# Infection Spread and Stability in Random graphs

**Ghurumuruhan Ganesan**
Theoretical Statistics and Mathematics Unit
Indian Statistical Institute, New Delhi
7 S. J. S. Sansanwal Marg
New Delhi 110016
INDIA

Thesis submitted to the Indian Statistical Institute in partial fulfilment of the requirements for the award of the degree of
Doctor of Philosophy

# Acknowledgements

I thank my advisor, Professor Rahul Roy, for his support and encouragement throughout the thesis. I thank him for his careful reading and crucial comments on the chapters in my thesis and research articles. He introduced me to various topics of research and I have benefitted immensely from discussions with him. I thank him for his invaluable advice.

I thank the referees for helpful comments in the review of my thesis. I thank the National Board of Higher Mathematics for providing the scholarship for my PhD studies. I thank the faculty of ISI for imparting knowledge to me through courses, seminars and personal communications. The staff at ISI have been very helpful to me and I take this opportunity to convey my thanks to them. I thank my friends for making my stay at ISI memorable. I thank my family for their love and support. I conclude by thanking my mother for her support and encouragement.

# Contents

# Chapter 1

# Introduction

## 1.1 Random Graphs

We briefly describe a few models of random graphs that arise in different applications.

### Erdös-Rényi Graphs

Consider $n$ fixed nodes $a_1, ..., a_n$ in the plane and for each $i$ and $j$, join $a_i$ and $a_j$ independently by an edge with probability $p_n$. The random graph so obtained was introduced by Erdös (1947) and is called the Erdös-Rényi (ER) binomial random graph. This is a slight variant of the uniform random graph studied by Erdös and Rényi (1960). The ER $(n, m)$-uniform random graph is defined on the sample space $\Omega_0$ consisting of the set of all graphs with $n$ vertices and $m$ edges and assigns equal probability for each graph in $\Omega_0$. ER graphs as described above are useful in the study of reliability in networks (see e.g. Janson, Luczak and Rucinski (2000)).

Various properties of the above graphs including emergence of giant component and diameter of the giant component have been studied (see e.g. Bollobás (1985)). Bollobás and Thomason (1987) have studied sharp threshold properties of ER random graphs and more generally, of arbitrary random subsets. Chromatic number of ER random graphs have also been extensively studied (see e.g. Janson, Luczak and Rucinski (2000)). Chromatic number of a graph is the minimum number of distinct colours needed to colour the vertices so that no two adjacent vertices share the same colour. Shamir and Spencer (1987) proved concentration of chromatic number for dense random graphs. Bollobas (1988) provided sharp estimates for chromatic number of dense random graphs using martingale inequalities.

We briefly describe giant component and connectivity regimes for ER binomial random graph. Analogous properties hold for ER uniform random graph (see Janson, Luczak and Rucinski (2000)). Suppose that $p_n = \frac{\lambda}{n}$ for some constant $\lambda > 0$. It is well-known (see e.g. Durrett (2006)) that if $\lambda < 1$, then the diameter of the largest cluster is less than $C_1 \log n$ with probability $1 - o(1)$ as $n \to \infty$, for some constant $C_1 > 0$. If $\lambda > 1$, then with probability $1 - o(1)$, there is a unique "giant component" containing roughly $\theta n$ nodes for some constant $\theta \in (0, 1)$. Moreover, every other component has less than $C_2 \log n$ nodes with probability $1 - o(1)$ as $n \to \infty$, for some constant $C_2 > 0$. Thus $\lambda = 1$ is the "critical" intensity beyond which the giant component begins to appear.

On the other hand, suppose that edges are present with probability $p_n = \frac{a \log n}{n}$ for some constant $a > 0$. We know that (see e.g. Durrett (2006)) if $a > 1$, the graph is connected with probability $1 - o(1)$ as $n \to \infty$. If $a < 1$, the graph is disconnected with probability $1 - o(1)$ as $n \to \infty$. This determines the critical value for the "connectivity regime". This dual critical behaviour is typical of such random graphs and is also present in random geometric graphs described below.

## Random Geometric Graphs

Closely related to ER graphs is the random geometric graph (RGG). The study of random geometric graphs originated with the modelling of communication networks (see e.g. Gilbert (1961)). In RGGs $n$ nodes are spatially distributed in a unit square each independently according to a certain density and two nodes $u$ and $v$ are joined to each other if their Euclidean distance between them is less than $r_n$. A slight variation of the above connectivity model is the Poisson Boolean model where the nodes are distributed according to a Poisson process and percolation properties of such models have also been studied (see Meester and Roy (1996)). RGGs are a particular case of the more general random connection model where two nodes $u$ and $v$ are joined to each other with probability $p(u, v)$. See Meester and Roy (1996) and references therein for results on the general random connection model.

Giant component regime and connectivity regime for RGGs have been extensively studied; see Sarkar (1995), Gupta and Kumar (1998) and Penrose (2003). We briefly summarize the pertinent results. Consider $n$ nodes independently distributed in the unit square $S = \left[-\frac{1}{2}, \frac{1}{2}\right]^2$ each according to a certain density $f$ satisfying

$$0 < \inf_{x \in S} f(x) \leq \sup_{x \in S} f(x) < \infty. \tag{1.1}$$

Connect two nodes $u$ and $v$ by an edge $e$ if the Euclidean distance $d(u,v)$ between them is less than $r_n$. The resulting random geometric graph (RGG) is denoted as $G = G(n, r_n, f)$.

**Theorem.** *(Penrose (2003)) If* $r_n^2 = \frac{c_1}{n}$ *for some constant* $c_1 > 0$ *sufficiently large and the density* $f(.)$ *is uniform, then:*
*(a) There exists a constant* $\epsilon = \epsilon(c_1) > 0$ *so that*

$$\mathbb{P}(G \text{ contains a component } C_G \text{ such that } \#C_G \geq \epsilon n) \longrightarrow 1$$

*and*

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

*as* $n \to \infty$. *If* $r_n^2 = c_2 \frac{\log n}{n}$ *for some constant* $c_2 > 0$ *and the density* $f(.)$ *satisfies (2.1), we have:*
*(b) If* $c_2$ *is sufficiently large, then* $\mathbb{P}(G \text{ is connected}) \longrightarrow 1$ *as* $n \to \infty$.
*(c) If* $c_2$ *is sufficiently small, then* $\liminf_n \mathbb{P}(G \text{ is not connected}) > 0$.

Also, $\epsilon(c_1) \to \frac{1}{2}$ as $c_1 \to \infty$; (see Chapters 9 and 11 of Penrose (2003)). 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 Gupta and Kumar (1998). The proof for non-uniform $f$ satisfying (2.1) is analogous (see e.g. Penrose (2003)). Part (a) and related results are discussed in Chapter 2 of Sarkar (1995) and Chapter 11 of Penrose (2003).

We remark that though the critical values for connectivity and giant component regime look similar for ER graphs and RGGs, the proofs are different. Our thesis concerns with infection spread on RGGs and rate of convergence for functionals of random graphs based on Poisson processes. The next chapter determines the size of giant component when $r_n$ is below the connectivity regime but $nr_n^2 \longrightarrow \infty$. The third chapter deals with infection spread in

RGGs and we use results from the second chapter to estimate the size of the total infected set. The final chapter studies convergence rate of locally determinable Poisson functionals.

## 1.2 Size of the giant component in RGGs

Our main aim in this chapter is to estimate the size of the giant component in the random geometric graph and study related properties.

In Chapter 2 we study the structure of giant component in the intermediate range i.e., when

$$c_1 \leq nr_n^2 \leq c_2 \log n \quad \text{and} \quad nr_n^2 \longrightarrow \infty, \tag{1.2}$$

for some positive constants $c_1, c_2$ and obtain estimates on the size and diameter of non-giant components. In our main result of this chapter, we show that if (1.2) is satisfied, then the giant component of $G$ contains at least $n - ne^{-\beta nr_n^2}$ nodes with probability at least $1 - o(1)$ as $n \to \infty$, for some constant $\beta > 0$. We also obtain estimates on the diameter and number of the non-giant components of $G$. The advantage of our approach is that it can also be used to study related problems in RGGs. The results of this chapter are also used to study infection spread in RGGs discussed in the next chapter.

## 1.3 Infection spread in RGGs

We consider the random geometric graph $G = G(n, r_n, f)$ as described above. To study the spread of infection in $G$, we equip each edge $e$ of $G$ with a passage time $t(e)$ that is exponentially distributed with unit mean (Durrett and Liu (1988), Gopalan et al (2011)). The passage times of distinct edges are independent. At time $t = 0$, the node $x_0$ closest to the origin in $S$ is infected. Any node $x_1$ that shares an edge $e$ with $x_0$ is infected after time $t(e)$. This process continues and infected nodes stay in that state forever. What is the minimum time elapsed after which no new nodes are infected? How many nodes are ultimately infected by the above process? In this paper we provide sharp bounds for the above two questions. The main tool we use to describe our results is the speed of infection spread.

We define the infection process on the probability space $(\Psi, \mathcal{H}, \mathbb{P})$. For any set $A \subseteq \mathbb{R}^2$ and $\alpha > 0$, define $\alpha A = \cup_{x \in A}\{\alpha x\}$ to be the dilation of $A$

by factor $\alpha$. At time $t = 0$, the node $x_0$ closest to the origin is infected. Let $G(x_0)$ denote the connected cluster of nodes in $G$ containing $x_0$. Let $I(t)$ be the set of nodes of $G(x_0)$ infected up to time $t$.

We say that infection spreads at speed at least $v_{n,low}$ if there exists functions $0 \leq a(x) = o(x)$ and $0 \leq g(x) = o(x)$ as $x \to \infty$ such that

$$\mathbb{P}\left(\bigcap_{a(r_n^{-1})\leq m\leq r_n^{-1}-g(r_n^{-1})}\left\{\left(G(x_0)\setminus I\left(\frac{m}{v_{n,low}}\right)\right)\bigcap mr_nS = \phi\right\}\right) = 1 - o(1) \tag{1.3}$$

as $n \to \infty$. In other words, we want all nodes of $G(x_0)$ contained in $mr_nS$ to be infected within time $\frac{m}{v_{n,low}}$. This must happen for "nearly all" indices $m$. We say that the speed is at most $v_{n,up}$ if there exists functions $0 \leq a(x) = o(x)$ and $0 \leq g(x) = o(x)$ as $x \to \infty$ such that

$$\mathbb{P}\left(\bigcap_{a(r_n^{-1})\leq m\leq r_n^{-1}-g(r_n^{-1})}\left\{I\left(\frac{m}{v_{n,up}}\right)\subseteq mr_nS\right\}\right) = 1 - o(1).$$

In our main result of Chapter 3 we prove that if (1.2) is satisfied, then the infection spreads with speed at least $D_1 n r_n^2$ and at most $D_2 n\sqrt{n}\log n$ for some positive constants $D_1$ and $D_2$. This is unlike regular lattices (like e.g. $\mathbb{Z}^2$) where the speed of infection spread is a constant.

