# A Sparse Graph Almost as Good as the Complete Graph on Points in $K$ Dimensions 

Pravin M. Vaidya*<br>AT\&T Bell Laboratories, Murray Hill, NJ 07974, USA


#### Abstract

A set $V$ of $n$ points in $k$-dimensional space induces a complete weighted undirected graph as follows. The points are the vertices of this graph and the weight of an edge between any two points is the distance between the points under some $L_{p}$ metric. Let $\varepsilon \leq 1$ be an error parameter and let $k$ be fixed. We show how to extract in $O\left(n \log n+\varepsilon^{-k} \log (1 / \varepsilon) n\right)$ time a sparse subgraph $G=(V, E)$ of the complete graph on $V$ such that: (a) for any two points $x, y$ in $V$, the length of the shortest path in $G$ between $x$ and $y$ is at most $(1+\varepsilon)$ times the distance between $x$ and $y$, and (b) $|E|=O\left(\varepsilon^{-k} n\right)$.


## 1. Introduction

A set $V$ of $n$ points in $k$-dimensional space induces a complete weighted undirected graph as follows. The points are the vertices of this graph and the weight of an edge between any two points is the distance between the points under some $L_{p}$ metric. Note that the $L_{p}$ distance between $x=\left(x_{1}, x_{2}, \ldots, x_{k}\right)$ and $y=\left(y_{1}, y_{2}, \ldots, y_{k}\right)$ is given by ( $\left.\sum_{i=1}^{k}\left|x_{i}-y_{i}\right|^{p}\right)^{1 / p}$. We study the problem of compactly representing some of the information represented by this graph. Specifically, given a set $V$ of $n$ points in $k$-dimensional space, we show how to extract a sparse subgraph $G=(V, E)$ of the complete graph on $V$ with the following properties:
(a) Let $d(x, y)$ denote the distance between $(x, y)$, and $D_{G}(x, y)$ denote the length of the shortest path in $G$ between $x$ and $y$. (By convention, $D_{G}(x, x)=0$.) Then, for any pair of points $x, y$ in $V, D_{G}(x, y) \leq(1+\varepsilon) d(x, y)$.
(b) $|E|=O\left(2^{k}\left(3+12 c_{p} / \varepsilon\right)^{k} n\right)$ where $c_{p}=k^{1 / p}$ for the $L_{p}$ metric. Note that $c_{p}$ is the length of the diagonal of a unit box in the $L_{p}$ metric.

[^0]Furthermore, we show how to construct $G$ in $O\left(n \log n+\varepsilon^{-k} \log (1 / \varepsilon) n\right)$ time for fixed $k$. Specifically, it is shown that $G$ can be obtained from $V$ in $O\left(4^{k} n \log n+k \log (k / \varepsilon) 4^{k}\left(3+12 c_{p} / \varepsilon\right)^{k} n\right)$ time. (Note that $G$ can be obtained in $O(n \log n)$ time for fixed $k$ and $\varepsilon$.)

To motivate the problem we briefly discuss two applications. The first application is mentioned in [1] and concerns the design of communication networks. One way to design a communication network on a set of $n$ points $V$ (in say two or three dimensions) is to use the complete graph on $V$; this minimizes the transmission distance between two points but there are $\Omega\left(n^{2}\right)$ links which could be too many. Using $G$ instead of the complete graph reduces the number of links to $O\left(\varepsilon^{-k} n\right)$ at the expense of increasing the transmission distance by a factor of $(1+\varepsilon)$.

The second applications is in finding approximate minimum spanning trees [11] in the complete graph on $V$. Note that the length of a spanning tree is the sum of the lengths of the edges in the spanning tree, and a minimum spanning tree is a spanning tree of minimum length. Since $D_{G}(x, y) \leq(1+\varepsilon) d(x, y)$ for a pair of points $x, y$ in $V$, it follows that there exists a spanning tree in $G$ whose length is at most $(1+\varepsilon)$ times the length of a minimum spanning tree in the complete graph on $V$. Thus a minimum spanning tree in $G$ is a good approximate minimum spanning tree in the complete graph on $V$. The graph $G$ could also be used to find quickly a perfect matching on $V$ whose length is close to the minimum (for details, see [10]).

The problem of approximating the complete graph on a set $V$ of $n$ points in the plane has been studied in [1]. In [1] it is shown that there is a planar subgraph of this complete graph such that the length of the shortest path in the subgraph between any two points $x$ and $y$ in $V$ is at most $\sqrt{10}$ times the euclidean distance between $x$ and $y$. Specifically, the $L_{1}$ Delaunay triangulation [1], [7] of $V$ is such a subgraph. Note that since the subgraph is planar it has $O(n)$ edges. In [4] a similar result is proved about the $L_{2}$ Delaunay triangulation [7] of $V$; specifically, that the length of the shortest path between $x$ and $y$ in the $L_{2}$ Delaunay triangulation is at most 5.08 times the euclidean distance between $x$ and $y$. For a set of points $V$ in the plane the graph $G$ described in this paper is asymptotically a better approximation to the complete graph on $V$ than the Delaunay triangulation, since $D_{G}(x, y) \leq$ $(1+\varepsilon) d(x, y)$ for any pair of points $x$ and $y$ in $V$, and any given $L_{p}$ metric. Also note that the graph $G$ provides a good approximation to the complete graph on points in any fixed dimensional space rather than just for the complete graph on points in the plane, and that the Delaunay triangulation is not useful for $k>2$ because it may not be sparse. Furthermore, if $\varepsilon$ is fixed, then the time for constructing $G$ is asymptotically the same as the time for constructing the Delaunay triangulation of a set of $n$ points in the plane; the construction of each of them requires $\Theta(n \log n)$ time.

It is worth noting that the simple algorithm given below, suggested by T. Feder and N . Nisan [5], will find a subgraph $H$ of the complete graph with distance and sparseness properties similar to those of $G$.

