*Comput. Math. Applic.* Vol. 15, No. 1, pp. 53–64, 1988
Printed in Great Britain. 

0097-4943/88 $3.00 + 0.00


# THE EXPECTED SIZE OF SOME GRAPHS IN COMPUTATIONAL GEOMETRY

L. DEVROYE
School of Computer Science, McGill University, Montreal, Quebec H3A 2K6, Canada

(*Received* 18 *August* 1987)

Communicated by E. Y. Rodin

**Abstract**—We consider $n$ independent points with a common but arbitrary density $f$ in $R^d$. Two points $(X_i, X_j)$ are joined by an edge when a certain set $S(X_i, X_j)$ does not contain any other data points. The expected number $E(N)$ of edges in the graph depends upon $n$, $f$ and the definition of $S$. Examples include rectangles, spheres and loons; these lead to the graph of all dominance pairs, the Gabriel graph and the relative neighborhood graph, respectively. Other graphs covered by our analysis include the nearest neighbor graph and the directional nearest neighbor graph. In all cases, we obtain asymptotic lower bounds that do not depend upon $f$ (and are hence useful in all applications involving these graphs, since we usually do not know $f$). For sparse graphs, exact asymptotic constants are obtained for $E(N)$ that are valid for all densities.

## 1. INTRODUCTION

We consider a sample $X_1, \ldots, X_n$ of independent points with a common but arbitrary density $f$ in $R^d$. These points form the vertices of a graph. Two points $(X_i, X_j)$ are joined by an edge when a certain set $S(X_i, X_j)$ does not contain any other data points. The definition of $S$ is not affected by the other points. The number of edges in the graph is $N$. In this paper we seek relationships between $E(N)$ and $S$, $n$ and $f$. To save space, we will call all graphs created in this manner *proximity graphs*, even though this term could be misleading in some cases.

*Example 1*

*Direct dominance pairs.* When $S(X_i, X_j)$ is the rectangle with $X_i$ and $X_j$ as vertices, $X_i$ and $X_j$ are said to define a direct dominance pair. The problem of determining whether a pair is a dominance pair has applications in rectangle enclosure problems [1, 2]. Algorithms for reporting all direct dominance pairs are given in Gutting *et al.* [3]. Klein [4] has shown that the expected number of direct dominance pairs for any density that is a product of $d$ marginal densities is asymptotic to $n \log^{d-1} n/(d-1)!$ ■

*Example 2*

*Gabriel graph.* The Gabriel graph [5] is obtained when $S$ is the sphere centered at $\frac{1}{2}(X_i + X_j)$, with $X_i$ and $X_j$ at opposite poles. It has been used extensively in geographic variation analyses in biology (for a list of references, see Matula and Sokal [6]). Algorithms for finding the Gabriel graph in $R^2$ are discussed in Matula and Sokal [6]. It is also shown there that for the uniform distribution in square, $E(N) \sim 2n$. We will see that for *all* densities, $\liminf E(N)/n \geqslant 2^{d-1}$ and that for most densities, $E(N) \sim 2^{d-1} n$. ■

*Example 3*

*The relative neighborhood graph.* The RNG, or relative neighborhood graph, is obtained by joining all pairs whose loon is empty, where the loon defined by a pair is the intersection of two spheres of equal radius, each having one point as center and the other point on its surface [7, 8]. It is a subgraph of the Gabriel graph. Supowit [9] has obtained an $O(n \log(n))$ algorithm for finding the RNG in two dimensions. We will see that for all densities, $E(N)/n \geqslant C_d + o(1)$, where $C_d$ is a constant depending upon $d$ only ($C_2$ is about 1.27). ■

*Example 4*

*The Delaunay triangulation.* In the Delaunay triangulation, two points are joined by an edge if they are Delaunay triangulation neighbors, i.e. when some sphere with the two points on its surface and center somewhere on the hyperplane that forms the locus of all points at equal distance from both points is empty, i.e. contains no other points. Delaunay triangulations are ubiquitous in computational geometry [10], yet it is still unclear how $E(N)$ is related to $f$. We know of course that $N \leqslant 3n - 6$. Unfortunately, the Delaunay triangulation is not a special case of the kinds of graphs studied here, because the definition of $S$ involves more than two points. However, since the Gabriel graph is a subgraph of the Delaunay triangulation, it is easy to see that the lower bounds derived for the Gabriel graph are applicable to the Delaunay triangulation as well. In particular, $E(N)/n \geqslant (2^{d-1} + o(1))$.

For a general discussion of proximity graphs and their applications, we refer to the survey papers by Toussaint [7, 8, 11]. For example, in Toussaint [7, 8], it is shown that the minimal spanning tree is a subgraph of the RNG, which is a subgraph of the Gabriel graph, which in turn is a subgraph of the Delaunay triangulation. ■

*Example 5*

*Infinite strip graph.* As an example of a more exotic graph, consider the graph formed when $S(X_i, X_j)$ is the infinite strip defined by two parallel hyperplanes through $X_i$ and $X_j$ that are perpendicular to $X_i - X_j$. This graph contains all dominance pairs as a subgraph, and cannot therefore possibly have a linear number of edges on the average. Since edges are created based upon a decision that is not "local", it cannot truly be called a proximity graph. For the same reason, the expected time analysis requires a different collection of tools. Its properties will be studied elsewhere. ■