The traditional subadditive methods (see e.g. Smythe and Wierman (2008)) of first passage percolation are not directly suitable in our scenario and we develop new techniques to establish the bound above. Finally we use results from Chapter 2 to obtain sharp bounds on the eventual size of the infected set.

## 1.4 Convergence rate of locally determinable Poisson functionals

Let $\mathcal{N}$ be a Poisson process with intensity measure $\Lambda(.)$ in $\mathbb{R}^d$ and place an independent mark $t_x$ on each point $x$ of $\mathcal{N}$. Let $\mathcal{N}_M$ be the resulting marked process and let $f(x) = f(x, \mathcal{N}_M), x \in \mathcal{N}$ be a 'locally determinable function' (for a more formal definition see (i)-(iv) in Chapter 4). Letting

$W = \left[-\frac{1}{2}, \frac{1}{2}\right]^d$, we evaluate the rate of convergence of

$$\frac{1}{\Lambda(nW)} \sum_{x \in \mathcal{N} \cap nW} f(x)$$

to its mean as $n \to \infty$, in terms of the decay rate of the radius of determinability.

Broadly speaking, there are at least two approaches towards determining convergence rate: regularity and stability. Regularity of a functional essentially requires that the value of the functional does not change much upon adding or removing a few points in the configuration. For graph functionals of binomial point processes, estimating such a change allows one to directly apply McDiarmid type concentration inequalities thereby obtaining exponential decay (see e.g., Chapter 3, Steele (1987), Baccelli and Bordenave (2005)). For Poisson point processes, one needs an additional step of conditioning on the number of points. Significant work, however, involving the geometry of the graph may be needed in establishing regularity. We seek to obtain concentration estimates (that are possibly weaker) with properties that are in a sense, easy to identify and calculate.

Penrose and Yukich (2003), Baryshnikov and Yukich (2005) study weak convergence of functionals using stability as a criterion. Roughly speaking, a functional is said to be stabilizing if local changes to the configuration does not affect the value of the functional far from the origin. The above works study weak convergence of such functionals via the objective method that essentially approximates inhomogenous Poisson processes by locally homogenous Poisson processes.

In a certain sense, stability highlights the locally determinable property of the functional and is quantified by the radius of determinability. Our principal aim in this chapter is to study how convergence rate varies with the decay rate of the radius of determinability. We illustrate using functionals of the Poisson Voronoi tessellation and the Poisson Boolean model.

# Chapter 2

# Size of the Giant Component in a Random Geometric Graph

## 2.1 Introduction

Consider $n$ nodes independently distributed in the unit square $S = \left[-\frac{1}{2}, \frac{1}{2}\right]^2$ each according to a certain density $f$ satisfying

$$0 < \inf_{x \in S} f(x) \leq \sup_{x \in S} f(x) < \infty. \tag{2.1}$$

Connect two nodes $u$ and $v$ by an edge $e$ if the Euclidean distance $d(u, v)$ between them is less than $r_n$, where

$$c_1 \leq nr_n^2 \leq c_2 \log n \quad \text{and} \quad nr_n^2 \longrightarrow \infty \tag{2.2}$$

for some positive constants $c_1$ and $c_2$ and as $n \to \infty$. The resulting random geometric graph (RGG) is denoted as $G = G(n, r_n, f)$. Here and henceforth all constants are independent of $n$.

Our main aim in this chapter is to estimate the size of the giant component in $G$ and study its related properties. The results of this chapter are of independent interest and are also used to determine the size of infected set in the spread of infection described in Chapter 3. We briefly describe the notation. The diameter of any subgraph $H$ of $G$ is defined as

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

where $d_H(u,v)$ represents the graph distance between the nodes $u$ and $v$ 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

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

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

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

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

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

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

**Theorem 2.1.** *Consider the graph $G = G(n, r_n, f)$, where the density $f(x)$ satisfies (2.1) and the radius $r_n$ satisfies (2.2) for some fixed positive constants $c_1$ and $c_2$. Let $U_n$ and $W_n$ be events as defined above and fix $\delta > 1$. There exists a positive constant $\beta = \beta(\delta)$ sufficiently small so that:*
*(i) $\mathbb{P}(U_n) \geq 1 - e^{-\beta n^{1-1/\delta}}$ and*
*(ii) $\mathbb{P}(W_n) \geq 1 - e^{-\beta n r_n^2}$, for all $n \geq 1$.*

The above result essentially says whenever $r_n$ is in the intermediate range as in (2.2), a giant component of $G$ exists with very high probability and moreover it contains nearly all the nodes.

## 2.2 Proof of Theorem 2.1

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 $\Delta$ 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.

### 2.2.1 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 $t = \left(\frac{\Delta}{r_n}\right)^2$. 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

$$p_k = 1 - k \inf_i \int_{S_i} f(x)dx \leq 1 - \beta_0 k r_n^2 \leq e^{-\beta_0 k r_n^2} \tag{2.3}$$

and $\beta_0 = \frac{1}{25}\inf_{x \in S} f(x) > 0$. Thus using the inequality $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$, we have

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

Setting $k = ete^{-\theta n r_n^2}$ 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

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

where we use $t = \Delta^2 r_n^{-2}$ and $\Delta \geq 4$, respectively, in obtaining the last two inequalities. In what follows, the constants $\{\beta_i\}_{i\geq 1}$ are not necessarily same in each occurrence. Let $\delta > 1$ be any constant. Since $r_n^2 \leq c_2\frac{\log n}{n}$ for some $c_2 > 0$ (see (2.2)), we choose $\theta$ sufficiently small so that

$$\theta nr_n^2 \leq \theta c_2 \log n \leq \frac{1}{\delta}\log n.$$

This implies that

$$\mathbb{P}\left(\sum_{i=1}^{t} Z_i \geq ete^{-\theta nr_n^2}\right) \leq 2\exp\left(-16e\beta_1 n^{1-1/\delta}\right).$$

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 $t = \frac{\Delta^2}{r_n^2} \leq \frac{25}{r_n^2}$, we get from the above equation that

$$\mathbb{P}\left(\#\mathcal{T}_G \geq 2025er_n^{-2}e^{-\theta nr_n^2}\right) \leq 2\exp\left(-16e\beta_1 n^{1-1/\delta}\right)$$

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.)

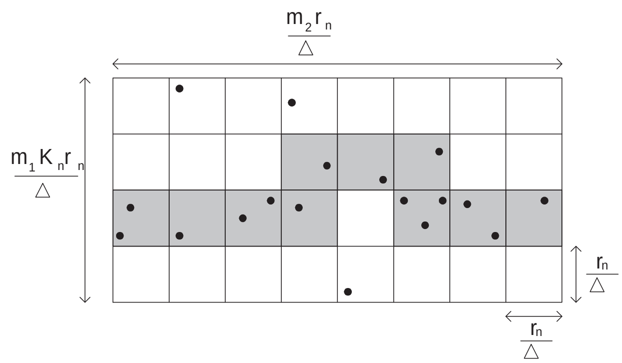



Figure 2.1: Occupied left-right crossing in the rectangle $R$ for some $\Delta \geq 4$.

From (2.2), we have that $K_n \geq \frac{1}{c_2}$. For positive integers $m_1, m_2$, 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 $L = (Y_0, Y_1, ..., Y_T)$ such that:
(a) For every $i$, the square $Y_i \in \{S_k\}_k$ and $Y_i$ and $Y_{i+1}$ share an edge.
(b) $Y_0$ intersects the left face of $R$ and $Y_T$ intersects the right face.
If every square in $L$ is occupied, we say that $L$ is an occupied left-right crossing. We define analogously top-bottom crossings and vacant crossings. The only difference in the definition of vacant crossings is that edge in (a) is replaced by corner. Sarkar (1995), Penrose (2003) and Franceschetti et al (2007) also use the concept of left-right crossings in different contexts with varying definitions.

Figure 3.1 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 $R$.

**Lemma 2.2.** *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*

$$1 - \frac{m_2}{n^{m_1 \delta_1}} \tag{2.4}$$

*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 2.2. The proof is independent of the rest of the proof of Theorem 2.1.

*Proof of Lemma 2.2*: To prove (2.4), 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.

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., [11] or [22]). 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 top-bottom crossing of length $k$ starting from $x$ (see Figure 2.2 for illustration). The area enclosed by the corresponding crossing $\Pi_1$ in $\mathbb{R}^2$ 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 $\Pi_1$ is (see also (2.3))

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

where $\beta_0 = \frac{1}{25}\inf_{x\in S} f(x) > 0$. Therefore the probability that the crossing $\Pi_1$ is vacant is less than

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

Since the number of 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 crossing of $k$ squares starting from the square $S_x$ with centre $x$ and contained in $R$ is bounded above by $8^k e^{-kn\beta_0 r_n^2}$. Any top-bottom crossing from starting from $S_x$ must necessarily contain at least $m_1K_n$ and no more than $m_1m_2K_n$ squares. Therefore the probability that there exists a vacant crossing starting from $S_x$ and contained in $R$ is bounded above by

$$\sum_{k=m_1K_n}^{m_1m_2K_n} 8^k e^{-k\beta_0 nr_n^2} \leq (e^{-\beta_1 nr_n^2})^{m_1K_n}$$

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 $8e^{-\beta_0 nr_n^2} < e^{-\beta_1 nr_n^2}$ for all $n$ sufficiently large. Since there are $m_2$ possibilities for $S_x$, the probability that there exists a vacant top-bottom crossing of $R$ is bounded above by

$$m_2(e^{-\beta_1 nr_n^2})^{m_1K_n} = m_2 e^{-\beta_1 m_1 \log n} = m_2\left(\frac{1}{n^{\beta_1}}\right)^{m_1}$$

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

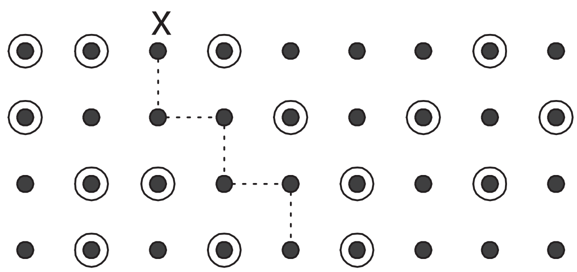



Figure 2.2: Vacant top-bottom crossing of a $4 \times 9$ rectangle in $\mathbb{Z}^2$ from the site $x$. Circled sites correspond to occupied squares.

