Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2013, Vol. 49, No. 4, 1130–1140
DOI: 10.1214/12-AIHP498

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Size of the giant component in a random geometric graph

Ghurumuruhan Ganesan

Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, New Delhi, 7 S. J. S. Sansanwal Marg, New Delhi 110016, India.
E-mail: guru9r@isid.ac.in

Received 16 April 2012; revised 1 June 2012; accepted 5 June 2012

Abstract. In this paper, we study the size of the giant component C_G in the random geometric graph $G = G(n, r_n, f)$ of n nodes independently distributed each according to a certain density $f(\cdot)$ in $[0, 1]^2$ satisfying $\inf_{x\in[0,1]^2} f(x) > 0$. If $\frac{c_1}{n} \le r_n^2 \le c_2\frac{\log n}{n}$ for some positive constants c_1, c_2 and $nr_n^2 \longrightarrow \infty$ as $n \to \infty$, we show that the giant component of G contains at least $n - o(n)$ nodes with probability at least $1 - e^{-\beta n r_n^2}$ for all n and for some positive constant β. We also obtain estimates on the diameter and number of the non-giant components of G.

Résumé. Dans cet article nous étudions la composante principale C_G dans le graphe géométrique aléatoire $G = G(n, r_n, f)$ avec n nœuds indépendants, chacun étant distribué selon une densité $f(\cdot)$ dans $[0, 1]^2$ telle que $\inf_{x\in[0,1]^2} f(x) > 0$. Si $\frac{c_1}{n} \le r_n^2 \le c_2\frac{\log n}{n}$ pour des constantes positives c_1, c_2 et $nr_n^2 \longrightarrow \infty$ quand $n \to \infty$, nous montrons que la composante principale de G contient au moins $n - o(n)$ nœuds avec probabilité minorée par $1 - e^{-\beta n r_n^2}$ pour tout n et pour une constante positive β. Nous obtenons aussi des estimations sur les diamètres et sur le nombre des plus petites composantes de G.

MSC: Primary 60D05; secondary 60C05

Keywords: Random geometric graphs; Size of giant component; Number of components

1. Introduction

Consider n nodes independently distributed in $S = [0, 1]^2$ each according to a certain density $f(\cdot)$ and say two nodes $u = (x_u, y_u)$, $v = (x_v, y_v) \in \mathbb{R}^2$ are connected to each other if the Euclidean distance $d(u, v)$ between them is less than r_n. We denote the resulting random geometric graph (RGG) as $G = G(n, r_n, f)$. Throughout the paper we assume the density f on $[0, 1]^2$ satisfies

$$0 < \inf_{x\in[0,1]^2} f(x) \le \sup_{x\in[0,1]^2} f(x) < \infty. \tag{1}$$

Random graphs as described above are important in many applications and properties like emergence of giant component, connectivity and area coverage have been studied before [2,4–6] for a variety of random graphs.

For the case of RGGs, we recollect the pertinent results below for convenience.

Theorem [4,6]. *If $r_n^2 = \frac{c_1}{n}$ for some constant $c_1 > 0$ sufficiently large and the density $f(\cdot)$ satisfies* (1), *then*:

(a) *There exists a constant $\varepsilon = \varepsilon(c_1) > 0$ so that*

$$\mathbb{P}(G \textit{ contains a component } C_G \textit{ such that } \#C_G \ge \varepsilon n) \longrightarrow 1$$

and

$$\frac{\#C_G}{n} \longrightarrow 2\varepsilon \text{ in probability}$$

as $n \to \infty$. *If* $r_n^2 = c_2 \frac{\log n}{n}$ *for some constant* $c_2 > 0$ *and the density* $f(\cdot)$ *satisfies* (1), *we have*:

(b) *If* c_2 *is sufficiently large, then* $\mathbb{P}(G$ *is connected*$) \longrightarrow 1$ *as* $n \to \infty$.

(c) *If* c_2 *is sufficiently small, then* $\liminf_n \mathbb{P}(G$ *is not connected*$) > 0$.

Here and henceforth any constant will always be independent of n and $\#C_G$ denotes the number of nodes in C_G. Part (a) of the above result describes the size of the giant component C_G of G. Parts (b) and (c) describe the behaviour of G in the densely connected regime. Indeed when the density f is uniform, parts (b) and (c) are proved in Corollary 3.1 and Corollary 2.1, respectively, of [4]. The proof for non-uniform f satisfying (1) is analogous. Part (a) and related results are discussed in Chapter 11 of [6]. Also, it is known that $\varepsilon(c_1) \longrightarrow \frac{1}{2}$ as $c_1 \to \infty$ (see Chapters 9 and 11 of [6]).

Not much is known about the graph for intermediate values of r_n. The size of the giant component is not known as a function of r_n. Our main contribution in this paper is developing techniques to analyze the structure of giant component in the intermediate range i.e., when $\frac{c_1}{n} \le r_n^2 \le c_2 \frac{\log n}{n}$ for sufficiently large positive constants c_1, c_2 and obtain estimates on the size and diameter of non-giant components (Theorem 1). The advantage of our approach is that it can also be used to study related problems in RGGs.

