Submitted to the Annals of Applied Probability

HAMILTON CYCLES IN RANDOM GEOMETRIC GRAPHS

By József Balogh[*], Béla Bollobás[†], and Mark Walters

University of Illinois, University of Cambridge and Univeristy of Memphis, and Queen Mary University of London

We prove that, in the Gilbert model for a random geometric graph, almost every graph becomes Hamiltonian exactly when it first becomes 2-connected. This proves a conjecture of Penrose.

We also show that in the k-nearest neighbour model, there is a constant κ such that almost every κ-connected graph has a Hamilton cycle.

1. Introduction. In this paper we mainly consider one of the frequently studied models for random geometric graphs, namely the Gilbert Model. Suppose that S_n is a $\sqrt{n} \times \sqrt{n}$ box and that $\mathcal{P}$ is a Poisson process in it with density 1. The points of the process form the vertex set of our graph. There is a parameter r governing the edges: two points are joined if their (Euclidean) distance is at most r.

Having formed this graph we can ask whether it has any of the standard graph properties, such as connectedness. As usual, we shall only consider these for large values of n. More formally, we say that $G = G_{n,r}$ has a property *with high probability* (abbreviated to whp) if the probability that G has this property tends to one as n tends to infinity.

Penrose [10] proved that the threshold for connectivity is $\pi r^2 = \log n$. In fact he proved the following very sharp result: suppose $\pi r^2 = \log n + \alpha$ for some constant α. Then the probability that $G_{n,r}$ is connected tends to $e^{-e^{-\alpha}}$.

He also generalised this result to find the threshold for κ-connectivity: namely $\pi r^2 = \log n + (\kappa - 1) \log \log n$. Moreover, he found the "obstruction" to κ-connectivity. Suppose we fix the vertex set (i.e., the point set in S_n) and "grow" r. This gradually adds edges to the graph. For a monotone graph

---

[*] This material is based upon work supported by NSF CAREER Grant DMS-0745185 and DMS-0600303, UIUC Campus Research Board Grants 09072 and 08086, and OTKA Grant K76099.

[†] Research supported in part by NSF grants DMS-0505550, CNS-0721983 and CCF-0728928, and ARO grant W911NF-06-1-0076.

AMS 2000 subject classifications: 05C80, 60D05, 05C45

Keywords and phrases: Hamilton cycles, random geometric graphs

property P let $\mathcal{H}(P)$ denote the smallest r for which the graph on this point set has the property P. Penrose showed that

$$\mathcal{H}(\delta(G) \geq \kappa) = \mathcal{H}(\text{connectivity}(G) \geq \kappa)$$

whp: that is, as soon as the graph has minimum degree κ it is κ-connected whp.

He also considered the threshold for G to have a Hamilton cycle. Obviously a necessary condition is that the graph is 2-connected. In the normal (Erdős-Rényi) random graph this is also a sufficient condition in the following strong sense. If we add edges to the graph one at a time then the graph becomes Hamiltonian exactly when it becomes 2-connected (see [14],[9],[8] and [5]).

Penrose, conjectured that the same is true for a random geometric graph. In this paper we prove the following theorem proving the conjecture.

Theorem 1. *Suppose that* $G = G_{n,r}$ *the two-dimensional Gilbert Model. Then*

$$\mathcal{H}(G \textit{ is 2-connected}) = \mathcal{H}(G \textit{ has a Hamilton cycle})$$

whp.

Combining this with the earlier result of Penrose we see that, if $\pi r^2 = \log n + \log\log n + \alpha$ then the probability that G has a Hamilton cycle tends to $e^{-e^{-\alpha}}$.

Some partial progress has been made on this conjecture previously. Petit [13] showed that if $\pi r^2/\log n$ tends to infinity then G is, whp, Hamiltonian, and Díaz, Mitsche and Pérez [7] proved that if $\pi r^2 > (1+\varepsilon)\log n$ for some $\varepsilon > 0$ then G is Hamiltonian whp. (Obviously, G is not Hamiltonian if $\pi r^2 < \log n$ since whp G is not connected!) Finally in [4] Balogh, Kaul and Martin proved Penrose's conjecture for the special case of the ℓ_∞ norm in two dimensions.

Our proof generalises to higher dimensions, and to other norms. The Gilbert Model makes sense with any norm and in number of dimensions: we let S_n^d be the d-dimensional hypercube with volume n. We prove the analogue of Theorem 1 in this setting.

Theorem 2. *Suppose that the dimension* d *and* $\|\cdot\|$*, a* p*-norm for some* $1 \leq p \leq \infty$*, are fixed. Let* $G = G_{n,r}$ *be the resulting Gilbert Model. Then*

$$\mathcal{H}(G \textit{ is 2-connected}) = \mathcal{H}(G \textit{ has a Hamilton cycle})$$

whp.

The proof is very similar to that of Theorem 1. However, there are some significant extra technicalities.

To give an idea why these occur consider connectivity in the Gilbert Model in the cube S_n^3 (with the Euclidean norm). Let A be the volume of a sphere of radius r. We count the expected number of isolated points in the process which are away from the boundary of the cube. The probability a point is isolated is e^{-A} so the expected number of such points is ne^{-A}, so the threshold for the existence of a central isolated point is about $A = \log n$.

However, consider the probability that a point near a face of the cube is isolated: there are approximately $n^{2/3}$ such points and the probability that they are isolated is about $e^{-A/2}$ (since about half of the sphere about the point is outside the cube S_n^3). Hence, the expected number of such points is $n^{2/3}e^{-A/2}$, so the threshold for the existence of an isolated point near a face is about $A = \frac{4}{3}\log n$. In other words isolated points are much more likely near the boundary. These boundary effects are the reason for many of the extra technicalities.

The k-nearest neighbour model. We also consider a second model for random geometric graphs: namely the k-nearest neighbour graph. In this model the initial setup is the same as in the Gilbert model: the vertices are given by a Poisson process of density one in the square S_n, but this time each vertex is joined to its k nearest neighbours (in the Euclidean metric) in the box. This naturally gives rise to a k-regular directed graph, but we form a simple graph $G = G_{n,k}$ by ignoring the direction of all the edges. It is easily checked that this gives us a graph with degree between k and $6k$.