Choose a tolerance angle $\theta$ (related to $\varepsilon$ ) and process edges in increasing order of length as follows. For each edge $(x, y)$, if $H$ does not contain an edge $(x, z)$ (or $(y, z)$ ) such that the angle $z x y$ (or $z y x$ ) is less than $\theta$, then add $(x, y)$ to $H$.
$H$ has bounded degree (for fixed $k$ ) but computing it takes $O\left(n^{2} \log n\right)$ time. The subgraph $G$ on the other hand can be obtained in $O(n \log n)$ time. We also note that in [3] Clarkson describes how to extract a subgraph of the complete graph on points in three-dimensions, with properties similar to those of $G$, in $O\left(n^{2}\right)$ time; his technique utilizes narrow cones and is quite different from ours.

In Section 2 we give some notation and definitions. In Section 3 we define the graph $G=(V, E)$ and in Section 4 we show that $G$ has the properties (a) and (b) mentioned at the beginning of this section. In Section 5 we describe how to extract the graph $G$ from the set of points $V$ in $O\left(n \log n+\varepsilon^{-k} \log (1 / \varepsilon) n\right)$ time for fixed $k$.

## 2. Notation and Definitions

We give some notation useful in defining the graph $G$. We define a box $b$ to be the product $J_{1} \times J_{2} \times \cdots \times J_{k}$ of $k$ intervals (either closed, semiclosed, or open), or, equivalently, the set of those points $x=\left(x_{1}, x_{2}, \ldots, x_{k}\right)$ such that $x_{i}$ lies in the interval $J_{i}$, for $i=1, \ldots, k$. A box is a cubical box iff all the $k$ intervals defining the box are of identical length. The size of cubical box $b$ equals the length of each of the intervals defining $b$ and is denoted by $\operatorname{size}(b)$. Note that the maximum $L_{p}$ distance between a pair of points in a cubical box $b$ is $c_{p} \operatorname{size}(b)$. We only deal with cubical boxes in this paper and box means cubical box. The centre $\psi(b)$ of a box $b$ is defined to be the point $\left(\psi_{1}(b), \ldots, \psi_{k}(b)\right)$ where $\psi_{i}(b)$ is the centre of the $i$ th interval defining the box $b$, for $i=1, \ldots, k$. Let $h_{i}(b)$ be the hyperplane defined by $h_{i}(b)=\left\{x: x_{i}=\right.$ $\left.\psi_{i}(b)\right\}$, let $L\left(h_{i}(b)\right)$ be the left open half-space $\left\{x: x_{i}<\psi_{i}(b)\right\}$, and let $R\left(h_{i}(b)\right)$ be the right closed half-space $\left\{x: x_{i} \geq \psi_{i}(b)\right\}$.

Let Immediate-successors(b) be the set of boxes defined by
Immediate-Successors(b)

$$
=\left\{b^{\prime}: b^{\prime}=b \cap f_{1} \cap \cdots \cap f_{k} \text {, where } f_{i}=L\left(h_{i}(b)\right) \text { or } f_{i}=R\left(h_{i}(b)\right), 1 \leq i \leq k\right\} .
$$

Immediate-Successors $(b)$ is the set of $2^{k}$ boxes obtained upon cutting up $b$ by $k$ mutually orthogonal hyperplanes passing through the center of $b$, each plane being perpendicular to one of the $k$ coordinate axes.

Corresponding to a cubical box $b$, let shrunk( $b$ ) be a cubical box such that:

1. If $|b \cap V| \leq 1$, then $\operatorname{shrunk}(b)=b \cap V$.
2. If $|b \cap V| \geq 2$, then (i) $\operatorname{shrunk}(b) \subseteq b$, (ii) $\operatorname{shrunk}(b) \cap V=b \cap V$, and (iii) the maximum $L_{\infty}$ distance between a pair of points in $\operatorname{shrunk}(b) \cap V$ equals the size of $\operatorname{shrunk}(b)$.

Intuitively, obtaining shrunk(b) corresponds to shrinking $b$ as much as possible without destroying its cubical shape or forcing outside any point in $V$ that is within it.

For a cubical box $b$ such that $|b \cap V| \geq 2$, let Successors( $b$ ) be the set of boxes defined as
$\operatorname{Successors}(b)=\left\{b^{\prime}: b^{\prime}=\operatorname{shrunk}\left(b^{\prime \prime}\right), b^{\prime \prime} \in \operatorname{Immediate}\right.$-Successors $\left.(b),\left|b^{\prime \prime} \cap V\right| \geq 1\right\}$.

Whereas if $|b \cap V| \leq 1$, then Successors $(b)=\varnothing$. Each box in Successors $(b)$ has been shrunk as much as possible so that further shrinkage would either destroy the cubical shape of the box or push out of the box a point in $V$ that was originally in the box.

Let Box-Tree be a rooted tree of boxes defined as follows:
Each node in the Box-Tree is a cubical box, the root is a smallest cubical box containing all the points in $V$, and the children of each nonleaf box (node) $b$ in the Box-Tree are the boxes in Successors(b).

Note that each nonleaf node in the Box-Tree has at least two children. Furthermore, each leaf is a singleton set containing exactly one point in $V$, and the leaf boxes partition the set $V$. The Box-tree is similar to the cell-tree [2] and the quadtree [6].

## 3. The Graph $G=(V, E)$

We now describe the graph $G=(V, E)$. Let $B$ be the set of all the boxes (nodes) in the Box-Tree. Let father ( $b$ ) denote the father or box $b$ in the Box-Tree. For a box $b \in B$, let $r e p(b)$ be a representative point in $b \cap V$. (Note that for a leaf box $b, r e p(b)$ is the unique point in $b \cap V$.) Some of the edges in $E$ incident to $r e p(b)$ will be defined in terms of a list of boxes in the "neighborhood" of $b$ which will be denoted by Near $(b)$. Let $d_{\text {min }}\left(b, b^{\prime}\right)$ denote the minimum distance between a point in $b$ and a point in $b^{\prime}$. For a box $b \in B, N e a r(b)$ is defined as

$$
\begin{aligned}
& \operatorname{Near}(b)=\left\{b^{\prime}: b^{\prime} \in B, \operatorname{size}\left(b^{\prime}\right)<\operatorname{size}(b), \text { size }\left(\text { father }\left(b^{\prime}\right)\right)\right. \\
&\left.\geq \operatorname{size}(b), d_{\min }\left(b, b^{\prime}\right) \leq \frac{6 c_{p}}{\varepsilon} \operatorname{size}(b)\right\}
\end{aligned}
$$

Note that Successors $(b) \subseteq \operatorname{Near}(b)$ and that the boxes in Near $(b)$ are disjoint. Also, observe that the boxes in $B$ are (partially) ordered by containment. Near(b) is obtained by taking the set of all boxes smaller than $b$ whose minimum distance from $b$ is less than (or equal to) a certain threshold ( $\left(6 c_{p} / \varepsilon\right) \operatorname{size}(b)$ ), and then picking the maximal boxes in this set.