### 2.2.2 Proof of (ii)

Tile the square $S$ horizontally into a set of rectangles $\mathcal{R}_H$ each of size $1 \times \frac{Mr_nK_n}{\Delta}$ and also vertically into rectangles each of size $\frac{Mr_nK_n}{\Delta} \times 1$ for some fixed integer constant $M \geq 1$ to be determined later. The argument below is for a perfect tiling as in Figure 2.3(a). Otherwise we perform an analogous analysis with tiling as in Figure 2.3(b). Let $R$ be a fixed $1 \times \frac{MK_nr_n}{\Delta}$ rectangle in the tiling $\mathcal{R}_H$ and let $\delta > 1$ be a fixed constant. From (2.4), we know that $R$ contains an occupied left-right crossing $L = (Y_0, Y_1, ..., Y_T)$ with probability at least

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

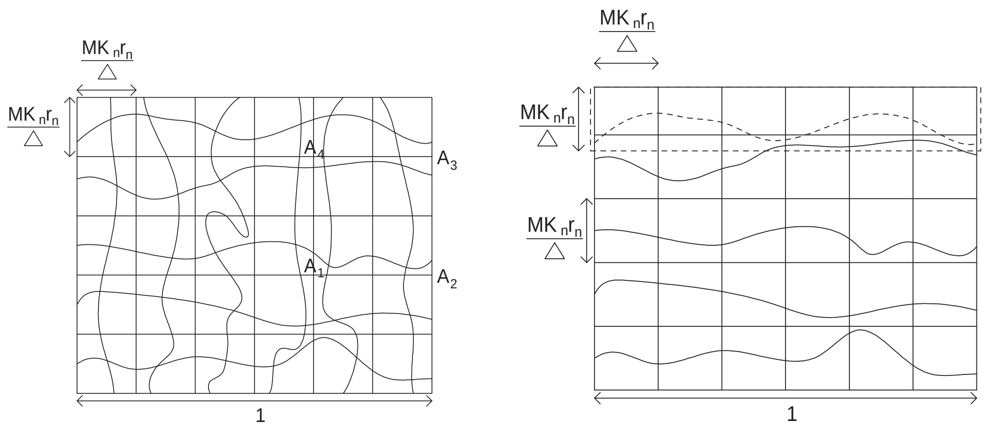
if $M$ is sufficiently large. Fix such an $M$. The first inequality above is because $r_n^2 \geq \frac{c_1}{n}$ for some constant $c_1$ (see (2.2)). Let $E_n^H$ denote the event that every rectangle in $\mathcal{R}_H$ contains an occupied left-right crossing. The number of rectangles in $\mathcal{R}_H$ is less than

$$\frac{\Delta}{Mr_nK_n} \leq \frac{\Delta}{Mr_n}\frac{1}{c_2} \leq \frac{\Delta}{Mc_2}\frac{\sqrt{n}}{\sqrt{c_1}} \leq D_1\sqrt{n}$$

for some constant $D_1 > 0$. In evaluating the above we again use (2.2). The first inequality is because $K_n = \frac{\log n}{nr_n^2} \geq \frac{1}{c_2}$ by our choice of $r_n$ in (2.2) and the second inequality follows because $r_n^2 \geq \frac{c_1}{n}$. It follows that

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

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




(a) The event $E_n$ in the unit square. Each wavy line is an occupied left-right crossing of $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ squares as in Figure 3.1.

(b) The tiling obtained when $\frac{\Delta}{MK_n r_n}$ is not an integer. The two topmost $1 \times \frac{MK_n r_n}{\Delta}$ rectangles in the tiling overlap.

Figure 2.3: Construction of the backbone.

analogous event $E_n^V$, 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

$$\mathbb{P}(E_n) \geq 1 - \frac{2}{n^{\delta+1}}. \tag{2.5}$$

In Figure 2.3(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 in Figure 2.3(b). A similar situation occurs in the vertical tiling. Following an analogous analysis as above, we obtain (2.5) 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 Figure 2.3(a). Therefore, if there exists a connected component $C$ of $G$ distinct from $C_0$, it must

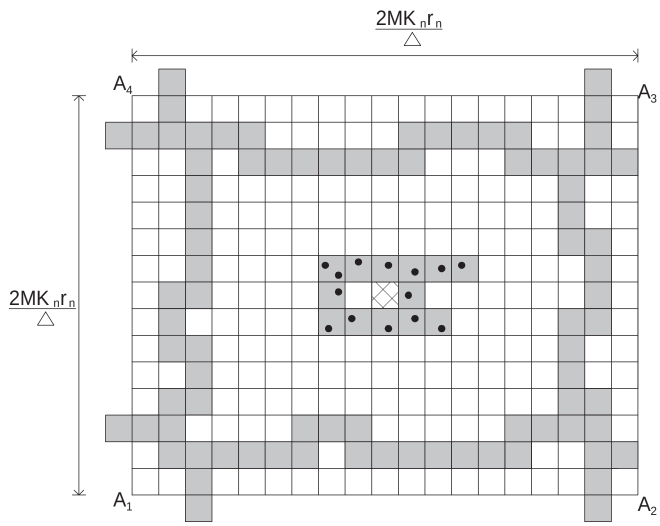



Figure 2.4: The square $A_1A_2A_3A_4$ in Figure 2.3(a) is magnified to show a component not attached to the backbone.

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 Figure 2.4, the square $A_1A_2A_3A_4$ of Figure 2.3(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 $\left(\frac{2MK_n}{\Delta}\right)^2 < (2MK_n)^2$ and therefore $diam(C) < (2MK_n)^2$. 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

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

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

**Lemma 2.3.** *We have*

$$\mathbb{P}(F_n) \geq 1 - e^{-\theta_1 n r_n^2} \tag{2.6}$$

*for some positive constants $\theta$ and $\theta_1$.*

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 $n - ne^{-\theta n r_n^2}$ 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 (2.5) and (2.6) that the event $E_n \cap F_n$ occurs with probability

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

for all $n$ sufficiently large. In the above estimate, we have used the fact (2.2) that $nr_n^2 \leq c_2 \log n$ for some positive constant $c_2$. This proves (ii) and hence Theorem 2.1. The proof of Lemma 2.3 is independent of the proof of Theorem 2.1 and is provided below. $\square$

*Proof of Lemma 2.3*: 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 Figure 2.4). 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 $\partial_1, ..., \partial_T$ be its disjoint boundaries. Each $\partial_i$ is a circuit of edges $(e_{i,1}, ..., e_{i,L_i})$ (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,

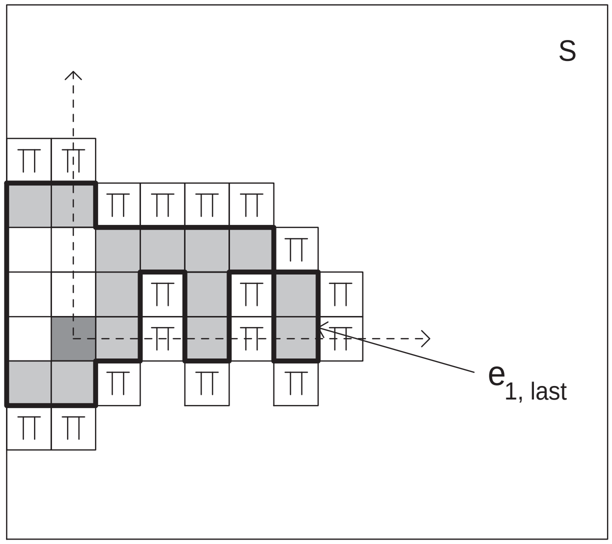



Figure 2.5: The occupied cluster $\mathcal{C}_i$ and the set of vacant squares $\pi_1$ (marked by the symbol Π) are shown for the square $S_i$ that is denoted by the dark square.

say $\partial_1$, contains all squares of $\mathcal{C}_i$ and all the other boundaries in its interior. Also, any square $S_{1,j}$ that has an edge $e_{1,j} \in \partial_1$ and not contained in $\mathcal{C}_i$ is necessarily vacant.

Let $\pi_1$ denote the set of distinct vacant squares that contain some edge in $\partial_1$. The path $\partial_1$ 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 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 Figure 2.5, the dark grey square is $S_i$ and the grey squares form $\mathcal{C}_i$. The set of vacant squares $\pi_1$ is shown by the squares marked Π and the curve of thick lines represents $\partial_1$.

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 $e_{1,last}$ be the "last" edge in $\partial_1$ that intersects the $X-$axis at $(x_{last}, 0)$. In other words, if an edge $e_{1,j}$ in $\partial_1$ intersects the $X-$axis at

$(x_j, 0)$, then $x_{last} > x_j$. In Figure 2.5, the edge $e_{1,last}$ is also shown. Clearly, there are at most $L_1$ possibilities for the location of edge $e_{1,last}$. Also, the number of choices for $\partial_1$ starting from $e_{1,last}$ is less than $4^{L_1}$.

Now, the total area of squares in $\pi_1$ 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 $\partial_1$, with probability at least $\frac{L_1}{8}\beta_0 r_n^2$ a particular node is present in $\pi_1$ where $\beta_0 = \frac{1}{25}\inf_{x\in S} f(x) > 0$ is as in (2.3). Therefore with probability at most

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

none of the $n$ nodes are present in $\pi_1$.

If $\mathcal{C}_i$ contains $k$ squares, then the number of edges $L_1$ in $\partial_1$ 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 $\partial_1$ has less than $\frac{\sqrt{k}}{4}$ edges. It is then necessary that $\partial_1$ 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 $\partial_1$ 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

$$\begin{aligned}\mathbb{P}(\{\#\mathcal{C}_i = k\} \cap \{X_i \neq 0\}) &\leq \sum_{\frac{\sqrt{k}}{4} \leq l \leq 4k} e^{-l\beta_0 n r_n^2/8} l 4^l \\ &\leq 4k \sum_{\frac{\sqrt{k}}{4} \leq l \leq 4k} \left(4e^{-\beta_0 n r_n^2/8}\right)^l \\ &\leq k e^{-\theta_0 n r_n^2 \sqrt{k}} \qquad (2.7)\end{aligned}$$

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 $4e^{-\beta_0 n r_n^2/8} < e^{-5\theta_0 n r_n^2}$ for all sufficiently large $n$. Letting $N(A)$ denote the number of nodes in the set $A$, we therefore have that

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

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

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

$I_2 = \mathbb{E}X_i - I_1$ 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 $\mathcal{C}$ containing $k$ squares, we estimate $\mathbb{P}(N(\mathcal{C}) \geq 2ek\delta_0 nr_n^2)$ first. Analogous to (2.3), we have a particular node is present in $\mathcal{C}$ with probability at most $q_k = k\delta_0 r_n^2$. Therefore

$$\begin{aligned}
\mathbb{P}(N(\mathcal{C}) \geq 2enq_k) &\leq \sum_{2enq_k \leq j \leq n} \binom{n}{j} q_k^j \\
&\leq \sum_{2enq_k \leq j \leq n} \left(\frac{ne}{j}\right)^j q_k^j \\
&\leq \sum_{2enq_k \leq j \leq n} \left(\frac{ne}{2enq_k}\right)^j q_k^j \\
&\leq \sum_{j \geq 2enq_k} \left(\frac{1}{2}\right)^j \\
&\leq e^{-2\beta_2 knr_n^2} \qquad (2.8)
\end{aligned}$$

for some positive constant $\beta_2$ independent of $k, i$ and the choice of $\mathcal{C}_0$. In the third inequality above, we have used the estimate $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$. Also, the expected number of nodes in any square $S_i$ is bounded above by

$$\sup_j \mathbb{E}N(S_j) = n \sup_j \int_{S_j} f(x)dx \leq n \sup_{x \in S} f(x) \frac{r_n^2}{\Delta^2} \leq D_1 nr_n^2 \qquad (2.9)$$

for some positive constant $D_1$ since $\sup_{x\in S} f(x) < \infty$ (see (2.1)) and $\Delta \geq 4$. Analogously,

$$\sup_j \mathbb{E}N(S_j)^2 \leq D_2(nr_n^2)^2 \qquad (2.10)$$

for some positive constant $D_2$.