Xue and Kumar [15] showed that there are constants c_1, c_2 such that if $k < c_1 \log n$ then the graph $G_{n,k}$ is, whp, not connected, and that if $k > c_2 \log n$ then $G_{n,k}$ is, whp, connected. Balister, Bollobás, Sarkar and Walters [1] proved reasonably good bounds on the constants: namely $c_1 = 0.3043$ and $c_2 = 0.5139$, and later [3] proved that there is some critical constant c such that if $k = c' \log n$ for $c' < c$ then the graph is disconnected whp, and if $k = c' \log n$ for $c' > c$ then it is connected whp. Moreover, in [2], they showed that in the latter case the graph is s-connected whp for any fixed $s \in \mathbb{N}$.

We would like to prove a sharp result like the above: i.e., that as soon as the graph is 2-connected it has a Hamilton cycle. However, we prove only the weaker statement that some (finite) amount of connectivity is sufficient. Explicitly, we show the following.

THEOREM 3. *Suppose that $k = k(n)$, that $G = G_{n,k}$ is the k-nearest neighbour graph, and that G is κ-connected for $\kappa = 5 \cdot 10^7$ whp. Then G has* *a Hamilton cycle whp.*

2. Proof of Penrose's Conjecture. We divide the proof into five parts: first we tile the square S_n with small squares in a standard tessellation argument. Secondly we identify "difficult" subsquares. Roughly, these will be squares containing only a few points, or squares surrounded by squares containing only a few points. Thirdly we prove some lemmas about the structure of the difficult subsquares. In stage 4 we deal with the difficult subsquares. Finally we use the remaining easy subsquares to join everything together.

Stage 1: Tessellation.. Let $r_0 = \sqrt{(\log n)/\pi}$ (so $\pi r_0^2 = \log n$), and let r be the random variable $\mathcal{H}(G$ is 2-connected). Let $s = r_0/c = c'\sqrt{\log n}$ where c is a large constant to be chosen later (1000 will do). We tessellate the box S_n with small squares of side length s. Whenever we talk about distances between squares we will always be referring to the distance between their centres. Moreover, we will divide all distances between squares by s, so, for example, a square's four nearest neighbours all have distance one.

By the result of Penrose [11] mentioned in the introduction we may assume that $(1 - 1/2c)r_0 < r < (1 + 1/2c)r_0$: formally the collection of point sets which do not satisfy this has measure tending to zero as n tends to infinity, and we ignore this set.

Hence points in squares at distance $\frac{r-\sqrt{2}s}{s} \geq \frac{r_0-2s}{s} = (c-2)$ are always joined and points in squares at distance $\frac{r+\sqrt{2}s}{s} \leq \frac{r_0+2s}{s} = (c+2)$ are never joined.

Stage 2: The "difficult" subsquares.. We call a square *full* if it contains at least M points for some M to be determined later (10^7 will do), and *non-full* otherwise. Let N_0 be the set of non-full squares. We say two non-full squares are joined if their ℓ_∞ distance is at most $4c - 1$ and define $\mathcal{N}$ to be the collection of non-full components.

First we bound the size of the largest component of non-full squares.

Lemma 4. *For any M, the largest component of non-full squares in the above tesselation has size at most*

$$U = \lceil \pi(c+2)^2 \rceil$$

whp.

Also, the largest component of non-full squares including a square within c of the boundary of S_n has size at most $U/2$ whp. Finally, there is no non-full square within distance Uc of a corner whp.

Proof. We shall make use of the following simple result: suppose that G is any graph with maximal degree Δ and v is a vertex in G. Then the number of connected subsets of size n of G containing v is at most $e\Delta^n$ (see e.g., Problem 45 of [6]).

Hence, the number of potential components of size U containing a particular square is at most $(e(8c)^2)^U$ so, since there are less than n squares, the total number of such potential components is at most $n(e(8c)^2)^U$. The probability that a square is non-full is at most $2s^{2M}e^{-s^2}/M!$. Hence, the expected number of components of size at least U is at most

$$n(2s^{2M}e^{-s^2}(e(8c)^2)/M!)^U \leq n\left(2\,(\log n)^M\,\frac{e(8c)^2}{M!}\right)^U \exp\left(-\frac{(c+2)^2\log n}{c^2}\right)$$

which tends to zero as n tends to infinity: i.e., whp, no such component exists.

For the second part there are at most $4c\sqrt{n}$ squares within distance c of the boundary of S_n and the result follows as above.

Finally, there are only $4U^2c^2$ squares within distance c of a corner. Since the probability that a square is non-full tends to zero we see that there is no such square whp. □

Note that this is true independently of M which is important since we will want to choose M depending on U.

In the rest of the argument we shall assume that there is no non-full component of size greater than U, no non-full component of size $U/2$ within c of an edge and no non-full square within Uc of a corner.

Between these components of non-full squares there are numerous full squares. To define this more precisely let $\widehat{G}$ be the graph with vertex set the small squares, and where each square is joined to all others within $(c-2)$ of this square (in the Euclidean norm). The graph $\widehat{G} \setminus N_0$ has one very large component $\widetilde{A}$ which we call the *sea*. We call the squares in this sea the *easy* squares and the remaining squares the *difficult* squares.

The idea is that it is trivial to find a cycle visiting every point of the process in a square in the sea $\widetilde{A}$, and that we can extend this cycle to a Hamilton cycle by adding each non-full component (and any full squares cut off by it) one at a time. However, it is easier to phrase the argument by starting with the difficult parts and then using the sea of full squares.

