Connectivity of a Gaussian Network

P. Balister
Department of Mathematical Sciences,
University of Memphis, Memphis TN 38152, USA
E-mail: pbalistr@msci.memphis.edu

B. Bollobás
Department of Mathematical Sciences,
University of Memphis, Memphis TN 38152, USA
E-mail: bollobas@msci.memphis.edu
and
Trinity College,
University of Cambridge, Cambridge CB2 1TQ, UK

A. Sarkar*
Department of Mathematical Sciences,
University of Memphis, Memphis TN 38152, USA
E-mail: asarkar@memphis.edu
* Corresponding author

M. Walters
Peterhouse,
University of Cambridge, Cambridge, CB2 1RD, UK
E-mail: mjw1009@cam.ac.uk

Abstract: Following Etherington, Hoge and Parkes, we consider a network consisting of (approximately) N transceivers in the plane $\mathbb{R}^2$ distributed randomly with density given by a Gaussian distribution about the origin, and assume each transceiver can communicate with all other transceivers within distance s. We give bounds for the distance from the origin to the furthest transceiver connected to the origin, and that of the closest transceiver that is not connected to the origin.

Keywords: Wireless sensor network; Gaussian distribution; Transceiver; Gilbert model; Continuum percolation.

Biographical notes: Béla Bollobás currently holds the Jabie Hardin Chair of Excellence in Graph Theory and Combinatorics at the University of Memphis and is a Fellow of Trinity College, Cambridge, England. He received PhDs from Eotvos University, Budapest, Hungary and the University of Cambridge, England. He works in extremal and probabilistic combinatorics, polynomials of graphs, combinatorial probability and percolation theory. He has written eight books, including Extremal Graph Theory, Random Graphs and Modern Graph Theory, and over 300 research papers. He has supervised over forty doctoral students, many of whom hold leading positions in prominent universities throughout the world.
Paul Balister is a Professor at the University of Memphis, where he has been since 1998. He obtained his PhD from the University of Cambridge, England in 1992 and subsequently held positions at Harvard and Cambridge. He has worked in algebraic number theory, combinatorics, graph theory and probability theory.
Amites Sarkar is a Visiting Assistant Professor at the University of Memphis. He obtained his PhD from the University of Cambridge in 1998, and has been at Memphis since 2003. He has worked in combinatorics and probability theory.
Mark Walters is a Teaching Fellow at Peterhouse College, Cambridge. He obtained his PhD from the University of Cambridge in 2000, and was a Research Fellow of Trinity College, Cambridge. He has worked in combinatorics and probability theory.

1 INTRODUCTION

In 1961, E.N. Gilbert defined and studied the following model of a random geometric graph, known as the *disc model* or *Gilbert model* (4). Let $\mathcal{P}$ be a Poisson process in the plane of intensity one, and join every point of $\mathcal{P}$ to every other point of $\mathcal{P}$ within distance r, for some fixed $r > 0$. For small r, most points are isolated, that is, not connected to any other points. However, as r increases, the points form small connected clusters, which then connect up to each other (as r increases still further), eventually forming a (connected) *giant component*, which contains a positive fraction of points in any large region. Loosely speaking, Gilbert derived upper and lower bounds for the smallest value of r such that the probability of the last eventuality (also known as *percolation*) is one.

In recent years there has been renewed interest in such graphs, which are now being used to model sensor networks and wireless ad-hoc networks in general. However, for some applications, it is desirable that the initial distribution of sensors is non-uniform. One such model was recently proposed by Etherington, Hoge and Parkes (3), in which the locations of the sensors are modeled as a Poisson process whose intensity is given by a two dimensional Gaussian distribution. Such a network might arise, for instance, if the sensors were dropped from an aircraft.

Specifically, Etherington, Hoge and Parkes defined the following random geometric graph $G = G(N, \sigma, s)$. We start with a Poisson process in $\mathbb{R}^2$ with intensity at radius r given by

$$\rho(r) = \tfrac{N}{2\pi\sigma^2} e^{-r^2/2\sigma^2},$$

for some constant σ. In addition, place a point at the origin, so that the expected total number of points is $N+1$. Then connect each point to every other point at distance less than s. We now ask two questions. First, what is the largest value of r such that every point in $D_r(0)$, the (open) disc of radius r centered at the origin, is joined to every other point of $D_r(0)$? Second, what is the smallest value of r such that $D_r(0)$ contains all the points of $C = C(0)$, the component of G containing the origin? ($C(0)$ consists of all the points of G that can be connected to the origin.) Formally, we define

$$r_- = \sup\{r : D_r(0) \cap V(G) \subseteq V(C)\},$$

and

$$r_+ = \inf\{r : V(C) \subseteq D_r(0)\}.$$

In this paper we derive lower and upper bounds for r_- and r_+, for various ranges of values of s. Since r_- and r_+ are random variables, our bounds will only hold *with high probability* (**whp**), that is, with probability tending to one as $N \to \infty$. By a uniform scaling of $\mathbb{R}^2$, we may without loss of generality fix $\sigma^2 = 1/2$, so that the density of points is given by

$$\rho(r) = \tfrac{N}{\pi} e^{-r^2}.$$

We imagine N to be very large, and consider various functions $s = s(N)$.

First we give a heuristic argument which yields the asymptotically correct value of r_- for a large range of values of s. Assuming that the density of points in a small disc of radius s is approximately constant, the probability that a point at distance r from the origin is an isolated vertex in G can be approximated by

$$e^{-\rho(r)\pi s^2} = e^{-Ns^2 e^{-r^2}}.$$

Therefore, provided $Rs \ll 1$, the expected number of isolated vertices in $D_R(0)$ is approximately

$$\int_0^R 2Nre^{-r^2} e^{-Ns^2 e^{-r^2}} dr = s^{-2}\Big(e^{-Ns^2 e^{-R^2}} - e^{-Ns^2}\Big). \quad (1)$$

When $R \le r_-$, there will be no isolated vertices in $D_R(0)$. On the other hand, it seems reasonable to suppose, by analogy with other similar problems, that the closest point to the origin not belonging to C is, with probability tending to one as $N \to \infty$, an isolated vertex in G. Therefore we may assume that if $R > r_-$, there will be at least one isolated vertex in $D_R(0)$. Consequently, if we choose R so that the expected number of isolated vertices in $D_R(0)$ is one, we might expect that $r_- \approx R$. Hence, if $N^{-1/2} \ll s \ll (\log N)^{-1/2}$,

$$r_-^2 \approx \log(Ns^2/\log(1/s^2)). \quad (2)$$

One approach to proving rigorous bounds for r_- is to apply methods and results for the related *disc model* $G_s(A)$, described above for the case $A = \mathbb{R}^2$. Here, given a region $A \subset \mathbb{R}^2$, we consider a Poisson process of intensity one in A, and join each point to all other points within a radius s to obtain $G_s(A)$. We have the following result of Gupta and Kumar (5), which was proved for the square S_N of area N by Penrose (6).