*Example 6*

*Nearest neighbor graph.* Consider the graph obtained by connecting each point with its nearest neighbor. The nearest neighbor graph is a subgraph of the Euclidean minimum spanning tree, and plays a role in closest point problems [10, pp. 180–181]. It is obvious that we have between $n/2$ and $n$ edges in such a graph. We will prove that for all densities, $E(N)/n \to 0.689\ldots$. ■

*Example 7*

*The sphere of influence graph.* For each $X_i$ in the plane, let $C_i$ be the circle centered at $X_i$ with the nearest neighbor of $X_i$ on its surface. If $C_i$ and $C_j$ have a nonempty intersection, $X_i$ and $X_j$ are connected. The corresponding graph is known as the sphere on influence graph. Avis and Horton [12] have studied the sphere of influence graph, and have shown that it has at most $29n$ edges. They also report that El-Gindy has pointed out that it can be found in $O(n \log(n))$ time by an algorithm of Bentley and Ottmann [13]. Unfortunately, there seems to be no inclusion property between any of the graphs discussed so far and the sphere of influence graph, which can often have several connected components. Also, the result of this paper do not apply directly to these graphs. ■

## 2. A USEFUL INEQUALITY

Most of the results in this paper are simple corollaries of an inequality provided in Theorem 1. That is why we forge directly ahead into a rather technical section. Some restrictions have to be put on $S$.

*Definition of a regular set*

In the halfplane $\{(u, v): u \in R, v \geqslant 0\}$, we consider a fixed bounded set $T$ ($T$ stands for "target set"). $T$ is symmetric about $u = 1/2$. A regular set $S$ is any set for which membership can be determined based upon some $T$, according to the procedure explained below. To determine whether $z \in S(x, y)$, we rotate and translate the space rigidly so that $x$ coincides with the origin, and $y$

coincides with $(\| x - y \|, 0, 0, 0, \ldots, 0)$. Then we shrink the space by a factor $\| x - y \|$ (shrinking by $\lambda$ means that $w$ gets mapped to $\lambda w$). Finally, we rotate around the first axis such that the transformed $z$ ends up in the positive halfplane (i.e. its second coordinate is nonnegative, and all coordinates from the third up are zero). The new location of $z$ after these three operations should fall in $T$. Note that because of the symmetry in $T$, $z \in S(x, y)$ if, and only if, $z \in S(y, x)$. ■

It is easy to see that if $z \in S(x, y)$ for some regular set $S$, than all the points at the same distance of the line $xy$, and with the same projection on that line are also in $S(x, y)$; in other words, we have rotational symmetry about $xy$. Examples include the loon defining the RNG, and the sphere defining the Gabriel graph.

The Gabriel graph, the RNG and the nearest neighbor graph are based on regular sets. The analysis of the direct dominance graph will be carried out elsewhere. The thrust of Theorem 1 is that the expected number of edges in a proximity graph based upon a regular set is virtually independent of $f$; the only factor truly influencing this expected number if the volume of $S(x, y)$ when $\| x - y \| = 1$. It is curious that all such $S$'s with the same volume, regardless of their shape, give rise to the same expected number of edges, asymptotically speaking.

*Theorem 1*

Let $N$ be the number of edges in a graph defined on the basis of a regular set $S$ defined on the basis of a target set $T$. Then, for any density $f$,

$$\liminf_{n \to \infty} \frac{E(N)}{n} \geqslant \frac{V_d}{2\lambda(T)},$$

where $V_d$ is the volume of the unit sphere in $R^d$, and $\lambda(T)$ is the $d$-dimensional volume of the $d$-dimensional set obtained by rotating $T$ about the first axis [equivalently, it is the volume of $S(x, y)$ when $\| x - y \| = 1$].

Furthermore, for almost all $x$,

$$\lim_{n \to \infty} E(N(X_1) | X_1 = x) = \frac{V_d}{\lambda(T)},$$

where $N(X_1)$ is the number of neighbors of $X_1$.

*Proof of Theorem 1*

We let $N(X_i)$ be the number of neighbors of $X_i$, and observe that

$$N = \frac{1}{2} \sum_{i=1}^{n} N(X_i).$$

Assume that for almost all $x_1$ (with respect to $f$),

$$\liminf_{n \to \infty} E(N(x_1)) \geqslant \frac{V_d}{\lambda(T)},$$

where $N(x_1)$ refers to the sample $X_2, \ldots, X_n$. Then, applying Fatou's lemma,

$$\begin{aligned} \liminf_{n \to \infty} \frac{E(N)}{n} &= \liminf_{n \to \infty} \int \frac{E(N(x_1)) f(x_1)}{2n} \, dx_1 \\ &\geqslant \int \liminf_{n \to \infty} \frac{E(N(x_1)) f(x_1)}{2} \, dx_1 \\ &= \frac{V_d}{2\lambda(T)}. \end{aligned}$$

Thus, we need only show the pointwise result. We will use the symbol $\mu(\cdot)$ to denote the probability measure of a set, i.e. the integral of $f$ over the set in question. Also, $\lambda(\cdot)$ denotes Lebesgue measure, the $\epsilon$ is an arbitrary small positive number.