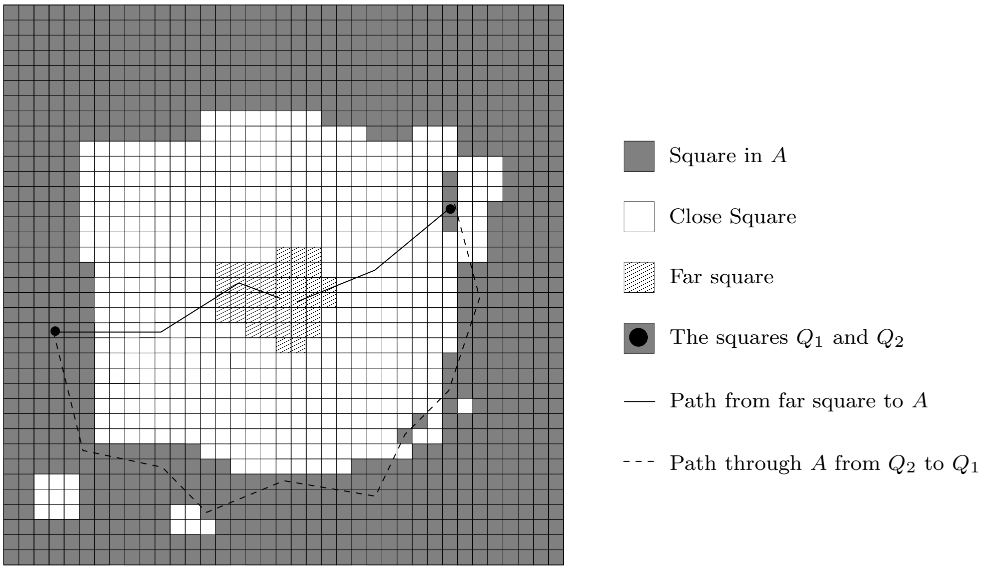

Fig 1. *A small part of S_n containing the non-full component N and the corresponding set A, far squares and close squares. It also shows the two vertex disjoint paths from the far squares to A and the path joining Q_2 to Q_1 (see stage 4).*

Stage 3: The structure of the difficult subsquares.

Consider one component $N \in \mathcal{N}$ of the non-full squares, and suppose that it has size u. By Lemma 4 we know $u < U$. We will also consider N_{2c}: the $2c$-blow-up of N: that is the set of all squares with ℓ_∞ distance at most $2c$ from a square in N.

Now some full squares may be cut off from the rest of the full squares by non-full squares in N. More precisely the graph $\widehat{G} \setminus N$ has one giant component $A = A(N)$; we call A^c the *cutoff* squares. Note that A is a superset of the sea $\widetilde{A}$, since we have only removed one component of the non-full squares. We shall see later (Corollary 10) that the sea is $\bigcap_{N\in\mathcal{N}} A(N)$.

We split the cutoff squares into two classes: those with a neighbour in A (in $\widehat{G}$) which we think of as being "close" to A, and the rest, which we shall call *far* squares. All the close squares must be in N (since otherwise they would be part of A). However, we do not know anything about the far squares: they may be full or non-full. See Figure 1 for a picture.

Lemma 5. *No two far squares are more than ℓ_∞ distance $c/10$ apart.*

Remark: this does not say whp since we are assuming this non-full component has size at most U.

Proof. Suppose not.

Suppose, first, that no point of N is within c of the edge of S_n, and that the two far squares are at horizontal distance at least $c/10$. Then consider the left most far square. All squares which are to the left of this and with distance to this square less than $(c-2)$ must be close and thus in N. Similarly with the right most far square. Also at least $(c-2)$ squares (in fact nearly $2(c-2)$) in each of at least $c/10$ columns between the orginal two far squares must be in N. This is a total of about $\pi(c-2)^2+(c-2)c/10 > U$ which is a contradiction (provided we chose c reasonably large).

If there is a point of N within c of the boundary then the above argument gives more than $U/2$ non-full squares. Indeed, either it gives half of each part of the above construction, or it gives all of one end and all the side parts. This contradicts the second part of our assumption about the size of non-full components.

We do not need to consider a component near two sides: it cannot be large enough to be near two sides. It also cannot go across a corner, since no square within distance Uc of a corner is non-full. □

This result can also be deduced from a result of Penrose as we do in the next section. We have the following instant corollary.

Corollary 6. *The graph $\widehat{G}$ restricted to the far squares is complete.* □

Corollary 7. *The set $\Gamma(A^c)$ of neighbours in $\widehat{G}$ of the cutoff squares A^c is contained in N_{2c}.*

Proof. We prove that $A^c \subset N_c$ the c-blow-up of N. Indeed, suppose not. Let x be a square in $A^c \setminus N_c$. First, x cannot be a neighbour of any square in A or x would also be in A; i.e., x is a far square.

Now, let y be any square with ℓ_∞ distance $c/5$ from x. The square y cannot be in N since then x would be in N_c. Therefore, y cannot be a neighbour of any square in A since then it would be in A and, since x and y are joined in $\widehat{G}$, x would be in A; i.e., y is also a far square. Hence, x and y are both far squares with ℓ_∞ distance $c/5$ which contradicts Lemma 5.

□

In particular, Corollary 7 tells us that the sets of squares cutoff by different non-full components and all their neighbours are disjoint (obviously the $2c$-blow-ups are disjoint).

The final preparation we need is the following lemma.

Lemma 8. *The set $N_{2c} \cap A$ is connected in $\widehat{G}$.*

Since the proof will be using a standard graph theoretic result, it is convenient to define one more graph $\widehat{G}_1$: again the vertex set is the set of small squares, but this time each square is joined only to its four nearest neighbours: i.e., $\widehat{G}_1$ is the ordinary square lattice. We need two quick definitions. First, for a set $E \in \widehat{G}_1$ we define the *boundary* $\partial_1 E$ of E to be set of vertices in E^c that are neighbours (in $\widehat{G}_1$) of a vertex in E. Secondly, we say a set E in $\widehat{G}_1$ is *diagonally connected* if it is connected when we add the edges between squares which are diagonally adjacent (i.e. at distance $\sqrt{2}$) to $\widehat{G}$. The lemma we need is the following; since its proof is short we include it here for completeness.

Lemma 9. *Suppose that E is any subset of $\widehat{G}_1$ with E and E^c connected. Then $\partial_1 E$ is diagonally connected: in particular, it is connected in $\widehat{G}$.*

Proof. Let F be the set of edges of $\widehat{G}_1$ from E to E^c, and let F' be the corresponding set of edges in the dual lattice. Consider the set F' as a subgraph of the dual lattice. It is easy to check that every vertex has even degree except vertices on the boundary of $\widehat{G}_1$. Thus we can decompose F' into pieces each of which is either a cycle or a path starting and finishing at the edge of $\widehat{G}_1$. Any such cycle splits $\widehat{G}_1$ into two components, and we see that one of these must be exactly E and the other E^c. Thus F' is a single component in the dual lattice, and it is easy to check that implies that $\partial_1 E$ is diagonally connected. $\square$

Proof of Lemma 8. Consider $\widehat{G}_1 \setminus N_{2c}$. This splits into components $B_1, B_2, \ldots, B_m$. By definition each B_i is connected. Moreover, each B_i^c is also connected. Indeed, suppose $x, y \in B_i^c$. Then there is an xy path in $\widehat{G}_1$. If this is contained in B_i^c we are done. If not then it must meet N_{2c}, but N_{2c} is connected. Hence we can take this path until it first meets N_{2c} go through N_{2c} to the point where the path last leaves N_{2c} and follow the path on to y. This gives a path in B_i^c.