The set of edges $E$ (and thus $G=(V, E)$ ) is defined as follows:

$$
E=E_{1} \cup E_{2},
$$

where

$$
E_{1}=\left\{\left(r e p(b), r e p\left(b^{\prime}\right)\right): b \in B, b^{\prime} \in \operatorname{successors}(b), r e p(b) \neq r e p\left(b^{\prime}\right)\right\}
$$

and

$$
E_{2}=\left\{\left(r e p(b), \operatorname{rep}\left(b^{\prime}\right)\right): b \in B, b^{\prime} \in B, b^{\prime} \in \operatorname{Near}(\operatorname{father}(b))\right\}
$$

## 4. $G=(V, E)$ is a Good Approximation to the Complete Graph on $V$

We show that $G=(V, E)$ has the following properties.
(a) For any two points $x, y$ in $V, D_{G}(x, y) \leq(1+\varepsilon) d(x, y)$ where the $D_{G}(x, y)$ denotes the length of the shortest path in $G$ between $x$ and $y$.
(b) $|E|=O\left(2^{k}\left(3+12 c_{p} / \varepsilon\right)^{k} n\right)$ where $c_{p}=k^{1 / p}$ for the $L_{p}$ metric.

To prove property (a) we require the following lemma.
Lemma 1. Let $x$ be a point in $V$ and let $b$ be $a$ box in $B$ such that $x \in b$. Then $D_{G}(x, \operatorname{rep}(b)) \leq 2 c_{p} \operatorname{size}(b)$.

Proof. Let $\left(\{x\}=b_{0}, b_{1}, \ldots, b_{r}=b\right)$ be the path from $\{x\}$ to $b$ in the Box-Tree. Then $b_{i} \in \operatorname{Successors}\left(b_{i+1}\right)$ and $\operatorname{size}\left(b_{i}\right) \leq \frac{1}{2} \operatorname{size}\left(b_{i+1}\right), 1 \leq i<r$. Moreover, for any of the $L_{p}$ metrics there is an edge in $E_{1}$ (and thereby in $E$ ) of length at most $c_{p} \operatorname{size}\left(b_{i+1}\right)$ between $\operatorname{rep}\left(b_{i}\right)$ and $\operatorname{rep}\left(b_{i+1}\right)$. Thus

$$
D_{G}(x, r e p(b)) \leq \sum_{i=0}^{r-1} D_{G}\left(r e p\left(b_{i}\right), \operatorname{rep}\left(b_{i+1}\right)\right) \leq \sum_{i=0}^{r-1} c_{p} \operatorname{size}\left(b_{i+1}\right) \leq 2 c_{p} \operatorname{size}(b) .
$$

We now show that $G$ has property (a) described above.
Lemma 2. For a pair of points $x, y$ in $V, D_{G}(x, y) \leq(1+\varepsilon) d(x, y)$.
Proof. For a point $z \in V$, let $\operatorname{Path}(z, r)$ be the set of all boxes $b \in B$ such that $b$ is on the path from $\{z\}$ to the root in the Box-Tree and size $(b)>r$. Note that, for each box $b \in \operatorname{Path}(z, r), z \in b$ if $r \geq 0$. Let $b_{x}$ be the smallest box in $\operatorname{Path}\left(x, \varepsilon d(x, y) / 6 c_{p}\right)$, and let $b_{y}$ be the smallest box in $\operatorname{Path}\left(y, \varepsilon d(x, y) / 6 c_{p}\right)$. Let $b_{x}$ be the box in Successors $\left(b_{x}\right)$ such that $x \in b_{x}^{\prime}$. (It could be that $b_{x}^{\prime}=\{x\}$.) Similarly, let $b_{y}^{\prime}$ be the box in Successors $\left(b_{y}\right)$ such that $y \in b_{y}^{\prime}$.

We have

$$
\begin{equation*}
D_{G}(x, y) \leq D_{G}\left(x, \operatorname{rep}\left(b_{x}^{\prime}\right)\right)+D_{G}\left(\operatorname{rep}\left(b_{x}^{\prime}\right), \operatorname{rep}\left(b_{y}^{\prime}\right)\right)+D_{G}\left(y, \operatorname{rep}\left(b_{y}^{\prime}\right)\right) . \tag{i}
\end{equation*}
$$

We bound $D_{G}\left(x, \operatorname{rep}\left(b_{x}^{\prime}\right)\right), D_{G}\left(y, \operatorname{rep}\left(b_{y}^{\prime}\right)\right)$, and $D_{G}\left(\operatorname{rep}\left(b_{x}^{\prime}\right), \operatorname{rep}\left(b_{y}^{\prime}\right)\right)$ in terms of $\varepsilon$ and $d(x, y)$. From Lemma 1 it follows that $D_{G}\left(x, \operatorname{rep}\left(b_{x}^{\prime}\right)\right) \leq 2 c_{p} \operatorname{size}\left(b_{x}^{\prime}\right)$ and $\left.D_{G}\left(y, \operatorname{rep}\left(b_{y}^{\prime}\right)\right) \leq 2 c_{p} \operatorname{size}\left(b_{y}^{\prime}\right)\right)$. Since $b_{x}$ is the smallest box in $\operatorname{Path}\left(x, \varepsilon d(x, y) / 6 c_{p}\right)$, and $\quad b_{x}^{\prime} \in \operatorname{Successors}\left(b_{x}\right), \quad \operatorname{size}\left(b_{x}^{\prime}\right) \leq \varepsilon d(x, y) / 6 c_{p}$, and hence $D_{G}\left(x, \operatorname{rep}\left(b_{x}^{\prime}\right)\right) \leq$ $(\varepsilon / 3) d(x, y)$. Similarly, we can show that $D_{G}\left(y, \operatorname{rep}\left(b_{y}^{\prime}\right)\right) \leq(\varepsilon / 3) d(x, y)$. We show below that $D_{G}\left(\operatorname{rep}\left(b_{x}^{\prime}\right), \operatorname{rep}\left(b_{y}^{\prime}\right)\right) \leq(1+\varepsilon / 3) d(x, y)$. Then from (i) above it would follow that $D_{G}(x, y) \leq(1+\varepsilon) d(x, y)$.