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

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

for some positive constant $D_3$ independent of $i$. In obtaining the final estimate, we use (2.10) and the notation $\sum_{\#\mathcal{C}=k}$ refers to the sum over all clusters $\mathcal{C}$ containing $k$ squares of which one of them is $S_i$. Since the number of such clusters is less than $8^k$, we get

$$I_1 \le D_3 n r_n^2 \sum_{k\ge 1} k 8^k e^{-k\beta_2 n r_n^2} \le D_4 n r_n^2 e^{-\beta_3 n r_n^2}$$

for some positive constants $D_4$ and $\beta_3$, independent of $i$.

To evaluate $I_2$ we write

$$\begin{aligned} I_2 &= \mathbb{E}\sum_{k\ge 1}\sum_{\#\mathcal{C}=k}\sum_{S_j\in\mathcal{C}} N(S_j)\mathbf{1}(\mathcal{C}_i=\mathcal{C})\mathbf{1}(N(\mathcal{C})\le 2ek\delta_0 n r_n^2)\mathbf{1}(X_i\neq 0)\\ &\le 2e\delta_0 n r_n^2 \mathbb{E}\sum_{k\ge 1} k \sum_{\#\mathcal{C}=k}\sum_{S_j\in\mathcal{C}} \mathbf{1}(\mathcal{C}_i=\mathcal{C})\mathbf{1}(X_i\neq 0)\\ &= 2e\delta_0 n r_n^2 \mathbb{E}\sum_{k\ge 1} k^2 \sum_{\#\mathcal{C}=k} \mathbf{1}(\mathcal{C}_i=\mathcal{C})\mathbf{1}(X_i\neq 0)\\ &= 2e\delta_0 n r_n^2 \sum_{k\ge 1} k^2 \mathbb{P}(\{\#\mathcal{C}_i=k\}\cap\{X_i\neq 0\})\\ &\le 2e\delta_0 n r_n^2 \sum_{k\ge 1} k^3 e^{-\theta_0 n r_n^2\sqrt{k}}\\ &\le D_5 n r_n^2 e^{-\beta_5 n r_n^2} \end{aligned}$$

for some positive constants $D_5$ and $\beta_5$ independent of $i$, where the second inequality follows from the estimate (2.7). From the estimates of $I_1$ and $I_2$, we therefore have that

$$\mathbb{E}X_i \le D_6 n r_n^2 e^{-\beta_6 n r_n^2} \tag{2.11}$$

for some positive constants $D_6$ and $\beta_6$ 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

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{\Delta^2 r_n^{-2}} X_i \ge n e^{-\theta n r_n^2}\right) &\le \frac{\sum_i \mathbb{E}X_i}{n} e^{\theta n r_n^2}\\ &\le (\Delta^2 r_n^{-2})\frac{D_6 n r_n^2 e^{-\beta_6 n r_n^2}}{n} e^{\theta n r_n^2}\\ &\le D_7 e^{-\theta_1 n r_n^2} \end{aligned}$$

for some positive constants $\theta_1$ and $D_7$, if $\theta$ is sufficiently small. Since $F_n = \left\{\sum_i X_i < n e^{-\theta n r_n^2}\right\}$, this proves the lemma. $\square$

# Chapter 3

# Infection Spread in Random Geometric Graphs

## 3.1 Introduction

We consider the random geometric graph $G = G(n, r_n, f)$ in the unit square $S = \left[-\frac{1}{2}, \frac{1}{2}\right]^2$ as described in Chapter 2. To study the spread of infection in $G$, we equip each edge $e$ of $G$ with a passage time $t(e)$ that is exponentially distributed with unit mean. The passage times of distinct edges are independent. At time $t = 0$, the node $x_0$ closest to the origin in $S$ is infected. Any node $x_1$ that shares an edge $e$ with $x_0$ is infected after time $t(e)$. This process continues and infected nodes stay in that state forever.

We define the infection process on the probability space $(\Theta, \mathcal{H}, \mathbb{P})$. We describe the construction briefly in Section 3.2. For any set $A \subseteq \mathbb{R}^2$ and $\alpha > 0$, define $\alpha A = \cup_{x \in A}\{\alpha x\}$ to be the dilation of $A$ by factor $\alpha$. At time $t = 0$, the node $x_0$ closest to the origin is infected. Let $G(x_0)$ denote the connected cluster of nodes in $G$ containing $x_0$. Let $I(t)$ be the set of nodes of $G(x_0)$ infected up to time $t$. We say that infection spreads at speed at least $v_{n,low}$ if there exists functions $0 \leq a(x) = o(x)$ and $0 \leq g(x) = o(x)$ as

$x \to \infty$ such that

$$\mathbb{P}\left(\bigcap_{a(r_n^{-1}) \le m \le r_n^{-1} - g(r_n^{-1})} \left\{ \left( G(x_0) \setminus I\left(\frac{m}{v_{n,low}}\right)\right) \bigcap mr_nS = \phi \right\}\right) = 1 - o(1) \tag{3.1}$$

as $n \to \infty$. In other words, we want all nodes of $G(x_0)$ contained in $mr_nS$ to be infected within time $\frac{m}{v_{n,low}}$. This must happen for "nearly all" indices $m$. Throughout, we use the standard terminology $o(.)$ and $O(.)$ in the regime $n \to \infty$. We say that the speed is at most $v_{n,up}$ if there exists functions $0 \le a(x) = o(x)$ and $0 \le g(x) = o(x)$ as $x \to \infty$ such that

$$\mathbb{P}\left(\bigcap_{a(r_n^{-1}) \le m \le r_n^{-1} - g(r_n^{-1})} \left\{ I\left(\frac{m}{v_{n,up}}\right) \subseteq mr_nS \right\}\right) = 1 - o(1).$$

We have the main result of the chapter.

**Theorem 3.1.** *Consider the graph $G = G(n, r_n, f)$ where $f$ and $r_n$ satisfy (2.1) and (2.2), respectively. There exists positive constants $D_1$ and $D_2$ such that*

$$D_1 n r_n^2 \le v_{n,low} \le v_{n,up} \le D_2 n \sqrt{n} \log n. \tag{3.2}$$

Theorem 3.1 above is analogous to the shape theorem for infected set in regular lattices like $\mathbb{Z}^2$ (see e.g. Grimmett (1999)). The main difference is that the speed of infection spread in RGGs grows with $n$ whereas it is bounded in regular lattices.

Let $T_{elap}$ denote the time taken to infect all nodes of $G(x_0)$ and let $N_{inf} = \#G(x_0)$ denote the number of nodes that remain infected in $S$ after time $T_{elap}$. We have the following corollary regarding $T_{elap}$ and $N_{inf}$.

**Corollary 3.2.** *We have that*

$$\mathbb{P}\left(\frac{r_n^{-1}}{D_3 n \sqrt{n} \log n} \le T_{elap} \le \frac{r_n^{-1}}{D_4 n r_n^2}\right) = 1 - o(1), \text{ as } n \to \infty \tag{3.3}$$

*and*

$$\frac{r_n^{-1}}{D_3 n \sqrt{n} \log n} \le \mathbb{E}T_{elap} \le \frac{r_n^{-1}}{D_4 n r_n^2} \tag{3.4}$$

*for some positive constants $D_3$ and $D_4$. Also, as $n \to \infty$,*

$$\mathbb{P}(N_{inf} \ge n - ne^{-\theta n r_n^2}) = 1 - o(1) \tag{3.5}$$

*for some positive constant $\theta$.*

Thus with high probability, infection starting from the node closest to the origin eventually spreads to nearly all nodes.

From a practical point of view, it is very important to study infection spread in RGGs due to its applications in various fields. The fundamental difficulty, however, is the fact that it is a dense graph and unlike bounded degree graphs, the speed of infection spread is not bounded (see Theorem 3.1). Hence traditional subadditive techniques developed for first passage percolation (see e.g. Smythe and Wierman (2008)) cannot be directly applied for RGGs. Also, it is not known how many nodes will be ultimately infected if infection starts from a randomly chosen node.

Our proof technique is general and holds for a wide range of radius $r_n$. We also have some auxiliary results in the course of our proof that are of independent interest.

The chapter is organized as follows. In Section 3.2, we state and prove the geometric results regarding RGGs that are needed for analysis of infection spread. In Section 3.3, we prove lower bound on the speed in Theorem 3.1. In Section 3.4, we prove the upper bound and finally, in Section 3.5, we prove Corollary 3.2.

## 3.2 Preliminaries

We briefly describe the probability space in little more detail. We define the point process on the probability space $(\Omega, \mathcal{F}, \mu)$ and following a construction analogous to Chapter 1 of Meester and Roy (1996), we define the infection on the probability space on $(\Theta, \mathcal{H}, \mathbb{P})$, where $\Theta = \Xi \times \Omega$ and $\mathbb{P} = \nu_p \times \mu$ is a product measure. For any event $A \in \mathcal{H}$, we then have that

$$\mathbb{P}(A) = \int_\Omega \nu_p(A_\omega)\mu(d\omega) \tag{3.6}$$

where $A_\omega = \{\xi \in \Xi : (\omega, \xi) \in A\}$. In other words, $\nu_p(A_\omega)$ is the probability that $A$ occurs for a fixed configuration of points $\omega$.

In what follows, we collect a couple of geometric results (see Propositions 3.3 and 3.4) regarding RGGs and a result concerning Poissonization (Lemma 3.5) that are required for studying infection spread and are also of independent interest. At time $t = 0$, infection starts from the node $x_0$ closest to the origin. To trace the spread of infection, we first establish that there exists a path of edges starting from $x_0$ and reaching close to the boundary of $S$. We proceed as follows.

