The Diameter of Nearest Neighbor Graphs

David Eppstein

Department of Information and Computer Science
University of California, Irvine, CA 92717

Tech. Report 92-76

July 7, 1992

Abstract

Any connected plane nearest neighbor graph has diameter $\Omega(n^{1/6})$. This bound generalizes to $\Omega(n^{1/3d})$ in any dimension d.

For any set of n points in the plane, we define the *nearest neighbor* graph by selecting a unique nearest neighbor for each point, and adding an edge between each point and its neighbor. This is a directed graph with outdegree one; thus it is a pseudo-forest. Each component of the pseudo-forest is a tree, with a length-two directed cycle at the root.

As with minimum spanning trees, the maximum degree in a nearest neighbor graph is five. Monma and Suri [1] showed that, conversely, any tree with vertex degree at most five is the minimum spanning tree of some point set; thus minimum spanning tree topologies are exactly characterized by their degrees. Paterson and Yao [2] considered the corresponding question for nearest neighbor graphs. They showed that a tree with depth D can have at most $O(D^9)$ vertices. Thus unlike minimum spanning trees, nearest neighbor graphs can not be too bushy: a tree with many vertices must contain a long path. Paterson and Yao also constructed an example of a nearest neighbor graph with m points and $\Omega(D^5)$ vertices. There remains a large gap between D^5 and D^9, and we are left with the question of the exact relation between depth and size of nearest neighbor graphs.

In this paper we tighten this gap, by demonstrating that a nearest neighbor graph with diameter m can have at most $O(D^6)$ points. Equivalently, a nearest neighbor graph with n points must have diameter $\Omega(n^{1/6})$.

It is possible that the insight into this problem provided by our proof can remove the remaining gap, by showing what a point set must look like if its nearest neighbor graph is to have diameter $O(n^{1/6})$, or alternately by leading the way to a proof that the diameter must be $\Omega(n^{1/5})$.

Our proof that there can be $O(D^6)$ points in a nearest neighbor graph with diameter D follows the same general outline as Paterson and Yao's proof of their $O(D^9)$ bound, so we summarize that outline here.

Paterson and Yao partition the plane into an infinite sequence of similar annuli, centered on the origin which is chosen to lie on the root of the nearest neighbor tree. The outer radius of each annulus is some suitably large constant C times the inner radius. Each point is assigned to the annulus containing the outermost point on the path from the point to the origin in the nearest neighbor graph. (Most points are assigned to the annulus containing them, but a point may be assigned to a larger annulus if its nearest neighbor path goes away from the origin before returning.) They then count the number of points that can be assigned to any one annulus, and the number of possible annuli.

Lemma 1 (Paterson and Yao [2]). *At most $O(C^2D^2)$ points can be assigned to any annulus.*

Proof: Suppose a point is assigned to an annulus with inner radius r. Then the path from that point to the origin has length at least r. There are at most D edges in the path, and the edge lengths decrease as the path nears the origin, so the first edge must have length at least r/D. Thus each such point is contained in an empty disk with radius r/D. If we halve the radii to $r/2D$, no two disks meet. All such disks are contained in a large disk with radius $(Cr + r/2D)$, which has area $O(C^2r^2)$. If there were more points than the stated bound, the areas of the disks around them would add to more than this. □

Lemma 2 (Paterson and Yao [2]). *If C is sufficiently large, $O(D^7)$ annuli can be assigned points.*

Proof: Each such annulus contains a point assigned to it, namely the outermost on the path from any of its assignees. Thus we need only consider points that are contained in the annuli to which they are assigned, and edges between such points. Define the *depth* of such an edge to be the distance in the nearest neighbor graph from the origin, where distance along a single edge is measured as one. Then each edge has a depth between one and D. We label each annulus A by the seven-tuple $L(A)$ of the smallest depths of edges that begin inside the inner radius of A, and end either within A or outside its outer radius. Paterson and Yao show that, if one annulus has a larger radius than another, its label is larger in lexicographic order; therefore each of the $O(D^7)$ labels is used at most once. □

As in Paterson and Yao's proof, we divide the plane into similar annuli. Our proof retains Lemma 1, but modifies Lemma 2 to show that there can in fact only be $O(D^4)$ annuli to which points are assigned.

Lemma 3. *Let edges e and f pass entirely across a given annulus. Then angle ef is at least $60° - 2 \arcsin 1/C$.*

Proof: Let $|f| > |e|$. The outer endpoint of e must be outside a $60°$ wedge based on f. The points outside the annulus view the circle inside the annulus as having an angular radius of $2 \arcsin 1/C$, and any point outside the wedge will then form an angle of at least $60° - 2 \arcsin 1/C$ with f. □

Lemma 4. *Let e and f pass entirely across a given annulus, and let the outer endpoints of e and f be in two non-adjacent annuli. Then angle ef is at least* $90° - 4 \arcsin 1/C$.