We now show that $D_{G}\left(\operatorname{rep}\left(b_{x}^{\prime}\right), \operatorname{rep}\left(b_{y}^{\prime}\right)\right) \leq(1+\varepsilon / 3) d(x, y)$. We assume that the boxes $b_{x}^{\prime}$ and $b_{y}^{\prime}$ are distinct. (Otherwise, $D_{G}\left(r e p\left(b_{x}^{\prime}\right), r e p\left(b_{y}^{\prime}\right)\right)=0$.) Suppose that $\operatorname{size}\left(b_{x}\right) \leq \operatorname{size}\left(b_{y}\right)$. (The case where $\operatorname{size}\left(b_{y}\right) \leq \operatorname{size}\left(b_{x}\right)$ is similar.) Then $b_{y}^{\prime} \in N e a r\left(b_{x}\right)$. This follows from the following three observations and the definition of $\operatorname{Near}\left(b_{x}\right)$. First, since $\operatorname{size}\left(b_{x}\right)>\varepsilon d(x, y) / 6 c_{p}$, we have

$$
d_{\min }\left(\mathrm{b}_{\mathrm{x}}, b_{y}^{\prime}\right) \leq d(x, y) \leq \frac{6 c_{p}}{\varepsilon} \operatorname{size}\left(b_{x}\right) .
$$

Second,

$$
\operatorname{size}\left(b_{y}^{\prime}\right) \leq \frac{\varepsilon d(x, y)}{6 c_{p}}<\operatorname{size}\left(b_{x}\right)
$$

Third, $\operatorname{size}\left(\operatorname{father}\left(b_{y}^{\prime}\right)\right)=\operatorname{size}\left(b_{y}\right) \geq \operatorname{size}\left(b_{x}\right)$.
Since father $\left(b_{x}^{\prime}\right)=b_{x}$ and $b_{y}^{\prime} \in \operatorname{Near}\left(b_{x}\right),\left(\operatorname{rep}\left(b_{x}^{\prime}\right), \operatorname{rep}\left(b_{y}^{\prime}\right)\right) \in E_{2}$ and thus $\left(r e p\left(b_{x}^{\prime}\right)\right.$, $\left.\operatorname{rep}\left(b_{y}^{\prime}\right)\right) \in E$. Then, for any of the $L_{p}$ metrics,

$$
\begin{aligned}
D_{G}\left(r e p\left(b_{x}^{\prime}\right), r e p\left(b_{y}^{\prime}\right)\right) & =d\left(r e p\left(b_{x}^{\prime}\right), r e p\left(b_{y}^{\prime}\right)\right) \\
& \leq d(x, y)+d\left(x, \operatorname{rep}\left(b_{x}^{\prime}\right)\right)+d\left(y, \operatorname{rep}\left(b_{y}^{\prime}\right)\right) \\
& \leq d(x, y)+c_{p} \operatorname{size}\left(b_{x}^{\prime}\right)+c_{p} \operatorname{size}\left(b_{y}^{\prime}\right) \\
& \leq d(x, y)+c_{p} \frac{\varepsilon d(x, y)}{6 c_{p}}+c_{p} \frac{\varepsilon d(x, y)}{6 c_{p}} \\
& \leq\left(1+\frac{\varepsilon}{3}\right) d(x, y)
\end{aligned}
$$

To bound the number of edges in $G$ we require a bound on the size of $N e a r(b)$ for each $b \in B$. Such a bound is provided by the following lemma.

Lemma 3. For $b \in B$, let $A(b)$ be a subset of $B$ satisfying the following two conditions:

1. For each $b^{\prime} \in A(b), d_{\text {min }}\left(b, b^{\prime}\right) \leq r \delta$, and size $\left(\right.$ father $\left.\left(b^{\prime}\right)\right) \geq \delta \geq \operatorname{size}(b)$.
2. For any pair of boxes $b^{\prime}, b^{\prime \prime}$ in $A(b), b^{\prime} \cap b^{\prime \prime}=\varnothing$.

Then $|A(b)| \leq 2^{k}(3+2 r)^{k}$.

Before giving a proof of Lemma 3 we give an intuitive sketch. First, observe that a box $b^{\prime}$ in $A(b)$ is obtained by shrinking a box of size at least $\delta / 2$ in Immediatesuccessors $\left(\right.$ father $\left.\left(b^{\prime}\right)\right)$; thus the parameter $\delta$ is a measure of the size of the empty space around a box in $A(b)$. Based on this observation we can construct a set of disjoint boxes of the same cardinality as $A(b)$, with each box in the set of size $\delta / 2$ and contained in a box of size approximately $2 r \delta$ around $b$. A bound on $|A(b)|$ may then be obtained by noting that at most $O(v r)^{k}, v$ fixed, disjoint boxes of size $\delta / 2$ may be packed in a box of size $2 r \delta$.

Proof of Lemma 3. Recall that Immediate-Successors $\left(b^{\prime}\right)$ is the set of $2^{k}$ boxes obtained by cutting up $b^{\prime}$ by $k$ mutually orthogonal hyperplanes passing through the center of $b^{\prime}$, each plane being perpendicular to one of the $k$ coordinate axes. Let

$$
A_{f}(b)=\left\{b^{\prime}: b^{\prime} \in \operatorname{Immediate-Successors}\left(\text { father }\left(b^{\prime \prime}\right)\right), b^{\prime} \supseteq b^{\prime \prime}, b^{\prime \prime} \in A(b)\right\}
$$

We have

$$
\forall b^{\prime} \in A_{f}(b), \quad \operatorname{size}\left(b^{\prime}\right) \geq \delta / 2 \geq \operatorname{size}(b) / 2
$$