Before we state the main result, we define the diameter of a graph. The diameter of any subgraph H of G is defined as

$$\operatorname{diam}(H) = \sup_{u,v} d_H(u,v),$$

where $d_H(u,v)$ represents the graph distance between the nodes u and v in H and the supremum is taken over all pairs u, v belonging to the vertex set of H. We state the main result of the paper below. Let $\mathcal{T}_G$ denote the collection of all components of G. For a fixed $\beta > 0$ we define the following events: Let

$$U_n = U_n(\beta) = \left\{\#\mathcal{T}_G \le \frac{1}{r_n^2} \mathrm{e}^{-\beta n r_n^2}\right\}$$

denote the event that the number of components of G is less than $\frac{1}{r_n^2}\mathrm{e}^{-\beta n r_n^2}$,

$$V_n = V_n(\beta) = \left\{\text{there exists } C_0 \in \mathcal{T}_G \text{ such that } \#C_0 \ge n - n\mathrm{e}^{-\beta n r_n^2}\right\}$$

denote the event that there exists a (giant) component C_0 in $\mathcal{T}_G$ whose size is at least $n - n\mathrm{e}^{-\beta n r_n^2}$ and

$$W_n = W_n(\beta) = V_n \cap \left\{ \sup_{C \in \mathcal{T}_G \setminus C_0} \operatorname{diam}(C) \le \frac{1}{\beta}\left(\frac{\log n}{n r_n^2}\right)^2 \right\}$$

denote the event that the diameter of every component of G other than the giant component C_0 is less than $\frac{1}{\beta}(\frac{\log n}{n r_n^2})^2$.

Theorem 1. *Consider the graph* $G = G(n, r_n, f)$, *where the density* $f(x)$ *satisfies* (1) *and the radius* r_n *satisfies*

$$\frac{c_1}{n} \le r_n^2 \le \frac{c_2 \log n}{n} \tag{2}$$

for some fixed positive constants c_1 *and* c_2. *Let* U_n *and* W_n *be events as defined above and fix* $\delta > 1$. *If* $n r_n^2 \longrightarrow \infty$ *as* $n \to \infty$, *there exists a positive constant* $\beta = \beta(\delta)$ *sufficiently small so that*:

(i) $\mathbb{P}(U_n) \ge 1 - \mathrm{e}^{-\beta n^{1-1/\delta}}$ *and*

(ii) $\mathbb{P}(W_n) \ge 1 - \mathrm{e}^{-\beta n r_n^2}$, *for all* $n \ge 1$.

The above result essentially says whenever r_n is in the intermediate range as in (2), a giant component of G exists with very high probability and moreover it contains nearly all the nodes.

2. Proof of Theorem 1

Divide the unit square S into small $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ closed squares $\{S_i\}_{i\geq 1}$ and choose $\Delta = \Delta_n \in [4, 5]$ so that $\frac{\Delta}{r_n}$ is an integer. We choose such a Δ so that nodes in adjacent squares can be joined by an edge in G. Define S_i to be occupied if it has at least one node and vacant otherwise.

2.1. *Proof of* (i)

We first count the number of vacant squares in the set $\{S_i\}_i$. We then use the fact that for each vacant square S_j, the $\frac{8r_n}{\Delta} \times \frac{8r_n}{\Delta}$ square with same centre as S_j intersects at most 81 distinct components of G to prove (i). The choice of 8 is not crucial and any integer larger than 2 suffices since we only need to estimate the number of components "associated" with S_j. The total number of squares is $t = (\frac{\Delta}{r_n})^2$. To obtain an estimate on the total number of vacant squares, we let $\{Z_i\}_{1\leq i\leq t}$ be Bernoulli random variables taking values either zero or one. We set $Z_i = 1$ if and only if the square S_i is vacant which happens if and only if none of the n nodes are in S_i.

We note that the sum $\sum_i Z_i$ equals k if and only if there are exactly k vacant squares. Since the random variables $\{Z_i\}_i$ are not independent, we cannot evaluate the probability that $\sum_i Z_i = k$ using standard binomial estimates. We therefore proceed as follows. The number of ways of choosing k squares from a total of t squares is $\binom{t}{k}$. The total area of the k squares is $k\frac{r_n^2}{\Delta^2} \geq \frac{kr_n^2}{25}$ since $\Delta \leq 5$. All the k squares chosen are empty with probability at most p_k^n, where

$$p_k = 1 - k \inf_i \int_{S_i} f(x)\,\mathrm{d}x \leq 1 - \beta_0 k r_n^2 \leq \mathrm{e}^{-\beta_0 k r_n^2} \tag{3}$$

and $\beta_0 = \frac{1}{25}\inf_{x\in[0,1]^2} f(x) > 0$. Thus using the inequality $\binom{n}{k} \leq (\frac{n\mathrm{e}}{k})^k$, we have

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{t} Z_i \geq k\right) &\leq \sum_{j=k}^{t}\binom{t}{j} p_j^n \leq \sum_{j=k}^{t}\left(\frac{t\mathrm{e}}{j}\right)^j p_j^n \\ &\leq \sum_{j=k}^{t}\left(\frac{t\mathrm{e}}{j}\right)^j \mathrm{e}^{-j\beta_0 n r_n^2} \leq \sum_{j=k}^{t}\left(\frac{t\mathrm{e}}{k}\right)^j \mathrm{e}^{-j\beta_0 n r_n^2}.\end{aligned}$$

Setting $k = \mathrm{e}t\mathrm{e}^{-\theta n r_n^2}$ for some constant $\theta < \beta_0$ to be determined later and letting $\beta_1 = \beta_0 - \theta$, we get for all sufficiently large n that

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{t} Z_i \geq \mathrm{e}t\mathrm{e}^{-\theta n r_n^2}\right) &\leq \sum_{j=k}^{t}\mathrm{e}^{-j\beta_1 n r_n^2} \\ &\leq \frac{\mathrm{e}^{-k\beta_1 n r_n^2}}{1-\mathrm{e}^{-\beta_1 n r_n^2}} \leq 2\mathrm{e}^{-k\beta_1 n r_n^2} \\ &= 2\exp\left(-\mathrm{e}t\mathrm{e}^{-\theta n r_n^2}\beta_1 n r_n^2\right) \\ &= 2\exp\left(-\beta_1 \mathrm{e}\Delta^2 n\mathrm{e}^{-\theta n r_n^2}\right) \\ &\leq 2\exp\left(-16\mathrm{e}\beta_1 n\mathrm{e}^{-\theta n r_n^2}\right),\end{aligned}$$