Hence, by Lemma 9, we see that each $\partial_1 B_i$ is connected in $\widehat{G}$ for each i (where ∂_1 denotes the boundary in $\widehat{G}_1$). Obviously $\partial_1 B_i \subset N_{2c}$.

As usual, for a set of vertices V let $\widehat{G}[V]$ denote the graph $\widehat{G}$ restricted to the vertices in V.

Claim. *Any two vertices in $\cup_{i=1}^m \partial_1 B_i$ are connected in $\widehat{G}[A \cap N_{2c}]$.*

Proof. Suppose not. Without loss of generality assume that, for some $k < m$, $\widehat{G}[\cup_{i=1}^k \partial_1 B_i]$ is connected and that no other $\partial_1 B_i$ is connected via a path to it. Pick $x \in B_1$ and $y \in B_m$. Both x and y are in A (since they are not in N_{2c} and $A^c \subset N_{2c}$ by Corollary 7).

Hence there is a path from x to y in A. Consider the last time it leaves $\cup_{i=1}^k B_i$. The path then moves around in N_{2c} before entering some B_j with $j > k$. This gives rise to a path in $A \cap N_{2c}$ from a point in $\cup_{i=1}^k \partial_1 B_i$ to a point in ∂B_j, contradicting the choice of k. $\square$

We now complete the proof of Lemma 8. To avoid clutter we shall say that two points are *joined* if they are connected by a path. Suppose that $x, y \in A \cap N_{2c}$. Since A is connected there is a path in A from x to y. If the path is contained in N_{2c} we are done. If not, consider the first time the path leaves N_{2c}. It must enter one of the B_i, crossing the boundary $\partial_1 B_i$. Hence x is joined to some $w \in \partial_1 B_i$ in $A \cap N_{2c}$. Similarly, by considering the last time the path is not in N_{2c} we see that y is joined to some $z \in \partial_1 B_j$ for some j. However, since the claim showed that w and z are joined in $A \cap N_{2c}$, we see that x and y are joined in $A \cap N_{2c}$. $\square$

COROLLARY 10. *The set of sea squares* $\widetilde{A} = \bigcap_{N \in cN} A(N)$. *In particular, for any* $N \in \mathcal{N}$ *we have* $\widetilde{A} \cap N_{2c} = A(N) \cap N_{2c}$.

PROOF. We prove that the set $A' = \bigcap_{N \in \mathcal{N}} A(N)$ is connected in $\widehat{G}$.

Given two squares x, y in A' pick a path in $\widehat{G}$ from x to y. Now for each non-full component N in turn do the following. If the path misses N_{2c} do nothing. Otherwise let w be the first point on the path in N_{2c} and z be the last point in N_{2c}. Replace the xy path by the path xw, any path wz in $A(N) \cap N_{2c}$ and then the path zy.

At each stage the modification ensured that the path now lies in $A(N)$. Also, the only vertices added to the path are in N_{2c} which is disjoint from all the previous N'_{2c}, and thus from all previous sets $A(N')$. Hence, when we have done this for all non-full components the path lies in every $A(N')$, i.e., in A'. Hence, A' is connected. Since it contains $\widetilde{A}$ we must have $A' = \widetilde{A}$.

Finally, by Corollary 7, $A(N') \supset N_{2c}$ for all $N' \neq N$, so $\widetilde{A} \cap N_{2c} = A(N) \cap N_{2c}$. $\square$

Stage 4: Dealing with the difficult subsquares.

We deal with each non-full component $N \in \mathcal{N}$ in turn. Fix one such component N.

Let us deal with the far squares first. There are three possibilities: the far squares contain no points at all, they contain one point in total, or they contain more than one point. In the first case, do nothing and proceed to the next part of the argument.

In the second case, by the 2-connectivity of G, we can find two vertex disjoint paths from this single vertex to points in squares in A. In the third case pick two points in the far squares. Again by 2-connectivity we can find vertex disjoint paths from these two vertices to points in squares in A.

Suppose that one of these paths meets A in square Q_1 and the other in square Q_2. Let P_1, P_2 be the squares containing the previous points on these paths. No two points in squares at (Euclidean) distance $(c+2)$ are joined we see that P_1 is within $(c+2)$ of Q_1. Since $P_1 \notin A$ we have that some square on a shortest P_1Q_1 path in $\widehat{G}_1$ is in N and thus that $Q_1 \in N_{2c}$. Similarly $Q_2 \in N_{2c}$. Combining we see that both Q_1 and Q_2 are in $N_{2c} \cap A$. By Lemma 8, we know that $N_{2c} \cap A$ is connected in $\widehat{G}$ so we can find a path from Q_1 to Q_2 in $N_{2c} \cap A$ in $\widehat{G}$. This "lifts" to a path in G going from the point in Q_2 to a point in Q_1 using at most one vertex in each subsquare on the way and never leaving N_{2c}.

Place the path from the point in Q_1 to the far vertex; round all points in the far region finishing back at the second chosen vertex and then to Q_2 and then through the sea as above back to Q_1. Since $Q_1 \in A \cap N_{2c}$, by Corollary 10 we have that $Q_1 \in \widetilde{A}$. Combining, we have a path starting and finishing in the same subsquare of the sea (i.e., Q_1) containing all the vertices in the far region.

Next we deal with the close squares: we deal with each close square P in turn. Since P is a close square we can pick $Q \in A$ with PQ joined in $\widehat{G}$.

If the square P has no point in it we ignore it. If it has one point in it then join that point to two points in Q.

If it has two or more points in it then pick two of them x, y: and pick two points uv in Q. Place the path formed by the edge ux round all the remaining unused vertices in the cutoff square finishing at y and back to the square Q with the edge yv in the cycle we are constructing.