Since the boxes in $A(b)$ are disjoint, the boxes in $A_{f}(b)$ are also disjoint and thus

$$
\left|A_{f}(b)\right|=|A(b)| .
$$

We bound $\left|A_{f}(b)\right|$. Note that, for each $b^{\prime} \in A_{f}(b), d_{\text {min }}\left(b, b^{\prime}\right) \leq r \delta$ and $\operatorname{size}\left(b^{\prime}\right) \geq$ $\delta / 2 \geq \operatorname{size}(b) / 2$. Shrink each box $b^{\prime} \in A_{f}(b)$ to obtain a box $b^{\prime \prime}$ such that $\operatorname{size}\left(b^{\prime \prime}\right)=$ $\delta / 2$ and $d_{\text {min }}\left(b, b^{\prime \prime}\right)=d_{\text {min }}\left(b, b^{\prime}\right)$; let $A_{f}^{\mathrm{s}}(b)$ be the set obtained by shrinking the boxes in $A_{f}(b)$ in this manner. It is clear that

$$
\left|A_{f}^{\mathrm{s}}(b)\right|=\left|A_{f}(b)\right|
$$

Let $\hat{b}$ be a box such that $\psi(\hat{b})=\psi(b)$ (i.e., $b$ and $\hat{b}$ have the same center) and $\operatorname{size}(\hat{b})=(3+2 r) \delta$. This choice for $\operatorname{size}(\hat{b})$ together with the condition that $d_{\text {min }}\left(b, b^{\prime \prime}\right) \leq r \delta$ for each $b^{\prime \prime} \in A_{f}^{\mathrm{s}}(b)$, guarantees that each box in $A_{f}^{\mathrm{s}}(b)$ is a subset of $\hat{b}$.

Since, for each box $b^{\prime \prime} \in A_{f}^{\mathrm{s}}(b), b^{\prime \prime} \subseteq \hat{b}$, and as boxes in $A_{f}^{\mathrm{s}}(b)$ are disjoint we have

$$
\left|A_{f}^{\mathrm{s}}(b)\right| \leq \frac{\text { volume of } \hat{b}}{\text { volume of a box in } A_{f}^{s}(b)} \leq 2^{k}(3+2 r)^{k} .
$$

Thus

$$
|A(b)|=\left|A_{f}(b)\right|=\left|A_{f}^{\mathrm{s}}(b)\right| \leq 2^{k}(3+2 r)^{k} .
$$

The number of edges in $E$ is bounded as follows. The number of edges in $E_{1}$ is upper bounded by the number of nodes (boxes) in the Box-tree, and so $\left|E_{1}\right| \leq 2 n$. There is an edge in $E_{2}$ between $\operatorname{rep}(b)$ and $\operatorname{rep}\left(b^{\prime}\right)$ only if $b \in \operatorname{Near}\left(f a t h e r\left(b^{\prime}\right)\right)$ or $b^{\prime} \in \operatorname{Near}($ father $(b))$. Thus

$$
\left|E_{2}\right| \leq \sum_{b \in B}|N e a r(b)| .
$$

Since the boxes in $\operatorname{Near}(b)$ are disjoint, we can apply Lemma 3 with $A(b)=$ $\operatorname{Near}(b), \delta=\operatorname{size}(b)$, and $r=6 c_{p} / \varepsilon$, and conclude that

$$
|\operatorname{Near}(b)| \leq 2^{k}\left(3+\frac{12 c_{p}}{\varepsilon}\right)^{k}
$$

Hence

$$
\left|E_{2}\right| \leq \sum_{b \in B}|N e a r(b)|=O\left(2^{k}\left(3+\frac{12 c_{p}}{\varepsilon}\right)^{k} n\right)
$$

Finally,

$$
|E| \leq\left|E_{1}\right|+\left|E_{2}\right|=O\left(2^{k}\left(3+\frac{12 c_{p}}{\varepsilon}\right)^{k} n\right)
$$

We have shown that $G$ has the property (b) mentioned at beginning of the section.

## 5. Speedily Constructing $G=(V, E)$

We see that the Box-Tree can be constructed in $O\left(4^{k} n \log n\right)$ time. We also show that once the Box-Tree is available, the sets Near $(b)$ can be obtained for all the boxes $b \in B$ in $O\left(n \log n+k \log (k / \varepsilon) 4^{k}\left(3+12 c_{p} / \varepsilon\right)^{k} n\right)$ time. The representatives in all the boxes in $B$ can be chosen in $O(n)$ time by starting with the leaf boxes in the Box-Tree, and, for a nonleaf box $b$, letting rep(b) equal rep( $b^{\prime}$ ) for some $b^{\prime} \in$ Successors(b). It then follows that $G$ can be obtained in $O\left(4^{k} n \log n+\right.$ $\left.k \log (k / \varepsilon) 4^{k}\left(3+12 c_{p} / \varepsilon\right)^{k} n\right)$ time.

The Box-Tree can be obtained in $O\left(4^{k} n \log n\right)$ time as a byproduct of the All-Nearest-Neighbors Algorithm in [9]. We start with a tree consisting of just the root box, and grow the tree by splitting a leaf box in the current tree that has the largest volume among all the leaf boxes in the current tree. A box $b$ is split by $k$ mutually perpendicular hyperplanes through its center to give the boxes in Immediate-Successors( $b$ ) and the boxes in Immediate-Successors( $b$ ) are then suitably shrunk to obtain the boxes in Successors( $b$ ). The leaf boxes in the current tree partition the points in $V$. For each leaf box $b$ in the current tree, $k$ sorted lists of the points in $b \cap V$ are maintained, each list containing the points ordered on one of the $k$ coordinates. The ordered lists enable efficient splitting of boxes. For details the reader may refer to [9].