Theorem 1. *Let A_N be a disc of area N, and let $s = s(N)$ satisfy $\pi s^2 = \log N + \omega(1)$. Then* **whp** *$G_s(A_N)$ is connected.*

However, it turns out that applying this result yields only the weak bound

$$r_-^2 \gtrsim \log(Ns^2/\log N). \quad (3)$$

A heuristic explanation is as follows. For the result of Gupta and Kumar, the obstruction to connectivity is the existence of isolated vertices, and it seems reasonable to suppose that this is also true for our model. If we choose s so that the probability that a vertex is isolated is $o(N^{-1})$, then the expected number of isolated vertices is $o(1)$, so **whp** there are no isolated vertices and (by the above fact) **whp** $G_s(A)$ is connected. For our model, even if we choose r so that the probability that a vertex at distance r from the origin is isolated is $\Theta(N^{-1})$, as in (3), then the only vertices in $D_r(0)$ which have probability $\Theta(N^{-1})$ of being isolated lie very close to the boundary of $D_r(0)$, and there are far fewer than N such vertices. Consequently, it is likely that r_- is somewhat larger than suggested by (3), and we will argue directly in the proof of Lemma 5 to show that indeed (2) is much closer to the truth. Nevertheless, we will use Theorem 1 in deriving a lower bound for r_- when s is very small.

Next we turn to r_+. For the disc model $G_s(\mathbb{R}^2)$, with an additional point at the origin, results from (2) (see also (4)) show that there is a critical density $\gamma \approx 4.512$ so that if $\pi s^2 < \gamma$ then $C(0)$ (defined as above) is finite with probability one, while if $\pi s^2 > \gamma$ there is a non-zero probability that $C(0)$ is infinite. This suggests that **whp**

$$\gamma \approx \pi s^2 \rho(r_+) = N s^2 e^{-r_+^2},$$

so that **whp**

$$r_+^2 = O(\log(Ns^2)), \tag{4}$$

and we will show in Section 2 that in fact $r_+^2 = \Theta(Ns^2)$ for a large range of values of s.

For $Rs \gg 1$, the above arguments fail as the density of points in a disc of radius s at distance R from the origin is far from constant. Of particular interest is the value of s for which G becomes connected. We show that this occurs when

$$2s\sqrt{\log N} \approx \log\log N - \tfrac{1}{2}\log\log\log N.$$

The main obstacle to connectivity is the presence of isolated vertices that are among the furthest points from the origin. These vertices are at distance $R \approx \sqrt{\log N}$, so $Rs \gg 1$.

2 Precise results

Let $r_{\min}$ and $r_{\max}$ be the distance of the nearest (respectively furthest) point from the origin.

Theorem 2. *Define $G = G(N,s)$, r_- and r_+ as above. Then the following statements hold with high probability.*

1. *If $2s\sqrt{\log N} \geq \log\log N - \frac{1}{2}\log\log\log N + \omega(1)$ then G is connected.*

2. *If $2s\sqrt{\log N} = \log\log N - \frac{1}{2}\log\log\log N + O(1)$ then each of the following events has probability bounded below by some positive constant:*

   (a) *G is connected,*

   (b) *G is connected except for the furthest point from 0 which is isolated,*

   (c) *G is connected except for one isolated point that is not furthest from 0,*

   (d) *G has more than three components.*

3. *If $2s\sqrt{\log N} \leq \log\log N - \frac{1}{2}\log\log\log N - \omega(1)$ then G is disconnected and $r_+ < r_{\max}$* **whp**. *Moreover, if $2s\sqrt{\log N} \leq C\log\log N$ for some $C < 1$ and $Ns^2 \geq \log N$,*

$$\begin{aligned} r_-^2 &= \log(Ns^2/\log(1/s^2)) + 2s\sqrt{\log N} \\ &\quad - \tfrac{3}{2}\log\max\{1, s\sqrt{\log N}\} + O(1) \\ r_+^2 &= \log(Ns^2) + \Theta((s\sqrt{\log N}\log\log N)^{1/2}) \\ &\quad + O(1). \end{aligned}$$

   [If $2s\sqrt{\log N} = o(1/\log\log N)$, these simplify to

$$\begin{aligned} r_-^2 &= \log(Ns^2/\log(1/s^2)) + O(1) \\ r_+^2 &= \log(Ns^2) + O(1).] \end{aligned}$$

4. *If $Ns^2 = C\log N$ for some constant $C > 0$ then*

   (a) *if $C > 1$, $r_-^2 = \Theta(1)$,*

   (b) *if $C = 1$, $r_-^2 = (1 + o(1))\log\log N/\log N$,*

   (c) *if $C < 1$, $r_-^2 = N^{C-1+o(1)}$.*

   In all cases $r_+^2 = \log(Ns^2) + O(1) = \log\log N + O(1)$ as above.

5. *If $Ns^2 \to \infty$ then $r_-/s \to \infty$, while $r_+^2 = \log(Ns^2) + O(1)$ as above.*

6. *If $Ns^2 = C > 0$, then r_-/s has a limiting distribution.*

   (a) *If $C > \gamma$, where γ is the critical density for disc percolation, then with positive probability $r_+^2 = (1 + o(1))\log(C/\gamma)$. Conditional on this not occurring r_+/s has a limiting distribution.*

   (b) *If $C \leq \gamma$ then r_+/s has a limiting distribution.*

   In both cases there is a positive probability that $r_+ = 0$.

7. *If $Ns^2 = o(1)$ then* **whp** *the origin is isolated, so that $r_- = r_{\min}$ and $r_+ = 0$.*

Part 1 of the theorem is proved as part of Lemma 5 in Section 2.1, and part 2 follows from the remarks following the proof of Lemma 5. Part 3 is contained in Lemmas 5, 7, 8, and 9, except for the assertion relating to $r_{\max}$, which follows from the remarks following the proof of Lemma 5. The remainder of the theorem is proved in Section 3.

3 General bounds

First we prove two easy bounds on $r_{\max}$ giving an idea of the scale of this distribution.