where we use $t = \Delta^2 r_n^{-2}$ and $\Delta \ge 4$, respectively, in obtaining the last two expressions. We use the fact that $nr_n^2 \longrightarrow \infty$ to obtain the third inequality. In what follows, the constants $\{\beta_i\}_{i\ge1}$ are not necessarily same in each occurrence. Let $\delta > 1$ be any constant. Since $r_n^2 \le c_2 \frac{\log n}{n}$ for some $c_2 > 0$ (see (2)), we choose θ sufficiently small so that

$$\theta n r_n^2 \le \theta c_2 \log n \le \frac{1}{\delta} \log n.$$

This implies that

$$\mathbb{P}\left(\sum_{i=1}^{t} Z_i \ge \mathrm{e} t \mathrm{e}^{-\theta n r_n^2}\right) \le 2\exp\bigl(-16\mathrm{e}\beta_1 n^{1-1/\delta}\bigr).$$

Also, for each vacant square S_j, the $\frac{8r_n}{\Delta} \times \frac{8r_n}{\Delta}$ square with same centre as S_j intersects at most 81 distinct components of G. Since $t = \frac{\Delta^2}{r_n^2} \le \frac{25}{r_n^2}$, we get from the above equation that

$$\mathbb{P}\bigl(\#\mathcal{T}_G \ge 2025\mathrm{e} r_n^{-2} \mathrm{e}^{-\theta n r_n^2}\bigr) \le 2\exp\bigl(-16\mathrm{e}\beta_1 n^{1-1/\delta}\bigr)$$

and (i) follows.

The rest of the proof is devoted to establishing (ii). The idea is to tile S horizontally and vertically into rectangles and show that each rectangle contains a crossing of edges in the longer direction with high probability. We then join together these crossings to form a "backbone" and show that it forms a part of the giant component. Throughout, we define $K_n = \frac{\log n}{nr_n^2}$ and allow K_n to be an integer. (Later, we show that the tiling is (slightly) modified if K_n is not an integer without any change in the argument.)

From (2), we have that $K_n \ge \frac{1}{c_2}$. For positive integers m_1 and m_2, let R be any $\frac{m_2 r_n}{\Delta} \times \frac{m_1 K_n r_n}{\Delta}$ rectangle contained in S which contains exactly $m_1 m_2 K_n$ of the squares from $\{S_i\}_i$. We define a left–right crossing in R to be any set of distinct squares $L = (S_{j_1}, \dots, S_{j_l})$ such that:

(a) For every i, the squares S_{j_i} and $S_{j_{i+1}}$ share an edge.
(b) S_{j_1} intersects the left face of R and S_{j_l} intersects the right face.

If every square in L is occupied, we say that L is an occupied left–right crossing. We define analogously occupied top–bottom and vacant crossings of R. The only difference in the definition of vacant crossings is that "edge" in condition (a) above is replaced by "corner". Figure 1 illustrates an occupied left–right crossing in a $\frac{m_2 r_n}{\Delta} \times \frac{m_1 K_n r_n}{\Delta}$ rectangle R. The nodes in the rectangle are illustrated as dark dots and the sequence of grey squares form an occupied left–right crossing in R. We need the following estimate on the probability of occurrence of an occupied left–right crossing in R.

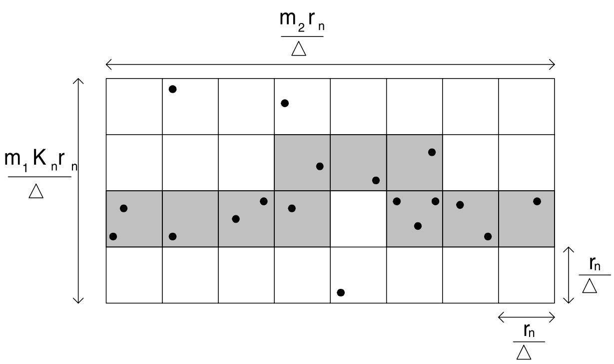


Fig. 1. Occupied left–right crossing in the rectangle R for some $\Delta \ge 4$.

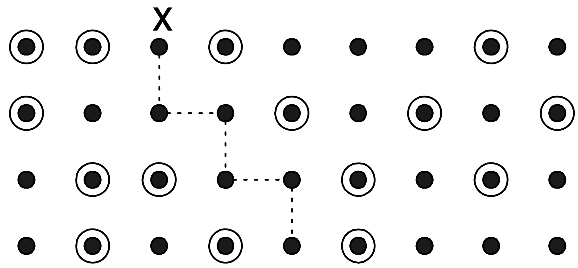


Fig. 2. Vacant top–bottom crossing of a 4×9 rectangle in $\mathbb{Z}^2$ from the site x. Circled sites are occupied.

Lemma 2. *For $n \geq N_0$ (independent of the choices of m_1 and m_2), the event that an occupied left–right crossing occurs in R has probability at least*

$$1 - \frac{m_2}{n^{m_1 \delta_1}} \tag{4}$$

for some constant $\delta_1 > 0$ (independent of the choices of m_1 and m_2).

We use the above estimate to construct a "backbone" of G and thus prove (ii). Before we do so, we prove Lemma 2. The proof is independent of the rest of the proof of Theorem 1.

Proof of Lemma 2. To prove (4), we identify the centre of each square S_i contained in R with a vertex in $\mathbb{Z}^2$ in the natural way. Thus the rectangle R has an equivalent rectangle $\tilde{R}$ consisting of sites in $\mathbb{Z}^2$. Say that a site is occupied if the corresponding square S_i is occupied and vacant otherwise. Analogous to crossings in R, define occupied and vacant crossings in $\tilde{R}$.