We now describe how to obtain the sets Near(b) for the boxes in $B$. The boxes in $B$ are processed in non-increasing order of size; $B_{\mathrm{p}}$ denotes the set of processed boxes in $B$. Let $B_{\mathrm{s}}$ be the set of boxes given by

$$
B_{\mathrm{s}}=\left\{b: \exists b^{\prime} \in B_{p} \text { s.t. } b \in \operatorname{Successors}\left(b^{\prime}\right), b \notin B_{\mathrm{p}}\right\} .
$$

For each box $b \in B_{s}$, we maintain two sets of boxes, $\alpha(b)$ and $\beta(b)$, where

$$
\alpha(b)=\left\{b^{\prime}: b^{\prime} \in B_{\mathbf{s}}, d_{\min }\left(b, b^{\prime}\right) \leq \frac{6 c_{p}}{\varepsilon} \operatorname{size}(b)\right\}
$$

and

$$
\beta(b)=\left\{b^{\prime}: b \in \alpha\left(b^{\prime}\right)\right\} .
$$

The set $\alpha(b)$ is eventually used to obtain the set $N e a r(b)$.
At each step the largest box in $B_{\mathrm{s}}$, denoted by $b_{\mathrm{L}}$, is processed. The set $\operatorname{Near}\left(b_{\mathrm{L}}\right)$ is obtained from $\alpha\left(b_{\mathrm{L}}\right), b_{\mathrm{L}}$ is moved from $B_{\mathrm{s}}$ to $B_{\mathrm{p}}$ and the boxes in $\operatorname{Successors}\left(b_{\mathrm{L}}\right)$ are added to $B_{\mathrm{s}}$. The $\alpha$ and $\beta$ sets are created for the boxes in $\operatorname{Successors}\left(b_{\mathrm{L}}\right)$, and suitably updated for the boxes in $\alpha\left(b_{\mathrm{L}}\right) \cup \beta\left(b_{\mathrm{L}}\right)$. We note that the $\alpha$ sets of only the boxes in Successors $\left(b_{\mathrm{L}}\right) \cup \beta\left(b_{\mathrm{L}}\right)$ are affected during a step, and the $\beta$ sets of only the boxes in Successors $\left(b_{\mathrm{L}}\right) \cup \alpha\left(b_{\mathrm{L}}\right)$ are affected during a step.

To compute $\operatorname{Near}\left(b_{\mathrm{L}}\right)$ quickly during a step we rely on the following lemma which is proved later.

Lemma 4. At the start of each step,

$$
\operatorname{Near}\left(b_{\mathrm{L}}\right) \subseteq\left(\alpha\left(b_{\mathrm{L}}\right) \cup\left(\bigcup_{b \in \alpha\left(b_{\mathrm{L}}\right)} \operatorname{Successors}(b)\right)\right.
$$

To create and update the $\alpha, \beta$ sets efficiently we utilize the following observations:

1. A box $b$ is added to or deleted from $\beta\left(b^{\prime}\right)$ whenever $b^{\prime}$ is added to or deleted from $\alpha(b)$.
2. For each $b \in \operatorname{Successors}\left(b_{\mathrm{L}}\right), \alpha(b) \subseteq\left(\alpha\left(b_{\mathrm{L}}\right) \cup \operatorname{Successors}\left(b_{\mathrm{L}}\right)\right)$ and $\beta(b) \subseteq$ $\left(\beta\left(b_{\mathrm{L}}\right) \cup \operatorname{Successors}\left(b_{\mathrm{L}}\right)\right)$.
3. For each $b \in \alpha\left(b_{\mathrm{L}}\right)$ (resp. $b \in \beta\left(b_{\mathrm{L}}\right)$ ), only a box in $\operatorname{Successors}\left(b_{\mathrm{L}}\right)$ can get added to $\beta(b)(r e s p . \alpha(b))$.

We now give the algorithm for computing the sets $\operatorname{Near}(b)$. Initially, $B_{p}$ is empty and $B_{\mathrm{s}}$ is the singleton set containing the root of the Box-Tree.

## Procedure Construct-Near-sets

Begin
/* Initialize */
$B_{\mathrm{p}}:=\varnothing ; B_{\mathrm{s}}:=\{$ root $\} ; b_{\mathrm{L}}:=$ root $;$
$\alpha($ root $):=\{$ root $\} ; \beta($ root $):=\{$ root $\} ;$
While $\operatorname{size}\left(b_{\mathrm{L}}\right)>0$ do
Begin
/* Process $b_{\mathrm{L}}$ */
/* Compute Near $\left(b_{\mathrm{L}}\right)$ from $\alpha\left(b_{\mathrm{L}}\right)^{*}$ /
$\operatorname{Near}\left(b_{\mathrm{L}}\right):=\alpha\left(b_{\mathrm{L}}\right) \cup\left(\bigcup_{b \in \alpha\left(b_{\mathrm{L}}\right)}\right.$ Successors $\left.(b)\right)$;
Delete from Near $\left(b_{\mathrm{L}}\right)$ each box $b$ such that
$\operatorname{size}(b) \geq \operatorname{size}\left(b_{\mathrm{L}}\right)$ or $\operatorname{size}\left(\operatorname{father}(b)<\operatorname{size}\left(b_{\mathrm{L}}\right)\right.$;
Delete from Near $\left(b_{\mathrm{L}}\right)$ each box $b$ such that $d_{\text {min }}\left(b_{\mathrm{L}}, b\right)>\left(6 c_{p} / \varepsilon\right) \operatorname{size}\left(b_{\mathrm{L}}\right)$; $/$ * Update $\alpha, \beta$ sets and $B_{\mathrm{p}}, B_{\mathrm{s}}{ }^{*} /$