The square Q is a neighbour of $P \in A^c$ so, by Corollary 7 is in N_{2c}. Since Q is also in A we see, by Corollary 10 as above, that $Q \in \widetilde{A}$.

When we have completed this construction we have placed every vertex in a cutoff square on one of a collection of paths each of which starts and finishes at the same square in the sea (although different paths may start and finish in different squares in the sea).

We use at most $2U+2$ vertices from any square in $A = A(N)$ when doing this so, provided that $M > 2U + 2 + (2c+1)^2$, there are at least $(2c+1)^2$ unused vertices in each square of A when we finish this. Moreover, obviously

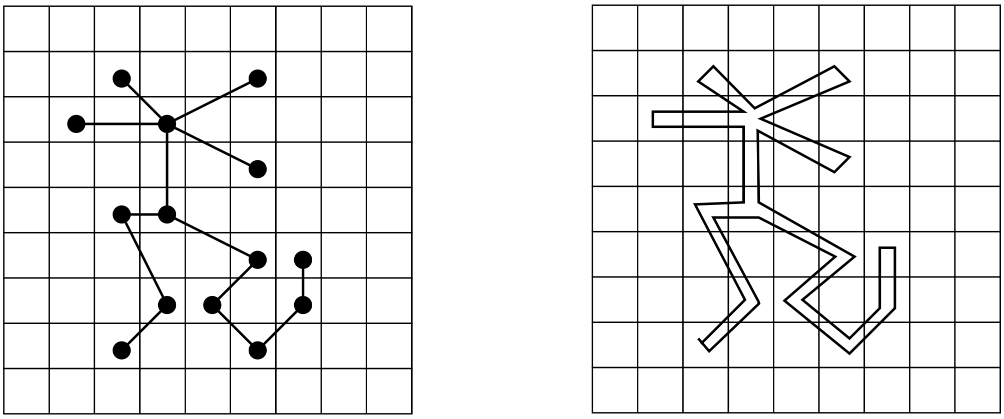

FIG 2. *A tree of subsquares and its corresponding tree cycle.*

the only squares touched by this construction are in N_{2c} and for distinct non-full components these are all disjoint. Hence, when we have done this for every non-full component $N \in \mathcal{N}$ there are at least $(2c+1)^2$ unused vertices in each square of the sea $\widetilde{A}$.

Stage 5: Using the subsquares in the sea to join everything together.

It just remains to string everything together. This is easy. Take a spanning tree for $\widetilde{A}$, the sea of squares. By doubling each edge we can think of this as a cycle, as in Figure 2. This cycle visits each square at most $(2c+1)^2$ times. (In fact, by choosing a spanning tree such that the sum of the edge lengths is minimal we could assume that it visits each vertex at most six times but we do not need this.) Convert this into a Hamilton cycle as follows. Start at an unused vertex in a square of the sea. Move to any (unused) vertex in the next square in the tree cycle. Then, if this is the last time the tree cycle visits this square visit all remaining vertices and join in all the paths constructed in the first part of the argument, then leave to the next square in the tree cycle. If it is not the last time the tree cycle visits this square then move to any unused vertex in the next square in the tree cycle. Repeat until we complete the tree cycle. Then join in any unused vertices and paths to this square constructed earlier before closing the cycle. □

3. Higher Dimensions.

We generalise the proof in the previous section to higher dimensions. Much of the argument is the same, in particular, essentially all of stages four and five. We include details of all differences but refer the reader to the previous section where the proof is identical.

Stage 1: Tessellation.

We work in the d-dimensional hypercube S_n^d of volume n (for simplicity we will abbreviate hypercube to cube in the following). As mentioned in the introduction, we no longer have a nice formula for the critical radius: the boundary effects dominate.

Instead, we consider the expected number of isolated vertices $E = E(r)$. We need a little notation: let A_r denote the set $\{x : d(x, A) \leq r\}$ and $|\cdot|$ denote Lebesgue measure.

We have $E = \int_{S_n^d} \exp(-|\{x\}_r|)dx$. Let $r_0 = r_0(n)$ be such that $E(r_0) = 1$. As before fix c a large constant to be determined later, and let $s = r_0/c$. It is easy to see that $r_0^d = \Theta(\log n)$ and $s^d = \Theta(\log n)$. We tile the cube S_n^d with small cubes of side length s.

As before, let $r = \mathcal{H}(G$ is 2-connected). By Penrose [11],[12] the probability that $r \notin [r_0(1 - 1/2c), r_0(1 + 1/2c)]$ tends to zero and we ignore all these points sets. This time any two points in cubes at distance $\frac{r-s\sqrt{d}}{s} \geq \frac{r_0-ds}{s} = (c-d)$ are joined, and no points in cubes at distance $\frac{r+s\sqrt{d}}{s} \leq \frac{r_0+ds}{s} = (c+d)$ are joined.

Stage 2: The "difficult" subcubes.

Exactly as before we define non-full cubes to be those containing at most M points, and we say two are joined if they have ℓ_∞ distance at most $4c-1$.

We wish to prove a version of Lemma 4. However, we have several possible boundaries: for example, in three dimensions we have the centre, the faces, the edges and the corners. We call a non-full component containing a cube Q *bad* if it consists of at least $(1 + 1/c)|Q_{r_0}|/s^d$ cubes. (Note a component can be bad for some cubes and not others).

LEMMA 11. *The expected number of bad components tends to zero as n tends to infinity. In particular there are no bad components whp.*

PROOF. The number of components of size U containing a particular cube is at most $(e(8c)^d)^U$. The probability that a cube is non-full is at most $2s^{dM}e^{-s^d}/M!$.

Since $\min\{|Q_{r_0}| : \text{cubes } Q\} = \Theta(\log n)$ and $s^d = \Theta(\log n)$, the expected number of bad components is at most

$$\begin{aligned}
&\sum_{\text{cubes } Q} (2s^{dM}e^{-s^d}(e(8c)^d)/M!)^{(1+1/c)|Q_{r_0}|/s^d} \\
&\qquad = \sum_{\text{cubes } Q} (2s^{dM}(e(8c)^d)/M!)^{(1+1/c)|Q_{r_0}|/s^d} \exp(-(1+1/c)|Q_{r_0}|) \\
&\qquad = o(1) \sum_{\text{cubes } Q} \exp(-|Q_{r_0}|) \\
&\qquad \le o(1) \int_{S_n^d} \exp(-|\{x\}_{r_0}|) dx \\
&\qquad = o(1)E(r_0) \\
&\qquad = o(1).
\end{aligned}$$

□

(Again, note that this is true independently of M.)