We now use the fact that either a left–right occupied crossing or a top–bottom vacant crossing must always occur in $\tilde{R}$ but not both (see e.g., [1] or [3]). To evaluate the probability of a vacant top–bottom crossing, we fix a point x in the top face of $\tilde{R}$ and consider a vacant crossing of length k starting from x (see Fig. 2 for illustration). The area enclosed by the corresponding vacant top–bottom crossing Π_1 in $R \subset \mathbb{R}^2$ is $\frac{kr_n^2}{\Delta^2} \geq \frac{kr_n^2}{25}$, since $\Delta \leq 5$. The probability that a particular node is present in Π_1 is (see (3))

$$\int_{\Pi_1} f(x)\,\mathrm{d}x \geq k\beta_0 r_n^2,$$

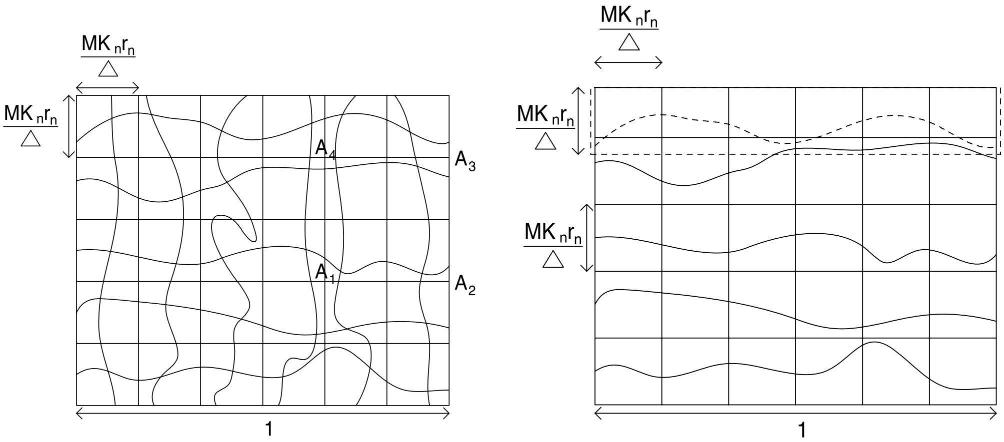
where $\beta_0 = \frac{1}{25}\inf_{x\in[0,1]^2} f(x) > 0$. Therefore the probability that Π_1 is vacant is less than

$$\left(1 - k\beta_0 r_n^2\right)^n \leq \mathrm{e}^{-kn\beta_0 r_n^2}.$$

Since the number of vacant top–bottom crossings of length k starting from x is less than 8^k (at each step no more than eight choices are possible), the probability that there exists a vacant top–bottom crossing of k squares starting from the square S_x with centre x and contained in R is bounded above by $8^k \mathrm{e}^{-kn\beta_0 r_n^2}$. Any top–bottom crossing starting from S_x must necessarily contain at least $m_1 K_n$ and no more than $m_1 m_2 K_n$ squares. Therefore the probability that there exists a vacant top–bottom crossing starting from S_x and contained in R is bounded above by

$$\sum_{k=m_1 K_n}^{m_1 m_2 K_n} 8^k \mathrm{e}^{-k\beta_0 n r_n^2} \leq \left(\mathrm{e}^{-\beta_1 n r_n^2}\right)^{m_1 K_n}$$

for a fixed constant $0 < \beta_1 < \beta_0$ and all $n \geq N_0$, for some constant N_0 independent of the choices of m_1 and m_2. In the above, we use the fact that $nr_n^2 \longrightarrow \infty$ and therefore that $8\mathrm{e}^{-\beta_0 n r_n^2} < \mathrm{e}^{-\beta_1 n r_n^2}$ for all n sufficiently large. Since



(a) (b)

Fig. 3. Construction of the backbone. (a) The event E_n in the unit square. Each horizontal wavy line is an occupied left–right crossing of $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ squares as in Fig. 1. (b) Start horizontal tiling from below. The two topmost $1 \times \frac{MK_n r_n}{\Delta}$ rectangles in the tiling overlap. Perform a vertical tiling analogously.

there are m_2 possibilities for S_x, the probability that there exists a vacant top–bottom crossing of R is bounded above by

$$\begin{aligned} m_2\left(\mathrm{e}^{-\beta_1 n r_n^2}\right)^{m_1 K_n} &= m_2 \mathrm{e}^{-\beta_1 m_1 \log n} \\ &= m_2\left(\frac{1}{n^{\beta_1}}\right)^{m_1} \end{aligned}$$

since $K_n = \frac{\log n}{n r_n^2}$. $\square$

2.2. *Proof of* (ii)

Tile the square S horizontally into a set of rectangles $\mathcal{R}_H$ each of size $1 \times \frac{M r_n K_n}{\Delta}$ and also vertically into rectangles each of size $\frac{M r_n K_n}{\Delta} \times 1$ for some fixed integer constant $M \geq 1$ to be determined later. The argument below is for a perfect tiling as in Fig. 3(a). Otherwise we perform an analogous argument with tiling as in Fig. 3(b). Let R be a fixed $1 \times \frac{MK_n r_n}{\Delta}$ rectangle in the tiling $\mathcal{R}_H$ and let $\delta > 1$ be a fixed constant. From (4), we know that R contains an occupied left–right crossing with probability at least

$$1 - \frac{\Delta}{r_n}\frac{1}{n^{M\delta_1}} \geq 1 - \frac{\Delta}{\sqrt{c_1}}\frac{\sqrt{n}}{n^{M\delta_1}} \geq 1 - \frac{1}{n^{\delta+2}}$$

if M is sufficiently large. Fix such an M. The first inequality above is because $r_n^2 \geq \frac{c_1}{n}$ for some constant c_1 (see (2)). Let E_n^H denote the event that every rectangle in $\mathcal{R}_H$ contains an occupied left–right crossing in G satisfying (a)–(b) described above. The number of rectangles in $\mathcal{R}_H$ is less than