We observe that

$$E(N(x_1)) = (n-1)\int (1-\mu(S(x_1,x_2)))^{n-2} f(x_2)\,dx_2$$

$$\geq (n-1)\int_{\|x_2-x_1\|\leq\delta} (1-(1+\epsilon)f(x_1)\lambda(S(x_1,x_2)))^{n-2} f(x_2)\,dx_2,$$

where $\delta > 0$ is so small that

$$\left|\frac{\mu(S(x_1,x_2))}{\lambda(S(x_1,x_2))} - f(x_1)\right| \leq \epsilon f(x_1)$$

and

$$\left|\frac{\mu(B(x_1,x_2))}{\lambda(B(x_1,x_2))} - f(x_1)\right| \leq \epsilon f(x_1),$$

for all $\|x_2-x_1\| < \delta$, where $B(x_1,x_2)$ is the ball of radius $\|x_2-x_1\|$ centered at $x_1$. The fact that this can be done is a consequence of the fixed structure of $S$ ($S$ can only be translated, rotated and shrunk uniformly in all directions), the boundedness of $S$, and the Lebesgue density theorem [14, pp. 108–109]. It should be noted that $\delta$ depends upon $x_1$. A point $x_1$ with $f(x_1) > 0$ and $\delta > 0$ for every $\epsilon > 0$ will be called a Lebesgue point. The Lebesgue density theorem states that almost all points (with respect to $f$) are Lebesgue points. We assume that $x_1$ is a Lebesgue point.

Next, we introduce the nonincreasing funtion

$$\Psi(r) \triangleq (1-(1+\epsilon)\lambda(T)f(x_1)r^d)^{n-2} I_{r\leq\delta},$$

where $r > 0$. We note that the volume of $S(x,y)$ is $\lambda(T)$ times $\|x-y\|^d$. Thus, the lower bound for $E(N(x_1))$ is

$$(n-1)\int_{\|x_2-x_1\|\leq\delta} (1-(1+\epsilon)f(x_1)\lambda(T)\|x_1-x_2\|^d)^{n-2} f(x_2)\,dx_2$$

$$= (n-1)E(\Psi(\|X_2-x_1\|))$$

$$= (n-1)\int_0^1 P(\Psi(\|X_2-x_1\|) > t)\,dt$$

$$= (n-1)\int_0^1 \left(\int_{x_2:\Psi(\|x_2-x_1\|)>t} f(x_2)\,dx_2\right) dt$$

$$= (n-1)\int_0^1 \left(\int_{x_2:\|x_2-x_1\|\leq\Psi^{-1}(t)} f(x_2)\,dx_2\right) dt$$

$$\geq (n-1)\int_{t:0<t<1;\Psi^{-1}(t)\leq\delta} (1-\epsilon)V_d(\Psi^{-1}(t))^d f(x_1)\,dt$$

where we once again applied the Lebesgue density theorem; the $\delta > 0$ introduced here is the same as the one used in the definition of $\Psi$. We note that

$$(\Psi^{-1}(t))^d = \frac{1-t^{1/(n-2)}}{(1+\epsilon)\lambda(T)f(x_1)},$$

when $t \geq (1-(1+\epsilon)\lambda(T)f(x_1)\delta^d)^{n-2}$. Resubstitution of this in the last expression gives us yet another lower bound (because we integrate over fewer $t$'s):

$$(n-1)\int_{t:>t\geq\xi} (1-\epsilon)V_d f(x_1)\frac{1-t^{1/(n-2)}}{(1+\epsilon)\lambda(T)f(x_1)}\,dt$$

$$\text{(where } \xi \triangleq (1-(1+\epsilon)\lambda(T)f(x_1)\delta^d)^{n-2})$$

$$= (n-1)\frac{1-\epsilon}{1+\epsilon}\frac{V_d}{\lambda(T)}\int_{t:1>t\geqslant\xi}(1-t^{1/(n-2)})\,\mathrm{d}t$$

$$\geqslant (n-1)\frac{1-\epsilon}{1+\epsilon}\frac{V_d}{\lambda(T)}\left(\frac{1}{n-1}-\xi\right)$$

$$= (1+o(1))\frac{1-\epsilon}{1+\epsilon}\frac{V_d}{\lambda(T)},$$

because $\xi \downarrow 0$. Recalling that $\epsilon$ was arbitrary, this concludes the proof of the lower bound for $E(N)$, and for the lower bound for $E(N(x_1))$ at all Lebesgue points $x_1$.

What follows is simply an upper bound for $E(N(x_1))$ at Lebesgue points $x_1$.

We observe that

$$E(N(x_1)) \leqslant (n-1)\int \exp[-(n-2)\mu(S(x_1,x_2))]f(x_2)\,\mathrm{d}x_2$$

$$\leqslant (n-1)\int_{\|x_2-x_1\|\leqslant\delta} \exp[-(n-2)\mu(S(x_1,x_2))]f(x_2)\,\mathrm{d}x_2$$

$$+ (n-1)\int_{\|x_2-x_1\|>\delta} \exp[-(n-2)\mu(S(x_1,x_2))]f(x_2)\,\mathrm{d}x_2,$$

where $\delta > 0$ is as defined above. Thus,

$$E(N(x_1)) \leqslant (n-1)\int_{\|x_2-x_1\|\leqslant\delta} \exp[-(n-2)\lambda(T)f(x_1)(1-\epsilon)\|x_2-x_1\|^d]f(x_2)\,\mathrm{d}x_2$$

$$+ (n-1)\int_{\|x_2-x_1\|>\delta} \exp[-(n-2)\lambda(T)f(x_1)(1-\epsilon)\|x_2-x_1\|^d]f(x_2)\,\mathrm{d}x_2$$

$$\leqslant n\int_{\|x_2-x_1\|\leqslant\delta} \exp[-(n-2)\lambda(T)f(x_1)(1-\epsilon)\|x_2-x_1\|^d]f(x_2)\,\mathrm{d}x_2$$

