The Annals of Applied Probability
1997, Vol. 7, No. 2, 340–361

THE LONGEST EDGE OF THE RANDOM MINIMAL SPANNING TREE

BY MATHEW D. PENROSE

University of Durham

For n points placed uniformly at random on the unit square, suppose M_n (respectively, M'_n) denotes the longest edge-length of the nearest neighbor graph (respectively, the minimal spanning tree) on these points. It is known that the distribution of $n\pi M_n^2 - \log n$ converges weakly to the double exponential; we give a new proof of this. We show that $P[M'_n = M_n] \to 1$, so that the same weak convergence holds for M'_n.

1. Introduction. Suppose n bushes are randomly scattered in the unit square, and a disease (or fire) then appears at one of them. Once sick, a bush never recovers, and passes on the disease to every other bush within a distance r. Eventually, all the bushes become sick, except for those which are insulated by a zone of radius r containing no bushes that ever become sick. After a long period of time (relative to the time scale of the spread of the disease), all the sick bushes die, leaving behind any insulated bushes. If a sufficient number of such bushes remain, there will be a chance for the forest to regrow. We are here interested in the question: for which values of r is there likely to be one or more such insulated bushes?

The geometry of this question can be reformulated in terms of the minimal spanning tree (MST), an object much studied in combinatorial optimization. The Euclidean MST on a set of n points (denoted $\eta_1, \dots, \eta_n$) in $\mathbf{R}^\nu$ is the connected graph with these points as vertices and with minimum total edge-length. In the present paper, we take the η_i to be *random*, independently uniformly distributed on the unit cube $B = (-1/2, 1/2]^\nu$, and write $\mathscr{X}_n$ for the point process $\{\eta_1, \dots, \eta_n\}$. Various authors have studied this random MST, starting with Beardwood, Halton and Hammersley [7]. For a survey, see [28] or [19].

We shall derive the asymptotic distribution of the *maximum* of these edge-lengths, denoted M_n. By known properties of the MST [see (12) below], $M_n < r$ if and only if for every pair of points η_i, η_j there is a sequence of points of $\mathscr{X}_n$, starting with η_i and ending with η_j, with each pair of successive points in the sequence separated by a distance less than r. In terms of the ecological model of the opening paragraph, the statement $M_n \geq r$ is equivalent to the existence of an insulated bush. Note that if the objective function is the maximum rather than the sum of the edge-lengths, the MST remains optimal, although it may not be the unique optimum.

Received September 1995; revised November 1996.
AMS 1991 *subject classifications*. Primary 60D05, 60G70; secondary 05C05, 90C27.
Key words and phrases. Geometric probability, minimal spanning tree, nearest neighbor graph, extreme values, Poisson process, Chen–Stein method, continuum percolation.

Two useful simplifications to the model are the Poisson and toroidal assumptions. In the *toroidal* model, instead of the Euclidean metric $d(i, j) = |\eta_i - \eta_j|$, we use the metric $d(i, j) = \min_{z \in Z^\nu} |\eta_i - \eta_j - z|$, which eliminates boundary effects. In the *Poisson* model, instead of $\mathscr{X}_n$ we consider the point process $\mathscr{P}_n := \{\eta_1, \dots, \eta_{N_n}\}$, where N_n is a Poisson variable with mean n, independent of $\{\eta_i\}$. So $\mathscr{P}_n$ is simply a homogeneous Poisson process on the cube of rate n. The independence properties of $\mathscr{P}_n$ simplify the analysis; also, as argued in [14], the Poisson model is sometimes more realistic.

Set $\pi_\nu = \pi^{\nu/2}/\Gamma((\nu/2)+1)$, the volume of the unit ball in ν dimensions. In its simplest form, the basic result of this paper is that for the toroidal model with $\nu \geq 2$ or the Euclidean model with $\nu = 2$, if M_n is the maximum edge length in the MST on either $\mathscr{P}_n$ or $\mathscr{X}_n$, then the distribution of $n\pi_\nu M_n^\nu - \log n$ converges weakly to the double exponential distribution:

$$\lim_{n\to\infty} P[n\pi_\nu M_n^\nu - \log n \leq \alpha] = \exp(-e^{-\alpha}), \qquad \alpha \in \mathbf{R}. \tag{1}$$

Our first step will be to look at the *k-nearest neighbor graph* (k-NNG), which is important in its own right. For k a fixed integer, the k-NNG on $\mathscr{X}_n$ is the graph in which each point of $\mathscr{X}_n$ is connected by an edge to its kth nearest neighbor out of the other points of $\mathscr{X}_n$, and the k-NNG on $\mathscr{P}_n$ is defined likewise. We write simply NNG for 1-NNG. Note that the NNG is a subgraph of the MST, as can be seen directly or from (12) below.

It is known (see below) that if M_n denotes the maximum edge-length in the NNG (rather than the MST) on $\mathscr{X}_n$, then (1) holds. Thus, to prove (1) for the MST, it suffices to prove that with the obvious notation,

$$\lim_{n\to\infty} P[M_n(MST) = M_n(NNG)] = 1. \tag{2}$$

This key comparison is achieved by Theorem 1 below.

In the ecological model, one may wish to record the number of insulated bushes and their positions, rather than simply whether or not such a bush exists. Alternatively, one may wish to record the length and location of the edge of the MST that is longest, second longest, and so on; similarly for the k-NNG. This takes us into the realm of weak convergence of $(\nu + 1)$-dimensional point processes, which is the setting of our most general results (Theorems 2 and 3), which include (1) as a special case.

The generalization of (1) to the $(k+1)$-NNG, with longest edge again denoted M_n, is

$$\lim_{n\to\infty} P[n\pi_\nu M_n^\nu - \log n - k\log(\log n) + \log k! \leq \alpha] = \exp(-e^{-\alpha}). \tag{3}$$

Henze [17] proves a related result by an argument he says is "long and tedious." Steele and Tierney [30] observe that this can be modified to prove (3) for $k = 0$ (i.e., (1) for the NNG), for the toroidal model with $\nu \geq 2$ or the Euclidean model with $\nu = 2$. Here we use a completely different argument based on Stein's method, to prove a more general point-process result (Theorem

2) with (3) as a special case. The basic idea is quite simple; see Section 4. The method can be used to give explicit error bounds in (3). We consider the toroidal model for all $k \geq 0$ and all ν, and the Euclidean model for $k = 0$ and $\nu \leq 2$. The results for $(k+1)$-NNG hold for $\nu = 1$; our arguments for MST apply only for $\nu \geq 2$.

The MST and k-NNG have applications in computer science, the physical sciences and in biology; see the references in [29] and [30]. Statisticians have used the MST and k-NNG on n random points in ν dimensions, representing multivariate observations, as a means of imposing a structure on these points. For example, the MST is bound up with the so-called single linkage algorithm for partitioning the points of $\mathscr{X}_n$ into clusters, as described in, for example, [15, 16]. The single linkage clusters "at level r" are the components of the MST when edges of length greater than r are removed, and M_n is the level above which there is just one cluster.

The edges of the MST have been used as a multivariate analogue to the interpoint spacings for one-dimensional data. For example, Rohlf [26] proposes the use of longest edges of the MST as a means of detecting multivariate outliers. For a recent use of this method, see [13]; for criticisms see [9]. Part of the problem is that the distribution of M_n for the MST has not been well understood. The results in this paper are a step towards rectifying this situation.

We briefly mention some other results in the field. Appel and Russo [3, 4] derive strong laws for M_n for both the NNG and MST, complementing the weak limits given here. Dette and Henze [10] look at M_n for the NNG in the Euclidean model with $\nu \geq 3$, which is not considered here. Other functionals besides M_n for which weak limits have been derived are the total edge length (see [6] for the k-NNG, [21] for the MST), and the minimum edge length ([27]). Jaillet [18] derives a bound on the probability that M_n is large for the MST, which he uses to compare Euclidean and toroidal models, with regard to total edge length. Also related are results of Hall [14] and Janson [20] concerning the coverage of the cube B by small balls, for example, the probability that every point of B is covered by at least k balls of radius r centered at $\mathscr{X}_n$. The statement that $M_n \leq r$ for the k-NNG is the statement that every point of $\mathscr{X}_n$ is covered by at least $k+1$ such balls. An application of Stein's method to a coverage problem is given in [1].

Qualitatively, the meaning of (1) is that (i) for the NNG, the asymptotics for M_n are as if the nearest-neighbor distances were independent, and (ii) the longest edge is likely to be the same for the MST as for the NNG. It is reasonable to expect this description to be valid for other distributions of the η_i, besides the uniform case considered here. In Penrose [24] the description is shown to hold for normally distributed η_i.

2. Statement of results. For $\alpha \in \mathbf{R}$, we shall say that an edge (i, j) of the MST or NNG is *α-long* if $n\pi_\nu(d(i, j))^\nu - \log n > \alpha$. Thus, (1) says the probability that no α-long edge exists tends to $\exp(e^{-\alpha})$. Since (1) holds for the NNG, our first theorem gives us the comparison (2) between MST and NNG.

THEOREM 1. *Consider the toroidal model with $\nu \geq 2$ or the Euclidean model with $\nu = 2$. Let $\alpha \in \mathbf{R}$. Then with probability approaching* 1 *as $n \to \infty$, every α-long edge of the MST on $\mathscr{P}_n$ or on $\mathscr{X}_n$ is also in the corresponding NNG, and moreover every such edge has an end at a leaf of the MST, that is, a vertex of degree* 1.

Our most general results are given in terms of point processes, a natural setting for the study of extreme values. To state them, we first need to give some definitions. Given a separable metric space E, a point process $\mathscr{Y}$ on E is a random set of points in E that is at most countable. We write $\mathscr{Y}(A)$ for the number of points of $\mathscr{Y}$ in a set A. The particular spaces of interest here are (1) $E = B = (-0.5, 0.5]^\nu$, with the Euclidean or toroidal metric, and (2) $E = \mathbf{R} \times B$.