$$\frac{\Delta}{M r_n K_n} \leq \frac{\Delta}{M r_n}\frac{1}{c_2} \leq \frac{\Delta}{M c_2}\frac{\sqrt{n}}{\sqrt{c_1}} \leq D_1\sqrt{n}$$

for some constant $D_1 > 0$. In evaluating the above we again use (2). The first inequality is because $K_n = \frac{\log n}{n r_n^2} \geq \frac{1}{c_2}$ by our choice of r_n in (2) and the second inequality follows because $r_n^2 \geq \frac{c_1}{n}$. It follows that

$$\mathbb{P}\left(E_n^H\right) \geq 1 - \frac{D_1\sqrt{n}}{n^{\delta+2}} \geq 1 - \frac{1}{n^{\delta+1}}$$

for all n sufficiently large. Following an analogous analysis for the vertically tiled rectangles described in the first paragraph of the proof and defining an analogous event E_n^V with occupied top–bottom crossings, we have that $\mathbb{P}(E_n^V) \geq 1 - \frac{1}{n^{\delta+1}}$. Thus if $E_n = E_n^H \cap E_n^V$, we have that

$$\mathbb{P}(E_n) \geq 1 - \frac{2}{n^{\delta+1}}. \tag{5}$$

In Fig. 3(a), we depict the occurrence of the event E_n. We see that the event E_n results in a connected set of $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ squares $\mathcal{B} \subseteq \{S_i\}_i$ forming a "backbone" of crossings in S. Let C_0 denote the component of G containing nodes in $\mathcal{B}$.

In the above, we have assumed that $K_n = \frac{\log n}{nr_n^2}$ is an integer. If not, we set $K_n = \lceil \frac{\log n}{nr_n^2} \rceil$ and starting from the base of the square S, we perform an analogous horizontal tiling as above. The only difference is that the two topmost rectangles could overlap as seen in Fig. 3(b). A similar situation occurs in the vertical tiling. Following an analogous analysis as above, we obtain (5) and a corresponding backbone. The rest of the argument below remains unchanged.

We note that the tiling of S into vertical and horizontal rectangles induces a tiling of S into $\frac{Mr_nK_n}{\Delta} \times \frac{Mr_nK_n}{\Delta}$ size squares $\{S_i'\}_i$. If the event E_n occurs, then the resulting backbone $\mathcal{B}$ (and hence the component C_0) intersects each square S_i' "vertically" and "horizontally" as shown in Fig. 3(a). Therefore, if there exists a connected component C of G distinct from C_0, it must necessarily be contained in a $\frac{2MK_n}{\Delta} \times \frac{2MK_n}{\Delta}$ square with centre at some $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ square S_i. In Fig. 4, the square $A_1A_2A_3A_4$ of Fig. 3(a) is magnified and a component C distinct from C_0 is shown. The centre of the hatched $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ square is also the centre of $A_1A_2A_3A_4$.

Clearly in such a component C, the minimum number of edges traversed in going from any node u to any other node v is at most $(\frac{2MK_n}{\Delta})^2 < (2MK_n)^2$ and therefore $\mathrm{diam}(C) < (2MK_n)^2$. To summarize, so far we have proved that if event E_n occurs, then a backbone $\mathcal{B}$ and hence the component C_0 containing all the nodes in squares comprising the backbone and possibly other nodes exist. Moreover, any component of G distinct from C_0 has diameter less than $(2MK_n)^2$. Recall that $\mathcal{T}_G$ is the set of all components of G and for $\theta > 0$ let

$$F_n = F_n(\theta) = \left\{ \sum_{C \in \mathcal{T}_G : \mathrm{diam}(C) < (2MK_n)^2} \#C < n\mathrm{e}^{-\theta nr_n^2} \right\}$$

denote the event that the sum of sizes of components whose diameter does not exceed $(2MK_n)^2$ is less than $n\mathrm{e}^{-\theta nr_n^2}$. We have the following estimate on probability of occurrence of the event F_n.

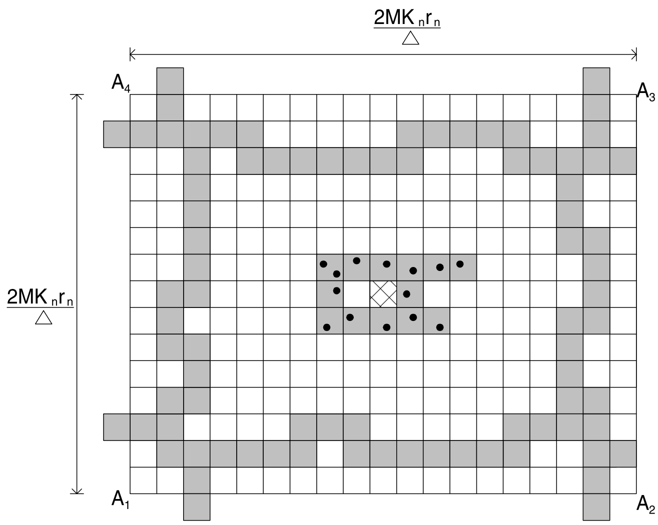


Fig. 4. The square $A_1A_2A_3A_4$ in Fig. 3(a) is magnified to show a component not attached to the backbone.

Lemma 3. *We have*

$$\mathbb{P}(F_n) \geq 1 - \mathrm{e}^{-\theta_1 n r_n^2} \tag{6}$$

for some positive constants θ *and* θ_1.