$$+ n\exp[-(n-2)\lambda(T)f(x_1)(1-\epsilon)\delta^d].$$

The second term in the upper bound is $o(1)$. The first term is handled as in the lower bound. It can be written as

$$n\,E(\Psi(\|X_2-x_1\|))$$

$$\text{(where}\Psi(r) \triangleq I_{r\leqslant\delta}\exp[-(n-2)\lambda(T)f(x_1)(1-\epsilon)r^d])$$

$$= n\int_0^1 P(\Psi(\|X_2-x_1\|)>t)\,\mathrm{d}t$$

$$= n\int_0^1\int_{x_2:\Psi(\|x_2-x_1\|)>t} f(x_2)\,\mathrm{d}x_2\,\mathrm{d}t$$

$$= n \int_0^1 \int_{x_2 : \|x_2 - x_1\| \leqslant \Psi^{-1}(t)} f(x_2)\, dx_2\, dt$$

$$\leqslant n \int_{t : \Psi^{-1}(t) > \delta} dt + n \int_{t : 0 < t < 1;\, \Psi^{-1}(t) \leqslant \delta} \int_{x_2 : \|x_2 - x_1\| \leqslant \Psi^{-1}(t)} (1+\epsilon) f(x_1)\, dx_2\, dt$$

(Lebesgue density theorem)

$$= n\,\Psi(\delta) + n \int_{t : \Psi(\delta) \leqslant t < 1} (1+\epsilon) f(x_1)\, V_d [\Psi^{-1}(t)]^d\, dt$$

$$= o(1) + n \int_{\Psi(\delta)}^1 \log\left(\frac{1}{t}\right) dt \frac{1+\epsilon}{1-\epsilon} \frac{V_d}{(n-2)\lambda(T)}$$

$$= o(1) + n\,[1 - \Psi(\delta) + \Psi(\delta)\log(\Psi(\delta))] \frac{1+\epsilon}{1-\epsilon} \frac{V_d}{(n-2)\lambda(T)}$$

$$= o(1) + (1 + o(1)) \frac{1+\epsilon}{1-\epsilon} \frac{V_d}{\lambda(T)},$$

since $n\,\Psi(\delta) \to 0$. This concludes the proof of Theorem 1. ■

## 3. DISCUSSION OF THEOREM 1

It is noteworthy that the lower bound depends upon the volume induced by $T$ only. It is applicable to all densities $f$.

Perhaps equally interesting is the fact that for almost all $x$, conditional on $X_1 = x$, the expected number of neighbors of $X_1$ is $V_d/\lambda(T) + o(1)$. This is due to the fact that locally, every density, no matter how pathological, is "almost" uniform in a small neighborhood of almost all points. Hence, since the decision to include an edge or not is virtually always based upon the points in a small neighborhood of the candidate vertices, the expected degree of each vertex is roughly as for the uniform density.

One should not conclude from Theorem 1 that all the properties of $N$ or $N(X_1)$ are distribution-free. Indeed, the rate of convergence of the various quantities in Theorem 1 to their asymptotic values depends very much on $f$. Thus, it is certainly *not* possible to argue as follows:

$$\limsup_{n \to \infty} \frac{E(N)}{n} \leqslant \int \frac{1}{2} \limsup_{n \to \infty} E((N(X_1) | X_1 = x) f(x)\, dx = \frac{V_d}{2\lambda(T)},$$

by Theorem 1. One can bring the limit supremum under the integral only under certain conditions. One such condition is that the integrand, a function of $n$ and $x$ here, can be uniformly bounded from above in $n$ by an integrable function. The lower bound of Theorem 1 can be attained however for some distributions, as we will see below.

It is known that planar graphs cannot have more than $3n - 6$ edges. Thus, results like Theorem 1 can be used to prove possible nonplanarity of certain graphs; indeed, if $E(N) \geqslant (\alpha + o(1))n$ for some $\alpha > 3$ and all densities, it is easy to see that the proximity graph defined by that particular $S$ must have some configuration for which it cannot possibly be planar.

## 4. THE GABRIEL GRAPH

For the Gabriel graph, $\lambda(T) = V_d/2^d$. Thus, we conclude that

$$\liminf_{n \to \infty} \frac{E(N)}{n} \geqslant 2^{d-1},$$

for all $f$. In addition, at almost all $x$, $E(N(X_1) | X_1 = x)$ tends to $2^d$. This generalizes some results of Matula and Sokal [6]. They have shown that the Gabriel graph in the plane is planar and has

at most $3n - 8$ edges. They also showed that the expected number of edges is asymptotic to $2n$ for the uniform distribution on the unit square. We will see in Theorem 3 that the latter result remains valid for nearly all densities of interest to users.

## 5. THE RELATIVE NEIGHBORHOOD GRAPH

Since the loon defining the RNG contains the set $S$ defining the Gabriel graph, it is clear that the RNG is contained in the Gabriel graph. To obtain exact information on how $E(N)$ varies with $n$, we need to compute the area of the unit loon generated by $T$ with some care. We have, for $d = 2$,

$$\frac{\lambda(T)}{V_2} = \frac{2}{\pi}\left(\frac{\pi}{3} - \frac{3^{1/2}}{4}\right)$$

$$= \frac{2}{3} - \frac{3^{1/2}}{2\pi} = 0.3910022190\ldots.$$

Therefore, the expected degree at almost all $x$ is $2.557530243\cdots + o(1)$. Also, the expected number of edges is at least $n(1.2787651215\cdots + o(1))$ for any density.