From now on we assume that there is no bad component.

Stage 3: The structure of the difficult subcubes.

In this stage we will need one extra geometric result of Penrose, a case of Proposition 2.1 of [11] or Proposition 5.15 of [12].

PROPOSITION 12. *Suppose d is fixed and that $\|\cdot\|$ is a p-norm for some $1 \le p \le \infty$. Then there exists $\eta > 0$ such that if $F \subset O^d$ (the positive octant in $\mathbb{R}^d$) with ℓ_∞ diameter at least $r/10$ and x is the point of F with the smallest l_1 norm; then $|F_r| \ge |F| + |\{x\}_r| + \eta r^d$.*

We begin this stage by proving Lemma 5 for this model.

LEMMA 13. *No two far cubes are more than ℓ_∞ distance $c/10$ apart.*

PROOF. Suppose not. Then let F be the set of far cubes, let x be the point of F closest to a corner in the l_1 norm and let Q be the cube containing x. We know that all the cubes within $(c-d)$ of a far cube are not in A. Hence all such cubes which are not far must be close, and thus non-full.

The number of close cubes is at least

$$\begin{aligned}\frac{|F_{(c-2d)s}\setminus F|}{s^d} &\geq \frac{|\{x\}_{(c-2d)s}|+\eta((c-2d)s)^d}{s^d} && \text{by Proposition 12}\\ &\geq \frac{|Q_{(c-3d)s}|+\eta r_0^d/2}{s^d} && \text{provided } c \text{ is large enough}\\ &\geq \frac{|Q_{(1-3d/c)r_0}|+\eta r_0^d/2}{s^d}\\ &\geq \frac{(1-3d/c)^d|Q_{r_0}|+\eta r_0^d/2}{s^d}\\ &> \frac{(1+1/c)|Q_{r_0}|}{s^d}, && \text{provided } c \text{ is large enough.}\end{aligned}$$

This shows that the component is bad which is a contradiction. ☐

Corollaries 6 and 7 hold exactly as before. Lemma 8 also holds, we just need to replace Lemma 9 by the following higher dimensional analogue. Note that, even in higher dimensions we say two squares are diagonally connected if their centres have distance $\sqrt{2}$.

LEMMA 14. *Suppose that E is any subset of $\widehat{G}_1$ with E and E^c connected. Then $\partial_1 E$ is diagonally connected: in particular, it is connected in $\widehat{G}$.*

PROOF. Let I be a (diagonally connected) component of $\partial_1 E$. We aim to show the $I = \partial_1 E$ and, thus, that $\partial_1 E$ is diagonally connected.

CLAIM. *Suppose that C is any circuit in $\widehat{G}_1$. Then the number of edges of C with one end in E and the other end in I is even.*

PROOF OF CLAIM. We say that a circuit is contractible to a single point using the following operations. First, we can remove an out and back edge. Secondly, we can do the following two dimensional move. Suppose that two consecutive edges of the circuit form two sides of a square; then we can replace them by the other two sides of the square keeping the rest of the circuit the same. For example, we can replace $(x, y+1, \vec{z}) \to (x+1, y+1, \vec{z}) \to (x+1, y, \vec{z})$ in the circuit by $(x, y+1, \vec{z}) \to (x, y, \vec{z}) \to (x+1, y, \vec{z})$.

Next we show that C is contractible. Let $w(C)$ denote the weight of the circuit: that is the sum of all the coordinates of all the vertices in C. We show that, if C is non-trivial, we can apply one of the above operations and reduce w. Indeed, let v be a vertex on C with maximal coordinate sum, and suppose that v_- and v_+ are the vertices before and after v on the circuit. If $v_- = v_+$ then we can apply the first operation removing v and v_+ from the circuit which obviously reduces w. If not, then both v_- and v_+ have strictly smaller coordinate sums than v and we can apply the second operation reducing w by two. We repeat the above until we reach the trivial circuit.

Now, let J be the number of edges of C with an end in each of E and I. The first operation obviously does not change the parity of J. A simple finite check yields the same for the second operation. Indeed, assume that we are changing the path from $(x, y+1), (x+1, y+1), (x+1, y)$ to $(x, y+1), (x, y), (x+1, y)$. Let F be the set of these four vertices. If no vertex of I is in F then obviously J does not change. If there is a vertex of I in F then, by the definition of diagonally connected, $F \cap I = F \cap \partial_1 E$. Hence the parity of J does not change. (It is even if $(x, y+1)$ and $(x+1, y)$ are both in E or both in E^c and odd otherwise.) $\square$

Suppose that there is some vertex $v \in \partial_1 E \setminus I$ and that $u \in E$ is a neighbour of v. Let $y \in I$ and $x \in E$ be neighbours. Since E and E^c are connected we can find paths P_{xu} and P_{vy} in E and E^c respectively. The circuit P_{xu}, uv, P_{vy}, yx contains a single edge from E to I which contradicts the claim. $\square$

To complete this stage observe that Corollary 10 holds as before.

Stage 4: Dealing with the difficult subcubes, and Stage 5: Using the subcubes in the sea to join everything together.

These two stages go through exactly as before (with one trivial change: replace $(2c+1)^2$ by $(2c+1)^d$). This completes the proof of Theorem 2. $\square$

4. Proof of Theorem 3.

In this section we prove Theorem 3. Once again, the proof is very similar to that in Section 2. We shall outline the key differences, and emphasise why we are only able to prove the weaker version of the result.

Stage 1: Tessellation.

The tessellation is similar to before, but this time some edges may be much longer than some non-edges.

Let $k = \mathcal{H}(G$ is κ-connected) be the smallest k that $_{n,k}$ is κ-connected. Since G is connected we may assume that $0.3 \log n < k < 0.52 \log n$ (see [1]). Let r_- be such that any two points at distance r_- are joined whp: e.g., Lemma 8 of [1] implies that this is true provided $\pi r_-^2 \leq 0.3 e^{-1-1/0.3} \log n$, so we can take $r_- = 0.035\sqrt{\log n}$.