Lemma 3. *$\mathbb{P}(r_{\max}^2 \leq \log N + \alpha) = e^{-e^{-\alpha}}$, so* **whp** *$\log N - \omega(1) < r_{\max}^2 < \log N + \omega(1)$.*

Proof. The number of points outside radius R is Poisson distributed with mean Ne^{-R^2}, and thus the probability that there are no points further than R from the origin is $\exp(-Ne^{-R^2})$. The result follows. $\square$

Next we prove some bounds on r_-. Before we do this, we first prove a simple lemma concerning the mean number of points in a disc, which takes into account the variation in density across the disc.

Lemma 4. *Fix a point z of $\mathbb{R}^2$ at distance $r \geq s$ from the origin. Then the expected number $E_{r,s}$ of vertices of G lying in $D_s(z)$ is given by*

$$\begin{aligned} E_{r,s} &= \mathbb{E}|V(G) \cap D_s(z)| \\ &= Ne^{-(r-s)^2} f(r,s)\theta(r,s) \end{aligned}$$

where

$$f(r,s) = \min\left\{\tfrac{1}{2}, s^2, \tfrac{1}{\sqrt{4\pi(r-s)}}, \sqrt{\tfrac{s}{4\pi(r-s)^3}}\right\}$$

and $c \leq \theta(r,s) \leq 1$ for some $c > 0$ independent of r and s.

Remark. Numerical calculations show that we can take $c = 0.3055$.

Proof. We can calculate the expected number exactly as

$$\frac{N}{\pi}\int_{-s}^{+s}\int_{-\sqrt{s^2-x^2}}^{\sqrt{s^2-x^2}} e^{-(r+x)^2-y^2}\,dy\,dx.$$

We can estimate $\int_{-z}^{z} e^{-y^2}\,dy = \min\{\sqrt{\pi}, 2z\}\theta_1(z)$ where $0 < c_1 \leq \theta_1(z) \leq 1$. Writing $x = -s + \varepsilon$ the above expression becomes

$$\tfrac{N}{\pi}e^{-(r-s)^2}\int_0^{2s} e^{-2(r-s)\varepsilon-\varepsilon^2} m(\varepsilon, s)\theta_1(\sqrt{2\varepsilon s - \varepsilon^2})\,d\varepsilon,$$

where

$$m(\varepsilon, s) = \min\{\sqrt{\pi}, 2\sqrt{2\varepsilon s - \varepsilon^2}\}.$$

We can bound the above integral by

$$\int_0^\infty 2\sqrt{2\varepsilon s}\,e^{-2(r-s)\varepsilon}\,d\varepsilon = \sqrt{\frac{\pi s}{4(r-s)^3}}.$$

Since this integral is dominated by the contribution when $\varepsilon \sim (r-s)^{-1}$, this will give the correct order of magnitude when $r - s \gg 1, s^{-1}, s$. Similarly, we can bound the integral by $\int_0^\infty \sqrt{\pi}\,e^{-2(r-s)\varepsilon}\,d\varepsilon = \frac{\sqrt{\pi}}{2(r-s)}$, and this gives the correct order of magnitude when $s \gg r - s \gg 1$. We can bound the integral by $\int_0^\infty e^{-\varepsilon^2}\sqrt{\pi}\,d\varepsilon = \frac{\pi}{2}$ and this gives the right order when $s \gg 1 \gg r - s$. Finally we can bound the integral by $\int_0^{2s} 2\sqrt{2\varepsilon s - \varepsilon^2}\,d\varepsilon = \pi s^2$, and this gives the correct order when $1 \gg s$ and $s^{-1} \gg r - s$. Thus we have bounded $E_{r,s}$ as required, and shown that the bound is of the right order except on a compact region in $\mathbb{R}^2_+$, where the result follows by continuity (see Figure 1). □

3.1 Bounds for r_-

With the aid of Lemma 4, we can obtain fairly tight bounds for r_-.

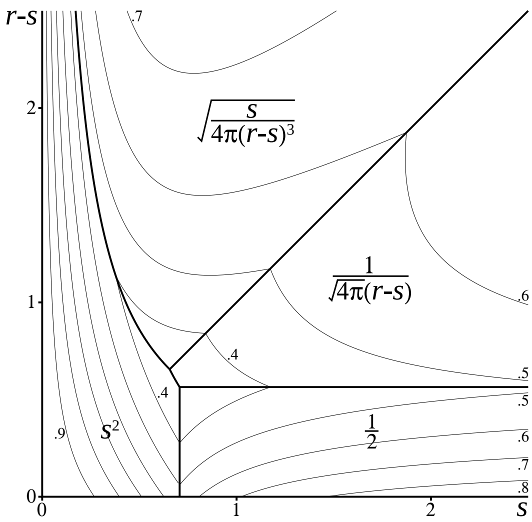


Figure 1: The function $\theta(r,s)$. Formulae shown are values of the minimum in the expression in Lemma 4.

Lemma 5. *Suppose that $s = s(N) = o(1)$ and $s \geq \sqrt{(\log N)/N}$. Then* **whp**

$$\begin{aligned} r_-^2 &\geq \log(Ns^2/\log(1/s^2)) + 2s\sqrt{\log N} \\ &\quad - \tfrac{3}{2}\log\max\{1, 3s\sqrt{\log N}\} - 3. \end{aligned}$$

Moreover, if

$$2s\sqrt{\log N} \geq \log\log N - \frac{1}{2}\log\log\log N + \omega(1)$$

then G is connected **whp**.

Proof. For any point x at distance $r \geq s$ from 0, let the region $A(x)$ be the intersection of the ball $D_s(x)$ and $D_r(0)$. Note that all of $A(x)$ is closer to 0 than x is. Let E_R be the expected number of vertices x within R of 0 with no points in the region $A(x)$. If $E_R \to 0$ then **whp** $r_- \geq R$ since **whp** we can find a sequence of points joining any point $x \in D_R(0)$ to 0.