For $d > 2$, we have

$$\lambda(T) = \int_0^{3^{1/2}/2} 2((1-r^2)^{1/2} - \tfrac{1}{2})\, \mathrm{d}(V_d r^d),$$

which yields the formula

$$\frac{V_d}{\lambda(T)} = \frac{1}{\displaystyle\int_0^{3^{1/2}/2} 2((1-r^2)^{1/2} - \tfrac{1}{2})\, dr^{d-1}\, \mathrm{d}r}$$

for the asymptotic degree at almost all $x$.

## 6. THE NEAREST NEIGHBOR GRAPH

Let $S(x, y)$ be the union of the spheres of radius $\| x - y \|$ centered at $x$ and $y$, respectively. If $S(x, y)$ contains no data points, then $x$ and $y$ are each other's nearest neighbors, hence they correspond to a double edge in the nearest neighbor graph. In $2d$, the area $\lambda(T)$ is equal to $2V_2$ minus the area of the loon of the RNG, i.e.

$$\frac{\lambda(T)}{V_2} = \frac{4}{3} + \frac{3^{1/2}}{2\pi} = 1.6089977809\ldots.$$

Thus, if $E(N)$ is the expected number of double edges,

$$\liminf_{n\to\infty} \frac{E(N)}{n} \geqslant \frac{1}{\dfrac{8}{3} + \dfrac{3^{1/2}}{\pi}}.$$

Since the number of edges in the nearest neighbor graph is $n$ minus the number of double edges, we have, for all densities $f$,

$$\limsup_{n\to\infty} \frac{E\,(\text{number of edges in nearest neighbor graph})}{n} \leqslant \frac{\dfrac{5}{3} + \dfrac{3^{1/2}}{\pi}}{\dfrac{8}{3} + \dfrac{3^{1/2}}{\pi}}$$

$$= \frac{5\pi + 27^{1/2}}{8\pi + 27^{1/2}}$$

$$= 0.6892475516\ldots.$$

We will see below (Theorem 2) that for *all* densities, the expected number of edges in the nearest neighbor graph is $(c + o(1))n$, where $c = 0.6892475516\ldots$ is the constant defined above.

For uniform distributions in the plane (more precisely, homogeneous Poisson processes in the plane), many statistical properties of nearest neighbor graphs are well-known; for a tour of these results, one can consult Getis and Boots [15], where references to applications in geography and biology are given. For example, it is known that the probability that $X_1$ will form a reciprocal nearest neighbor pair with its nearest neighbor (i.e. $X_1$ is the nearest neighbor of its nearest neighbor) is asymptotic to

$$c \triangleq \frac{6\pi}{8\pi + 27^{1/2}} = 0.6215\ldots,$$

see Refs [16–18]. This implies that the expected number of edges in the nearest neighbor graph is asymptotic to

$$\left(\frac{c}{2} + (1-c)\right)n,$$

where the first contribution comes from the double edges, and the second term from the single edges. We verify easily that

$$\frac{c}{2} + (1-c) = \frac{5\pi + 27^{1/2}}{8\pi + 27^{1/2}}.$$

This corresponds to what we found to be true for all densities.

## 7. DIRECTIONAL NEAREST NEIGHBOR GRAPHS

Flinchbaugh and Jones [19] studied the directional nearest neighbor graph in $R^2$, obtained by connecting each point with its nearest neighbor in one of $r$ fixed divisions. The divisions are obtained by positioning the origin at a point, and partitioning the space into infinite slices of a pie, each with angle $2\pi/r$. The number of edges grows at most linearly in $n$, so that Theorem 3 below applies, but unfortunately, a crucial symmetry condition on $S$ used by us is violated.

Nevertheless, we can get some idea of the expected number of edges in similar graphs obtained as follows: join $X_i$ and $X_j$ if in the cone of angle $\alpha$ centered at $X_i$ with $X_j$ on its bisector, $X_j$ is the nearest point to $X_i$.

Here too, we proceed first by computing the expected number of double edges. We note in passing that for $\alpha = 2\pi/3$, every double edge corresponds to a RNG edge. For $\alpha \leqslant 2\pi/3$, the double edges define a supergraph of the RNG. A simple computation shows that in Theorems 1 and 3, which are both applicable here,

$$\lambda(T) = 2\frac{\alpha}{2} - \frac{1}{2}\tan\left(\frac{\alpha}{2}\right).$$

Hence, for all densities of Theorem 3, and $0 < \alpha \leqslant 2\pi/3$,

$$E(N)/n \to \frac{\pi}{2\alpha + \tan\left(\frac{\alpha}{2}\right)}.$$

For the RNG ($\alpha = 2\pi/3$), we obtain the limit value $1/(4/3 - 3^{1/2}/\pi) = 1.2787651215\ldots$. Let us now consider a graph in which we associate with each edge one or two directions; an edge with two directions is said to be a double edge; and $X_i$ points to $X_j$ whenever $X_j$ is the nearest neighbor of $X_i$. Above, we have already counted the expected number of double edges. The total expected number of directions attached to edges is easily seen to be asymptotic to $n \times V_2/\lambda(T)$, where $T$ now stands for the cone of angle $\alpha$, intersected with the unit circle. Thus, the expected number of directions is asymptotic to $2n\pi/\alpha$. The expected number of edges is obtained by subtracting from this the expected number of double edges. This yields the result (valid for $0 < \alpha \leqslant 2\pi/3$)

$$\frac{E(N)}{n} \to \frac{2\pi}{\alpha}\left(1 - \frac{1}{4 + \frac{2}{\alpha}\tan\left(\frac{\alpha}{2}\right)}\right),$$

which is valid for all densities of Theorem 3. It is interesting to observe that this limit varies as $8\pi/(5\alpha)$ and $\alpha \downarrow 0$, so that by controlling $\alpha$, we have in fact full control on the sparseness of the graph. For small $\alpha$, these graphs cannot possibly be planar.

