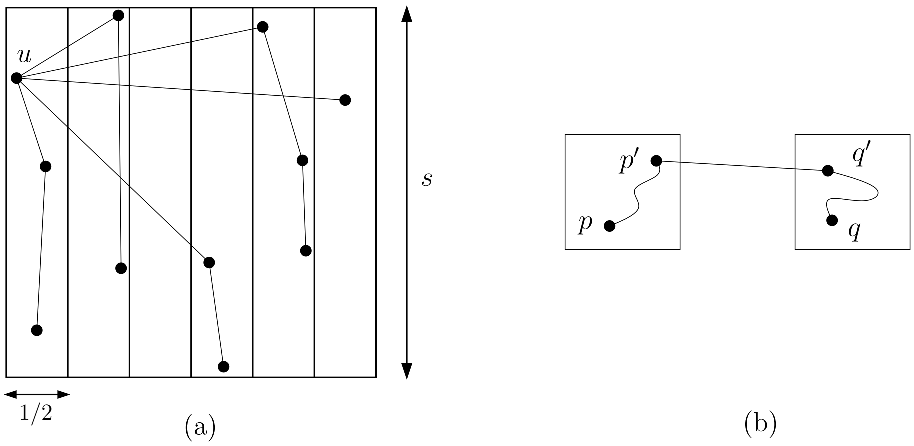

Fig. 4. (a) For every point set with spread s there is a spanning tree with dilation $O(s)$. (b) If p and q lie in different subsquares then there is a path via p' and q' of length $O(s/m \cdot d(p,q))$.

Theorem 7. *For any n and any k with $0 \leqslant k < n$,*

$$\delta(n,s,k) = \Omega(s/\sqrt{k+1}),$$

where $s = O(n/\sqrt{k})$.

Proof. We will construct a set of n points S with spread s such that any graph with $n-1+k$ edges has dilation $\Omega(s/\sqrt{k+1})$. As above set $m := \sqrt{k+1}$.

Consider a square arrangement of $m \times m$ unit circles with centers at $(4i, 4j)$, where $0 \leqslant i, j < m$. Place $\Theta(s/m)$ evenly spaced points on each circle, thus sm points in total. If $sm < n$ then the remaining points can be evenly distributed within each circle such that the smallest interpoint distance is m/s. Since $s = \Omega(\sqrt{n})$ and $s = O(n/m)$ this guarantees that the total number of placed points is n, as illustrated in Fig. 5. This set constitutes the set S.

The spread of S is $\Theta(s)$ since the smallest interpoint distance is m/s and the largest interpoint distance is approximately $4\sqrt{2}m$. It remains to prove that any network with $n-1+k$ edges on S has dilation $\Omega(s/\sqrt{k+1})$.

In the case when $k = 0$ we can apply Theorem 1 to one of the unit circles. The number of points on the circle is s, since $m = \sqrt{k+1} = 1$, which implies that any tree on this set has dilation $\Omega(s)$.

For $k > 0$ consider a minimum dilation graph G on S with $n-1+k$ edges. The graph G can be seen as a collection of m^2 subgraphs where each subgraph G' is induced by the points lying inside or on a common circle and the edges of G' is the edges connecting two points in G'. Since G must be connected G can be seen as a spanning tree of S with k extra edges and by the pigenhole principle at least one of the subgraphs, say G', will contain at most $\frac{n}{m^2} - 1 = \frac{n}{k+1} - 1$ edges.

Using Theorem 1 it holds that the dilation of G' is $\Omega(s/m)$. More specifically, following the proof of Theorem 1 it holds that there are two adjacent points p and q on the convex hull of G' with interpoint distance $O(m/s)$ such that the shortest path between p and q within G' has length $\Omega(m/s)$. Of course a shortest path between two points may use edges of G that are not in G', however, such a path would have length at least 2. As a result the dilation of G would again be $\Omega(m/s)$ which completes the proof. $\square$

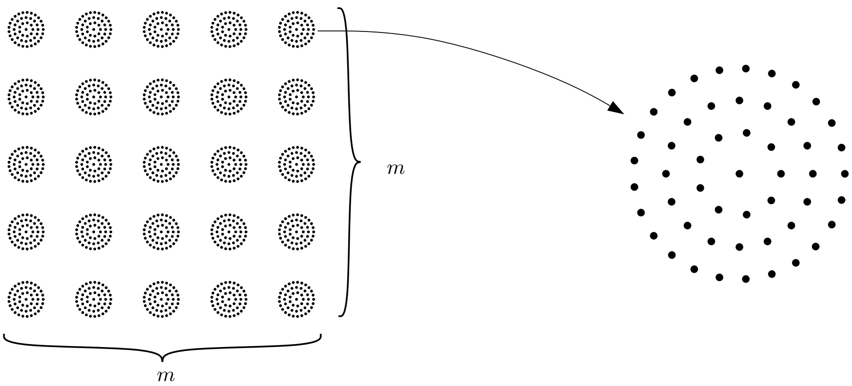

Fig. 5. Illustrating the lower bound construction used in the proof of Theorem 7.

Corollary 2. *Let S be a set of points forming a $(\sqrt{n} \times \sqrt{n})$-grid it holds that $\delta(n,k) = \Theta(n/\sqrt{k+1})$.*

The corollary follows from the fact that a regular grid has spread $s = \Theta(\sqrt{n})$.