Divide $S$ into small $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ squares $\{S_k\}_{k\geq 1}$ where $\Delta = \Delta_n \in [4,5]$ is such that $\frac{\Delta}{r_n}$ is an integer. We choose such a $\Delta$ so that nodes in adjacent squares are joined together by an edge. Let $S_0$ denote the square in $\{S_k\}_k$ containing the origin. Say that $\Gamma(x_0)$ occurs if $x_0 \in S_0$ and there exists a path of edges $(e_0, e_1, ..., e_{fin})$ such that:
(i) $e_0$ contains $x_0$ as one of its endvertex and
(ii) there exists an endvertex $z_{fin} \in G$ of $e_{fin}$ such that $d(z_{fin}, \partial S) \leq \frac{r_n}{2}$.
The following result estimates the probability of occurrence of $\Gamma(x_0)$.

**Proposition 3.3.** *There exists a constant $\theta_1 > 0$ so that*

$$\mathbb{P}(\Gamma(x_0)) \geq 1 - e^{-\theta_1 n r_n^2} \tag{3.7}$$

*for all $n \geq 1$.*

The proof of this geometric result is analogous to Lemma 2.3 of Chapter 2 and we briefly sketch the proof here. Also, in this proof and throughout, we repeatedly use the following concept of denseness of squares.

*For a fixed $i$, let $10\sigma_i$ be the mean number of nodes in the $\dfrac{r_n}{\Delta} \times \dfrac{r_n}{\Delta}$ square $S_i$.* (3.8)

Using (2.1) and $\Delta \in [4,5]$, we have for all $i$ that

$$10\beta_1 n r_n^2 \leq 10\sigma_i \leq 10\beta_2 n r_n^2 \tag{3.9}$$

where $\beta_1 = \frac{1}{25}\inf_{x\in S} f(x) > 0$ and $\beta_2 = \frac{1}{16}\sup_{x\in S} f(x) < \infty$. Define $S_i$ to be *dense* if it has more than $\sigma_i$ nodes and *sparse* otherwise. The definition of a dense square is slightly stronger than the definition of an occupied square in Chapter 2.

*Proof of Proposition 3.3*: Say that a set of squares $\mathcal{C} \subseteq \{S_i\}_i$ is a cluster if they form a connected set in $\mathbb{R}^2$. Define $\mathcal{C}$ to be dense if each square in $\mathcal{C}$ is dense.

Let $S_{or}$ denote the $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ square containing the origin. We claim that

$$\mathbb{P}(S_{or} \text{ is dense}) \geq 1 - e^{-\theta_2 n r_n^2} \tag{3.10}$$

for some constant $\theta_2 > 0$. Indeed, a particular node is present in $S_{or}$ with probability $p_n = \int_{S_{or}} f(x)dx$. Without loss of generality, we assume that $np_n =: 10\sigma_0$ is an integer; the argument below holds with minor modifications otherwise. We have by the unimodality property of the binomial distribution (see e.g. Alam (1972)) that

$$\mathbb{P}(S_{or} \text{ is sparse}) = \sum_{k \le \sigma_0} \binom{n}{k} p_n^k (1-p_n)^{n-k} \le \sigma_0 \binom{n}{\sigma_0} p_n^{\sigma_0} (1-p_n)^{n-\sigma_0}.$$

Since $r_n^2 \longrightarrow 0$ and $\Delta \in [4,5]$, it is easy to check that $p_n \longrightarrow 0$ as $n \to \infty$ and in particular, $1 - p_n \ge \frac{1}{2}$ for all $n$ sufficiently large. Thus using $\binom{n}{k} \le \left(\frac{ne}{k}\right)^k$, the final term in the previous equation can be bounded above as

$$\sigma_0 \left(\frac{ne}{\sigma_0}\right)^{\sigma_0} p_n^{\sigma_0}(1-p_n)^{n-\sigma_0} = \sigma_0(1-p_n)^n \left(\frac{np_ne}{\sigma_0(1-p_n)}\right)^{\sigma_0} \le \sigma_0 e^{-np_n} \left(\frac{2enp_n}{\sigma_0}\right)^{\sigma_0}$$

where we use $1 - x < e^{-x}$ in obtaining the final inequality. Again using $np_n = 10\sigma_0$ and (3.9) and the fact that $nr_n^2 \longrightarrow \infty$, we get

$$\mathbb{P}(S_{or} \text{ is sparse}) \le \sigma_0 e^{-10\sigma_0} (20e)^{\sigma_0} \le e^{-5\sigma_0}$$

for all $n$ sufficiently large. By (3.9), we get (3.10).

Suppose that $S_{or}$ is dense and suppose $\mathcal{C}_{or}$ denotes the maximal dense cluster containing $S_{or}$. If $S_{or}$ is dense and $\Gamma(x_0)$ does not occur, then necessarily $\mathcal{C}_{or}$ must be surrounded by a circuit of sparse squares contained in $S$. Our aim is to show that such an event is very unlikely. This is because sparse squares occur with probability at most $e^{-\theta_2 nr_n^2}$ and consequently a large number of vacant squares cannot be "attached to" $\mathcal{C}_{or}$. Following an analysis analogous to the proof of Lemma 2.3 of Chapter 2, we have for $k \ge 1$ that

$$\mathbb{P}(\{\#\mathcal{C}_{or} = k\} \cap \Gamma^c(x_0) \cap \{S_{or} \text{ is dense}\}) \le ke^{-2\theta_0 nr_n^2\sqrt{k}}$$

for a fixed positive constant $\theta_0$ and all $n \ge N_0$, where $N_0$ is a constant that does not depend on $k$. This implies that

$$\mathbb{P}(\Gamma^c(x_0) \cap \{S_{or} \text{ is dense}\}) = \mathbb{P}(\{\#\mathcal{C}_{or} \ge 1\} \cap \Gamma^c(x_0) \cap \{S_{or} \text{ is dense}\}) \le e^{-\theta_0 nr_n^2}$$

for all $n$ sufficiently large. From the (3.10), we then get (3.7). $\square$

Before we go further we point out the differences in the analysis here from Chapter 2. The concept of dense left-right crossings defined below is slightly stronger than occupied left-right crossings defined in Chapter 2. The estimate on the probability of occurrence of dense left-right crossings in Proposition 3.4 with a bound on the length is however a non-trivial extension of Lemma 2.2 of Chapter 2. The construction of the backbone using dense left-right crossings in Section 3.3 and the subsequent analysis is fundamentally different from Chapter 2.

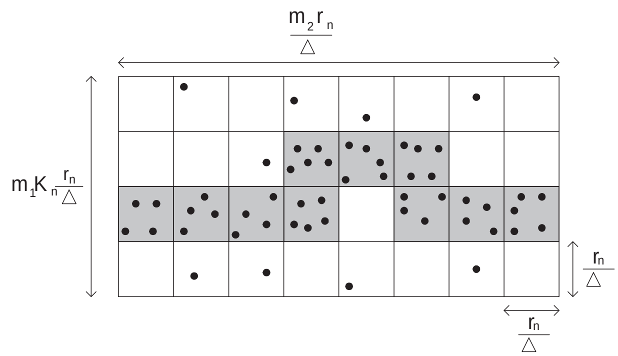



Figure 3.1: A dense left-right crossing in the rectangle $R$ in $\mathbb{R}^2$.

## Left-right Crossings

From Proposition 3.3, we know that with high probability, there exists a path starting from $x_0$ and crossing $mr_nS$ for each $1 \leq m \leq r_n^{-1}-1$. To estimate the time taken for infection to cross the boundary of $mr_nS$, we need to find paths whose edges have low passage time. Left-right crossings described below are useful in that aspect.

Let $K_n = \frac{\log n}{nr_n^2}$ and for positive integers $m_1, m_2$, let $R$ be any $m_2\frac{r_n}{\Delta} \times m_1K_n\frac{r_n}{\Delta}$ rectangle in the unit square $S$ that contains exactly $m_1m_2K_n$ of the squares in $\{S_i\}_i$ and intersects none others. Without loss of generality we allow $K_n$ to be an integer throughout and with minor modifications the argument presented below holds for general $K_n$. We define a left-right crossing in $R$ to be any sequence of squares $L = (Y_1, Y_2, ..., Y_T)$ such that:
(a) For every $i$, the squares $Y_i$ and $Y_{i+1}$ share an edge.
(b) $Y_1$ intersects the left side of $R$ and $Y_T$ intersects the right side.
(c) For every $i \neq 1, T$, neither the left edge nor the right edge of $R$ intersects the square $Y_i$.
If every square in $L$ is dense, we say that $L$ is a dense left-right crossing. Figure 3.1 illustrates a dense left-right crossing in $R$. In Section 3.3, we use dense left-right crossings to obtain paths with low passage times.

We have the following result regarding the probability of occurrence of a dense left-right crossing. Let $R_1$ be the $m\frac{r_n}{\Delta} \times MK_n\frac{r_n}{\Delta}$ rectangle containing exactlr $mMK_n$ squares from $\{S_k\}_k$ and let $E_n(R_1)$ denote the event that there exists a dense left-right crossing of $R_1$ containing less than $10Mm$ squares.

**Proposition 3.4.** *There exists positive constants* $C_1$ *and* $M$ *so that for all* $n \geq 1$ *and* $m \geq n^{1/9}$ *we have*

$$\mathbb{P}(E_n(R_1)) \geq 1 - \frac{C_1}{n^9}. \tag{3.11}$$

This is a stronger result than Lemma 2.2 of Chapter 2 since we also control the length of the left-right crossing here. The fact there exists a dense crossing with less than $10Mm$ squares plays a crucial role in obtaining path of edges with the desired low passage times (see also Remark 3.7 following Lemma 3.6).

To prove the above Proposition, we employ Poissonization and assume that the nodes are distributed according to a Poisson process with intensity function $nf(.)$. Defining $\mathbb{P}_o$ to be the probability measure under the Poissonized system, we have the following result.

**Lemma 3.5.** *For any measurable event* $A$, *we have*

$$\mathbb{P}(A) \geq 1 - C_1\sqrt{n}(1 - \mathbb{P}_o(A)), \tag{3.12}$$

*for some absolute constant* $C_1$ *independent of* $A$.

*Proof of Lemma 3.5*: To prove (3.12), we note that in the Poisson case, the number of nodes $N$ in the unit square $S$ is a Poisson random variable with mean $n$; and therefore by Stirling's formula, $\mathbb{P}_o(N = n) = e^{-n}\frac{n^n}{n!} \geq \frac{C_2}{\sqrt{n}}$ for some positive constant $C_2$. Since

$$\begin{aligned}
\mathbb{P}_o(A^c) &= \sum_{k=0}^{\infty} \mathbb{P}_o(A^c|N = k)\mathbb{P}_o(N = k) \\
&\geq \mathbb{P}_o(A^c|N = n)\mathbb{P}_o(N = n) \\
&= \mathbb{P}(A^c)\mathbb{P}_o(N = n)
\end{aligned}$$

we get (3.12). □

*Proof of Proposition 3.4*: We note that $R_1$ is the $m\frac{r_n}{\Delta} \times MK_n\frac{r_n}{\Delta}$ rectangle with centre as origin, where $K_n = \frac{\log n}{nr_n^2} \leq \log n$, $m \geq 4n^{1/9}$ and $M \geq 1$ is a constant.