## 8. ASYMPTOTICS FOR GRAPHS WITH BOUNDED MAXIMAL DEGREE

We have hinted at the fact that the lower bound on $E(N)$ given in Theorem 1 can be attained for some densities.

In some proximity graphs, the maximal degree of each vertex is bounded by a number depending upon $d$ only; for example, for any value of $d$, is is known that the nearest neighbor graph has maximal vertex degree bounded by a constant $C$ depending upon $d$ only. For all such graphs, we have a very general property, valid for all densities (Theorem 2): the lower bound of Theorem 1 is attained for all densities.

In the next section, we will consider proximity graphs that are *worst-case sparse*, i.e. for any $x_1, \ldots, x_n$, the number of edges does not exceed $Cn$ for some constant $C$. It suffices to note that this condition is satisfied for all planar *graphs*: hence, for $d = 2$, it holds for most of the graphs discussed above, including the RNG, the graph formed by double edges in the nearest neighbor graph, and the Gabriel graph. For worst-case sparse graphs, possibly having unbounded maximal vertex degree, the lower bound of Theorem 1 is attained under some (mild) conditions on the underlying distribution (Theorem 3).

*Theorem 2*

Consider a proximity graph based on a regular set $S$, and assume that the maximal vertex degree is bounded by a constant $C$ depending upon $d$ only. Then, for all densities,

$$\lim_{n\to\infty} \frac{E(N)}{n} = \frac{V_d}{2\lambda(T)}.$$

*Proof of Theorem 2*

For every $\epsilon, \delta > 0$, we can partition $R^d$ into a set $G_{\epsilon,\delta}$, and its complement, $B_{\epsilon,\delta}$. The "$G$" stands for "good" and the "$B$" stands for "bad". $G_{\epsilon,\delta}$ is the collection of all $x$ for which

$$\left|\frac{\mu(S(x,y))}{\lambda(S(x,y))} - f(x)\right| \leqslant \epsilon f(x)$$

and

$$\left|\frac{\mu(B(x,y))}{\lambda(B(x,y))} - f(x)\right| \leqslant \epsilon f(x),$$

for all $y$ with $\|x-y\| < \delta$, where $B(x,y)$ is the ball centered at $x$ with radius $\|x-y\|$, and $\|x\| \leqslant 1/\delta$. By the Lebesgue density theorem, it is possible to find $\delta < 0$ depending upon $\epsilon$, such that $\mu(B_{\epsilon,\delta}) < \epsilon$. We pick $\delta$ in this manner, and write $G_\epsilon$ and $B_\epsilon$ from here onwards.

The data points are partitioned into two sets, according to membership in $G_\epsilon$ or its complement. The number of edges $N$ can be written as $N_G + N_B$, where $N_G$ refers to the edges in which both vertices are in $G_\epsilon$. $N_B$ refers to the other edges. The expected number of $N_B$ is bounded by $CE$(cardinality of $B_\epsilon$) (because every vertex has degree $C$ in the worst case). This is $Cn\mu(B_\epsilon) \leqslant Cn\epsilon$. When divided by $n$, this is as small as desired by our choice of $\epsilon$.

What follows is simply an upper bound for $E(N_G)$.

We observe that

$$E(N_G)/(n/2) \leqslant (n-1) \iint_{x_1,x_2 \in G_\epsilon} \exp[-(n-2)\mu(S(x_1,x_2))] f(x_1) f(x_2)\, dx_2\, dx_1$$

$$\leqslant (n-1) \iint_{x_1,x_2 \in G_\epsilon;\, \|x_2-x_1\| \leqslant \delta} \exp[-(n-2)\mu(S(x_1,x_2))] f(x_1) f(x_2)\, dx_2\, dx_1$$

$$+(n-1)\iint\limits_{x_1,x_2\in G_\epsilon;\|x_2-x_1\|>\delta} \exp[-(n-2)\mu(S(x_1,x_2))]f(x_1)f(x_2)\,dx_2\,dx_1.$$

$$\leqslant (n-1)\iint\limits_{x_1,x_2\in G_\epsilon,\|x_2-x_1\|\leqslant\delta} \exp[-(n-2)\lambda(T)f(x_1)(1-\epsilon)\|x_2-x_1\|^d]f(x_1)f(x_2)\,dx_2\,dx_1$$

$$+(n-1)\iint\limits_{x_1\in G_\epsilon,\|x_2-x_1\|>\delta} \exp[-(n-2)\lambda(T)f(x_1)(1-\epsilon)\delta^d]f(x_1)f(x_2)\,dx_2\,dx_1$$

$$\leqslant n\iint\limits_{x_1,x_2\in G_\epsilon,\|x_2-x_1\|\leqslant\delta} \exp[-(n-2)\lambda(T)f(x_1)(1-\epsilon)\|x_2-x_1\|^d]f(x_1)f(x_2)\,dx_2\,dx_1$$

$$+n\int_{x_1\in G_\epsilon} \exp[-(n-2)\lambda(T)f(x_1)(1-\epsilon)\delta^d]f(x_1)\,dx_1.$$