Now $|A(x)| \geq |D_s(x)|/3$, and since the density of points is higher in $A(x)$ than in $D_s(x) \setminus A(x)$, $\mathbb{E}|V(G) \cap A(x)| \geq \frac{1}{3}E_{r,s}$. By Lemma 4, the probability that $A(x)$ is empty is at most

$$\exp\left(-\tfrac{1}{3}E_{r,s}\right) \leq \exp\left(-N\mu e^{-(r-s)^2}\right),$$

provided $\mu \leq \frac{c}{3}\min\left\{s^2, \sqrt{\frac{s}{4\pi(r-s)^3}}\right\}$, the other two terms in the minimum in Lemma 4 being redundant when $s = o(1)$. For $r < 3s$ we can take $\mu = \frac{c}{3}s^2$. Then

$$\begin{aligned} E_{3s} &\leq \int_s^{3s} 2Nre^{-r^2}e^{-N\mu e^{-(r-s)^2}}\,dr \\ &\leq \int_s^{3s} 2Nre^{-N\mu e^{-4s^2}}\,dr \\ &\leq 8Ns^2 e^{-(c/3-o(1))Ns^2}. \end{aligned}$$

By assumption $Ns^2 \to \infty$, so $E_{3s} \to 0$. For $3s \le r \le R$ set $\mu = \frac{c}{3}s^2 \min\{1, (3Rs)^{-3/2}\}$. Writing $z = N\mu e^{-(r-s)^2} + \log\mu + 2rs - s^2$ (note that z is decreasing in r) we have $\frac{dz}{dr} = -2(r-s)N\mu e^{-(r-s)^2} + 2s$. Thus we have $|\frac{dz}{dr}| > rN\mu e^{-(r-s)^2}$ provided $r \ge 3s$ and $N\mu e^{-(r-s)^2} > 2$. Then, writing

$$z_0 = N\mu e^{-(R-s)^2} + \log\mu + 2Rs - s^2, \tag{5}$$

we have

$$\begin{aligned} E_R - E_{3s} &\le \int_{3s}^{R} 2Nre^{-r^2} e^{-N\mu e^{-(r-s)^2}}\, dr \\ &\le \int_{z_0}^{\infty} 2\mu^{-1} e^{(r-s)^2 - r^2 - N\mu e^{-(r-s)^2}}\, dz \\ &= \int_{z_0}^{\infty} 2e^{-z}\, dz = 2e^{-z_0}, \end{aligned}$$

so it is enough that $z_0 \to \infty$, and

$$N\mu e^{-(R-s)^2} > 2. \tag{6}$$

Set $(R-s)^2 = \log(N\mu/\log(1/s^2)) - \alpha$, $\alpha > 0$. Then $N\mu e^{-(R-s)^2} = e^{\alpha}\log(1/s^2) > 2$ for small s, and

$$z_0 = (e^{\alpha} - 1)\log(1/s^2) + \log(\mu/s^2) + 2Rs - s^2.$$

Now $\log(\mu/s^2) = -\frac{3}{2}\log\max\{1, 3Rs\} + \log(c/3)$, so $\log(\mu/s^2) + 2Rs$ is bounded below. Since $\alpha > 0$ and $s \to 0$, we have $z_0 \to \infty$ as required.

If $Rs > 1$, then $R^2 = (1 + o(1))\log N$, so $\log(\mu/s^2) \ge -\frac{3}{2}\log\max\{1, 3s\sqrt{\log N}\} - 2.5$. Thus, taking $\alpha = 0.1$,

$$\begin{aligned} R^2 &= \log(N\mu/\log(1/s^2)) + 2Rs - s^2 - 0.1 \\ &\ge \log(Ns^2/\log(1/s^2)) + 2s\sqrt{\log N} \\ &\quad - \tfrac{3}{2}\log\max\{1, 3s\sqrt{\log N}\} - 3, \end{aligned}$$

and $r_- \ge R$ since $E_R \to 0$. If $Rs \le 1$ then the same bound applies since $2s\sqrt{\log N} < 0.1$ unless $R^2 = (1 + o(1))\log N$.

For the last part, set $R^2 = \log N + \alpha$ where α is constant. Assume $2s\sqrt{\log N} = \log\log N - \frac{1}{2}\log\log\log N + \beta$, where $\beta \to \infty$, but $\beta = o(\log\log N)$. Then $2Rs = (1 + o(1))\log\log N$, so $\mu = c'(2Rs)^{1/2}R^{-2} = \Theta(\sqrt{\log\log N}/\log N)$. Thus $z_0 \ge \log\mu + 2Rs - s^2 = \beta + O(1) \to \infty$. Similarly $N\mu e^{-(R-s)^2} = e^{\log\mu + 2Rs - \alpha + o(1)} = e^{\beta - \alpha + O(1)} \to \infty$. Hence $r_-^2 \ge \log N + \alpha$ **whp** for any fixed α. Thus $r_- > r_{\max}$, so G is connected **whp**. □

This establishes part 1 of Theorem 2. For part 2, suppose that

$$2s\sqrt{\log N} = \log\log N - \tfrac{1}{2}\log\log\log N + \beta,$$

where we will initially suppose that β is constant. The proof of Lemma 5 shows that there exists an α and C, depending only on β such that $E_R \le C$ when $R^2 = \log N + \alpha$. Since the expected number of points at distance greater than R from the origin is $e^{-\alpha}$, $E_\infty \le C'$ for some constant $C' \le C + e^{-\alpha}$ depending only on β. By dividing the plane into, say, $10C'$ sectors, we see that there is a large probability that all the points in any one particular sector can be connected to a point nearer the origin, since in any one sector the expected number of points which can't be connected to a point nearer the origin is at most 0.1. Since events in different sectors are almost independent, it follows (by, for example, the Lovász Local Lemma) that G is connected with probability bounded away from zero. On the other hand, let us estimate the probability $p(\alpha, \beta)$ that a vertex at distance R is isolated, where $R^2 = \log N + \alpha$ and α is constant. We have

$$\begin{aligned} p(\alpha, \beta) &= \exp\left\{-Ne^{-(R-s)^2}\sqrt{\tfrac{s}{4\pi(R-s)^3}}\,\theta(R, s)\right\} \\ &\ge \exp\left\{-c_0 N e^{-\log N - \alpha + 2Rs}\sqrt{\tfrac{s}{R^3}}\right\} \\ &\ge \exp\left\{-c_1 e^{-\alpha + 2s\sqrt{\log N}}\tfrac{\sqrt{\log\log N}}{(\log N)}\right\} \\ &\ge \exp\left\{-c_2 e^{\beta - \alpha}\right\} = c_3. \end{aligned}$$

Hence the probability that the furthest, or second furthest point from the origin is isolated is bounded below by a constant. Since these points are likely to be far from one another, they are isolated almost independently. This, together with the first observation, establishes part 2. Finally, if $\beta \to -\infty$ with α fixed, we see that $p(\alpha, \beta) \to 1$ and so $r_+ < r_{\max}$ **whp**, establishing the first statement of part 3.