Before we prove the above result, we complete the proof of (ii). Whenever $E_n \cap F_n$ occurs, the component C_0 contains at least $n - n\mathrm{e}^{-\theta n r_n^2}$ nodes and is therefore the giant component. Also, the diameter of any non-giant component is less than $(2MK_n)^2$. Choosing $\theta_1 > 0$ smaller if necessary, we have from (5) and (6) that the event $E_n \cap F_n$ occurs with probability

$$\mathbb{P}(E_n \cap F_n) \geq 1 - \mathrm{e}^{-\theta_1 n r_n^2} - \frac{2}{n^{\delta+1}} \geq 1 - 2\mathrm{e}^{-\theta_1 n r_n^2}$$

for all n sufficiently large. In the above estimate, we have used the fact (2) that $nr_n^2 \leq c_2 \log n$ for some positive constant c_2. This proves (ii) and hence Theorem 1. The proof of Lemma 3 is independent of the proof of Theorem 1 and is provided below.

Proof of Lemma 3. Say that a set of squares $\mathcal{C} \subseteq \{S_i\}_i$ is a cluster if they form a connected set in $\mathbb{R}^2$. We say that the cluster $\mathcal{C}$ is occupied if every square in the cluster is occupied.

Fix i and consider the square S_i. If S_i is occupied, denote $\mathcal{C}_i$ to be the maximal occupied cluster containing S_i. Set X_i to be the number of nodes in $\mathcal{C}_i$ if $\mathcal{C}_i$ is contained in the $2(2MK_n)^2 r_n \times 2(2MK_n)^2 r_n$ square S_i^{in} with same centre as S_i. Otherwise set X_i to be zero. Thus, $\sum_i X_i$ is an upper bound on the sum of size of components whose diameter is less than $2(2MK_n)^2$. In the beginning of the proof of (ii), we recall that to obtain the estimate $(2MK_n)^2$ on the diameter of a component not attached to the backbone, we had considered a $2MK_n \times 2MK_n$ square appropriately centred (like $A_1A_2A_3A_4$ in Fig. 4). In this subsection, however, we are not given any information regarding the backbone. Therefore, to obtain a bound on the size of a component whose diameter is less than $(2MK_n)^2$ the only information we have is that the component is enclosed in a (slightly bigger) $2(2MK_n)^2 \times 2(2MK_n)^2$ square.

We first estimate $\mathbb{P}(\{\#\mathcal{C}_i = k\} \cap \{X_i \neq 0\})$ for $k \geq 1$. Suppose that $X_i \neq 0$ and therefore that the cluster $\mathcal{C}_i$ is contained in the square S_i^{in}. Our aim now is to obtain a sufficiently large number of vacant squares "attached to" $\mathcal{C}_i$. Consider $\mathcal{C}_i$ as a set in $\mathbb{R}^2$ and let $\partial_1, \ldots, \partial_T$ be its disjoint boundaries. Each ∂_i is a circuit of edges $(e_{i,1}, \ldots, e_{i,L_i})$ (not necessarily self-avoiding) such that $e_{i,1}$ and e_{i,L_i} touch each other. Since $\mathcal{C}_i$ is connected, one of the boundaries, say ∂_1, contains all squares of $\mathcal{C}_i$ and all the other boundaries in its interior. Also, any square S_j that has an edge $e_{1,j} \in \partial_1$ and not contained in $\mathcal{C}_i$ is necessarily vacant.

Let π_1 denote the set of distinct vacant squares that contain some edge in ∂_1. The path ∂_1 contains $L_1 \geq 2$ edges of which at least $\frac{L_1}{2}$ of them have an endvertex in the interior of the unit square S. (Here we use the fact that the cluster $\mathcal{C}_i$ is contained in S_i^{in}. If we did not have such a bounding box for the cluster $\mathcal{C}_i$, the above statement will not hold; e.g. consider the event that each square in $\{S_k\}_k$ contains at least one node.) From the discussion in the previous paragraph, each such "interior" edge has a vacant square "attached" to it. Since each vacant square is counted at most four times (once for each of its four edges), this implies that $\#\pi_1 \geq \frac{L_1}{8}$. In Fig. 5, the dark grey square is S_i and the grey squares form $\mathcal{C}_i$. The set of vacant squares π_1 is shown by the squares marked Π and the curve of thick lines represents ∂_1.

To compute the probability that such a vacant set of squares occurs, we set the centre of S_i to be the origin and draw X- and Y-axes parallel to the sides of S_i. Let $e_{1,\text{last}}$ be the "last" edge in ∂_1 that intersects the X-axis at $(x_{\text{last}}, 0)$. In other words, if an edge $e_{1,j}$ in ∂_1 intersects the X-axis at $(x_j, 0)$, then $x_{\text{last}} > x_j$. In Fig. 5, the edge $e_{1,\text{last}}$ is also shown. Clearly, there are at most L_1 possibilities for the location of edge $e_{1,\text{last}}$. Also, the number of choices for ∂_1 starting from $e_{1,\text{last}}$ is less than 4^{L_1}.

Now, the total area of squares in π_1 is at least $\frac{L_1}{8}\frac{r_n^2}{\Delta^2} \geq \frac{L_1}{8}\frac{r_n^2}{25}$ since $\Delta \leq 5$. Given ∂_1, with probability at least $\frac{L_1}{8}\beta_0 r_n^2$ a particular node is present in π_1 where $\beta_0 = \frac{1}{25}\inf_{x\in[0,1]^2} f(x) > 0$ is as in (3). Therefore with probability at most

$$\left(1 - \frac{1}{8}\beta_0 L_1 r_n^2\right)^n \leq \mathrm{e}^{-\beta_0 L_1 n r_n^2/8}$$

none of the n nodes are present in π_1.

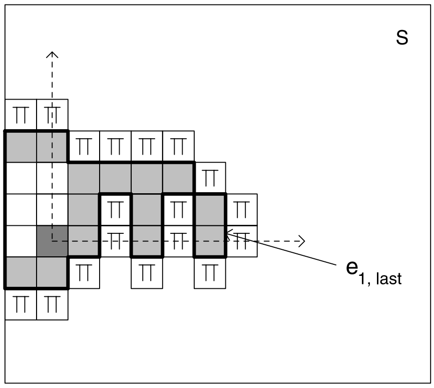


Fig. 5. The occupied cluster $\mathcal{C}_i$ and the set of vacant squares π_1 (marked by the symbol Π) are shown for the square S_i that is denoted by the dark square.