We refer to a finite point process in B as *nice* if the interpoint distances are a.s. all distinct. The empirical point processes $\mathscr{X}_n$ and $\mathscr{P}_n$ on B of this paper are nice (in either the Euclidean or toroidal metric), so that the MST and k-NNG are a.s. uniquely defined. For point processes on $\mathbf{R} \times B$, we have a different definition of niceness, which we now explain.

Suppose the points of a finite or countable set $\mathbf{y} \subset \mathbf{R} \times B$ can be listed as $\mathbf{y} = \{(t_m, \mathbf{x}_m),\ m \geq 1\}$, with $t_1 > t_2 > t_3 > \cdots$, and with $t_m \to -\infty$ as $m \to \infty$ in the case that $\mathbf{y}$ is infinite. We shall refer to this as the *canonical listing* of the points of $\mathbf{y}$.

Let $\mathscr{S}$ denote the semiring of subsets of $\mathbf{R} \times B$ of the form $[\alpha, \beta) \times A$, with $-\infty < \alpha \leq \beta \leq \infty$, and with $A \subset B$ being a product of intervals.

We shall say a point process $\mathscr{Y}$ in $\mathbf{R} \times B$ is *nice* if (1) it has a.s. a canonical listing, and (2) $P[\mathscr{Y}(\partial S) > 0] = 0$ for all $S \in \mathscr{S}$. Our notion of weak convergence of nice point processes on $\mathbf{R} \times B$ is given by the equivalent statements of the following lemma.

LEMMA 1. *Let $\mathscr{Y}_n$, $n \geq 0$ be a sequence of nice point processes on $\mathbf{R} \times B$, with $\mathscr{Y}_0$ being infinite almost surely. Then the following statements* (a) *and* (b) *are equivalent.*

(a) *For any collection of disjoint sets $R_1, \dots, R_K$ in $\mathscr{S}$,*

$$(\mathscr{Y}_n(R_1), \mathscr{Y}_n(R_2), \dots, \mathscr{Y}_n(R_K)) \to_d (\mathscr{Y}_0(R_1), \mathscr{Y}_0(R_2), \dots, \mathscr{Y}_0(R_K)) \quad \text{as } n \to \infty,$$

where $\to_d$ is convergence in distribution in $\mathbf{R}^\nu$;

(b) *There exist coupled point processes $\mathscr{Y}'_n$, $n = 0, 1, 2, \dots$, all on the same probability space, such that* (i) *$\mathscr{Y}'_n$ has the same distribution as $\mathscr{Y}_n$, for each n, and* (ii) *with $\mathscr{Y}'_n$ given by the canonical listing $\mathscr{Y}'_n = \{(T'_{n,m}, \mathbf{X}'_{n,m}), m \geq 1\}$, we have $(T'_{n,m}, \mathbf{X}'_{n,m}) \to (T'_{0,m}, \mathbf{X}'_{0,m})$ almost surely as $n \to \infty$, for each m.*

PROOF. Obviously (b) implies (a). Conversely, assume (a). Let $\{(T_{n,m}, \mathbf{X}_{n,m}), m \geq 1\}$ be the canonical listing of $\mathscr{Y}_n$. Let $S_1, \dots, S_M \in \mathscr{S}$. By re-expressing the following events in terms of the number of points of point processes in sets in $\mathscr{S}$, and using (a), we have

$$\lim_{n\to\infty} P\left[\bigcap_{m=1}^{M} \{(T_{n,m}, \mathbf{X}_{n,m}) \in S_m\}\right] = P\left[\bigcap_{m=1}^{M} \{(T_{0,m}, \mathbf{X}_{0,m}) \in S_m\}\right].$$

Therefore (see [8], page 19), $((T_{n,m}, \mathbf{X}_{n,m}),\ m \ge 1) \to_d ((T_{n,m}, \mathbf{X}_{n,m}),\ m \ge 1)$ in $(\mathbf{R} \times B)^\infty$ as $n \to \infty$, and (b) follows by the Skorohod representation theorem ([12], Theorem 3.1.8). □

DEFINITION. Given nice point processes $\mathscr{Y}_n$, $n \ge 0$, with $\mathscr{Y}_0$ almost surely infinite, we write $\mathscr{Y}_n \to_d \mathscr{Y}_0$ as $n \to \infty$ if either statement (a) or (b) in Lemma 1 holds. This is also equivalent to the convergence in distribution of $\mathscr{Y}_n$ to $\mathscr{Y}_0$ viewed as random elements of the space of point measures on $(-\infty, \infty] \times B$, with the vague topology (see [25], Chapter 3).

We now look at nearest neighbors. Given a nice point process $\mathscr{Y} = \{\eta_1, \dots, \eta_N\}$ in B, we define $R_{i,k}(\mathscr{Y})$ to be the distance from η_i to its kth nearest neighbor in $\mathscr{Y}$, using the Euclidean or toroidal metric according to the context. Write $R_i(\mathscr{Y})$ for $R_{i,1}(\mathscr{Y})$; that is,

$$R_i(\mathscr{Y}) = \min\{d(i, j);\ j \le N\}. \tag{4}$$

We define a point process on $\mathbf{R} \times B$, denoted $\mathscr{G}_{n,k}(\mathscr{Y})$, which records the (rescaled) lengths and locations of long edges of the $(k+1)$-NNG on $\mathscr{Y}$, as follows:

$$\mathscr{G}_{n,k}(\mathscr{Y}) = \{(n\pi_\nu(R_{i,k+1}(\mathscr{Y}))^\nu - \log n - k\log(\log n) + \log k!, \eta_i)\colon \eta_i \in \mathscr{Y}\}.$$

We write simply $\mathscr{G}_n(\mathscr{Y})$ for the point process $\mathscr{G}_{n,0}(\mathscr{Y}) = \{(n\pi_\nu(R_i(\mathscr{Y}))^\nu - \log n, \eta_i)\colon \eta_i \in \mathscr{Y}\}$.

Let $\mathscr{P}_\infty$ denote a nonhomogeneous Poisson point process on $\mathbf{R} \times B$ with mean measure $\mu(\cdot) = E[\mathscr{P}_\infty(\cdot)]$ given by $\mu(dt\, d\mathbf{x}) = e^{-t}\, dt\, d\mathbf{x}$. In the canonical listing $\mathscr{P}_\infty = \{(T_m, \mathbf{X}_m),\ m \ge 1\}$, the T_m are the points of a Poisson process on $\mathbf{R}$ with mean measure $e^{-t}\, dt$, arranged in decreasing order, and the $\mathbf{X}_i$ are independent and uniform on B. Our main result for the k-NNG has this point process as a weak limit, as follows.

THEOREM 2. *For the toroidal model with* $\nu \ge 1$ *or the nontoroidal Euclidean model with* $\nu = 1$ *or* $\nu = 2$,

$$\mathscr{G}_n(\mathscr{P}_n) \to_d \mathscr{P}_\infty \quad \text{as } n \to \infty, \tag{5}$$

and

$$\mathscr{G}_n(\mathscr{X}_n) \to_d \mathscr{P}_\infty \quad \text{as } n \to \infty. \tag{6}$$

Also for the toroidal model with $\nu \ge 1$, *and* $k \ge 0$,

$$\mathscr{G}_{n,k}(\mathscr{P}_n) \to_d \mathscr{P}_\infty \quad \text{as } n \to \infty, \tag{7}$$

and

$$\mathscr{G}_{n,k}(\mathscr{X}_n) \to_d \mathscr{P}_\infty \quad \text{as } n \to \infty. \tag{8}$$

In particular, if M_n denotes the length of the longest edge of the $(k+1)$-NNG, then

$$\lim_{n\to\infty} P[n\pi_\nu M_n^\nu - \log n - k\log(\log n) + \log k! \le \alpha] = \exp(-e^{-\alpha}). \tag{9}$$

For $\alpha \in \mathbf{R}$, we shall call an edge (i, j) of the $(k+1)$-NNG *α-long* if $n\pi_\nu(d(i, j))^\nu - \log n - k\log(\log n) + \log k! \ge \alpha$. The following result is intended to clarify the statement of Theorem 2; it says that the risk of an α-long edge being counted twice over by the point process $\mathscr{G}_n(\mathscr{P}_n)$ or $\mathscr{G}_n(\mathscr{X}_n)$ is negligible.

LEMMA 2. *Let $\alpha \in \mathbf{R}$. For the toroidal model with $\nu \ge 1$ and $k \ge 0$, or the Euclidean model with $\nu \le 2$ and $k = 0$, the number of α-long edges (i, j) of the $(k+1)$-NNG on $\mathscr{X}_n$ or $\mathscr{P}_n$ for which η_i is the $(k+1)$st nearest neighbor of η_j and η_j is the $(k+1)$st nearest neighbor of η_i, converges in probability to zero.*

Lemma 2 can be deduced from Theorem 2. In brief, take $\varepsilon_n \to 0$ so that $P[M_n > \varepsilon_n] \to 0$. Since a homogeneous Poisson process on B has no multiple points a.s., the probability that there exist $i, j \le N_n$ with $d(i, j) \le \varepsilon_n$, such that the edge from i to its $(k+1)$st nearest neighbor is α-long and likewise for j, converges to zero.

By Lemma 2, the number of α-long edges of the $(k+1)$-NNG on $\mathscr{P}_n$ has the same asymptotic distribution as the number of points $\mathscr{G}_n(\mathscr{P}_n)$ in $[\alpha, \infty) \times B$; since this set is in $\mathscr{S}$, by Theorem 2 this asymptotic distribution is Poisson with mean $\exp(-\alpha)$, and therefore (9) follows. Similarly, Theorem 2 gives us asymptotic formulas for the (joint) distributions of the second, third and so on, longest edges of the k-NNG, and also says that the locations of these edges are asymptotically independent and uniform on B.