For all $b \in \operatorname{Successors}\left(b_{\mathbf{L}}\right)$ do
$\alpha(b):=\alpha\left(b_{\mathrm{L}}\right) \cup \operatorname{Successors}\left(b_{\mathrm{L}}\right)-\left\{b_{\mathrm{L}}\right\} ;$
$\beta(b):=\beta\left(b_{\mathrm{L}}\right) \cup \operatorname{Successors}\left(b_{\mathrm{L}}\right)-\left\{b_{\mathrm{L}}\right\} ;$
For all $b \in\left(\beta\left(b_{\mathrm{L}}\right)-\left\{b_{\mathrm{L}}\right\}\right)$ do
$\alpha(b):=\alpha(b) \cup \operatorname{Successors}\left(b_{\mathrm{L}}\right)-\left\{b_{\mathrm{L}}\right\} ;$
For all $\left.b \in\left(\alpha\left(b_{\mathrm{L}}\right)\right)-\left\{b_{\mathrm{L}}\right\}\right)$ do
$\beta(b):=\beta(b) \cup$ Successors $\left(b_{\mathrm{L}}\right)-\left\{b_{\mathrm{L}}\right\} ;$
For all $b \in\left(\operatorname{Successors}\left(b_{\mathrm{L}}\right) \cup \beta\left(b_{\mathrm{L}}\right)-\left\{b_{\mathrm{L}}\right\}\right.$ do
Delete from $\alpha(b)$ each box $b^{\prime}$ such that $d_{\text {min }}\left(b, b^{\prime}\right)>\left(6 c_{p} / \varepsilon\right) \operatorname{size}(b)$ and whenever $b^{\prime}$ is deleted from $\alpha(b)$ also delete $b$ from $\beta\left(b^{\prime}\right)$;
$B_{\mathrm{p}}:=B_{\mathrm{p}} \cup\left\{b_{\mathrm{L}}\right\} ; B_{\mathrm{s}}:=B_{\mathrm{s}} \cup \operatorname{Successors}\left(b_{\mathrm{L}}\right)-\left\{b_{\mathrm{L}}\right\} ;$
$b_{\mathrm{L}}:=$ largest box in $B_{\mathrm{s}}$;
endwhile
For each leaf box $b \in B, N e a r(b):=\varnothing$;
end Construct-Near-sets
First, we show the correctness of the above procedure, and then bound the time requirement. An execution of the while loop in the above procedure is referred to as a step. It is easily seen that the $\alpha$ and $\beta$ sets are correctly updated at each step.

Lemma 5. Let $\Pi=\left(b_{0}, b_{1}, \ldots, b_{m}\right), m \geq 1$, be any path in the Box-Tree such that $b_{0}$ is a leaf, $b_{i+1}=$ father $\left(b_{i}\right), 0 \leq i<m$, and $b_{m}$ is the root. At the start of each step $\Pi$ satisfies the following condition:

There exists an $r \geq 0$ such that (1) $b_{r} \in B_{\mathrm{s}}$, (2) $b_{i} \in\left(B-\left(B_{\mathrm{p}} \cup B_{\mathrm{s}}\right)\right), 0 \leq i<r$, and (3) $b_{j} \in B_{\mathrm{p}}, r<j \leq m$.

Proof. By easy induction based on the observation that at the end of a step $b_{\mathrm{L}}$ is moved from $B_{\mathrm{s}}$ to $B_{\mathrm{p}}$ and all the boxes in Successors $\left(b_{\mathrm{L}}\right)$ are added to $B_{\mathrm{s}}$.

From Lemma 4 it follows that when the above procedure has terminated, $N e a r(b)$ has been correctly computed for each box in $B_{\mathrm{p}}$. From Lemma 5 it follows that when the procedure terminates, $B_{\mathrm{p}} \cup B_{\mathrm{s}}=B$, and each box in $B_{\mathrm{s}}$ is a leaf. So the procedure correctly computes $\operatorname{Near}(b)$ for all $b \in B$.

To bound the time requirement of the procedure we need to bound the sizes of the $\alpha$ and $\beta$ sets. The next lemma is useful in proving bounds on the sizes of the $\alpha$ and $\beta$ sets.

Lemma 6. At the start of each step the following statements hold:
(1) The boxes in $B_{\mathrm{s}}$ are disjoint.
(2) For each $b \in B_{\mathrm{s}}$, $\operatorname{size}($ father $(b)) \geq \operatorname{size}\left(b_{\mathrm{L}}\right)$.

Proof. (1) follows from Lemma 5. (2) follows by an easy induction based on the observation that during each step the largest box in $B_{\mathrm{s}}$ is processed.

We next bound the sizes of the $\alpha$ and $\beta$ sets.

Lemma 7. At the start of each step, for each box $b \in B_{\mathrm{s}}$ the following bounds hold:
(I) $|\alpha(b)| \leq 2^{k}\left(3+12 c_{p} / \varepsilon\right)^{k}$.
(II) $|\beta(b)| \leq 2^{k}\left(3+12 c_{p} / \varepsilon\right)^{k}$.

Proof. From the definition of $\alpha(b)$ and Lemma 6 it follows that the conditions of Lemma 3 are satisfied with $A(b)=\alpha(b), \delta=\operatorname{size}(b)$, and $r=6 c_{p} / \varepsilon$. (I) then follows by the application of Lemma 3.

Next note that, for each box $b^{\prime} \in \beta(b)$,

$$
d_{\min }\left(b, b^{\prime}\right) \leq \frac{6 c_{p}}{\varepsilon} \operatorname{size}\left(b^{\prime}\right) \leq \frac{6 c_{p}}{\varepsilon} \operatorname{size}\left(b_{\mathrm{L}}\right)
$$

Then from Lemma 6 it follows that the conditions of Lemma 3 are satisfied with $A(b)=\beta(b), \delta=\operatorname{size}\left(b_{\mathrm{L}}\right)$, and $r=6 c_{p} / \varepsilon$. (II) then follows by the application of Lemma 3.

We now show Lemma 4.

Lemma 4. At the start of each step,

$$
\operatorname{Near}\left(b_{\mathrm{L}}\right) \subseteq\left(\alpha\left(b_{\mathrm{L}}\right) \cup\left(\bigcup_{b \in \alpha\left(b_{\mathrm{L}}\right)} \operatorname{Successors}(b)\right) .\right.
$$

Proof. Note that
$\operatorname{Near}\left(b_{\mathrm{L}}\right)$

$$
=\left\{b: \operatorname{size}(b)<\operatorname{size}\left(b_{\mathrm{L}}\right), \operatorname{size}(\operatorname{father}(b)) \geq \operatorname{size}\left(b_{\mathrm{L}}\right), d_{\min }\left(b_{\mathrm{L}}, b\right) \leq \frac{6 c_{p}}{\varepsilon} \operatorname{size}\left(b_{\mathrm{L}}\right)\right\}
$$

We first show that

$$
\begin{equation*}
\operatorname{Near}\left(b_{\mathrm{L}}\right) \subseteq\left(B_{\mathrm{s}} \cup\left(\bigcup_{b \in B_{\mathrm{s}}} \operatorname{Successors}(b)\right)\right) \tag{i}
\end{equation*}
$$

From Lemmas 5 and 6 it follows that, for all $b \in B_{\mathrm{p}}$, $\operatorname{size}(b) \geq \operatorname{size}\left(b_{\mathrm{L}}\right)$. Thus

$$
\begin{equation*}
\operatorname{Near}\left(b_{\mathrm{L}}\right) \cap B_{\mathrm{p}}=\varnothing \tag{ii}
\end{equation*}
$$

From Lemma 5 we get that, for each $b \in B-\left(B_{\mathrm{p}} \cup B_{\mathrm{s}}\right), \operatorname{size}(b)<\operatorname{size}\left(b_{\mathrm{L}}\right)$. Hence,

$$
\begin{equation*}
\text { for each } b \in \operatorname{Near}\left(b_{\mathrm{L}}\right), \quad \text { father }(b) \in\left(B_{\mathrm{p}} \cup B_{\mathrm{s}}\right) \text {. } \tag{iii}
\end{equation*}
$$

Equation (i) then follows from (ii), (iii), and Lemma 5.