For the rest of this proof we work in the Poissonized system. Our first step is to translate the problem to $\mathbb{Z}^2$. We identify each $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ square $S_i$ with a vertex $z_i \in \mathbb{Z}^2$. The rectangle $R_1$ thus corresponds to a $m \times MK_n$

rectangle $R_1^{int}$ in $\mathbb{Z}^2$. Also, there is a one-one correspondence between left-right crossings of $R_1$ and of $R_1^{int}$ given by the nearest neighbour connection on the integer lattice. We now construct two i.i.d. Bernoulli site percolation measures $P_{p_1}$ and $P_{p_2}$ on $R_1^{int}$ as follows.

Recall from (3.9) that the mean number of nodes in the $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ square $S_i$ is $10\sigma_i$ and that $S_i$ is dense if it has more than $\sigma_i$ nodes and sparse otherwise. Analogously, we allow every site in $R_1^{int}$ to be in one of the two states, dense or sparse. In the first measure $P_{p_2}$, we set each site $z_i \in R_1^{int}$ to be dense with probability

$$p_2 = \inf_i \mathbb{P}_o(S_i \text{ contains more than } \sigma_i \text{ nodes}). \tag{3.13}$$

Since the Poisson process is disjoint on independent sets, we have that $P_{p_2}$ is an i.i.d. site percolation measure on the rectangle $R_1^{int}$. Using standard estimates on Poisson distribution (see e.g. Penrose (2003)), we have

$$p_2 \geq 1 - e^{-2\beta_{11} n r_n^2}. \tag{3.14}$$

for all $n$ sufficiently large, where $\beta_{11} > 0$ is some constant.

For $P_{p_1}$, we set $z_i$ to be dense with probability $p_1 = 1 - e^{-2\theta_1 n r_n^2}$ for some $\theta_1 \in (0, \beta_{11})$. This is done to ensure that

$$p_2 - p_1 \geq e^{-2\theta_1 n r_n^2} - e^{-2\beta_{11} n r_n^2} \geq e^{-4\theta_1 n r_n^2} \tag{3.15}$$

for all $n$ sufficiently large. Here we use $nr_n^2 \to \infty$.

Let $A$ denote the event that $R_1^{int}$ has a dense left-right crossing and let $I_r(A)$ denote the event that there are at least $r$ disjoint dense left-right crossings of $R_1^{int}$. From Theorem 2.45 of Grimmett (1999), we have that

$$P_{p_2}(I_r(A)) \geq 1 - \left(\frac{p_2}{p_2 - p_1}\right)^r (1 - P_{p_1}(A)).$$

Analogous to the proof of Lemma 2.2 of Chapter 2, we have that

$$P_{p_1}(A) \geq 1 - \frac{C_1 m}{n^{M\theta_1}}$$

for some positive constant $C_1$ and all $n \geq 1$. Thus letting $r = \frac{M}{10} K_n = \frac{M}{10} \frac{\log n}{n r_n^2}$, we have from (3.15) that

$$P_{p_2}(I_r(A)) \geq 1 - \exp\left(\frac{4M\theta_1}{10} \log n\right) \frac{C_1 m}{n^{M\theta_1}} \geq 1 - C_1 m n^{-\frac{3M\theta_1}{5}}$$

for all $n$ sufficiently large. Choosing the constant $M$ sufficiently large, we now have $P_{p_2}(I_r(A)) \geq 1 - \frac{1}{n^{10}}$. If $I_r(A)$ occurs in $R_1^{int}$, we then have at least $\frac{MK_n}{10}$ disjoint dense left-right crossings of the rectangle $R_1$ in the Poissonized system. Since $R_1$ is of size $m\frac{r_n}{\Delta} \times MK_n\frac{r_n}{\Delta}$, the sum of lengths of all the disjoint dense left-right crossings is less than $mMK_n$. But this implies that at least one of the dense crossing contains less than $10Mm$ squares and hence $E_n(R_1)$ occurs.

To relate this to our Poissonized system, we let $\mathbb{P}_{site}$ be the site percolation measure obtained the following way: a vertex $z_i \in R_1^{int}$ is dense if and only if the corresponding square $S_i$ is dense. By our choice of $p_2$ in (3.13), we then have that $P_{p_2} \leq_{st} \mathbb{P}_{site}$; i.e., $\mathbb{P}_{site}$ stochastically dominates $P_{p_2}$. We thus have that

$$\mathbb{P}_o(E_n(R_1)) \geq \mathbb{P}_{site}(I_r(A)) \geq P_{p_2}(I_r(A)) \geq 1 - \frac{1}{n^{10}}$$

and from (3.12) we get (3.11). $\square$

## 3.3 Proof of Theorem 3.1: Lower bound on speed

For obtaining the lower bound on speed, we choose $a(r_n^{-1}) = n^{1/9}$ as the starting index from which we trace the infection spread (see definition prior to Theorem 3.1). This suffices since $n^{1/9} = o(r_n^{-1})$ by (2.2). (In fact any $\alpha < \frac{1}{2}$ suffices since $n^{\alpha} = o(r_n^{-1})$ by (2.2).)

Fix $m \geq n^{1/9}$ and tile $m\frac{r_n}{\Delta}S$ horizontally into a set $\mathcal{R}_H$ of $m\frac{r_n}{\Delta} \times MK_n\frac{r_n}{\Delta}$ rectangles and also vertically into a set $\mathcal{R}_V$ of disjoint rectangles each of size $MK_n\frac{r_n}{\Delta} \times m\frac{r_n}{\Delta}$. Here and henceforth we fix the constant $M$ so that Proposition 3.4 holds. For now we allow $m$ to be a multiple of $MK_n$ and extend to the general case at the end. The first step is to construct a backbone of low passage time paths in each rectangle of $\mathcal{R}_H$ and $\mathcal{R}_V$.

The strategy of the proof is this: We obtain an explicit upper bound on the passage time of each path of the backbone. We then estimate the time taken for the infection to reach some node of this backbone starting from $x_0$. Our estimates hold for each $n^{1/9} \leq m \leq r_n^{-1} - (\log n)^2$ resulting in the lower bound on the speed.

For a vertical rectangle $R$ in $\mathcal{R}_V$, we define $E_n(R)$ to be the event that it contains a dense top-bottom crossing consisting of less than $10Mm$ squares. (A top-bottom crossing of $R$ is a left-right crossing of the rectangle $R_{rot}$ obtained by rotating $R$ by 90 degrees about its centre). Again Proposition 3.4 is applicable to each rectangle $R$ in $\mathcal{R}_V$ with left-right crossing replaced by top-bottom crossing. Defining $E_{n,tot} := \bigcap_{R \in \mathcal{R}_H \cup \mathcal{R}_V} E_n(R)$ and using the fact that the number of rectangles in the set $\mathcal{R}_V \cup \mathcal{R}_H$ is $O\left(\frac{\Delta}{r_n}\right) = O(\sqrt{n})$ (by (2.2)), we then have that

$$\mathbb{P}\left(E_{n,tot}\right) \geq 1 - O(\sqrt{n})\frac{1}{n^9} \geq 1 - \frac{1}{n^8} \tag{3.16}$$

for all $n$ large enough.

We henceforth assume that $E_{n,tot}$ occurs. Consider now the lowermost rectangle $R_2 \in \mathcal{R}_H$ and let $L(R_2) = (J_1, J_2, ..., J_q)$ be the bottommost dense left-right crossing of $R_2$ containing $q \leq 10Mm$ squares where each $J_i \in \{S_k\}_k$. Such a left-right crossing is obtained in an iterative manner as follows: Let $\mathcal{S}_1 = \{L'_i\}_{1\leq i\leq W} = \{(S_{i,1}, ..., S_{i,H_i})\}_{1\leq i\leq W}$ be the set of all dense left-right crossings of $R_2$ containing less than $10Mm$ squares. Let $y_{i,j}$ be the $y$-coordinate of the centre of $S_{i,j}$. For $j \geq 2$, we iteratively define

$$\mathcal{S}_j = \{L'_i \in \mathcal{S}_{j-1} : y_{i,j} = \min_{L'_k \in \mathcal{S}_{j-1}} y_{k,j}\}.$$

Thus $\mathcal{S}_2$ is the subset of crossings of $\mathcal{S}_1$ such that the centre of the first square has the least $y-$coordinate and so on. This procedure terminates after a finite number of steps resulting in a unique dense left-right crossing. Also, the final crossing obtained does not depend on the initial ordering of the left-right crossings.

Let $u_1$ be the node that is closest to the centre of $J_1$. For $1 \leq i \leq q-1$, we perform the following iteratively: Consider the set of all edges from $u_i$ that have an endvertex in $J_{i+1}$ and choose that edge $h_i$ with the minimal passage time. The endvertex of $h_i$ distinct from $u_i$ is set to be $u_{i+1}$. Let $L_h(R_2) = (h_1, ..., h_{q-1})$ be the resulting path of edges.

In Figure 3.2, the hatched sequence of squares is the crossing $L(R_2)$. The path $L_h(R_2)$ is also shown. Since every $J_i$ is dense, at each iteration we have chosen the minimum among at least $\beta_1 n r_n^2$ edges where $\beta_1 > 0$ is the constant in (3.9) and therefore we expect to get an edge with low passage time. The following result determines the overall passage time of the resulting path of edges. Define the passage time $T(R_2) = \sum_{i=1}^{q-1} t(h_i)$, where $L_h(R_2) = (h_0, ..., h_{q-1})$ is the path obtained as above. (For completeness, we define $T(R_2) = \infty$ if $E_{n,tot}$ does not occur).