The second term in the upper bound is $o(1)$. This can be seen as follows: using the fact that $ne^{-nu}\leqslant 1/(eu)$ for all $n\geqslant 1$ and all $u>0$, it is easy to see that the integrand is uniformly bounded in $n$ by an integrable function (here we also need the fact that $G_\epsilon$ vanishes outside a big square). Furthermore, the integrand tends to zero with $n$ for almost all $x_1$, so that we can apply the Lebesgue dominated convergence theorem. The first term is handled as in the proof of the lower bound (see Theorem 1): it can be written as

$$\int_{G_\epsilon}\left(n\int_{x_2\in G_\epsilon,\|x_2-x_1\|\leqslant\delta}\exp[-(n-2)\lambda(T)f(x_1)(1-\epsilon)\|x_2-x_1\|^d]f(x_2)\,dx_2\right)f(x_1)\,dx_1$$

$$\triangleq\int_{G_\epsilon}[E(\Psi(\|X_2-x_1\|)I_A(X_2))]f(x_1)\,dx_1,$$

where

$$A\triangleq\{x_2\cdot x_2\in G_\epsilon,\|x_2-x_1\|\leqslant\delta\}.$$

However, for fixed $x_1\in G_\epsilon$,

$$E(\Psi(\|X_2-x_1\|)I_A(X_2))=n\int_0^1 P(X_2\in A,\Psi(\|X_2-x_1\|)>t)\,dt$$

$$=n\int_0^1\int_{x_2:\Psi(x_2,x_1)>t,x_2\in A} f(x_2)\,dx_2\,dt$$

$$=n\int_0^1\int_{x_2:\|x_2-x_1\|\leqslant\Psi^{-1}(t),x_2\in A} f(x_2)\,dx_2\,dt$$

$$\leqslant n\int_{t:\Psi^{-1}(t)>\delta} dt+n\int_{t:0<t<1;\Psi^{-1}(t)\leqslant\delta}\int_{x_2:\|x_2-x_1\|\leqslant\Psi^{-1}(t)}(1+\epsilon)f(x_1)\,dx_2\,dt$$

$$(x_1\in G_\epsilon)$$

$$=n\Psi(\delta)+n\int_{t:\Psi(\delta)\leqslant t<1}(1+\epsilon)f(x_1)V_d(\Psi^{-1}(t))^d\,dt$$

$$=n\Psi(\delta)+\int_{\Psi(\delta)}^1\log\left(\frac{1}{t}\right)dt\,\frac{1+\epsilon}{1-\epsilon}\frac{V_d}{(n-2)\lambda(T)}$$

$$\leqslant n\,\Psi(\delta) + n\frac{1+\epsilon}{1-\epsilon}\frac{V_d}{(n-2)\lambda(T)}.$$

We have seen above that

$$\int_{G_\epsilon} n\,\Psi(\delta) f(x_1)\,dx_1 \to 0$$

as $n \to \infty$; hence the first term in the upper bound has an $o(1)$ contribution to $E(N)/n$ [note that $\Psi$ depends upon $x_1$, so it was not enough to just verify that $n\Psi(\delta) \to 0$ for all $x_1$]. The second term in the upper bound does not depend upon $x_1$, so that we can conclude, by the arbitrary nature of $\epsilon$, that lim sup $E(N)/n \leqslant V_d/(2\,\lambda(T))$. This, combined with the lower bound of Theorem 1, concludes the proof of Theorem 2. ■

## 9. ASYMPTOTICS FOR SPARSE GRAPHS

We now introduce a *regularity condition* for a density $f$. For every $\epsilon,\delta > 0$, let $B_{\epsilon,\delta}$ be the collection of all $x \in R^d$ for which

$$\inf_{y:\|y-x\|\leqslant\delta} \quad \inf_{z:z\in S(x,y) \text{ or } \|z-x\|\leqslant\delta} f(z) < (1-\epsilon) f(x)$$

or

$$\sup_{y:\|y-x\|\leqslant\delta} \quad \sup_{z:z\in S(x,y) \text{ or } \|z-x\|\leqslant\delta} f(z) > (1+\epsilon) f(x).$$

We demand that the boundary of $B_{\epsilon,\delta}$ have zero Lebesgue measure, where the boundary of a set is defined as the closure of a set minus the set itself. The boundary of a closed set is obviously empty. Examples of sufficient conditions follow.

*Example 8*

*Uniform density on a convex set.* When $f$ is the uniform density on a convex set $C$, and $\epsilon$ is small enough $B_{\epsilon,\delta}$ consists of all the points in $C$ that are within distance $\delta$ of the complement of $C$, and all the points of the complement of $C$ that are within distance $\delta$ of $C$. This in turn is the difference of two nested convex sets. Its boundary has zero Lebesgue measure. ■

*Example 9*

*The multivariate normal density.* For the multivariate normal density, the argument is rather simple. In fact, the regularity condition is satisfied for most unimodal radially symmetric densities. ■

*Theorem 3*

Consider a proximity graph based on a regular set $S$, and assume that it is worst-case sparse. Then, for all densities satisfying the regularity condition given above,

$$\lim_{n\to\infty} \frac{E(N)}{n} = \frac{V_d}{2\,\lambda(T)}.$$

*Proof of Theorem 3*

Let $\epsilon$, $\delta$, $G_\epsilon$ and $B_\epsilon$ be as in the proof of Theorem 2. The data points are partitioned into two sets, according to membership in $G_\epsilon$ or its complement. The number of edges $N$ can be partitioned into three collections, $N_{GG}$ (connecting vertices entirely within $G_\epsilon$), $N_{BB}$ (connecting vertices entirely in $B_\epsilon$), and $N_{BG}$ (connecting vertices from $G_\epsilon$ with vertices in $B_\epsilon$).