Let r_+ be such that no edge in the graph has length more than r_+. Then, again by [1], we have

$$\pi r_+^2 \le 4e(1+0.52) \le 17$$

whp, so we can take $r_+ = 2.3\sqrt{\log n} \le 66 r_-$.

From here on, we ignore all point sets with an edge longer than r_+ or a non-edge shorter than r_-.

Let $s = r_-/\sqrt{8}$. We tessellate the box S_n with small squares of side length s. (Since we are proving only this weaker result our tesselation does not need to be very fine.) By the choice of s and the bound on r_- any two points in neighbouring or diagonally neighbouring squares are joined in G. Also, by the bound on r_+ no two points in squares with centres at distance more than $(66\sqrt{5}+2)s < 150s$ are joined. Let $D = 10^4$; we have that no two points in squares with centres distance Ds apart are joined.

Stage 2: The "difficult" subsquares.

We call a square *full* if it contains at least $M = 10^9$ points and *non-full* otherwise. We say two non-full squares are joined if they are at ℓ_∞ distance at most $2D-1$.

First we bound the size of the largest component of non-full squares.

LEMMA 15. *The largest component of non-full squares has size less than 7000 whp.*

PROOF. The number of components of size 7000 containing a particular square is at most $(e(4D)^2)^{7000}$ so, since there are less than n squares, the total number of such components is at most $n(e(4D)^2)^{7000}$. The probability that a square is non-full is at most $2s^{2M}e^{-s^2}/M!$. Hence, the expected number of components of size at least 7000 is at most

$$n(2s^{2M}e^{-s^2}(e(4D)^2)/M!)^{7000}$$
$$\le n\left(2\left(\frac{(0.035)^2\log n}{5}\right)^M \frac{e(4D)^2}{M!}\right)^{7000} \exp\left(\frac{-7000(0.035)^2\log n}{8}\right)$$

which tends to zero as n tends to infinity (since $7000(-0.035)^2/8 > 1.07 > 1$): i.e., whp, no such component exists. □

In the rest of the argument we shall assume that there is no non-full component of size greater than 7000.

Stage 3: The structure of the difficult subsquares.

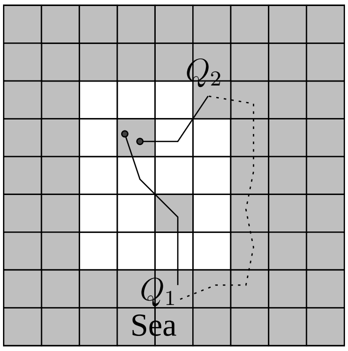

FIG 3. *Two paths from one cutoff square to the sea together with the path from the meeting point in Q_2 to the square Q_1.*

As usual we fix one component N of the non-full squares, and suppose that it has size u (so we know $u < 7000$). This time we define $\widehat{G}$ to be the graph on the small squares where each square is joined to its eight nearest neighbours (i.e., adjacent and diagonal). Let A be the giant component of $G\backslash N$, and again split the cutoff squares into close and far depending whether they have a neighbour (in $\widehat{G}$) in A.

By the vertex isoperimetric inequality in the square there are at most $u^2/2$ squares in $A^c \setminus N$ so $|A^c| \leq u^2/2 + u < 2.5 \cdot 10^7$.

Next we prove a result similar to Corollary 7.

LEMMA 16. *The set of cutoff squares A^c is in N_D (where $D = 10^4$ as above).*

PROOF. Suppose not, and that Q is a square in A^c not in N_D. Then all squares within ℓ_∞ distance of Q at most D are not in N. Hence they must be in A^c (since otherwise there would be a path from Q to a square in A not going through any square in N). Hence $A^c > D^2 = 10^8$ which contradicts Lemma 15. ☐

Finally, we need the analogue of Lemma 8 whose proof is exactly the same.

LEMMA 17. *The set $N_D \cap A$ is connected in $\widehat{G}$.* ☐

Stage 4: Dealing with the difficult subsquares.

Let us deal with these cutoff squares now. From each cutoff square that contains at least two vertices pick any 2 vertices and from each cutoff square that contains a single vertex pick that vertex with multiplicity two. We have picked at most $5 \cdot 10^7$ vertices, so since G is $\kappa = 5 \cdot 10^7$ connected we can simultaneously find vertex disjoint paths from each of our picked vertices to vertices in squares in A (two paths from those vertices that are repeated).

We remark that these are not just single edges: these paths may go through other cutoff squares.

Call the first point of such a path which is in the sea a *meeting point*, and the square containing it a *meeting* or *terminating* square.

Each cutoff square has two meeting points say in subsquares Q_1, Q_2. Since the longest edge is at most r_+ both Q_1 and Q_2 are in N_D. Since $A \cap N_D$ is connected in $\widehat{G}$ we construct a path in the squares in $A \cap N_D$ from the meeting point in Q_2 to a vertex in Q_1 using at most one vertex in each subsquare on the way, and missing all the other meeting points. This is possible since each full square contains at least $M = 10^9$ vertices.

Place the path from the terminating point in Q_1 to the cutoff square, round all the vertices in the cutoff square not used in the rest of the construction, followed by the path back to the meeting point in Q_2 and then through the sea as above back to Q_1. We have a path starting and finishing in the same subsquare (Q_1) containing all the (unused) vertices in the cutoff square.

Do this for every cutoff square. When doing this, in each square in the sea we use at most two vertices for each cutoff square. Moreover, obviously only squares in N_D are touched by this construction. Since non-full squares in distinct components are at distance at least $2D$ the squares touched by different non-full components are distinct. Thus in total we have used at most $4 \cdot 10^7$ vertices in any square in the sea, and since $M = 10^9$ there are many (we shall only need 8) unused vertices left in each full square in the sea.

Stage 5: Using the subsquares in the sea to join everything together.

This is exactly the same as before.

5. Comments on the *k*-nearest neighbour proof. We start by giving some reasons why the proof in the k-nearest neighbour model only yields the weaker Theorem 3. The first superficial problem is that we use squares in the tesselation which are of “large” size rather than relatively small as in the proof of Theorem 1, (in other words we did not introduce the constant c when setting s depending on r).