If $\mathcal{C}_i$ contains k squares, then the number of edges L_1 in ∂_1 satisfies $\frac{\sqrt{k}}{4} \leq L_1 \leq 4k$. The upper bound is clear. To see why the lower bound is true, suppose that ∂_1 has less than $\frac{\sqrt{k}}{4}$ edges. It is then necessary that ∂_1 is contained in the $\frac{\sqrt{k}}{2}\frac{r_n}{\Delta} \times \frac{\sqrt{k}}{2}\frac{r_n}{\Delta}$ square S_{pk} with the same centre as S_i. The square S_{pk} contains at most $\frac{k}{4}$ squares from $\{S_j\}_j$. This is a contradiction since the path ∂_1 contains $\mathcal{C}_i$ in its interior and $\mathcal{C}_i$ contains k squares. Thus for $k \geq 1$ we have from the above discussion that

$$\begin{aligned}
&\mathbb{P}\big(\{\#\mathcal{C}_i = k\} \cap \{X_i \neq 0\}\big) \\
&\quad \leq \sum_{\sqrt{k}/4 \leq l \leq 4k} \mathrm{e}^{-l\beta_0 n r_n^2/8} l 4^l \\
&\quad \leq 4k \sum_{\sqrt{k}/4 \leq l \leq 4k} \big(4\mathrm{e}^{-\beta_0 n r_n^2/8}\big)^l \\
&\quad \leq k\mathrm{e}^{-\theta_0 n r_n^2 \sqrt{k}}
\end{aligned} \tag{7}$$

for a fixed positive constant $\theta_0 < \frac{\beta_0}{40}$ and all $n \geq N_0$, where N_0 is a constant that does not depend on k. Here we use the fact that $nr_n^2 \longrightarrow \infty$ and hence that $4\mathrm{e}^{-\beta_0 n r_n^2/8} < \mathrm{e}^{-5\theta_0 n r_n^2}$ for some constant $\theta_0 > 0$ and for all sufficiently large n. Letting $N(A)$ denote the number of nodes in the set A, we therefore have that

$$\begin{aligned}
\mathbb{E}X_i &= \mathbb{E}\sum_{\mathcal{C}_0}\sum_{S_j \in \mathcal{C}_0} N(S_j)\mathbf{1}(\mathcal{C}_i = \mathcal{C}_0)\mathbf{1}(X_i \neq 0) \\
&= I_1 + I_2,
\end{aligned}$$

where the summation in the first line is over all clusters $\mathcal{C}_0$ that contain the square S_i and are contained in S_i^{in}. In the above equation,

$$I_1 = \mathbb{E}\sum_{\mathcal{C}_0}\sum_{S_j \in \mathcal{C}_0} N(S_j)\mathbf{1}(\mathcal{C}_i = \mathcal{C}_0)\mathbf{1}\big(N(\mathcal{C}_0) \geq 2\mathrm{e}k\delta_0 n r_n^2\big)\mathbf{1}(X_i \neq 0),$$

$I_2 = \mathbb{E}X_i - I_1$ and $\delta_0 = \frac{1}{16}\sup_{x \in S} f(x)$.

To evaluate I_1 and I_2, we need a couple of preliminary estimates. For a fixed $\mathcal{C}_0$ containing k squares, we estimate $\mathbb{P}(N(\mathcal{C}_0) \geq 2\mathrm{e}k\delta_0 n r_n^2)$ first. Indeed since a particular node is present in $\mathcal{C}_0$ with probability at most $q_k = k\delta_0 r_n^2$, we have that

$$\begin{aligned}\mathbb{P}\big(N(\mathcal{C}_0) \geq 2\mathrm{e}nq_k\big) &\leq \sum_{2\mathrm{e}nq_k \leq j \leq n} \binom{n}{j} q_k^j \\ &\leq \sum_{2\mathrm{e}nq_k \leq j \leq n} \left(\frac{n\mathrm{e}}{j}\right)^j q_k^j \\ &\leq \sum_{2\mathrm{e}nq_k \leq j \leq n} \left(\frac{n\mathrm{e}}{2\mathrm{e}nq_k}\right)^j q_k^j \\ &\leq \sum_{j \geq 2\mathrm{e}nq_k} \left(\frac{1}{2}\right)^j \\ &\leq \mathrm{e}^{-2\beta_2 knr_n^2} \end{aligned} \tag{8}$$

for some positive constant β_2 independent of k, i and the choice of $\mathcal{C}_0$. In the third inequality above, we have used the estimate $\binom{n}{k} \leq (\frac{n\mathrm{e}}{k})^k$. Also, the expected number of nodes in any square S_i is bounded above by

$$\sup_j \mathbb{E}N(S_j) = n \sup_j \int_{S_j} f(x)\,\mathrm{d}x \leq n \sup_{x\in[0,1]^2} f(x)\frac{r_n^2}{\Delta^2} \leq D_1 nr_n^2 \tag{9}$$

for some positive constant D_1 since $\sup_{x\in[0,1]^2} f(x) < \infty$ (see (1)) and $\Delta \geq 4$. Analogously,

$$\sup_j \mathbb{E}N(S_j)^2 \leq D_2\big(nr_n^2\big)^2 \tag{10}$$

for some positive constant D_2.