If s is not $o(1)$ then G is connected, so $r_- = \infty$ and $r_+ = r_{\max}$. If $s \le \sqrt{(\log N)/N}$, the bound for r_-^2 in Lemma 5 is negative, so trivially true. We shall show later that in this case $r_- = o(1)$.

We call a point $x \in V(G)$ *isolated* if $D_s(x) \cap V(G) = \{x\}$. Clearly any isolated point $x \ne 0$ cannot lie in $V(C)$, so must be at distance at least r_- from 0. To obtain an upper bound for r_-, we follow the proof of Lemma 5 but instead estimate E'_R, the expected number of isolated points within distance R of 0. We require the following lemma.

Lemma 6. *Let E'_R denote the expected number of isolated points in G within distance R of 0. If $E'_R \to \infty$ then $r_- \le R$* **whp**.

Proof. First note that any two isolated points must be at least distance s apart. Hence there is at most one isolated point in any square of side length $2s/3$. On the other hand, the existence of isolated points in two such squares is independent if the squares are at distance at least $2s$, since the event that a square contains an isolated point depends only on the process within distance s of that square. Tile $\mathbb{R}^2$ with squares $S_{ij} = [0, 2s/3]^2 + (2si/3, 2sj/3)$. Partition this collection into 16 classes according to the value of $(i \bmod 4, j \bmod 4) \in \mathbb{Z}_4^2$. Let X_{kl}, $k, l \in \mathbb{Z}_4$ be the number of isolated points within R of 0 and in one of the squares in class (k, l). The X_{kl} are dependent, but still $\sum_{k,l}\mathbb{E}X_{kl} = E'_R$. Hence there is at least one $X_{k,l}$, say $X_{k'l'}$, with $\mathbb{E}X_{k'l'} \ge E'_R/16$. Now $X_{k'l'}$ is a sum of *independent* bernoulli random variables with mean p_i, say.

Thus

$$\mathbb{P}(X_{k'l'}=0)=\prod_i(1-p_i)\le\prod_i\exp(-p_i)$$
$$=\exp(-\mathbb{E}X_{k'l'})\le\exp(-E'_R/16),$$

so if $E'_R\to\infty$, then **whp** there is an isolated point within distance R of 0, and $r_-\le R$ **whp**. ☐

Lemma 7. *Suppose that $s\ge\sqrt{(\log N)/N}$ and also that $2s\sqrt{\log N}\le(1-\varepsilon)\log\log N$ for some $\varepsilon>0$. Then* **whp**

$$r_-^2\le\log(Ns^2/\log(1/s^2))+2s\sqrt{\log N}$$
$$-\tfrac{3}{2}\log\max\{1,2s\sqrt{\log N}\}+\log(2/\varepsilon).$$

Moreover, if

$$2s\sqrt{\log N}\le\log\log N-\frac{1}{2}\log\log\log N-\omega(1)$$

then G is disconnected **whp**.

Proof. We estimate E'_R and use Lemma 6. The probability that $x\in V(G)$ is isolated is at least $\exp(-E_{r,s})\ge\exp(-N\mu e^{-(r-s)^2})$, where $\mu=\min\{s^2,\sqrt{\frac{s}{4\pi R'^3}}\}$, when x is distance at least $R'+s$ from the origin. Hence

$$E'_R\ge\int_{R'+s}^{R}2Nre^{-r^2}e^{-N\mu e^{-(r-s)^2}}\,dr$$
$$=\int_{R'}^{R-s}2N(z+s)e^{-z^2-2sz-s^2}e^{-N\mu e^{-z^2}}\,dz$$
$$\ge e^{-2Rs}\int_{R'}^{R-s}2Nze^{-z^2}e^{-N\mu e^{-z^2}}\,dz$$
$$=e^{-2Rs}\left[\mu^{-1}e^{-N\mu e^{-z^2}}\right]_{z=R'}^{R-s}$$
$$=\mu^{-1}e^{-2Rs}\left(e^{-N\mu e^{-(R-s)^2}}-e^{-N\mu e^{-R'^2}}\right).$$

Set $(R-s)^2=\log(N\mu/\log(1/s^2))-\alpha$, where $\alpha<0$ is constant. Consider the case when $s\sqrt{\log N}=O(1)$ first. Take $R'=0$, so $\mu=s^2$. Then $E'_R\ge e^{-2Rs-\log(1/s^2)(e^\alpha-1)}-s^{-2}e^{-Ns^2}$. By assumption on s, $s^{-2}e^{-Ns^2}\le 1/\log N\to 0$, and $Rs=O(1)$, so $E'_R\to\infty$ when $\alpha<0$. Now assume $s\sqrt{\log N}$ is large. Then $R^2=(1+o(1))\log N$. Take $R'=(1-\varepsilon)R$. Now $N\mu e^{-(R-s)^2}=e^\alpha\log(1/s^2)\to\infty$. Also $e^{R'^2}/e^{(R-s)^2}=o(1)$, so $\log E'_R\ge-\log\mu-2Rs-e^\alpha\log(1/s^2)+o(1)=(1-e^\alpha+o(1))\log\log N-2Rs+o(1)\ge(\varepsilon-e^\alpha+o(1))\log\log N\to\infty$ for $\alpha<\log\varepsilon$.

Now, if $(R-s)^2=\log(N\mu/\log(1/s^2))-\alpha$ then

$$R^2\le\log(Ns^2/\log(1/s^2))+2s\sqrt{\log N}$$
$$-\tfrac{3}{2}\log\max\{1,2s\sqrt{\log N}\}-\alpha+o(1),$$

and $r_-\le R$.

For the last part, set $R^2=\log N-\alpha$. Then $\log\mu+2Rs\to-\infty$. Thus $N\mu e^{-(R-s)^2}=e^{\log\mu+2Rs+\alpha+o(1)}\to 0$. Therefore, using the approximation $e^{-\theta}\approx 1-\theta$ for small θ,

$$E'_R\ge(1-o(1))\mu^{-1}e^{-2Rs}(N\mu e^{-R'^2}-N\mu e^{-(R-s)^2})$$
$$=(1-o(1))(e^{\log N-R'^2-2Rs}-e^{\alpha-s^2}).$$

If $R'^2=(1-\varepsilon)R^2$, then $\log N-R'^2-2Rs=\varepsilon\log N+O(\log\log N)$ so $E'_R\to\infty$ as required. Since this holds for all $\alpha>0$, $r_-<r_{\max}$ and G is disconnected **whp**. ☐

This completes the proof of the estimate for r_- in part 3 of Theorem 2.