**Lemma 3.6.** *There exists positive constants $D_1$ and $\delta_1$ such that*

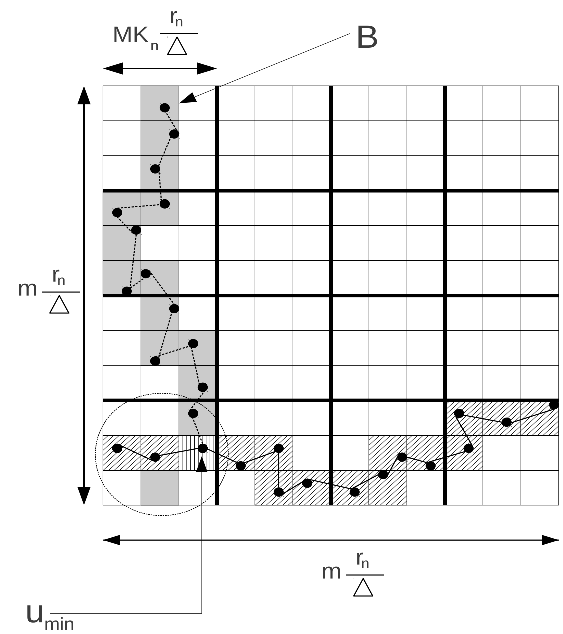
$$\mathbb{P}\left(\left\{T(R_2) \geq \frac{D_1 m}{nr_n^2}\right\} \bigcap E_{n,tot}\right) \leq e^{-\delta_1 m}$$

*for all $m \geq 4n^{1/9}$.*

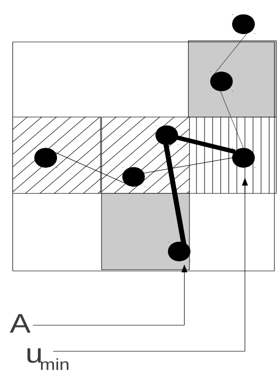
**Remark 3.7.** *If we did not have $q \leq 10Mm$, then the only upper bound on $q$ would be number of squares in $R_2$ which is $mMK_n$. Following the proof of Lemma 3.6, we would have obtained a bound of $\frac{D_1 m K_n}{nr_n^2}$ that is off the desired bound by a factor $K_n$.*

The above result (which is assumed for now and proved later) implies that if $E_{n,tot}$ occurs, then infection starting from some node in the path $L_h(R_2)$ spreads to all the nodes of $L_h(R_2)$ within time $\frac{D_1 m}{nr_n^2}$ with high probability. The factor $nr_n^2$ occurs essentially because we have chosen the minimum among $\beta_1 nr_n^2$ edges at each iteration above. This is the fundamental difference of RGGs from graphs with bounded degree where such an unbounded factor cannot appear. Recall that we continue to assume that $E_{n,tot}$ holds and therefore the path $L_h(R_2)$ is well-defined. Now, to determine the time taken for infection to reach some node of $L_h(R_2)$, we grow low passage time paths from $L_h(R_2)$ in the vertical direction. This is possible because the horizontal rectangle $R_2$ intersects each vertical rectangle $R \in \mathcal{R}_V$ and each such rectangle has a dense top-bottom crossing (due to the occurrence of the event $E_{n,tot}$).

Fix the leftmost vertical rectangle $R_l \in \mathcal{R}_V$ and consider the leftmost dense top-bottom crossing $TB(R_l) = (A_1, ..., A_s)$ of $R_l$ consisting of $s \leq 10Mm$ squares. This is obtained in an analogous iterative manner as for bottom most left-right crossings as described above. The dense left-right crossing $L(R_2)$ obtained above and the dense top-bottom crossing $TB(R_l)$ intersect in the sense that there exists a square $A_{l_0}$ with minimum index in $\{S_k\}_k$ that is present in both $TB(R_l)$ and $L(R_2)$. Here $1 \leq l_0 \leq s$ is a random index. In Figure 3.2(a), the set of grey squares constitute the dense crossing $TB(R_l)$. The vertically hatched square (which denotes $A_{l_0}$) and the hatched square to the left of it are common to $L(R_2)$ and $TB(R_l)$. Suppose that $A_{l_0} = J_{i_0} \in L(R_2)$ for some random index $1 \leq i_0 \leq q$. By construction, there exists an edge $h_{i_0}$ of $L_h(R_2)$ that has an endvertex $u_{i_0}$ in $J_{i_0}$. We now start from $u_{i_0}$ and perform the same iterative edge searching procedure that was used to obtain $L_h(R_2)$ above, on the latter part $(A_{l_0}, ..., A_s)$ of $TB(R_l)$.




(a)




(b)

Figure 3.2: Construction of backbone in the rectangles $R_2$ and $R_l$.

(In our figure this latter part is the vertically hatched square together with the set of grey squares lying above it and $u_{i_0}$ is the point marked $u_{min}$.)

Set $u'_{l_0} = u_{i_0}$. For each $l_0 \leq i \leq s-1$, we iteratively choose the edge $h'_i$ with minimal passage time that has one endvertex as $u'_i$ and one endvertex in $A_{i+1}$. The node thus obtained in $A_{i+1}$ is defined to be $u'_{i+1}$. The resulting path of edges starting from $u'_{l_0}$ and ending at some node $B$ of $G$ is called $(h'_{l_0}, h'_{l_0+1}, ..., h'_{s-1})$ (see Figure 3.2(a)). Similarly, starting from $i = l_0 - 1$ and for each $l_0 - 1 \geq i \geq 2$, we iteratively choose the edge $h'_i$ with minimal passage time that has one endvertex as $u'_i$ and one endvertex in $A_{i-1}$. We obtain a path of edges $(h'_{l_0-1}, ..., h'_1)$. In Figure 3.2(b), we have zoomed the circled part of Figure 3.2(a). The path of thick edges starting from $u'_{l_0}$ to $A$ constitute $(h'_{l_0-1}, ..., h'_1)$. We set $TB_h(R_l)$ to be the concatenation

$$TB_h(R_l) = (h'_1, ..., h'_{l_0-1}, h'_{l_0}, h'_{l_0+1}, ..., h'_{s-1})$$

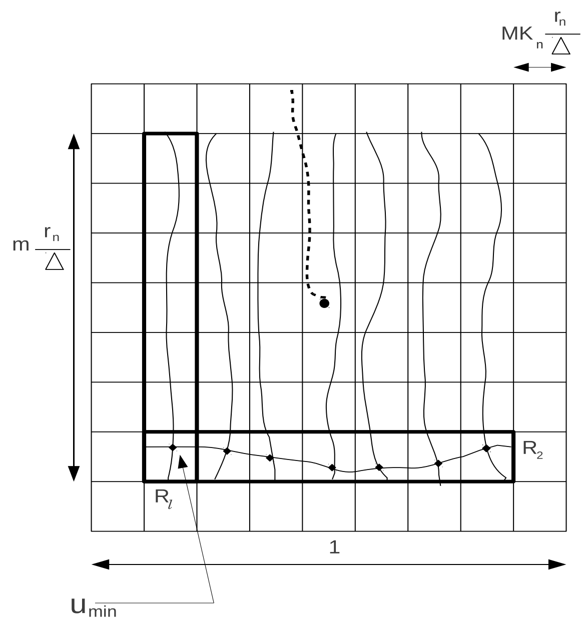
and define the passage time of the rectangle $R_l$ to be

$$T(R_l) = \sum_{i=1}^{s-1} t(h'_i).$$

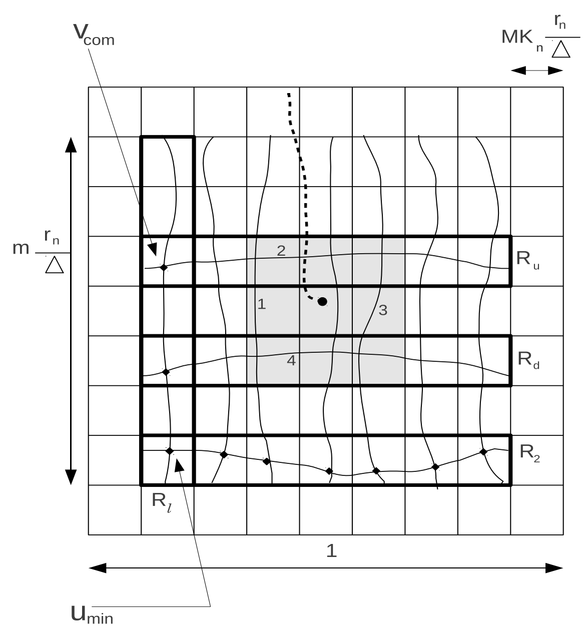
Repeat now the above procedure for each $R \in \mathcal{R}_V$ and obtain corresponding paths $TB_h(R)$. This results in a connected set of edges $\mathcal{P}_e$ that form a comb-like backbone as in Figure 3.3(a). The advantage of working with $\mathcal{P}_e$ is that we have an explicit bound on the passage time of each of its paths via Lemma 3.6. This is because even if the passage times of two distinct paths in $\mathcal{P}_e$ are not independent, Lemma 3.6 holds for each of their passage times individually with the *same* constants $D_1$ and $\delta_1$. This can then be used to estimate the time taken for infection to spread from some node of a path in $\mathcal{P}_e$ to the boundary.

Before we do so, we need to settle the following question: Does infection originally starting from node $x_0$ ever reach this backbone? Or equivalently, is $x_0$ is connected to $\mathcal{P}_e$? As we see from Figure 3.3(a), even if $\Gamma(x_0)$ occurs, the path $\pi_0$ from $x_0$ to the boundary of $S$ that is present due to the occurrence of the event $\Gamma(x_0)$ (see definition prior to Proposition 3.3) and the backbone $\mathcal{P}_e$ constructed above need not intersect. To remedy the situation, we "trap" paths starting from $x_0$ by adding horizontal paths to $\mathcal{P}_e$.

Let $R_0$ denote the rectangle in $\mathcal{R}_H$ containing $x_0$ and let $R_u$ and $R_d$ denote the rectangles in $\mathcal{R}_H$ sharing an edge with $R_0$ and lying above and below $R_0$, respectively. Since $E_{n,tot}$ occurs, each of the rectangles $R_u$ and $R_d$




(a)




(b)

Figure 3.3: Adding horizontal paths to the backbone to trap the path from the node $x_0$ denoted by the dark circle at the centre.

contain a dense left-right crossing with less than $10Mm$ squares. Consider the rectangle $R_u$ and let $L(R_u) = (W_1, ..., W_f)$ be the bottom most dense left-right crossing of $R_u$ containing $f \leq 10Mm$ squares. Further, let $(A_{z_0}, ..., A_{z_1})$ denote the segment of the top-bottom dense crossing $TB(R_l)$ of the vertical rectangle $R_l$, that is contained in $R_u$. Here $1 \leq z_0 \leq z_1 \leq s$ are random indices.