Turning to the MST, we define a point process recording the lengths and locations of long edges of the MST, as for the NNG. To do this, we specify the location of edge (i, j) by the midpoint of the geodesic from η_i to η_j. This midpoint, denoted $\mathbf{m}(i, j)$, is an element of B satisfying $d(\mathbf{m}(i, j), \eta_i) = d(\mathbf{m}(i, j), \eta_j) = (1/2)d(i, j)$. Set

$$\mathscr{M}_n(\mathscr{Y}) = \{(n\pi_\nu(d(i, j))^\nu - \log n, \mathbf{m}(i, j)) \colon (i, j) \in MST(\mathscr{Y})\}.$$

Our main result for the MST is the following.

THEOREM 3. *In the toroidal model with $\nu \ge 2$ or the Euclidean model with $\nu = 2$,*

$$\mathscr{M}_n(\mathscr{P}_n) \to_d \mathscr{P}_\infty \quad \text{as } n \to \infty \quad \text{and} \quad \mathscr{M}_n(\mathscr{X}_n) \to_d \mathscr{P}_\infty \quad \text{as } n \to \infty.$$

Theorem 3 can be deduced from Theorems 1 and 2 by a routine argument which we omit. The remaining sections are devoted to the proofs of Theorems 1 and 2. The restatements of parts of these theorems in the later sections are labelled as propositions.

3. The MST on the torus. In this section we prove Theorem 1 for the Poisson toroidal model (Proposition 1 below). First we prove a weaker version of that result.

LEMMA 3. *Let $\alpha \in \mathbf{R}$, and let $r_n = r_n(\alpha)$ be given by*

$$n\pi_\nu r_n^\nu - \log n = \alpha. \tag{10}$$

For the toroidal model, let $D_n(i, j)$ be the event that (i, j) is an edge of the MST on $\mathscr{P}_n$ and that $d(i, j) \geq r_n$, but $R_i(\mathscr{P}_n) < r_n$ and $R_j(\mathscr{P}_n) < r_n$. Then

$$\lim_{n\to\infty} P\left[\bigcup_{i<j\leq N_n} D_n(i, j)\right] = 0. \tag{11}$$

REMARKS. Lemma 3 suffices to prove interesting statements about the MST in the Poisson, toroidal model. Indeed, it is easy to deduce the basic result (1) for MST in the Poisson toroidal model from the corresponding result for NNG and Lemma 3. In terms of the ecological model with range of infection r_n, Lemma 3 says that with probability approaching 1, every insulated bush is isolated. That is, its r_n-neighborhood contains no other bush.

The proof uses ideas from continuum percolation. For $r > 0$, $x \in \mathbf{R}^\nu$ and any set of points S in $\mathbf{R}^\nu$, let the "r-cluster of x in S," denoted $C_r(x; S)$, be the union of $\{x\}$ and the set of of $y \in S$ such that there is a sequence $y_1, \dots, y_n = y$ of points of S with $d(y_i, y_{i-1}) < r$ for each i, with $y_0 = x$. This notation is relevant to the MST because of the following deterministic fact, given in Proposition 2.1 of Alexander [2]: for a nice point process $\mathscr{Y} = \{\eta_1, \dots, \eta_N\}$ in B,

$$(i, j) \in MST(\mathscr{Y}) \quad \text{iff } \eta_j \notin C_{d(i,j)}(\eta_i; \mathscr{Y}). \tag{12}$$

It is immediate from (12) that if $d(i, j) = R_i(\mathscr{Y})$, then $(i, j) \in \mathrm{MST}(\mathscr{Y})$, that is, the NNG is a subgraph of the MST.

Let $\mathscr{P}_\lambda$ be a homogeneous Poisson process of rate λ on $\mathbf{R}^\nu$. The proof of Lemma 3 is based on the fact that for large λ, the 1-cluster of 0 in $\mathscr{P}_\lambda$, if finite, is likely to be a singleton. This is a special case of Theorem 3 of Penrose [23].

LEMMA 4 ([23]). *Suppose $\nu \geq 2$. Then*

$$\lim_{\lambda\to\infty} \frac{P[\mathrm{card}(C_1(0; \mathscr{P}_\lambda)) < \infty]}{P[C_1(0; \mathscr{P}_\lambda) = \{0\}]} = 1.$$

PROOF OF LEMMA 3. Write C_i^n for the cluster $C_{r_n}(\eta_i; \mathscr{P}_n)$. By (12), $D_n(i, j)$ is contained in the event that C_i^n and C_j^n are distinct and are not singletons.

For any $S \subset B$, let $\mathrm{diam}(S)$ denote its diameter $\sup\{d(x, y): x, y \in S\}$. For $\rho > 0$, define events

$$E_n(\rho; i) = \{0 < \mathrm{diam}(C_i^n) < \rho r_n\} \tag{13}$$

and

$$F_n(\rho; i, j) = \{\mathrm{diam}(C_i^n) > \rho r_n\} \cap \{\mathrm{diam}(C_j^n) > \rho r_n\} \cap \{C_i^n \neq C_j^n\}. \tag{14}$$

Then

$$\bigcup_{i<j\le N_n} D_n(i,j) \subset \Big(\bigcup_{i\le N_n} E_n(\rho;i)\Big) \cup \Big(\bigcup_{i<j\le N_n} F_n(\rho;i,j)\Big). \tag{15}$$

Let $\mathscr{P}'_\lambda$ denote a homogeneous Poisson process on $\mathbf{R}^\nu$ of rate λ. By Palm theory for the Poisson process, spatial homogeneity of the torus and the scaling property of the Poisson process,

$$\begin{aligned} E\Big[\sum_{i\le N_n} \mathbf{1}\{0<\operatorname{diam}(C_i^n)<\rho r_n\}\Big] &= \int_B P[0<\operatorname{diam}(C_{r_n}(x;\mathscr{P}_n))<\rho r_n]n\,dx \\ &= nP[0<\operatorname{diam}(C_{r_n}(0;\mathscr{P}'_n))<\rho r_n] \\ &= nP[0<\operatorname{diam}(C_1(0;\mathscr{P}'_{nr_n^\nu}))<\rho]. \end{aligned} \tag{16}$$

Since $nr_n^\nu \to \infty$ as $n\to\infty$, it follows from Lemma 4 that

$$\lim_{n\to\infty} \frac{P[0<\operatorname{diam}(C_1(0;\mathscr{P}'_{nr_n^\nu}))<\rho]}{P[C_1(0;\mathscr{P}'_{nr_n^\nu})=\{0\}]} = 0. \tag{17}$$

By the definition of r_n, the denominator $P[C_1(0;\mathscr{P}'_{nr_n^\nu}) = \{0\}]$ is equal to $\exp(-\pi_\nu nr_n^\nu) = n^{-1}e^{-\alpha}$, so that the expression (16) converges to zero, and for any fixed $\rho>0$,

$$\lim_{n\to\infty} P\Big[\bigcup_{i\le N_n} E_n(\rho;i)\Big] = 0. \tag{18}$$

The proof of Lemma 3 is completed by applying the following result, along with (15) and (18).

LEMMA 5. *Let $F_n(\rho;i,j)$ be defined by* (14). *Then there exists $\rho\in(0,\infty)$ such that for the toroidal model*,

$$\lim_{n\to\infty} P\Big[\bigcup_{i<j\le N_n} F_n(\rho;i,j)\Big] = 0.$$

PROOF. We modify the proof of Lemma 2 of [23] to take into account the fact that we work on a finite region.

For $a>0$ and x in the torus B, let $B_a(x)$ be the closed ν-dimensional cube of side a centered at x, "wrapped around" toroidally so that $B_a(x)\subset B$.

Take $\delta=\delta(n)\in((9\nu)^{-1},(8\nu)^{-1})$, such that $1/(2\delta r_n)$ is an integer (this is possible for large n). Let T_n^ν denote the lattice torus $\mathbf{Z}^\nu\cap[-1/(2\delta r_n), 1/(2\delta r_n)]$ with opposite faces identified, made into a graph by connecting nearest-neighbor pairs as for the usual integer lattice.

Suppose $F_n(\rho;i,j)$ occurs. We construct a "path" (or "surface" if $\nu>2$) of boxes of side δr_n, which separates C_i^n from C_j^n, and which must be devoid of points of $\mathscr{P}_n$. Let W_i denote the union of the balls of radius $3r_n/4$ centered at the points of C_i^n; this set is connected. Let U_i denote the set of $z\in T_n^\nu$ such that $B_{\delta r_n}(\delta r_n z)$ has nonempty intersection with W_i; this is a connected subset of T_n^ν. Let $\partial_j U_i$ denote the exterior external boundary of U_i, that is, the set of $z \in T_n^\nu \backslash U_i$ such that z has a neighbor in U_i and such that $\delta r_n z$ and η_j lie in the same connected component of $B\backslash W_i$.

For each $z \in \partial_j U_i$, the cube $B_{\delta r_n}(\delta r_n z)$ lies near the boundary of W_i and by an application of the triangle inequality cannot contain any point of $\mathscr{P}_n$. Since $\delta r_n \partial_j U_i$ is exterior both to W_i and to W_j, it has diameter at least ρr_n; therefore $\mathrm{card}(\partial U_i) \geq \rho/\delta$. Finally, $\partial_j U_i$ is $*$-connected. (A set $A \subset \mathbf{Z}^\nu$ is said to be $*$-connected of for each z, $z' \in A$, there is a finite path (z_n) in A from z to z', with $\|z_n - z_{n-1}\|_\infty = 1$ for each z_n in the path; the modification from $\mathbf{Z}^\nu$ to the torus T_n^ν should be clear.) See, for example, Lemma 2.1 of [11].