Proof: Let $|f| > |e|$, and let the annulus containing the outer endpoint of e have radii r and Cr. The outer endpoint of f must be outside a circle of radius C^2r since we assume it is in a non-adjacent annulus to the endpoint of e. Then the outer endpoint of e must be outside a circle c with f as radius, centered on the outer endpoint of f (otherwise that endpoint would have the outer endpoint of e as a closer neighbor). The outer endpoint of f views circle Cr as having an angular radius of $2 \arcsin 1/C$, so circle c covers a wedge of annulus (r, Cr), based on f with angle $90° - 2 \arcsin 1/C$, which cannot contain the endpoints of e. The inner endpoints of e and f are within a small circle of angular diameter $2 \arcsin 1/C$ as viewed from the outer endpoint of e. Therefore the angle of e with f must be at least $90° - 4 \arcsin 1/C$. □

If we choose $C \geq 29$, $\arcsin 1/C$ will be less than $2°$. Then the angles will "look like" $60°$ and $90°$, in that the following inequalities hold:

- $7(60° - 2 \arcsin 1/C) > 360°$, so that no annulus can be crossed by seven edges.
- $5(60° - 2 \arcsin 1/C) + (90° - 4 \arcsin 1/C) > 360°$, so that if six edges cross, all angles are smaller than $90° - \epsilon$.
- $3(90° - 2 \arcsin 1/C) + 2(60° - 4 \arcsin 1/C) > 360°$, so that if there are five crossing edges, all but two angles are smaller than $90° - \epsilon$.

Theorem 1. *In any connected plane nearest neighbor graph with diameter D, there are at most $O(D^6)$ points.*

Proof: We consider two types of edge for each annulus A: those that cross the inner but not the outer radius, and those that cross both radii. Two edges of the second type are *related* in A if they have an angle less than $90° - 4 \arctan 1/C$, or if they are connected by a chain of related edges. There can be at most three equivalence classes of related edges, except for the single case that four edges cross the annulus at approximate right angles.

We label each annulus by the 4-tuple of the four smallest edge depths, only allowing a single depth from each equivalence class of related edges. If there are not enough edges to fill out the tuple, we fill the remaining positions with the value D. Unlike the labels used in Lemma 2, these labels need not increase in lexicographic order. However we show that we can find a subsequence of the label sequence, covering a constant fraction of the annuli, in which the labels do increase. $O(D^4)$ labels are possible in this subsequence, and therefore the entire sequence consists of $O(D^4)$ annuli.

Suppose we have constructed some such sequence out to some annulus A. Let S be the set of edges that cross both radii of A. We choose the next annulus A' as follows. We start by considering the next annulus beyond A, to which some point is assigned. Then as long as the annulus we are considering either has an endpoint of an edge in S, or has four edges crossing at approximate right angles, we move on to the next such annulus. Once we find an annulus A' satisfying neither of these conditions, we choose it as the next annulus in the sequence, and continue as before.

Suppose we encounter an annulus with four edges crossing at approximate right angles. No point can be within four circles having those edges as radii, which together cover all of the annulus except for a small region near the center. Any edge crossing the outer annulus boundary would view this region as having angular diameter $x = O(1) \arcsin 1/C$, and by an argument similar to Lemma 3 would have angles greater than $60° - x$ with all four crossing edges. But no such angles, and hence no such edges can exist. The next annulus containing any points contains the endpoint of one of the four crossing edges, and is not of this special form. So as we search from A for the next annulus A', at least every other annulus is not of this form.

The remaining annuli in the search from A for A' each contain an endpoint of an edge in S, so there can be at most six such annuli. Thus after we try $O(1)$ annuli we find the next one in the sequence, and the sequence we construct will contain a constant fraction of all the annuli.

Finally, we show that the label of A' is larger than that of A. Let S' be the set of edges that cross the inner radii of both A and A'. Any such edge is in S, and since A' contains no endpoint of S each such edge crosses the outer radius of A'. All its relatives in A are also in S', and are related in S'. Therefore $|S' \cap L(A')| \leq 3$, and $L(A')$ contains an edge not in S'.

Let e be such an edge with the smallest depth. As we follow the path from e to the tree root, we will eventually encounter an edge e' crossing the inner radius of A. The depth of e' is strictly smaller than that of e. We form a label L' by adding e' to the set $S' \cap L(A')$, and filling all remaining positions in the four-tuple by D. Then since e' has lower depth than all edges in $L(A') - L$, it follows that $L' < L(A')$ in the lexicographic order. On the other hand, all edges in $L' - e'$ are unrelated in A, and no relative of e' can be in S' or hence in L'. Thus L' is a potential label for A and the true label $L(A)$ is no higher in lexicographic order.

We have shown that $L(A) < L(A')$. The subsequence we construct, and thus the entire sequence of annuli, has at most $O(D^4)$ members. Each annulus is assigned $O(D^2)$ points, and the theorem is proved. $\square$

Paterson and Yao [2] generalize their $O(D^9)$ bound to $O(D^{d+\tau(d)+1})$ in any dimension d. The *kissing number* $\tau(d)$ gives the maximum cardinality of any set of vectors in which all angles are $\geq 60°$. The corresponding quantity for angles $\geq 90°$ is simply $2d$. Lemma 4 can be used to show that at most $2d$ subsets of crossing edges can be separated by angles of $90° - \epsilon$, and that this number is further reduced to $2d - 1$ except in the single case of $2d$ approximately orthogonal edges. Lemma 1 generalizes to a bound of $O(D^d)$ points per higher-dimensional annulus. Thus we can improve the bound above to $O(D^{3d})$. This replaces a double exponential ($\tau(d)$ is $\Omega(2^{.207d})$) by a single exponential. There are nearest neighbor graphs based on grid graphs with $\Omega(D^d)$ points, so we are close to the right exponent.

References

[1] C. Monma and S. Suri. Transitions in geometric spanning trees. 7th ACM Symp. Computational Geometry (1991) 239–249.

[2] M.S. Paterson and F.F. Yao. On nearest-neighbor graphs. 19th Int. Colloq. Automata, Languages, and Programming (1992) to appear.