Obviously we could have introduced this constant. The difficulty when trying to mimic the proof of Theorem 1 is the large difference between r_- and r_+, which corresponds to having a very large number of squares (many times πc^2) in our non-full component N. This means that we cannot easily prove anything similar to Lemma 5. Indeed, a priori, we could have two far squares with πc^2 non-full squares around each of them.

A different way of viewing this difficulty is that, in the k-nearest neighbour model, the graph $\widehat{G}$ on the small squares does not approximate the real graph G very well, whereas in the Gilbert Model it is a good approximation. Thus, it is not surprising that we only prove a weaker result.

This is typical of results about the k-nearest neighbour model: the results tend to be weaker than for the Gilbert Model. This is primarily because the obstructions tend to be more complex: for example, the obstruction for connectivity in the Gilbert Model is the existence of an isolated vertex. Obviously in the k-nearest neighbour model we never have an isolated vertex; the obstruction must have at least $k+1$ vertices.

Extensions of Theorem 3. When proving Theorem 3 we only used two facts about the random geometric graph. First, that any two points at distance $r_- = 0.035\sqrt{\log n}$ are joined whp. Secondly, that the ratio of r_+ (the longest edge) to r_- (the shortest non-edge) was at most 60 whp. Obviously, we could prove the theorem (with different constants) in any graph with $r_- = \Theta(\sqrt{\log n})$ and r_+/r_- bounded. This includes higher dimensions and different norms and to different shaped regions instead of S_n (e.g. to disks or toruses). Indeed, the only place we used the norm was in obtaining the bounds on r_+ and r_- in stage 1 of the proof.

Indeed, it also generalises to irregular distributions of vertices provided that the above bounds on r_- and r_+ hold. For example, it holds in the square S_n where the density of points in the Poisson Process decrease linearly from 10 to 1 across the square.

6. Closing Remarks and Open Questions. A related model where the result does not seem to follow easily from our methods is the directed version of the k-nearest neighbour graph. As mentioned above, the k-nearest neighbour model naturally gives rise to a directed graph and we can ask whether this has a directed Hamilton cycle. Note that this directed model is significantly different from the undirected: for example it is likely (see [1]) that the obstruction to directed connectivity (i.e., the existence of a directed path between any two vertices) is a single vertex with in-degree zero; obviously this cannot occur in the undirected case where every vertex has degree at least k. In some other random graph models a sufficient condition for the existence of a Hamilton cycle (whp) is that there are no vertices of in-degree or out-degree zero. Of course, in the directed k-nearest neighbour model every vertex has out-degree k so we ask the following question.

QUESTION. *Let* $\vec{G} = \vec{G}_{n,k}$ *be the directed* k*-nearest neighbour model. Is*

$$\mathcal{H}(\vec{G} \text{ has a Hamilton cycle}) = \mathcal{H}(\vec{G} \text{ has no vertex of in-degree zero})$$

whp?

It is obvious that the bound on connectivity in the k-nearest neighbour model can be improved, but the key question is "should it be two?" We make the following natural conjecture:

CONJECTURE. *Suppose that* $k = k(n)$ *such that the* k*-nearest neighbour graph* $G = G(k, n)$ *is a* 2*-connected whp. Then, whp,* G *has a Hamilton cycle.*

7. Acknowledgement. Some of the results published in this paper were obtained in June 2006 at the Institute of Mathematics of the National University of Singapore during the program 'Random Graphs and Real-world Networks'. We are grateful to the Institute for its hospitality.

REFERENCES

[1] P. Balister, B. Bollobás, A. Sarkar and M. Walters, *Connectivity of random k-nearest neighbour graphs.* Advances in Applied Probability. **37** (2005), 1–24.

[2] P. Balister, B. Bollobás, A. Sarkar and M. Walters, *Highly connected random geometric graphs.* Discrete Applied Maths. **157** (2009), 309–320.

[3] P. Balister, B. Bollobás, A. Sarkar and M. Walters, *A critical constant for the k-nearest neighbour model.* Advances in Applied Probability. (to appear)

[4] J. Balogh, H. Kaul and R. Martin, *A threshold for random geometric graphs with a Hamiltonian cycle.* Unpublished Manuscript.

[5] B. Bollobás, *The evolution of sparse graphs.* Graph Theory and Combinatorics. 35–57, Academic Press, London, (1984).

[6] B. Bollobás, The Art of Mathematics: Coffee Time in Memphis. Cambridge University Press, 2006.

[7] J. Díaz, D. Mitsche, X. Pérez, *Sharp threshold for Hamiltonicity of random geometric graphs.* SIAM Journal on Discrete Mathematics. **21** (2007), 57–65.

[8] J. Komlós and E. Szemerédi, *Limit distributions for the existence of Hamilton cycles in a random graph.* Discrete Mathemetics. **43** (1983) 55–64

[9] A. D. Korshunov, *A solution of a problem of P. Erdős and A. Rényi about Hamilton cycles in non-oriented graphs.* Metody Diskr. Anal. Teoriy Upr. Syst., Sb. Trudov Novosibirsk **31** (1977) 17–56 (in Russian)

[10] M.D. Penrose, *The longest edge of the random minimal spanning tree.* Annals of Applied Probability. **7** (1997), 340–361.

[11] M.D. Penrose, *On k-connectivity for a geometric random graph.* Random Structures and Algorithms. **15** (1999) 145–164.

[12] M.D. Penrose, Random Geometric Graphs. Oxford University Press, 2003.

[13] J. Petit, Layout Problems. PhD Thesis, Universitat Politècnica de Catalunya. March, 2001.

[14] L. Pósa, *Hamiltonian circuits in random graphs.* Discrete Mathematics. **14** (1976) 359–364.

[15] F. Xue and P.R. Kumar, *The number of neighbors needed for connectivity of wireless networks.* Wireless Networks. **10** (2004), 169–181.

Department of Mathematics,
University of Illinois,
Urbana,
IL 61801,
USA.
E-mail: jobal@math.uiuc.edu

Department of Pure Mathematics and Mathematical Statistics,
University of Cambridge,
Cambridge
CB3 0WB,
UK
E-mail: b.bollobas@dpmms.cam.ac.uk

School of Mathematical Sciences,
Queen Mary University of London,
London
E1 4NS,
UK
E-mail: M.Walters@qmul.ac.uk