Let $\mathscr{A}_{n,m}$ denote the set of $*$-connected sets $A \subset T_n^\nu$ of cardinality m. By the remarks in the previous paragraph,

$$
\begin{aligned}
P\Bigg[\bigcup_{i<j\leq N_n} F_n(\rho; i, j)\Bigg] &\leq \sum_{m \geq \rho/\delta} P\Bigg[\exists\, A \in \mathscr{A}_{n,m}; \mathscr{P}_n\Bigg(\bigcup_{z\in A} B_{\delta r_n}(\delta r_n z)\Bigg) = 0\Bigg] \\
&\leq \sum_{m\geq\rho/\delta} \mathrm{card}(\mathscr{A}_{n,m}) \exp(-mn\delta^\nu r_n^\nu).
\end{aligned}
\tag{19}
$$

By a Peierls argument (see [22], Lemma 3) there is a constant $\gamma = \gamma(\nu)$, such that the number of $*$-connected sets ("lattice animals") of cardinality m in T_n^ν containing the origin is bounded above by $e^{\gamma m}$, for all n, m. Therefore

$$
\mathrm{card}(\mathscr{A}_{n,m}) \leq (\delta r_n)^{-\nu} \exp(\gamma m).
$$

Also, if n is large, then $n\delta^d r_n^\nu \geq (\delta^\nu/2\pi_\nu)\log n$ and $\gamma < (\delta^d/4\pi_\nu)\log n$, so that

$$
\begin{aligned}
P\Bigg[\bigcup_{i<j\leq N_n} F_n(\rho; i, j)\Bigg] &\leq (\delta r_n)^{-\nu} \sum_{m\geq\rho/\delta} \exp((\gamma - n\delta^\nu r_n^\nu)m) \\
&\leq (\delta r_n)^{-\nu} \sum_{m\geq\rho/\delta} \exp\{-((\delta^\nu/4\pi_\nu)\log n)m\} \\
&\leq c(n/\log n)\exp(-((\delta^\nu/4\pi_\nu)\log n)\rho/\delta) \leq c' n^{1-\rho\delta^{\nu-1}/4\pi_\nu},
\end{aligned}
$$

where c, c' are positive constants. If the (fixed) value of ρ is suitably big, this converges to zero. □

PROPOSITION 1. *Let $\alpha \in \mathbf{R}$, and let $r_n = r_n(\alpha)$ be given by* (10). *Then for the toroidal model*,

$$
\lim_{n\to\infty} P[d(i, j) \geq r_n \text{ for some edge } (i, j) \in MST(\mathscr{P}_n)\backslash NNG(\mathscr{P}_n)] = 0. \tag{20}
$$

Moreover, with probability approaching 1 *as $n \to \infty$, every edge of the MST with length greater than r_n has one end at a leaf.*

PROOF. Let $R_{i,2}(\mathscr{P}_n)$ denote the distance from η_i to its second-nearest neighbor in $\mathscr{P}_n$. Then for any $\alpha < \beta$, setting $r_n = r_n(\alpha)$ and $s_n = r_n(\beta)$, and writing $U_r(x)$ for the ball of radius r centered at $x \in \mathbf{R}^\nu$,

$$
\begin{aligned}
&E[\operatorname{card}\{i \le N_n\colon r_n(\alpha) \le R_i(\mathscr{P}_n) \le R_{i,2}(\mathscr{P}_n) < r_n(\beta)\}] \\
&\quad = nP[\mathscr{P}_n(U_{r_n}(0)) = 0;\ \mathscr{P}_n(U_{s_n}(0)) \ge 2] \\
&\quad = n\exp(-n\pi_\nu r_n^\nu)(1 - \exp(-n\pi_\nu(s_n^\nu - r_n^\nu))(1 + n\pi_\nu(s_n^\nu - r_n^\nu))) \\
&\quad = n\exp(-n\pi_\nu r_n^\nu) - n\exp(-n\pi_\nu s_n^\nu)(1 + n\pi_\nu(s_n^\nu - r_n^\nu)) \\
&\quad = e^{-\alpha} - e^{-\beta}(1 + (\beta + \log n) - (\alpha + \log n)) \\
&\quad = e^{-\beta}(e^{\beta-\alpha} - 1 - (\beta - \alpha)) \qquad (21)\\
&\quad \le (\beta - \alpha)^2 e^{-\alpha}/2, \qquad (22)
\end{aligned}
$$

where the last inequality is from Taylor's theorem. If $\varepsilon > 0$ and $\alpha \in \mathbf{R}$, we can take $\alpha = \alpha_1 < \alpha_2 < \cdots < \alpha_K$, such that $\exp(-\alpha_K) < \varepsilon$ and such that

$$
\sum_{k=1}^{K-1} (\alpha_{k+1} - \alpha_k)^2 \exp(-\alpha_k)/2 < \varepsilon. \qquad (23)
$$

By (22) and (23), writing R_i for $R_i(\mathscr{P}_n)$, we have

$$
P\Bigg[\bigcup_{1\le k\le K}\ \bigcup_{i<j\le N_n} \{r_n(\alpha_k) \le \max(R_i, R_j) < d(i,j) < r_n(\alpha_{k+1})\}\Bigg] < 2\varepsilon. \qquad (24)
$$

Also, by (1), for large enough n,

$$
P\Bigg[\bigcup_{i\le N_n} \{R_i \ge \alpha_K\}\Bigg] < \exp(-\alpha_K) + \varepsilon < 2\varepsilon. \qquad (25)
$$

Third, by Lemma 3,

$$
\begin{aligned}
\lim_{n\to\infty} P\Bigg[\bigcup_{1\le k\le K}\ \bigcup_{i<j\le N_n} &\{\max(R_i, R_j) < r_n(\alpha_k) \le d(i,j); \\
&\qquad (i,j) \in MST(\mathscr{P}_n)\}\Bigg] = 0. \qquad (26)
\end{aligned}
$$

If $d(i,j) \ge r_n(\alpha)$ for some edge (i,j) of the MST on $\mathscr{P}_n$ that is not in the NNG, so that $d(i,j) > \max(R_i, R_j)$, then one of the three events described in (24), (25) and (26) must occur. So by combining these three estimates, we obtain (20), since ε is arbitrary.

We now prove the final sentence, that every α-long long edge of the MST is likely to end at a leaf. By the above, we may assume that all such edges are in the NNG, so that if (i,l) is in the MST with $d(i,l) \ge r_n$, then $d(i,l) = R_i$ or $d(i,l) = R_l$. Assuming the former, i could fail to be a leaf only if it were the nearest neighbor of some $j \ne l$, and therefore it now suffices to prove

$$
\lim_{n\to\infty} P\Bigg[\bigcup_{i\le N_n}\ \bigcup_{i\ne j\le N_n} \{r_n \le R_i < R_j = d(i,j)\}\Bigg] = 0. \qquad (27)
$$

By Palm theory for the Poisson process,

$$P\Big[\bigcup_{i\le N_n}\{R_i > 3r_n\}\Big] \le n\exp(-n\pi_\nu(3r_n)^\nu) \to 0.$$

Also, by the calculation in (33) below,

$$\lim_{n\to\infty} P\Big[\bigcup_{i<j\le N_n}\{R_i > r_n\}\cap\{R_j > r_n\}\cap\{d(i,j) < 3r_n\}\Big] = 0,$$

and these together yield (27).

4. The NNG on the torus. In this section we consider the NNG in the the Poisson toroidal model. Before proving the main point process limit (Proposition 5), we give a new proof of the basic formula (1). Let $\alpha \in \mathbf{R}$, and define $r_n = r_n(\alpha)$ by (10) above; that is, $n\pi_\nu r_n^\nu = \log n + \alpha$.

Partition the box $B = (-0.5, 0.5]^\nu$ into m^ν disjoint boxes of side m^{-1}, labelled $B_1, B_2, \dots, B_{m^\nu}$ and centered at $a_1, \dots, a_{m^\nu}$, respectively. Define the variable X_i to be the indicator of the event that there is a single point of $\mathscr{P}_n$ in B_i, that is, $\mathscr{P}_n(B_i) = 1$, and that $\mathscr{P}_n(B_j) = 0$ for all j with $0 < d(a_j, a_i) < r_n$. Define $p_i = E[X_i]$ and $p_{ij} = E[X_i X_j]$. Writing $a \sim_m b$ if $a/b \to 1$ as $m \to \infty$ (with n fixed), we have $p_i \sim_m (n/m^\nu)\exp(-n\pi_\nu r_n^\nu)$.