3.2 Bounds for r_+

Now we turn our attention to r_+. In this case we are interested in the existence of at least one point at distance R which is joined to the origin.

Lemma 8. *Suppose that $2s\sqrt{\log N}\le\log\log N$ and $r_-/s\to\infty$* **whp**. *Then,* **whp**,

$$r_+^2\ge\log(Ns^2)+\tfrac{1}{2}\sqrt{s\sqrt{\log N}\,\log\log N}$$
$$+O(s\sqrt{\log N}+1).$$

Proof. First assume $s\sqrt{\log N}\log\log N=O(1)$. Then the statement reduces to $r_+^2\ge\log(Ns^2)+O(1)$. Consider the disc $D_R(0)$ where R is given by

$$R^2=\log(Ns^2)-3.$$

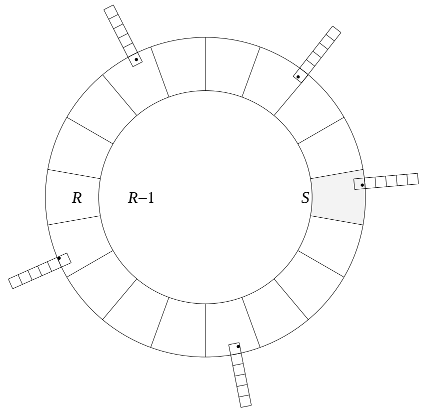
The Poisson process restricted to $D_R(0)$ stochastically dominates a Poisson process in $D_R(0)$ with constant intensity $\rho(R)$. Cover the disc with a square tessellation where the squares have side length $s/\sqrt{5}$. The number of points inside any of the squares which are wholly inside $D_R(0)$ dominates a Poisson distribution with mean $\lambda=\rho(R)s^2/5$. Substituting for ρ and R we get

$$\lambda=\tfrac{N}{5\pi}e^{-R^2}s^2=\tfrac{e^3}{5\pi}>1.27.$$

The probability that any such square contains no points is at most $e^{-\lambda}$.

We compare this process to a site percolation on $\mathbb{Z}^2$ by declaring a site to be open if its corresponding square contains at least one point. The site percolation dominates our process in the sense that percolation in the site model implies percolation in our model. Since $1-e^{-\lambda}>0.7>p_c(site)$ the probability that some square inside $D_{r_-}(0)$ is in an infinite component tends to one, and thus the probability that the origin in our process is in an infinite component tends to one. Thus, **whp**, the origin is joined to some point of the process at distance at least $R-s$ from the origin. Thus

$$r_+^2\ge(R-s)^2\ge R^2-2Rs$$
$$\ge\log(Ns^2)-3-2s\sqrt{\log N}$$
$$\ge\log(Ns^2)+O(1)$$



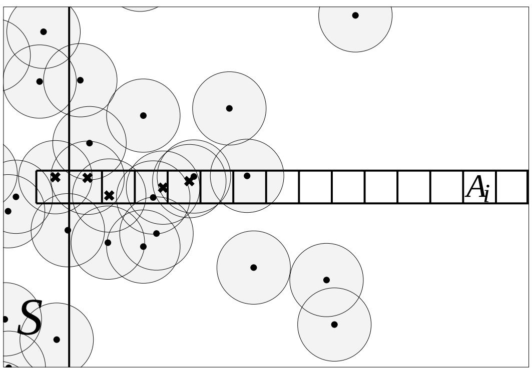


Figure 2: Regions S and A_i in the proof of Lemma 8. The discs shown have radius $s/2$, so that discs corresponding to adjacent vertices overlap. With high probability, at least one of the columns of small (approximate) squares contains a point in every square.

and the result follows.

Now assume $s\sqrt{\log N}\,\log\log N \to \infty$. We aim to show that **whp** $C(0)$ extends some distance beyond $D_R(0)$ in at least one narrow sector, where $R^2 = \log(Ns^2) - 3$ as before. By Lemma 5, $R^2 - r_-^2 = O(\log\log N)$, and so $R - r_- = o(1)$. Divide $D_R(0)$ into sectors of angle about $1/\sqrt{\log N}$. Consider the approximately square region formed by intersecting one of these sectors with an annulus with radii $R-1$ and R (see Figure 2). Subdivide this region S into sectors of angle $s/\sqrt{5\log N}$ and annuli of thickness $s/\sqrt{5}$. The region is thus subdivided into approximately square regions of side length at most $s/\sqrt{5}$. We aim to show that with a reasonable probability there is a point in $S \cap C(0)$ at distance at least $R - s/\sqrt{5}$ from 0. Each small square-like region contains at least one point with probability at least $p_0 = 1 - \exp(-s^2\rho(R)/5.1) > p_{\text{site}}$. Moreover, points in any two adjacent squares are within distance s, so are connected in G. Comparing this process with a site percolation with probability p_0, we see that with positive probability there is a square near the centre of S which is connected to the side $r = R$ of S, and hence there is a point within r_- of 0 joined to a point at least $R - s/\sqrt{5}$ of 0 in S.

Consider a sector of angle $s/\sqrt{5\log N}$ containing such a point. For $i = 0, 1, 2, \dots$ let $r_i = R + (i-1)s/\sqrt{5}$. Let A_i be the intersection of the sector with the annulus with inner radius r_i and outer radius r_{i+1}. Suppose that k is such that all the regions A_i for $1 \le i \le k$ contain at least one point of the process. Then, for $0 \le i < k$, points in adjacent regions A_i and A_{i+1} are adjacent in G, so we may conclude that r_+ is at least $R + (k-1)s/\sqrt{5}$.

The density in the region A_i is at least $\rho(r_{i+1})$. The condition on s implies that $R^2 = (1+o(1))\log N$. Thus, the area of the region A_i is at least $s^2/5.1$. Hence, setting $\lambda_0 = (s^2/5.1)\rho(R)$, the expected number λ_i of points in A_i satisfies $\lambda_i \ge \lambda_0 e^{-\alpha i}$ where $\alpha = 2r_{k+1}s/\sqrt{5}$ (the definition of R implies that $\lambda_0 = e^3/5.1\pi + o(1) > 1.25$).