To evaluate I_1, we now use Cauchy–Schwarz inequality to obtain that

$$\begin{aligned} I_1 &\leq \sum_{k\geq 1}\sum_{\#\mathcal{C}_0=k}\sum_{S_j\in\mathcal{C}_0} \mathbb{E}N(S_j)\mathbf{1}\big(N(\mathcal{C}_0) \geq 2\mathrm{e}k\delta_0 nr_n^2\big) \\ &\leq \sum_{k\geq 1}\sum_{\#\mathcal{C}_0=k}\sum_{S_j\in\mathcal{C}_0} \big(\mathbb{E}N^2(S_j)\big)^{1/2}\mathbb{P}\big(N(\mathcal{C}_0) \geq 2\mathrm{e}k\delta_0 nr_n^2\big)^{1/2} \\ &\leq D_3 nr_n^2 \sum_{k\geq 1}\sum_{\#\mathcal{C}_0=k}\sum_{S_j\in\mathcal{C}_0} \mathrm{e}^{-k\beta_2 nr_n^2} \end{aligned}$$

for some positive constant D_3 independent of i. In obtaining the final estimate, we use (8) and (10) and the notation $\sum_{\#\mathcal{C}_0=k}$ refers to the sum over all clusters $\mathcal{C}_0$ containing k squares of which one of them is S_i. Since the number of such clusters is less than 8^k, we get

$$I_1 \leq D_3 nr_n^2 \sum_{k\geq 1} k8^k \mathrm{e}^{-k\beta_2 nr_n^2} \leq D_4 nr_n^2 \mathrm{e}^{-\beta_3 nr_n^2}$$

for some positive constants D_4 and β_3, independent of i.

To evaluate I_2 we write

$$\begin{aligned} I_2 &= \mathbb{E}\sum_{k\geq 1}\sum_{\#\mathcal{C}_0=k}\sum_{S_j\in\mathcal{C}_0} N(S_j)\mathbf{1}(\mathcal{C}_i=\mathcal{C}_0)\mathbf{1}\big(N(\mathcal{C}_0) \leq 2\mathrm{e}k\delta_0 nr_n^2\big)\mathbf{1}(X_i \neq 0) \\ &\leq 2\mathrm{e}\delta_0 nr_n^2 \mathbb{E}\sum_{k\geq 1} k \sum_{\#\mathcal{C}_0=k}\sum_{S_j\in\mathcal{C}_0} \mathbf{1}(\mathcal{C}_i=\mathcal{C}_0)\mathbf{1}(X_i \neq 0) \end{aligned}$$

$$= 2\mathrm{e}\delta_0 n r_n^2 \mathbb{E} \sum_{k\geq 1} k^2 \sum_{\#\mathcal{C}_0 = k} \mathbf{1}(\mathcal{C}_i = \mathcal{C}_0)\mathbf{1}(X_i \neq 0)$$

$$= 2\mathrm{e}\delta_0 n r_n^2 \sum_{k\geq 1} k^2 \mathbb{P}\big(\{\#\mathcal{C}_i = k\} \cap \{X_i \neq 0\}\big)$$

$$\leq 2\mathrm{e}\delta_0 n r_n^2 \sum_{k\geq 1} k^3 \mathrm{e}^{-\theta_0 n r_n^2 \sqrt{k}} \leq D_5 n r_n^2 \mathrm{e}^{-\beta_5 n r_n^2}$$

for some positive constants D_5 and β_5 independent of i, where the second inequality follows from the estimate (7). From the estimates of I_1 and I_2, we therefore have that

$$\mathbb{E} X_i \leq D_6 n r_n^2 \mathrm{e}^{-\beta_6 n r_n^2} \tag{11}$$

for some positive constants D_6 and β_6 independent of i.

The number of squares in $\{S_i\}_i$ is $\Delta^2 r_n^{-2}$. By Markov inequality, we therefore have for $\theta > 0$ that

$$\mathbb{P}\left(\sum_{i=1}^{\Delta^2 r_n^{-2}} X_i \geq n\mathrm{e}^{-\theta n r_n^2}\right) \leq \frac{\sum_i \mathbb{E} X_i}{n} \mathrm{e}^{\theta n r_n^2}$$

$$\leq \big(\Delta^2 r_n^{-2}\big) \frac{D_6 n r_n^2 \mathrm{e}^{-\beta_6 n r_n^2}}{n} \mathrm{e}^{\theta n r_n^2} \leq D_7 \mathrm{e}^{-\theta_1 n r_n^2}$$

for some positive constants θ_1 and D_7, if θ is sufficiently small. Since $F_n = \{\sum_i X_i < n\mathrm{e}^{-\theta n r_n^2}\}$, this proves the lemma. □

Acknowledgments

I thank Professor Rahul Roy for a careful reading of the manuscript and a referee for comments which led to an improvement of the paper. The support from a National Board for Higher Mathematics scholarship and a Sandwich Ph.D. programme scholarship is gratefully acknowledged. I thank Kristina Mattson for help with the typesetting.

References

[1] B. Bollobas and O. Riordan. *Percolation*. Cambridge Univ. Press, New York, 2006. MR2283880

[2] M. Franceschetti, O. Dousse, D. N. C. Tse and P. Thiran. Closing gap in the capacity of wireless networks via percolation theory. *IEEE Trans. Inform. Theory* **53** (2007) 1009–1018. MR2302808

[3] G. Grimmett. *Percolation*, 2nd edition. *Grundlehren der Mathematischen Wissenschaften* **321**. Springer, Berlin, 1999. MR1707339

[4] P. Gupta and P. R. Kumar. Critical power for asymptotic connectivity in wireless networks. In *Stochastic Analysis, Control, Optimization and Applications* 547–566. *Systems Control Found. Appl.* Birkhäuser, Boston, MA, 1999.

[5] S. Muthukrishnan and G. Pandurangan. The bin-covering technique for thresholding random geometric graph properties. In *Proc. SODA* 989–998. ACM, New York, 2005. MR2298358

[6] M. Penrose. *Random Geometric Graphs*. Oxford Studies in Probability **5**. Oxford Univ. Press, Oxford, 2003. MR1986198