Let $v(r_n; r)$ be the volume of the union of two balls of radius r_n, with centers a distance r apart. Then $p_{ij} = 0$ if $d(a_i, a_j) < r_n$, and

$$p_{ij} \sim_m (n/m^\nu)^2 \exp(-nv(r_n; d(a_i, a_j))) \quad \text{on } r_n < d(a_i, a_j).$$

Define

$$Y_n^m = \sum_{i=1}^{m^\nu} X_i; \qquad Y_n = \lim_{m\to\infty} Y_n^m. \tag{28}$$

Then Y_n is the number of i for which $R_i(\mathscr{P}_n) > r_n$, where $R_i(\mathscr{P}_n)$ is the distance from η_i to its nearest neighbor in $\mathscr{P}_n$, and

$$E[Y_n] = \lim_{m\to\infty} E[Y_n^m] = n\exp(-n\pi_\nu r_n^\nu) = e^{-\alpha}. \tag{29}$$

To use the Chen–Stein method, as given in [5], we define a "neighborhood of influence" $\mathscr{N}_i$ for each $i \le m^\nu$ by

$$\mathscr{N}_i = \{j\colon d(a_i, a_j) \le 3r_n\} \tag{30}$$

and define the quantities

$$b_1 = \sum_i \sum_{j\in\mathscr{N}_i} p_i p_j, \qquad b_2 = \sum_i \sum_{i\ne j\in\mathscr{N}_i} p_{ij}. \tag{31}$$

Then

$$\lim_{m\to\infty} b_1 = (n\exp(-n\pi_\nu r_n^\nu))^2 \pi_\nu(3r_n)^\nu = e^{-2\alpha}\pi_\nu(3r_n)^\nu, \tag{32}$$

which converges to 0 as $n \to \infty$. Also,

$$
\begin{aligned}
\lim_{m\to\infty} b_2 &= n^2 \int_{r_n \le |x| \le 3r_n} \exp(-nv(r_n; |x|))\, dx \\
&\le n^2 \pi_\nu (3r_n)^\nu \exp(-(3/2)n\pi_\nu r_n^\nu) \\
&= 3^\nu n^2((\log n + \alpha)/n)(e^{-\alpha}/n)^{3/2},
\end{aligned} \tag{33}
$$

which converges to 0 as $n \to \infty$. Since X_i is independent of X_j for $j \notin \mathscr{N}_i$, it follows from Theorem 1 of [5] that the total variation distance between the distribution of Y_n and the Poisson with mean $e^{-\alpha}$ is at most $2\lim_{m\to\infty}(b_1 + b_2)$, and therefore Y_n converges in distribution to that Poisson distribution. Therefore if M_n is the maximum edge-length for the NNG on the Poisson toroidal model,

$$\lim_{n\to\infty} P[M_n \le r_n] = \lim_{n\to\infty} P[Y_n = 0] = \exp(-e^{-\alpha}).$$

In view of the definition (10) of r_n, this gives us (1).

We now prove (5) for the torus. Let $\mathscr{G}_n'$ denote the point process $\mathscr{G}_n(\mathscr{P}_n)$; that is, $\mathscr{G}_n' = \{(n\pi_\nu(R_i(\mathscr{P}_n))^\nu - \log n, \eta_i),\ 1 \le i \le N_n\}$.

PROPOSITION 2. *For the toroidal model,* $\mathscr{G}_n' \to_d \mathscr{P}_\infty$.

PROOF. Let K be a fixed positive integer, and let $S_1, \dots, S_K$ be disjoint subsets of $\mathbf{R} \times B$, with each S_i in the semiring $\mathscr{S}$. For $1 \le k \le K$, write $S_k = A_k \times [\alpha_k, \beta_k)$, with $A_k \subset B$ a ν-fold product of intervals. By Lemma 1, it suffices to prove that the K-dimensional random vector $(\mathscr{G}_n'(S_1), \dots, \mathscr{G}_n'(S_K))$ converges in distribution to $(Z_1, \dots, Z_K)$, where $Z_1, \dots, Z_K$ are independent Poisson variables with $E[Z_k] = (\exp(-\alpha_k) - \exp(-\beta_k))|A_k|$ for $1 \le k \le K$, with $|\cdot|$ denoting volume.

Divide B into cubes $B_1, \dots, B_{m^\nu}$ with B_i centered at a_i as before. Define

$$R_i^m := \min\{d(a_i, a_j): j \ne i,\ \mathscr{P}_n(B_j) \ge 1\}.$$

Let X_i^k be the indicator variable of the event $\{\mathscr{P}_n(B_i) = 1\} \cap \{a_i \in A_k\} \cap \{R_i^m \in [r_n(\alpha_k), r_n(\beta_k))\}$. Here $r_n(\alpha_k)$ and $r_n(\beta_k)$ are given by (10). That is, $n\pi_\nu(r_n(t))^\nu = \log n + t$.

Let $\alpha = \min(\alpha_1, \dots, \alpha_K)$. Let X_i be the indicator of the event $\{\mathscr{P}_n(B_i) = 1\} \cap \{R_i^m \ge r_n(\alpha)\}$. Since the regions $S_1, \dots, S_K$ are pairwise disjoint,

$$\sum_{k=1}^K X_i^k \le X_i \quad \text{a.s.} \tag{34}$$

Define $p_i^k = EX_i^k$ and $p_i = EX_i$. Also, define $p_{ij} = EX_{ij}$ and $p_{ij}^{kl} = E[X_i^k X_j^l]$. Define

$$Y_{n,k}^m = \sum_{i=1}^{m^\nu} X_i^k; \qquad Y_{n,k} = \lim_{m\to\infty} Y_{n,k}^m = \mathscr{G}_n'(S_k). \tag{35}$$

Let $\gamma_k = \beta_k$ if $\beta_k < \infty$ and $\gamma_k = \alpha_k$ if $\beta_k = \infty$. Then X_i^k is determined by the outcomes of $\mathscr{P}_n$ in those B_j with $|a_i - a_j| \leq r_n(\gamma_k)$. Set $\beta = \max(\gamma_1, \dots, \gamma_k)$.

Define $\mathscr{N}_i^k$ to be the set of (j, l) $(1 \leq j \leq m^2,\ 1 \leq l \leq K)$ such that $d(a_i, a_j) \leq 3r_n(\beta)$, so that X_i^k is independent of X_j^l for $(j, l) \notin \mathscr{N}_i^k$. Define the quantities

$$(36) \qquad b_1' = \sum_{(i,k)} \sum_{(j,l)\in\mathscr{N}_i^k} p_i^k p_j^l; \qquad b_2' = \sum_{(i,k)} \sum_{(i,k)\neq(j,l)\in\mathscr{N}_i^k} p_{ij}^{kl}.$$

Set $\mathscr{N}_i' = \{j: d(a_i, a_j) \leq 3r_n(\beta)\}$. By (34), $b_1' \leq b_1$ and $b_2' \leq b_2$, where b_1 and b_2 are as defined in (31), except that the sums are now over $\mathscr{N}_i'$. Therefore by a similar argument to (32) and (33), we obtain

$$(37) \qquad \lim_{n\to\infty} \limsup_{m\to\infty} b_1' = \lim_{n\to\infty} \limsup_{m\to\infty} b_2' = 0.$$

Also, $\lim_{m\to\infty} E[Y_{n,k}^m] = E[Z_k]$. By Theorem 2 of [5], the total variation distance between the distributions of $(Y_{n,1}, \dots, Y_{n,K})$ and $(Z_1, \dots, Z_K)$ are bounded by $4 \limsup_{m\to\infty}(b_1' + b_2')$. By (37), $(Y_{n,1}, \dots, Y_{n,K})$ converges in distribution to $(Z_1, \dots, Z_K)$ as $n \to \infty$. □

5. Boundary effects. We now drop the toroidal assumption for $\nu = 2$ (for $\nu \geq 3$, the resulting boundary effects dominate). First we look at the NNG.

PROPOSITION 3. *For the Euclidean model with* $\nu \leq 2$, $\mathscr{G}_n(\mathscr{P}_n) \to_d \mathscr{P}_\infty$ *as* $n \to \infty$.

PROOF. Let $\alpha > 0$ and let r_n be given by (10). Let Y_n^E denote the number of points of $\mathscr{P}_n$ whose nearest neighbor (in the Euclidean metric) is within a distance greater than r_n. The correction to the mean due to boundary effects is

$$(38) \qquad E[Y_n^E] - E[Y_n] = \int_{\{x\in B:\ d(x,\partial B)\leq r_n\}} n \exp(-n|U_{r_n}(x) \cap B|)\, dx + o(1),$$

where $U_r(x)$ denotes the r-neighborhood of x in B, and $|\cdot|$ denotes Lebesgue measure.

Let I_n be the contribution to the integral in (38) from values of $x = (x_1, \dots, x_\nu)$ with $|x_i - (1/2)| \leq r_n$ for just a single value of i, that is, x close to just one face of B. Then

$$(39) \qquad I_n = 2\nu(1 + o(1)) \int_0^{r_n} n \exp\{-n|U_{r_n}((-(1/2) + t, 0, \dots, 0)) \cap B|\}\, dt.$$

Let $g(r; t)$ denote the volume of the intersection of the ν-dimensional unit ball $U_r(0)$ with the slab $(0, t) \times \mathbf{R}^{\nu-1}$. Then

$$|U_{r_n}(-(1/2) + t, 0, \dots, 0) \cap B| = (\pi_\nu r_n^\nu/2) + g(r_n; t) = (\pi_\nu r_n^\nu/2) + r_n^\nu g(1; t/r_n),$$

so that by the change of variable $u = t/r_n$,

$$(40) \qquad I_n = 2\nu(1 + o(1)) n \exp(-n\pi_\nu r_n^\nu/2) \int_0^1 \exp(-n r_n^\nu g(1; u)) r_n\, du.$$

Since $nr_n^\nu \to \infty$ as $n \to \infty$, and $\int_0^1 \exp(-\theta g(1;u))\,du \sim (\pi_{\nu-1}\theta)^{-1}$ as $\theta \to \infty$,

$$I_n \sim 2\nu n r_n (e^{-\alpha}/n)^{1/2}/(\pi_{\nu-1} n r_n^\nu) = c(n^{1/2} r_n)(nr_n^\nu)^{-1}, \tag{41}$$

where c is a constant. Since $nr_n^\nu \to \infty$ logarithmically, $I_n \to 0$ for $\nu \le 2$.

For $\nu = 2$, the contribution to the integral in (38) from sites $x = (x_1, x_2)$ near the corners, that is, with $|x_i - (1/2)| \le r_n$ for $i = 1, 2$, is at most $4r_n^2 n(e^{-\alpha}/n)^{1/4}$, which converges to zero. Therefore $\lim_{n\to\infty} E[Y_n^E] = \exp(-\alpha)$ for $\nu \le 2$.

To show that the Chen–Stein method still gives Poisson limits, we need to check that the boundary contributions to the quantities b_1 and b_2 of (31) are negligible. The contribution to b_1 from regions near the edge but not near the corner is bounded by the expression

$$cr_n r_n^2 (n \exp(-n\pi r_n^2/2))^2 = c' r_n^3 n^{2-1}$$

which converges to zero.