5 Conclusions and open problems

We have shown that for any n-point set S in $\mathbb{R}^D$ and any parameter $0 \leqslant k < n$, there is a graph G with vertex set S, $n-1+k$ edges, degree at most five, and dilation $O(n/(k+1))$. We also proved a lower bound of $\Omega(n/(k+1))$ on the maximum dilation of such a graph. An interesting open problem is whether the degree can be reduced.

Minimum dilation graphs are not well understood yet. For instance, it is not known whether a minimum dilation tree for a given point set in the plane may self-intersect [7], not even if the point set is in convex position. (On the other hand, minimum dilation paths or tours can self–overlap, see Fig. 6) No efficient algorithm for computing the minimum dilation tree for a given point set is known. It would be interesting to either give such an algorithm, or show that the problem is NP-hard and look for algorithms that approximate the best possible dilation (instead of giving only a guarantee in terms of n and k, as we do).

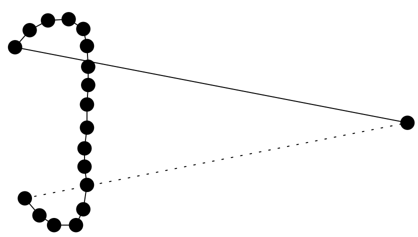

Fig. 6. A minimum dilation path that self-intersects (if the dotted edge is added we get a tour).

References

1. K. Alzoubi, X.-Y. Li, Y. Wang, P.-J. Wan, and O. Frieder. Geometric spanners for wireless ad hoc networks. *IEEE Transactions on Parallel and Distributed Systems*, 14(4):408–421, 2003.
2. P. B. Callahan and S. R. Kosaraju. A decomposition of multidimensional point sets with applications to k-nearest-neighbors and n-body potential fields. *Journal of the ACM*, 42:67–90, 1995.
3. B. Chandra, G. Das, G. Narasimhan, and J. Soares. New sparseness results on graph spanners. *International Journal of Computational Geometry and Applications*, 5:124–144, 1995.
4. G. Das and P. Heffernan. Constructing Degree–3 Spanners with Other Sparseness Properties. *International Journal of Foundations of Computer Science*, 7(2):121–136, 1996.
5. M. do Carmo. *Differential Geometry of Curves and Surfaces.* Prentice Hall, Englewood Cliffs, NJ, 1976.
6. Y. Emek and D. Peleg. Approximating minimum max-stretch spanning trees on unweighted graphs (full version). In *Proc. ACM-SIAM Symposium on Discrete Algorithms*, pages 261–270, 2004.
7. D. Eppstein. Spanning trees and spanners. In J.-R. Sack and J. Urrutia, editors, *Handbook of Computational Geometry*, pages 425–461. Elsevier Science Publishers, Amsterdam, 2000.
8. D. Eppstein and K. Wortman. Minimum dilation stars. In *Proc. ACM Symposium on Computational Geometry*, pages 321–326, 2005.
9. A. M. Farley, A. Proskurowski, D. Zappala, and K. J. Windisch. Spanners and message distribution in networks. *Discrete Applied Mathematics*, 137(2):159–171, 2004.
10. S. P. Fekete and J. Kremer. Tree spanners in planar graphs. *Discrete Applied Mathematics*, 108:85–103, 2001.
11. J. M. Keil and C. A. Gutwin The Delaunay triangulation closely approximates the complete Euclidean graph In *Proc. 1st Workshop on Algorithms and Data Structures*, pages 47-56, 1989.
12. J. M. Keil and C. A. Gutwin. Classes of graphs which approximate the complete Euclidean graph. *Discrete & Computational Geometry*, 7:13–28, 1992.
13. C. Levcopoulos and A. Lingas. There are planar graphs almost as good as the complete graphs and almost as cheap as minimum spanning trees. *Algorithmica*, 8:251–256, 1992.
14. X.-Y. Li. Applications of computational geomety in wireless ad hoc networks. In X.-Z. Cheng, X. Huang, and D.-Z. Du, editors, *Ad Hoc Wireless Networking*. Kluwer, 2003.
15. G. Navarro and R. Paredes. Practical construction of metric t-spanners. In *Proc. 5th Workshop Algorithm Engineering and Experimentation*, pages 69–81. SIAM Press, 2003.
16. G. Navarro, R. Paredes, and E. Chez. t-spanners as a data structure for metric space searching. In *Proc. 9th International Symposium on String Processing and Information Retrieval* , volume 2476 of *Lecture Notes in Computer Science*, pages 298–309. Springer-Verlag, 2002.
17. J. S. Salowe. Constructing multidimensional spanner graphs. *International Journal of Computational Geometry and Applications*, 1:99–107, 1991.
18. M. Smid. Closest point problems in computational geometry. In J.-R. Sack and J. Urrutia, editors, *Handbook of Computational Geometry*, pages 877–935. Elsevier Science Publishers, Amsterdam, 2000.
19. P. M. Vaidya. A sparse graph almost as good as the complete graph on points in K dimensions. *Discrete & Computational Geometry*, 6:369–381, 1991.
20. A. C. Yao. On constructing minimum spanning trees in k-dimensional spaces and related problems. *SIAM Journal on Computing*, 11, 1992.