The probability p_i that A_i contains at least one point is $1 - e^{-\lambda_i} \ge \frac{1}{e}e^{-\alpha i}$. Thus the probability that all the regions A_i for $0 \le i \le k$ contain at least one point is at least

$$\prod_{i=1}^{k} p_i \ge \exp(-\alpha\tbinom{k+1}{2} - k).$$

If

$$k = \left\lfloor \sqrt{0.98(\log\log N)/\alpha} \right\rfloor - 1$$

then this probability is at least $(\log N)^{-0.49-o(1)}$. (Note that $s\sqrt{\log N}\,\log\log N \to \infty$ ensures that $\alpha k \to \infty$.) However we have at least $R = \Theta((\log N)^{1/2})$ disjoint sectors so, **whp**, at least one of the sectors satisfies this. Thus **whp**,

$$\begin{aligned} r_+^2 \ge\, & R^2 + 2(k-1)Rs/\sqrt{5} \ge \log(Ns^2) \\ & + \tfrac{1}{2}\sqrt{s\sqrt{\log N}\log\log N} + O(s\sqrt{\log N} + 1). \end{aligned}$$

□

Lemma 9. *Suppose that $s = o(1)$ and $Ns^2 \to \infty$. Then,* **whp**,

$$\begin{aligned} r_+^2 \le\, & \log(Ns^2) + 2\sqrt{s\sqrt{\log N}\log\log N} \\ & + O(s\sqrt{\log N} + 1). \end{aligned}$$

Proof. Let the radius R be defined by $\pi s^2\rho(R) = 1/3$. We define a sequence of areas A_i. Let A_0 be $D_R(0)$, and for $i \ge 1$ let $A_i = D_{R+is}(0) \setminus D_{R+(i-1)s}(0)$. Let $V_1 = V \cap A_1$ and for $i \ge 2$ let V_i be the vertices in A_i joined to a vertex in V_1 wholly inside $D_{r+is}(0)$.

We want to bound the size of V_i. Let W_i be the points in A_i that are neighbours of vertices in V_{i-1}. Now, since the density in A_{i+1} is bounded above by $\rho(R+is)$ we have

$$\mathbb{E}(|W_{i+1}| \mid |V_i|) \le \pi s^2\rho(R+is)|V_i|.$$

Also, any vertex in V_i is either in W_i or is a descendant of a vertex in W_i. Moreover any vertex in $V_i \setminus W_i$ has an ancestor in W_i which can be reached from this vertex by a path using no vertex of any V_j for $j < i$. Note that these vertices may lie in A_j for $j < i$ but cannot lie in A_1 or A_0. However the expected number of descendants of a vertex in W_i which can be reached wholly outside $D_R(0)$ is at most $\sum_{j=1}^{\infty} 3^{-j} = 1/2$. Hence

$$\mathbb{E}(|V_i| \mid |W_i|) \le \tfrac{3}{2}|W_i|.$$

Combining these we see that, substituting for R,

$$\mathbb{E}(|V_{i+1}| \mid |V_i|) \le \tfrac{3}{2}\pi s^2 \rho(R+is)|V_i| \le \tfrac{1}{2}e^{-2Ris}|V_i|.$$

Also $|V_1|$ is at most the number of points outside $D_R(0)$, which is dominated by a Poisson distribution with mean $\frac{1}{3s^2}$. Hence

$$\mathbb{E}(|V_j|) \le 2^{-(j-1)} \exp(-2Rs\tbinom{j}{2})\tfrac{1}{3s^2}.$$

Thus if

$$j = \left\lceil \sqrt{(Rs)^{-1}\log(1/s^2)} \right\rceil + 1$$

then $\mathbb{E}(|V_j|) \le 2^{-j}e^{-Rs}$ (using $2\binom{j}{2} \ge (j-1)^2 + 1$ for $j \ge 2$). Now since $s = o(1)$, $(Rs)j^2 \to \infty$. Thus either j or Rs is large and so $\mathbb{E}(|V_j|) = o(1)$. Therefore **whp** $r_+ \le R + js$, so that **whp** $r_+^2 \le \log(Ns^2) + 2jRs + O(1)$. The result follows since we may assume $\log(1/s^2) \le \log\log N$ and $R \le \sqrt{\log N}$. □

This completes the proof of the estimate for r_+ in part 3 of Theorem 2.

Corollary 10. *If $s = o(\log\log N/\sqrt{\log N})$ and $Ns^2 \to \infty$, then* **whp**

$$r_{\max}^2 - r_+^2 \ge (1 + o(1))\log(1/s^2).$$

Proof. From Lemma 9, we have **whp**

$$\begin{aligned} r_+^2 \le& \log N - \log(1/s^2) + 2\sqrt{s\sqrt{\log N}\log\log N} \\ &+ O(s\sqrt{\log N} + 1) \\ \le& \log N - \log(1/s^2) + o(\log\log N). \end{aligned}$$

Thus for any $\varepsilon > 0$, **whp** $r_+^2 + (1-\varepsilon)\log(1/s^2) = \log N - \omega(1)$. The result now follows from Lemma 3. □

4 Small s

The results we have proved so far say very little about the case $Ns^2 = O(\log N)$. In this section we address this case.

Lemma 11. *Suppose that R is such that $\pi R^2\rho(R) \to \infty$, $Ns^2 = O(\log N)$, and*

$$Ns^2e^{-R^2} - \log N - \log R^2 \to \infty. \tag{7}$$

Then, **whp**, *$G|_{D_R(0)}$ is connected.*

Proof. The process restricted to $D_R(0)$ stochastically dominates a Poisson process of mean $\rho(R)$. The expected number of points of the Poisson process inside the disc is $\pi R^2\rho(R)$.

The hypotheses of the theorem imply that $R = O(1)$ (since if $R \to \infty$ then eventually $Ns^2e^{-R^2} - \log N - \log R^2 < -\frac{1}{2}\log N - \log R^2 \to -\infty$), so that $\rho(0) \le C\rho(R)$ for some constant C. We modify the original process inside $D_R(0)$ by keeping points of the process at distance $r \le R$ from the origin with probability $\rho(R)/\rho(r)$. Then the modified process has density exactly $\rho(R)$ in $D_R(0)$. Denote by G' the graph obtained from the modified process by joining two points at distance less than s.

Now suppose that $H = G|_{D_R(0)}$ is not connected. Then H will contain two vertices v_1 and v_2 which are not joined by a path in H. These vertices will be in $H' = G'|_{D_R(0)}$ with probability at least $1/C^2$, and they will certainly not be joined by a path in H'. Hence if H is not connected with probability at least p infinitely often as $N \to \infty$, then H' is not connected with probability at least p/C^2 infinitely often as $N \to \infty$. Thus if H' is connected **whp**, H is also connected **whp**.