For any pair (x, y) with $x, y \in B$, with x close to the left edge of B but not close to the corner, with $r_n \le |x - y| \le 3r_n$ and with x closer to the left edge of B than y, there exist a half-disk centered at x and a disjoint quarter-disk centered at y, both contained in B. Therefore the contribution to b_2 from regions near the edge but not near the corner of B is bounded by

$$cr_n r_n^2 n^2 \exp(-3n\pi r_n^2/4) = c' r_n^3 n^{2-(3/4)},$$

which converges to zero.

The contributions both to b_1 and to b_2 from regions near the corner are bounded by

$$c(r_n^2)^2 n^2 \exp(-n\pi r_n^2/4) = c' r_n^4 n^{2-(1/4)},$$

which also converges to zero. Therefore for $\nu = 2$, the arguments from Section 4 carry over to the Euclidean model, and so the statement (5) from Theorem 2 is also valid for the Euclidean model. □

Turning to the MST, we prove that the results of Section 3 carry over from the toroidal to the Euclidean model for $\nu = 2$.

PROPOSITION 4. *Let $\nu = 2$. Let $\alpha \in \mathbf{R}$, and let $r_n = r_n(\alpha)$ be given by* (10). *Then for the Euclidean model*,

$$\lim_{n\to\infty} P[d(i, j) > r_n \text{ for some } (i, j) \in MST(\mathscr{P}_n)\backslash NNG(\mathscr{P}_n)] = 0. \tag{42}$$

Also, with probability approaching 1, *every edge of the MST with length greater than r_n has one end at a leaf.*

To prove this, we shall require some analogous results to Lemma 4 for percolation on the half-space and quarter-space. Let $\mathbf{H}$ denote the half-space $[0, \infty) \times \mathbf{R}$, and let $\mathbf{Q}$ denote the quarter-space $[0, \infty) \times [0, \infty)$. Let $\mathscr{P}_\lambda^H$ (respectively, $\mathscr{P}_\lambda^Q$) denote the Poisson process of rate λ on $\mathbf{H}$ (respectively, $\mathbf{Q}$).

For $x \in \mathbf{R}^2$ and any set $\mathscr{A} \subset \mathbf{R}^2$, we write $L_r(x;\mathscr{A})$ for the event that x is the left-most point of $C_r(x;\mathscr{A})$, that is, the first coordinate of x is less than the first coordinate of any other point of $C_r(x;\mathscr{A})$.

LEMMA 6. *For any* $\rho > 0$,

$$\lim_{\lambda\to\infty} \sup_{x\in\mathbf{H}} \frac{P[0 < \operatorname{diam}(C_1(x;\mathscr{P}_\lambda^H)) < \rho; L_1(x;\mathscr{P}_\lambda^H)]}{P[C_1(x;\mathscr{P}_\lambda^H) = \{x\}]} = 0.$$

LEMMA 7. *For any* $\rho > 0$ *and any* $\varepsilon > 0$,

$$\lim_{\lambda\to\infty} \sup_{x\in\mathbf{Q}} \bigl(\exp\{\lambda((1/4)-\varepsilon)\}P[\operatorname{diam}(C_1(x;\mathscr{P}_\lambda^Q)) < \rho]\bigr) = 0.$$

We do not give detailed proofs of these results here. Lemma 6 can be proved by a similar argument to Lemmas 1 and 3 of [23]. Lemma 7 can be proved by a cruder version of the argument yielding Lemma 3 of [23]. Here the connection function $g(\cdot)$ of [23] is simply the indicator function of the unit circle, which simplifies the arguments somewhat.

LEMMA 8. *Let* $\nu = 2$, *let* $\alpha \in \mathbf{R}$ *and let* $r_n = r_n(\alpha)$ *be given by* (10). *For the Euclidean model, let* $D_n^E(i,j)$ *be the event that* (i,j) *is an edge of the MST on* $\mathscr{P}_n$, *and that* $d(i,j) \geq r_n$, *but* $R_i(\mathscr{P}_n) < r_n$ *and* $R_j(\mathscr{P}_n) < r_n$. *Then* $\lim_{n\to\infty} P[\bigcup_{i<j\leq N_n} D_n^E(i,j)] = 0$.

PROOF. For $\rho > 0$, let $F_n^E(\rho;i,j)$ denote the event that the (Euclidean) clusters $C_{r_n}(\eta_i;\mathscr{P}_n)$ and $C_{r_n}(\eta_j;\mathscr{P}_n)$ are distinct, and both of diameter at least ρr_n. For $x \in B$, let $G_n(\rho;x)$ denote the event that (i) $0 < \operatorname{diam}(C_{r_n}(x;\mathscr{P}_n)) < \rho r_n$, and (ii) x is the closest point to ∂B in $C_{r_n}(x;\mathscr{P}_n)$. Then for any $\rho > 0$,

$$\bigcup_{i<j\leq N_n} D_{ij}^E \subset \biggl(\bigcup_{i<j\leq N_n} F_n^E(\rho;i,j)\biggr) \cup \biggl(\bigcup_{i\leq N_n} G_n(\rho;\eta_i)\biggr). \tag{43}$$

The proof of Lemma 5 also works in the Euclidean setting; therefore we can take $\rho > 0$ such that

$$\lim_{n\to\infty} P\biggl[\bigcup_{i<j\leq N_n} F_n^E(\rho;i,j)\biggr] = 0. \tag{44}$$

Also, by Palm theory for the Poisson process,

$$E\biggl[\sum_{i\leq N_n} \mathbf{1}(G_n(\rho;\eta_i))\biggr] = n\int_B P[G_n(\rho;x)]\,dx. \tag{45}$$

We partition $B = [-1/2, 1/2]^2$ into three regions; a central region

$$B_n^1 = [-(1/2) + 2\rho r_n, (1/2) - 2\rho r_n]^2,$$

a corner region

$$B_n^2 = \{(x_1,x_2) \in B\colon |x_i - (1/2)| < 2\rho r_n,\ i = 1,2\},$$

and an edge region $B_n^3 = B \setminus (B_n^1 \cup B_n^2)$. By the proof of (18),

$$n \int_{B_n^1} P[G_n(\rho; x)]\, dx \to 0 \quad \text{as } n \to \infty.$$

Turning to the edge region, let $x = (x_1, x_2) \in B_n^3$ with $x_1 = -(1/2) + tr_n$ and $0 < t < \rho$. Then, setting $\mathbf{e} = (1, 0) \in \mathbf{R}^2$, we have

$$\begin{aligned} P[G_n(\rho; x)] &= P[0 < \operatorname{diam}(C_{r_n}(tr_n\mathbf{e}; \mathscr{P}_n^H)) < \rho r_n; L_{r_n}(tr_n\mathbf{e}; \mathscr{P}_n^H)] \\ &= P[0 < \operatorname{diam}(C_1(t\mathbf{e}; \mathscr{P}_{nr_n^2}^H)) < \rho; L_1(t\mathbf{e}; \mathscr{P}_{nr_n^2}^H)] \\ &\le P[C_1(t\mathbf{e}; \mathscr{P}_{nr_n^2}^H) = \{t\mathbf{e}\}] \\ &= P[C_{r_n}(x; \mathscr{P}_n)) = \{x\}], \end{aligned} \tag{46}$$

where the inequality holds uniformly in t, for large enough n, by Lemma 6. Therefore

$$n \int_{B_n^3} P[G_n(\rho; x)]\, dx \le n \int_{B_n^3} P[C_{r_n}(x; \mathscr{P}_n) = \{x\}], \tag{47}$$

which converges to 0 by the proof of Proposition 3.

Finally we deal with the integral over the corner region. By a similar argument to (46), using Lemma 7, we have for large n that

$$P[\operatorname{diam}(C_{r_n}(x; \mathscr{P}_n)) \le \rho r_n] \le \exp(-(1/8) n\pi r_n^2), \qquad x \in B_n^2.$$

Therefore for some constant c, the contribution to the integral in (45) from B_n^2 is bounded by $cnr_n^2 n^{-1/8}$, which converges to zero. Therefore the expression in (45) converges to 0. This, together with (44) and (43), gives us the result. □

PROOF OF PROPOSITION 4. For the Euclidean model, (21) holds up to a boundary correction which is $o(1)$ as $n \to \infty$. Therefore the proof in Section 3 of Proposition 1 can be adapted to the Euclidean model, using Lemma 8, to give us the result.

6. Non-Poisson models.

We now extend our results from $\mathscr{P}_n$ to $\mathscr{X}_n$. We do this by considering a Poisson process of slightly smaller intensity that is dominated by $\mathscr{X}_n$ with high probability.

LEMMA 9. *Define the function n^- of n by $n^- = n - n^{3/4}$. Then for the toroidal model with $\nu \ge 2$ or the Euclidean model with $\nu = 2$,*

$$\mathscr{G}_n(\mathscr{P}_{n^-}) \to_d \mathscr{P}_\infty \quad \textit{as } n \to \infty. \tag{48}$$

PROOF. Define the function h_n: $\mathbf{R} \to \mathbf{R}$ in such a way that $h_n(n^- \pi_\nu r^\nu - \log n^-) = n\pi r^\nu - \log n$; that is, define

$$h_n(t) = (t + \log(n^-))(n/n^-) - \log n.$$

Then $\mathscr{G}_n(\mathscr{P}_{n^-})$ is the image of $\mathscr{G}_{n^-}(\mathscr{P}_{n^-})$ under the mapping $(t, \mathbf{x}) \mapsto (h_n(t), \mathbf{x})$. Since $\mathscr{G}_{n^-}(\mathscr{P}_{n^-}) \to_d \mathscr{P}_\infty$ by results already proved, and since $h_n(t) \to t$ as

$n \to \infty$, locally uniformly in t, it follows by condition (b) in our definition of weak convergence that (48) also holds. □

PROPOSITION 5. *For the toroidal model with* $\nu \geq 2$ *or the Euclidean model with* $\nu = 2$,

$$\mathscr{G}_n(\mathscr{X}_n) \to_\nu \mathscr{P}_\infty \quad \text{as } n \to \infty. \tag{49}$$

PROOF. With n^- as above, write N_n^- for N_{n^-}, and $\mathscr{P}_n^-$ for $\mathscr{P}_{n^-}$. Given $\alpha \in \mathbf{R}$, let $r_n = r_n(\alpha)$ be given by (10) as before. Define the sets

$$S_n^- = \{i \leq N_n^-: R_i(\mathscr{P}_n^-) \geq r_n\}; \qquad S_n^f = \{i \leq n: R_i(\mathscr{X}_n) \geq r_n\}. \tag{50}$$

The superscript f stands for "fixed," referring to the fact that $\mathscr{X}_n$ has a nonrandom number of points.