If we consider the proximity graph formed on the basis of the points in $B_\epsilon$ alone, then the expected number of edges [and thus $E(N_{BB})$] is bounded by $CE$(cardinality of $B_\epsilon$) because the proximity graph is worst-case sparse. This is $Cn\mu(B_\epsilon) \leqslant Cn\epsilon$. When divided by $n$, this is as small as desired by our choice of $\epsilon$.

$E(N_{GG})$ is handled exactly as in the proof of Theorem 2 [it is called $E(N_G)$ there]. Using the notation of the proof of Theorem 2, we can conclude that

$$\limsup \frac{E(N_{GG})}{n} \leqslant \frac{(1+\epsilon)V_d}{2(1-\epsilon)\lambda(T)}.$$

Thus, we need only be concerned here with the smallness of $E(N_{BG})$. This too can be handled in the manner we handled the upper bound for $E(N_G)$ in the proof of Theorem 2, provided that we replace everywhere the event $X_2 \in G_\epsilon$ or the statement $x_2 \in G_\epsilon$ by the corresponding event and statement involving $B_\epsilon$. Furthermore, when breaking up the integral with respect to $dt$ into two pieces, we consider a breakpoint defined by the condition $\{t : \Psi^{-1}(t) > \rho\}$ for some $\rho \in (0, \delta)$. It can be verified that we have

$$\frac{E(N_{BG})}{n/2} \leqslant o(1) + \frac{(1+\epsilon)V_d}{2(1-\epsilon)\lambda(T)} \int_{x_1 \in G_\epsilon, \|x_2 - x_1\| \leqslant \rho \text{ for some } x_2 \in B_\epsilon} f(x_1)\, dx_1.$$

We are done if we can prove that the integral in the last upper bound can be made as small as desired by our choice of $\rho$, for every fixed $\epsilon$ and $\delta$. This is precisely where the regularity condition comes into play. Indeed, the function $f$ is integrable, and as $\rho \downarrow 0$, the set over which we integrate shrinks down to a set of zero Lebesgue measure (because the closure of $B_\epsilon$ intersected with $G_\epsilon$ has zero Lebesgue measure). Therefore, by the Lebesgue dominated convergence theorem, the integral can be made as small as desired by the choice of $\rho$. This, together with Theorem 1, concludes the proof of Theorem 3. ■

*Acknowledgement*—Research by the author was sponsored by NSERC Grant A3456 and by FCAC Grant EQ-1678.

## REFERENCES

1. H. Edelsbrunner and M. H. Overmars, On the equivalence of some rectangle problems. *Inform. Proc. Lett.* **14,** 124–127 (1982).
2. D. T. Lee and F. P. Preparata, An improved algorithm for the rectangle enclosure problem. *J. Algorithms* **3,** 218–224 (1982).
3. R. H. Gutting, O. Nurmi and T. Ottmann, The direct dominance problem. *Proc. Symp. on Computational Geometry*, pp. 81–88, ACM, New York (1985).
4. R. Klein, Direct dominance of points, *Int. J. Comput. Math.* **19,** 225–244 (1986).
5. K. R. Gabriel and R. R. Sokal, A new statistical approach to geographic variation analysis. *System. Zoology* **18,** 259–278 (1969).
6. D. W. Matula and R. R. Sokal, Properties of Gabriel graphs relevant to geographic variation research and the clustering of points in the plane. *Geograph. Anal.* **12,** 205–222 (1980).
7. G. T. Toussaint, Pattern recognition and geometrical complexity. *Fifth Int. Conf. on Pattern Recognition*, pp, 1324–1347 (1980).
8. G. T. Toussaint, The relative neighborhood graph of a finite planar set. *Pattern Recogn* **12,** 261–268 (1980).
9. K. J. Supowit, The relative neighborhood graph with an application to minimum spanning trees. *J. ACM*, **30,** 428–447 (1983).
10. F. P. Preparata and M. I. Shamos, *Computational Geometry. An Introduction.* Springer, New York (1985).
11. G. T. Toussaint, Computational geometric problems in pattern recognition. In *Pattern Recognition Theory and Applications* (Eds J. Kittler, K. S. Fu and L. F. Pau) pp. 73–91. Reidel, Holland (1982).
12. D. Avis and J. Horton, Remarks on the sphere of influence graph. In *Discrete Geometry and Convexity*, Vol. 440, pp. 323–327. New York Academy of Sciences, (1982).
13. J. L. Bentley and T. Ottmann, Algorithms for reporting and counting geometric intersections. *IEEE Trans. Comput.* **C-28,** 643–647 (1979).
14. R. L. Wheeden and A. Zygmund, *Measure and Integral.* Dekker, New York, (1977).
15. A. Getis and B. Boots, *Model of Spatial Processes.* Cambridge University Press, Cambridge (1978).
16. P. J. Clark and F. C. Evans, On some aspects of spatial pattern in biological populations. *Science* **121,** 397–398 (1955).
17. P. J. Clark, Grouping in spatial distributions. *Ecology* **35,** 445–453 (1956).
18. M. F. Dacey, Proportion of reflexive $n$th order neighbors in spatial distribution. *Geograph. Anal.* **1,** 385–388 (1969).