Now suppose that $b \in \operatorname{Near}\left(b_{\mathrm{L}}\right)$. From the definition of $N e a r\left(b_{\mathrm{L}}\right)$ we get that $d_{\text {min }}\left(b_{\mathrm{L}}, b\right) \leq\left(6 c_{p} / \varepsilon\right) \operatorname{size}\left(b_{\mathrm{L}}\right)$ and that $d_{\text {min }}\left(b_{\mathrm{L}}\right.$, father $\left.(b)\right) \leq\left(6 c_{p} / \varepsilon\right) \operatorname{size}\left(b_{\mathrm{L}}\right)$. Thus if $b \in B_{\mathrm{s}}$ we can conclude that $b \in \alpha\left(b_{\mathrm{L}}\right)$. So suppose $b \notin B_{\mathrm{s}}$. Then from (i) we get that father $(b) \in B_{\mathrm{s}}$ and hence father $(b) \in \alpha\left(b_{\mathrm{L}}\right)$. Thus we may conclude that, for each $b \in \operatorname{Near}\left(b_{\mathrm{L}}\right)$, either $b \in \alpha\left(b_{\mathrm{L}}\right)$ or father $(b) \in \alpha\left(b_{\mathrm{L}}\right)$. Lemma 4 then follows.

The running time of Procedure Construct-Near-sets is upper bounded as follows. By maintaining a heap [8] for the boxes in $B_{\mathrm{s}}, b_{\mathrm{L}}$ may be selected in $O(\log n)$ time per step; the total time for heap maintainence and selection of $b_{\mathrm{L}}$ is $O(n \log n)$. From Lemma 7 it follows that the sizes of $\alpha$ and $\beta$ sets are bounded by a constant (dependent on $k$ and $\varepsilon$ ) and hence there are a constant number of additions to $\alpha$ and $\beta$ sets during each step. Thus the total number of additions to and deletions from $\alpha$ and $\beta$ sets is $O(n)$. For a box $b$, we implement $\alpha(b)$ by a data structure which allows insertions and deletions to be performed in $O(\log (|\alpha(b)|))$ time, and allows access to a box $b^{\prime}$, with the largest value of the parameter $d_{\text {min }}\left(b, b^{\prime}\right)$ in $O(\log (|\alpha(b)|))$ time. The set $\beta(b)$ is also implemented by an identical data structure. Note that it suffices to implement $\alpha(b)$ and $\beta(b)$ as a heap or a 2-3 tree [8]. So the $\alpha$ and $\beta$ sets can be maintained in a total of $O(n)$ time. Furthermore, during a step $\operatorname{Near}\left(b_{\mathbf{L}}\right)$ is also obtained in constant time; so computing the Near sets also requires a total of $O(n)$ time. Explicitly evaluating the constants gives a running time of $O(n \log n+$ $\left.k \log (k / \varepsilon) 4^{k}\left(3+12 c_{p} / \varepsilon\right)^{k} n\right)$ for the Procedure Construct-Near-sets.

## 6. Remarks

The number of edges in $G$ have an exponential dependence on the dimension $k$ which could be as bad as $(c k)^{k}$ where $c$ is a constant that does not depend on $k$. An open question is whether this dependence on the dimension can be reduced to $\left(c^{\prime}\right)^{k}$ where $c^{\prime}$ does not depend on $k$; furthermore we would like $c^{\prime}$ to be small, say less than or equal to 4 . Better still is it possible to reduce the dependence on $k$ to a polynomial in $k$ at the expense of increasing the dependence on $n$ by a small factor, say $\log n$ ?

## References

1. P. Chew, There is a planar graph almost as good as the complete graph, Proc. 2nd Annual Symposium on Computational Geometry, 1986, pp. 169-177.
2. K. L. Clarkson, Fast algorithms for the all-nearest-neighbors problem, Proc. 24th Annual Symposium on Foundations Computer Science, 1983, pp. 226-232.
3. K. L. Clarkson, Approximation algorithms for shortest path motion planning, Proc. 19th Annual ACM Symposium Theory of Computing, 1987, pp. 56-65.
4. D. P. Dobkin, S. J. Friedman, and K. J. Supowit, Delaunay graphs are almost as good as complete graphs, Discrete Comput. Geom. 5 (1990), 399-408.
5. T. Feder, Personal communication, 1989.
6. R. A. Finkel and J. L. Bentley, Quad-trees: a data structure for retrieval on composite keys, Acta Inform. 4 (1974), 1-9.
7. F. P. Preparata and M. I. Shamos, Computational Geometry: An Introduction, Springer-Verlag, New York, 1985.
8. E. M. Reingold, J. Nievergelt, and N. Deo, Combinatorial Algorithms: Theory and Practice, Prentice Hall, Englewood Cliffs, NJ, 1977.
9. P. M. Vaidya, An $O(n \log n)$ algorithm for the all-nearest-neighbors problem, Discrete Comput. Geom. 4 (1989), 399-408.
10. P. M. Vaidya, Approximate minimum weight matching on points in $k$-dimensional space, Algorithmica, 4 (1989), 569-584.
11. A. C. Yao, On construction minimum spanning trees in $k$-dimensional space and related problems, SIAM J. Comput., 11 (1982), 721-736.

Received April 14, 1988, and in revised form November 26, 1990.


[^0]:    * Currently at the Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