The point processes $\mathscr{X}_n$ and $\mathscr{P}_n^-$ are coupled, since $\mathscr{X}_n$ is obtained from $\mathscr{P}_n^-$ by adding $n - N_n^-$ points to $\mathscr{P}_n$, if $N_n^- \leq n$, or by removing $N_n^- - n$ points from $\mathscr{P}_n$ if $N_n^- > n$; the latter case is exceptional since by Chebyshev's inequality, $P[N_n^- > n] \to 0$. In view of Lemma 9, to prove (49) it suffices to prove that for any given α,

$$\lim_{n\to\infty} P[S_n^f \neq S_n^-] = 0. \tag{51}$$

If $S_n^f \setminus S_n^-$ is nonempty, and $N_n^- \leq n$, then some point of $\mathscr{X}_n \setminus \mathscr{P}_n^-$ is added in the vacant region V_n defined by

$$V_n = \{x \in B: d(x, \eta_i) \geq r_n \text{ for all } i \leq N_n^-\},$$

with volume denoted $|V_n|$. Therefore,

$$P\big[S_n^f \setminus S_n^- \neq \varnothing\big] \leq P[|N_n^- - EN_n^-| > n^{3/4}] + P\left[\bigcup_{i=1}^{2n^{3/4}} \{\eta_i' \in V_n\}\right],$$

where $\eta_1', \eta_2', \ldots$ are independent and uniform on B, representing added points. By Chebyshev and Fubini, for the toroidal model,

$$\begin{aligned} P\big[S_n^f \setminus S_n^- \neq \varnothing\big] &\leq \frac{\operatorname{Var}(N_n^-)}{(n^{3/4})^2} + 2n^{3/4} E|V_n| \\ &\leq n^{-1/2} + 2n^{3/4}\exp(-\pi_\nu r_n^\nu (n - n^{3/4})) \\ &= n^{-1/2} + 2n^{3/4}(e^{-\alpha} n^{-1}(1 + o(1))) \to 0. \end{aligned} \tag{52}$$

For the Euclidean model with $\nu = 2$, it can be checked that $n^{3/4}E[|V_n|]$ still tends to 0; the corrections for boundary effects are negligible, by a similar calculation to the proof of Proposition 3.

Let W_n denote the union of those balls of radius r_n centered at points of $\mathscr{P}_n^-$ but devoid of other such points. If $S_n^- \setminus S_n^f$ is nonempty, and $N_n^- \leq n$, then some point of $\mathscr{X}_n \setminus \mathscr{P}_n^-$ is added in the region W_n. Therefore, if Y_n^- denotes the number of points of $\mathscr{P}_n^-$ whose nearest neighbor in $\mathscr{P}_n^-$ is at a distance of more than r_n, we have for each k that

$$
\begin{aligned}
P\big[S_n^- \backslash S_n^f \neq \varnothing; Y_n^- = k\big] &\le P[|N_n^- - EN_n^-| > n^{3/4}] \\
&\quad + P\Bigg[\bigcup_{i=1}^{2n^{3/4}} \{\eta_i' \in W_n; Y_n^- = k\}\Bigg] \\
&\le n^{-1/2} + 2n^{3/4} k(\pi_\nu r_n^\nu).
\end{aligned} \tag{53}
$$

The bounds in (52) and (53) converge to zero. Since the sequence $\mathscr{L}(Y_n^-)$ is tight (in fact, weakly convergent), (51) follows. □

Turning to the MST, we now prove Theorem 1 for $\mathscr{X}_n$.

PROPOSITION 6. *Let $\alpha \in \mathbf{R}$, and let $r_n = r_n(\alpha)$ be given by* (10). *Then for the toroidal or Euclidean model*,

$$
\lim_{n\to\infty} P[d(i, j) \ge r_n \text{ for some } (i, j) \in MST(\mathscr{X}_n) \backslash NNG(\mathscr{X}_n)] = 0. \tag{54}
$$

Moreover, with probability approaching 1 *as $n \to \infty$, every edge of the MST on $\mathscr{X}_n$ with length greater than r_n has one end at a leaf.*

PROOF. We proceed as in Lemma 3 from Section 3. Let $D_n^f(i, j)$ be the event that (i, j) is an edge of the MST on $\mathscr{P}_n$, and that $d(i, j) \ge r_n$, but $R_i(\mathscr{X}_n) < r_n$ and $R_j(\mathscr{X}_n) < r_n$. We prove that

$$
\lim_{n\to\infty} P\Bigg[\bigcup_{i<j\le n} D_n^f(i, j)\Bigg] = 0. \tag{55}
$$

Let $E_n^-(\rho; i)$ denote the event that $0 < \operatorname{diam}(C_{r_n}(\eta_i; \mathscr{P}_n^-)) < \rho r_n$, and let $E_n^f(\rho; i)$ denote the event that $0 < \operatorname{diam}(C_{r_n}(\eta_i; \mathscr{X}_n)) < \rho r_n$.

Suppose that $N_n^- \le n$, and that $E_n^f(\rho; j)$ occurs for some $j \le n$, but $\bigcup_{i\le N_n^-} E_n^-(\rho; i)$ does not. Then, since its diameter is less than ρr_n, the intersection of $C_{r_n}(\eta_j; \mathscr{X}_n)$ with $\mathscr{P}_n^-$ is either empty or consists of isolated points. In the first case, $(\mathscr{X}_n \backslash \mathscr{P}_n^-) \cap V_n$ is nonempty; in the second case, $(\mathscr{X}_n \backslash \mathscr{P}_n^-) \cap W_n$ is nonempty, with V_n and W_n defined in the proof of Proposition 5 above. Therefore

$$
\begin{aligned}
P\Bigg[\bigcup_{i\le n} E_n^f(\rho; i)\Bigg] &\le P\Bigg[\bigcup_{i\le N_n^-} E_n^-(\rho; i)\Bigg] + P[|N_n^- - EN_n^-| > n^{3/4}] \\
&\quad + P\Bigg[\bigcup_{i\le 2n^{3/4}} \{\eta_i' \in V_n\}\Bigg] + P\Bigg[\bigcup_{i\le 2n^{3/4}} \{\eta_i' \in W_n\}\Bigg].
\end{aligned} \tag{56}
$$

Suppose r_n^- is defined by (10) but using n^- instead of n. Since $r_n^- > r_n$ for large n, it follows from (18) in the toroidal case, or from the proof of Lemma 8 in the Euclidean case, that the first term in the right-hand side of (56) tends to zero. The other terms in (56) tend to zero by the estimates in (52) and (53).

Let $F_n^f(\rho; i, j)$ be the event that the clusters $C_{r_n}(\eta_i; \mathscr{X}_n)$ and $C_{r_n}(\eta_j; \mathscr{X}_n)$ are distinct, and are both of diameter greater than ρr_n. The proof of Lemma 5 also shows that $P[\bigcup_{i<j\le n} F_n^f(\rho; i, j)] \to 0$ for some ρ. Thus for suitable ρ,

$$\lim_{n\to\infty}\left(P\left[\bigcup_{i\le n} E_n^f(\rho; i)\right] + P\left[\bigcup_{i<j\le n} F_n^f(\rho; i, j)\right]\right) = 0,$$

which shows that (55) holds.

To complete the proof of (54), proceed as in the proof of Proposition 1 from Section 3 with $\mathscr{P}_n$ replaced by $\mathscr{X}_n$; the place of (21) in that proof is taken by

$$\lim_{n\to\infty} E[\mathrm{card}\{i \le n: \,: r_n(\alpha) \le R_i(\mathscr{X}_n) \le R_{i,2}(\mathscr{X}_n) < r_n(\beta)\}]$$
$$= e^{-\beta}(e^{\beta-\alpha} - 1 - (\beta - \alpha)),$$

which follows from a routine calculation for the multinomial distribution, which we omit.

The final sentence of Proposition 6 is verified by checking that (27) still holds with N_n replaced by n.