Provided that $\pi R^2\rho(R)$ tends to infinity as N tends to infinity, we can apply Theorem 1 to see that H' is connected, **whp**, if $\pi s^2\rho(R) \ge \log(\pi R^2\rho(R)) + \omega(1)$. Substituting for $\rho(R)$ and rearranging gives the result. □

Corollary 12. *Suppose $C > 1$ and $Ns^2 = C\log N$. Then* **whp**

$$r_-^2 \ge \log C - o(1).$$

Proof. Fix $\alpha > 0$ with $\alpha < \log C$. Substitute $R^2 = \log C - \alpha > 0$ in (7). Then $\pi R^2\rho(R) \to \infty$ and the left hand side of (7) becomes

$$e^{\alpha}\log N - \log N - \log R^2 = (e^{\alpha} - 1)\log N + O(1)$$

which tends to infinity. □

Lemma 13. *If $Ns^2 < (1-\varepsilon)\log N$ and $Ns^2 \to \infty$ then* **whp**

$$r_-^2 = N^{-1}\exp\{Ns^2(1+o(1))\}.$$

Proof. First we prove $r_-^2 \le N^{-1}\exp\{Ns^2(1+o(1))\}$. Let $R^2 = N^{-1}\exp\{Ns^2(1+\delta)\}$. Then if $0 < \delta < \varepsilon$, $R = o(1)$ and the expected number of isolated points in $D_R(0)$ is at least

$$\begin{aligned} N(1-e^{-R^2})\exp\{-Ns^2\} &\ge \tfrac{1}{2}NR^2\exp\{-Ns^2\} \\ &= \tfrac{1}{2}\exp(\delta Ns^2) \to \infty. \end{aligned}$$

The result follows from Lemma 6.

To prove $r_-^2 \ge N^{-1}\exp\{Ns^2(1+o(1))\}$, let $R^2 = N^{-1}\exp\{Ns^2(1-\delta)\}$ and apply Lemma 11. Then $\pi R^2\rho(R) \to \infty$ and we have

$$\begin{aligned} Ns^2e^{-R^2} - \log(NR^2) =& Ns^2(1-o(1)) - Ns^2(1-\delta) \\ &+ o(1) \to \infty. \end{aligned}$$

The result follows, since if $G|_{D_R(0)}$ is connected then $r_- \ge R$. □

Lemma 14. *If $Ns^2 = \log N$ then* **whp**

$$r_-^2 = (1+o(1))\frac{\log\log N}{\log N}.$$

Proof. For the lower bound we use Lemma 11. Take $R^2 = (1-\alpha)\log\log N/\log N$ for $\alpha > 0$. Certainly $\pi R^2\rho(R) = NR^2e^{-R^2} \to \infty$. Now

$$\begin{aligned} Ns^2e^{-R^2} &- \log N - \log R^2 \\ &\geq (e^{-R^2}-1)\log N - \log\left(\frac{(1-\alpha)\log\log N}{\log N}\right) \\ &\geq -R^2\log N - \log(1-\alpha) \\ &\quad - \log\log\log N + \log\log N \\ &= \alpha\log\log N + O(\log\log\log N) \to \infty \end{aligned}$$

so that **whp** $r_- \geq R$.

For the upper bound, we use the proof of Lemma 7. Take $R^2 = (1+\alpha)\log\log N/\log N$ for $\alpha > 0$. We may assume that $(R-s)^2 \geq (1+\frac{\alpha}{2})\log\log N/\log N$. Choose θ such that $\theta(1+\frac{\alpha}{2}) > 1$ and assume that N is large enough that $e^{-(R-s)^2} \leq 1 - \theta(R-s)^2$. Then with notation as in the proof of Lemma 7, it is easy to see that there is an absolute constant C such that

$$\begin{aligned} E'_R &\geq \frac{C}{s^2}e^{-\log N \cdot e^{-(R-s)^2}} \\ &\geq \frac{CN}{\log N}e^{-\log N\cdot e^{-\left(1+\frac{\alpha}{2}\right)\frac{\log\log N}{\log N}}}. \end{aligned}$$

Therefore,

$$\begin{aligned} \log E'_R &\geq \left\{1 - e^{-(1+\frac{\alpha}{2})\frac{\log\log N}{\log N}}\right\}\log N \\ &\quad + \log C - \log\log N \\ &\geq \left\{\theta(1+\tfrac{\alpha}{2})\frac{\log\log N}{\log N}\right\}\log N \\ &\quad + \log C - \log\log N \\ &= \left\{\theta(1+\tfrac{\alpha}{2}) - 1\right\}\log\log N + \log C \to \infty \end{aligned}$$

and so **whp** $r_- \leq R$ by Lemma 6. □

The last three lemmas together establish part 4 of Theorem 2. Part 7 is immediate, and for part 5, if $Ns^2 \to \infty$ then $r_-/s \to \infty$ by Lemma 13 and so Lemma 8 applies and it together with Lemma 9 give $r_+^2 = \log(Ns^2) + O(1)$. Part 6 follows from standard results on branching processes: we note only that if $Ns^2 = C > \gamma$, where γ is the critical density for disc percolation, and if $R^2 = \log C - \log\gamma$, then the probability that the origin belongs to the giant component in $G|_{D_R(0)}$ lies strictly between 0 and 1, and, conditional on this event occurring, $r_+ = (1+o(1))R$.

5 Conclusion

In this paper we have analyzed a model of a random geometric graph whose vertices are given by a Poisson process of Gaussian intensity: two such vertices are connected in the graph if they lie within distance s of each other. We have given precise bounds for two parameters which, loosely speaking, describe the width of the central component of such a graph. It is our hope that some of our methods will find application to other problems in this area.

REFERENCES

[1] B. Bollobás, Random Graphs, second edition, Cambridge University Press, 2001.

[2] P. Balister, B. Bollobás and M. Walters, *Continuum percolation with steps in the square or the disc*, Random Structures and Algorithms **26** (2005), 392–403.

[3] D.W. Etherington, C.K. Hoge and A.J. Parkes, *Global surrogates*, manuscript.

[4] E.N. Gilbert, *Random Plane Networks*, Journal of the Society for Industrial Applied Mathematics **9** (1961), 533–543.

[5] P. Gupta and P.R. Kumar, *Critical power for asymptotic connectivity in wireless networks*, Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming (W.M. McEneaney, G. Yin and Q. Zhang (Eds.)), 1998.

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

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