7. The k-NNG.

PROPOSITION 7. *For the toroidal model with $\nu \ge 1$ and $k \ge 0$, if M_n denotes the length of the longest edge of the $(k+1)$-NNG on $\mathscr{P}_n$ or $\mathscr{X}_n$, then* $\lim_{n\to\infty} P[n\pi_\nu M_n^\nu - \log n - k\log(\log n) + \log k! \le \alpha] = \exp(-e^{-\alpha})$. *More generally,* (7) *and* (8) *hold; that is, $\mathscr{G}_{n,k}(\mathscr{P}_n) \to_d \mathscr{P}_\infty$ and $\mathscr{G}_{n,k}(\mathscr{X}_n) \to_d \mathscr{P}_\infty$ as $n \to \infty$.*

PROOF. Let $\alpha > 0$. Define $s_n = s_n(\alpha, k)$ by $n\pi_\nu s_n^\nu = \log(n/k!) + k\log(\log n) + \alpha$, so that

$$\lim_{n\to\infty} n\exp(-n\pi_\nu s_n^\nu)(n\pi_\nu s_n^\nu)^k/k! = e^{-\alpha}. \tag{57}$$

Divide the torus B into disjoint boxes B_i centered at a_i, $1 \le i \le m^\nu$, as before. For this section, define X_i to be the indicator of the event that $\mathscr{P}_n(B_i) = 1$, and that $\mathrm{card}\{j: \mathscr{P}_n(B_j) > 0,\ 0 < d(a_i, a_j) < s_n\} = k$, and set

$$p_i = E[X_i] \sim_m (n/m^\nu)\exp(-n\pi_\nu s_n^\nu)(n\pi_\nu s_n^\nu)^k/k! \tag{58}$$

Also, set $p_{ij} = E[X_i X_j]$. Define $Y_n^m = \sum_{i=1}^{m^\nu} X_i$, and $Y_n = \lim_{m\to\infty} Y_n^m$. Thus, Y_n is the number of i for which $R_{i,k+1}(\mathscr{P}_n) > r_n > R_{i,k}(\mathscr{P}_n)$, and

$$E[Y_n] = \lim_{m\to\infty} E[Y_n^m] = n\exp(-n\pi_\nu s_n^\nu)(n\pi_\nu s_n^\nu)^k/k!, \tag{59}$$

so that $\lim_{n\to\infty} E[Y_n] = e^{-\alpha}$ by (57). Similarly,

$$\lim_{n\to\infty} E\mathrm{card}\{i \le N_n: R_{i,k}(\mathscr{P}_n) > s_n\} = 0,$$

so that the weak limit of Y_n is the same as the weak limit of $\mathrm{card}\{i \le N_n : R_{i,k+1}(\mathscr{P}_n) > r_n\}$.

As before, set $\mathcal{N}_i = \{j\colon d(a_i, a_j) \le 3r_n\}$, and set $b_1 = \sum_i \sum_{j\in\mathcal{N}_i} p_i p_j$ and $b_2 = \sum_i \sum_{i\ne j\in\mathcal{N}_i} p_{ij}$. Then

$$\lim_{m\to\infty} b_1 = (E[Y_n])^2 \pi_\nu (3s_n)^\nu,$$

which converges to 0 as $n\to\infty$.

Recall that $v(r;t)$ denotes the volume of the union of two balls of radius r, with centers a distance t apart. Let $v_1(r;t)$ denote the volume of the intersection of these two balls, and let $v_2(r;t) = (1/2)(v(r;t) - v_1(r;t))$ denote the volume that lies in the first ball but not the second. Then

$$\begin{aligned} \lim_{m\to\infty} b_2 = n^2 &\int_{|x|\le s_n} P[Z_1+Z_2 = Z_1+Z_3 = k-1]\,dx \\ &+ n^2\int_{s_n<|x|\le 3s_n} P[Z_1+Z_2=Z_1+Z_3=k]\,dx, \end{aligned} \tag{60}$$

where Z_1, Z_2, Z_3 are independent Poisson variables with mean $nv_1(s_n;|x|)$, $nv_2(s_n;|x|)$, $nv_2(s_n;|x|)$, respectively. The first term in the right-hand side of (60) is equal to

$$n^2\int_{t=0}^{3s_n}(\nu\pi_\nu t^{\nu-1}\,dt)\exp(-nv(s_n;t))\sum_{l=0}^{k-1}(nv_1(s_n;t))^{k-1-l}(nv_2(s_n;t))^{2l}.$$

Take the sum outside the integral and use the fact that $v(s_n;t) = \pi_\nu s_n^\nu + v_2(s_n;t)$ and $v_i(s_n;t) = s_n^\nu v_i(1;t/s_n)$. By the change of variable $u = t/s_n$ and then by the bound $v_1(1;u)\le \pi_\nu$, the lth term in the sum is

$$\begin{aligned} n^2 &\exp(-n\pi_\nu s_n^\nu)\int_{u=0}^{3}(\nu\pi_\nu s_n^\nu u^{\nu-1}du) \\ &\qquad\times \exp(-ns_n^\nu v_2(1;u))(ns_n^\nu v_1(1;u))^{k-1-l}(ns_n^\nu v_2(1;u))^{2l} \\ &\le \nu\pi_\nu^{k-l} n^2 s_n^\nu (ns_n^\nu)^{k+l-1}\exp(-n\pi_\nu s_n^\nu) \\ &\quad\times\int_0^3 \exp(-ns_n^\nu v_2(1;u))(v_2(1;u))^{2l}u^{\nu-1}\,du. \end{aligned} \tag{61}$$

Since $v_2(1;u)/u$ is bounded away from zero and from infinity on $0<u<3$ and since $\int_0^\infty e^{-\theta u}u^m\,du = \text{const.}\times\theta^{-(m+1)}$ as a function of θ, the expression in (61) is at most less than or equal to

$$c(ns_n^\nu)^{k+l-(2l+\nu)} n\exp(-n\pi_\nu s_n^\nu) \le c'(ns_n^\nu)^{-l-\nu}, \tag{62}$$

where the last inequality is from (57). This last bound converges to zero, and one can show similarly that the second sum in the right-hand side of (60) converges to zero. Therefore $\lim_{n\to\infty}\lim_{m\to\infty} b_2 = 0$, and by the result from [5], Y_n converges in distribution to a Poisson with mean $e^{-\alpha}$, giving us (9). The remainder of the proof of (7) is as spelled out in Section 4 for the special case $k = 0$. Likewise, (8) is proved by obvious modifications of the proof of Proposition 5 in Section 6.

REFERENCES

[1] ALDOUS, D. J. (1989). Stein's method in a two-dimensional coverage problem. *Statist. Probab. Lett.* **8** 307–314.

[2] ALEXANDER, K. S. (1995). Percolation and minimal spanning forests in infinite graphs. *Ann. Probab.* **23** 87–104.

[3] APPEL, M. J. B. and RUSSO, R. P. (1996). The connectivity of a graph on uniform points in $[0,1]^d$. Preprint, Univ. Iowa.

[4] APPEL, M. J. B. and RUSSO, R. P. (1996). The minimum vertex degree of a graph on uniform points in $[0,1]^d$. *Adv. in Appl. Probab.* To appear.

[5] ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.* **17** 9–25.

[6] AVRAM, F. and BERTSIMAS, D. (1993). On central limit theorems in geometrical probability. *Ann. Appl. Probab.* **3** 1033–1046.

[7] BEARDWOOD, J., HALTON, J. H. and HAMMERSLEY, J. (1959). The shortest path through many points. *Proc. Cambridge Philos. Soc.* **55** 299–327.

[8] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[9] CARONI, C. and PRESCOTT, P. (1995). On Rohlf's method for the detection of outliers in multivariate data. *J. Multivariate Anal.* **52** 295–307.

[10] DETTE, H. and HENZE, N. (1989). The limit distribution of the largest nearest-neighbour link in the unit d-cube. *J. Appl. Probab.* **26** 67–80.

[11] DEUSCHEL, J.-D. and PISZTORA, A. (1996). Surface order large deviations for high-density percolation. *Probab. Theory Related Fields* **104** 467–482.

[12] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

[13] HADI, A. S. and SIMONOFF, J. S. (1993). Procedures in the identification of multiple outliers in linear models. *J. Amer. Statist. Assoc.* **88** 1264–1272.

[14] HALL, P. (1985). Distribution of size, structure and number of vacant regions in a high-intensity mosaic. *Z. Wahrsch. Verw. Gebiete* **70** 237–265.

[15] HARTIGAN, J. A. (1975). *Clustering Algorithms.* Wiley, New York.

[16] HARTIGAN, J. A. (1981). Consistency of single linkage for high-density clusters. *J. Amer. Statist. Assoc.* **76** 388–394.

[17] HENZE, N. (1982). The limit distribution of the maxima of 'weighted' rth-nearest-neighbour distances. *J. Appl. Probab.* **19** 344–354.

[18] JAILLET, P. (1993). Cube versus torus models and the Euclidean minimal spanning tree constant. *Ann. Appl. Probab.* **3** 582–592.

[19] JAILLET, P. (1995). On properties of geometric random problems in the plane. *Annals of Operations Research* **61** 1–20.

[20] JANSON, S. (1986). Random coverings in several dimensions. *Acta Math.* **156** 83–118.

[21] KESTEN, H. and LEE, S. (1996). The central limit theorem for weighted minimal spanning trees on random points. *Ann. Appl. Probab.* **6** 495–527.

[22] KLARNER, D. A. (1967). Cell growth problems. *Canad. J. Math.* **19** 851–863.

[23] PENROSE, M. D. (1991). On a continuum percolation model. *Adv. in Appl. Probab.* **23** 536–556.

[24] PENROSE, M. D. (1996). Extremes for the minimal spanning tree on normally distributed points. Preprint.

[25] RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

[26] ROHLF, F. J. (1975). Generalization of the gap test for the detection of multivariate outliers. *Biometrics* **31** 93–101.

[27] SILVERMAN, B. and BROWN, T. (1978). Short distances, flat triangles and Poisson limits. *J. Appl. Probab.* **15** 815–825.

[28] STEELE, J. M. (1993). Probability and problems in Euclidean combinatorial optimization. *Statist. Sci.* **8** 48–56.

[29] STEELE, J. M., SHEPP, L. A. and EDDY, W. F. (1987). On the number of leaves of a Euclidean minimal spanning tree. *J. Appl. Probab.* **24** 809–826.

[30] STEELE, J. M. and TIERNEY, L. (1986). Boundary domination and the distribution of the largest nearest-neighbour link in higher dimensions. *J. Appl. Probab.* **23** 524–528.

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF DURHAM
SOUTH ROAD
DURHAM DH1 3LE
UNITED KINGDOM
E-MAIL: mathew.penrose@durham.ac.uk