Clearly, there exists a square $A_{z_2}$ with the least index in $\{S_k\}_k$ that is present in both $(A_{z_0}, ..., A_{z_1})$ and $L(R_a)$. Also, there exists a node $v_{z_2} \in A_{z_2}$ (shown as $v_{com}$ in Figure 3.3(b)) and an edge $h'_{z_2} \in TB_h(R_l) \subset \mathcal{P}_e$ that contains $v_{z_2}$ as one of its endvertex. Suppose $A_{z_2} = W_{t_0} \in L(R_a)$ for some random index $1 \leq t_0 \leq f$. As before, we consider the latter part $(W_{t_0}, W_{t_0+1}, ..., W_f)$ of the left-right crossing $L(R_u)$ and "grow" a path of edges iteratively starting from $v_{z_2}$ ending in $W_f$ and contained in $L(R_u)$. We choose the edge with minimum passage time at each iteration. Analogously, considering the former part $(W_1, ..., W_{t_0-1}, W_{t_0})$, we grow a path of edges with minimum passage times starting from $v_{z_2} \in W_{t_0}$ and ending in $W_1$. We call the concatenation of the two paths as $L_h(R_u)$ and define the passage time $T(R_u)$ as before. We perform an analogous procedure on $R_d$ and call the resulting path of edges as $L_h(R_d)$ and the corresponding passage time as $T(R_d)$.

Finally, we define

$$\mathcal{P} = L_h(R_2) \bigcup \bigcup_{R \in \mathcal{R}_V} TB_h(R) \bigcup L_h(R_u) \bigcup L_h(R_d) \tag{3.17}$$

as the backbone. The backbone is connected by construction. In Figure 3.3(b), the occurrence of the event $E_{n,tot} \cap \Gamma(x_0)$ and the resulting backbone of crossings in the square $m\frac{r_n}{\Delta}S$ are shown. The dark dot at the centre and the dotted line represents $x_0$, the node closest to the origin and the path due to the event $\Gamma(x_0)$, respectively. The dark dots at the junction of the paths signify intersection.

With the above backbone construction, we claim that if $E_{n,tot} \cap \Gamma(x_0)$ occurs, then there is a path of edges starting from $x_0$ and ending at some node of $\mathcal{P}$. We prove the claim as follows. First, the tiling of $m\frac{r_n}{\Delta}S$ into the set of rectangles $\mathcal{R}_V$ and $\mathcal{R}_H$ described above also tiles $m\frac{r_n}{\Delta}S$ into squares $\{S'_i\}_i$ each of size $MK_n\frac{r_n}{\Delta} \times MK_n\frac{r_n}{\Delta}$ as seen in Figure 3.3(b). Let $S(K_n)$ be the square in $\{S'_i\}_i$ that contains $S_{or}$, the square in $\{S_k\}_k$ containing the origin and let $S(3K_n)$ be the $3MK_n\frac{r_n}{\Delta} \times 3MK_n\frac{r_n}{\Delta}$ with the same centre as $S(K_n)$. Since $\Gamma(x_0)$ occurs, there exists a path $\pi_0$ of edges from $x_0$ that crosses

$S(3K_n)$. (If there is more than one such path, we choose that path whose sum of the length of edges is the least and call it $\pi_0$.)

In Figure 3.4(a), we have magnified the grey region $S(3K_n)$ of Figure 3.3(b) and shown the dense crossings containing the paths marked $1,2,3$ and 4. The dense crossings form a circuit around $x_0$ and therefore the path $\pi_0$ necessarily intersects the polygonal circuit shown in thick lines. Consequently $\pi_0$ must intersect some dense $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ square $S_\alpha$ marked $1,2,3$ or 4. Again by construction, $S_\alpha$ must contain some node $v$ of the backbone as shown in Figure 3.4(b). Since $\Delta \in [4,5]$, this implies that $u$ and $v$ are joined by an edge. Thus there is a path of edges from $x_0$ to some node $v$ of the backbone that is contained entirely in $S(3K_n)$. If there is more than one such node, we set $v$ to be that node which is closest in Euclidean distance to $x_0$.

To trace the infection starting from node $v$ of the backbone $\mathcal{P}$, we define

$$V_m = \bigcap_R \left\{ T(R) \leq \frac{D_1 m}{n r_n^2} \right\} \bigcap E_{n,tot},$$

where the intersection is taken over all rectangles $R$ present in the expression for $\mathcal{P}$ in (3.17) and $T(R)$ denotes the passage time (see Lemma 3.6) of the rectangle $R$. As mentioned before, even if the passage times of two distinct paths are not independent, Lemma 3.6 holds for each of them individually with the *same* constants $D_1$ and $\delta_1$. Thus from (3.16) and Lemma 3.6 we get that

$$\begin{aligned} \mathbb{P}(V_m^c) &= \mathbb{P}(E_{n,tot}^c) + \mathbb{P}\left( \bigcup_R \left\{ T(R) > \frac{D_1 m}{n r_n^2} \right\} \bigcap E_{n,tot} \right) \\ &\leq \mathbb{P}(E_{n,tot}^c) + \sum_R \mathbb{P}\left( \left\{ T(R) > \frac{D_1 m}{n r_n^2} \right\} \bigcap E_{n,tot} \right) \\ &\leq \frac{1}{n^8} + C_1 \sqrt{n} e^{-\delta_1 m} \end{aligned}$$

for some positive constant $C_1$. In obtaining the final estimate above, we use the fact that the number of rectangles in $\mathcal{R}_H \cup \mathcal{R}_V = O(r_n^{-1}) = O(\sqrt{n})$ by (2.2). Since $m \geq n^{1/9}$, we have for all $n \geq N_0$ sufficiently large and all $m \geq n^{1/9}$ that

$$\mathbb{P}(V_m) \geq 1 - \frac{2}{n^8}. \tag{3.18}$$

The following result estimates local passage times and is the final ingredient needed for the proof of lower bound. Let $m_1$ be the smallest integer that

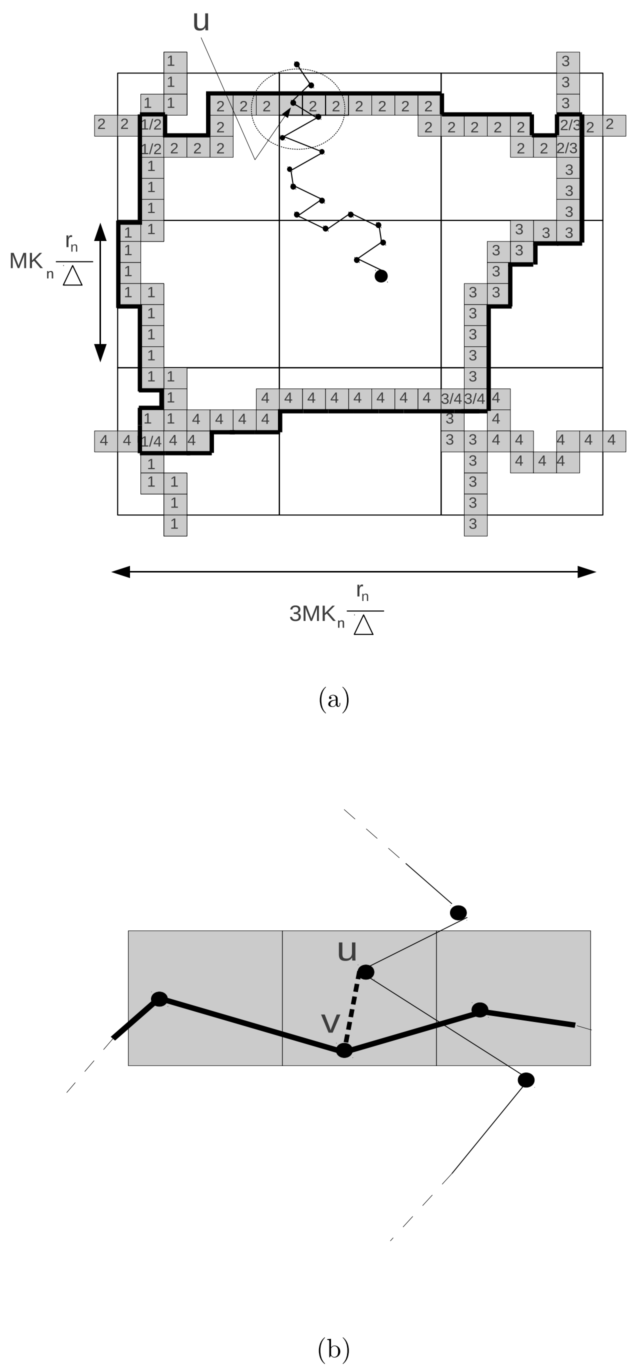



Figure 3.4: The path $\pi_0$ from $x_0$ necessarily intersects the circuit of dense squares with boundary denoted by thick line and hence the backbone.

is a multiple of $MK_n$ and such that $S \subseteq m_1 \frac{r_n}{\Delta} S$. The tiling of $m_1 \frac{r_n}{\Delta} S$ into the rectangles in $\mathcal{R}_H$ and $\mathcal{R}_V$ also tiles $m_1 \frac{r_n}{\Delta} S$ into $MK_n \frac{r_n}{\Delta} \times MK_n \frac{r_n}{\Delta}$ squares $\{S'_i\}_i$ as seen in Figure 3.3(a). Let $T_i$ denote the sum of passage times of the edges that have at least one endvertex in $S'_i$ and let $T_{max} = \max_i T_i$.

**Lemma 3.8.** *There exists a constant $C_1 > 0$ so that*

$$\mathbb{P}(T_{max} > (\log n)^8) \leq \frac{C_1}{n^9} \tag{3.19}$$

*for all $n \geq 1$.*

Assuming the above lemma (which is proved later), we now complete the proof of the lower bound on the speed. Fix $m \geq n^{1/9}$. If the event $V_m \cap \{T_{max} \leq (\log n)^8\} \cap \Gamma(x_0)$ occurs, then within time $(\log n)^8$ all nodes of $S(K_n)$ are infected and within time $2(\log n)^8$ all nodes of $S(3K_n)$ are infected. This necessarily implies that infection has reached some node of the backbone within time $2(\log n)^8$. From the backbone, the infection therefore reaches at least one node of each square $S'_i$ contained in $m\frac{r_n}{\Delta}S$ within time $2(\log n)^8 + \frac{4D_1 m}{nr_n^2}$.

Hence within time $2(\log n)^8 + \frac{4D_1 m}{nr_n^2} + (\log n)^8 \leq \frac{5D_1 m}{nr_n^2}$, the infection reaches all nodes of $G(x_0)$ in $m\frac{r_n}{\Delta}S$. In the final estimate, we use the fact that $m \geq n^{1/9}$ and therefore that $(\log n)^8 = o\left(\frac{m}{nr_n^2}\right)$ by virtue of (2.2). Summarizing, if $m \geq n^{1/9}$ and $V_m \cap \{T_{max} \leq (\log n)^8\} \cap \Gamma(x_0)$ occurs, then

$$\left(G(x_0) \setminus I\left(\frac{5D_1 m}{nr_n^2}\right)\right) \bigcap m\frac{r_n}{\Delta}S = \phi,$$

which is nearly what we want to prove.

So far we have assumed that $m$ is a multiple of $MK_n$ and estimated the time taken to cross the boundary of $m\frac{r_n}{\Delta}S$. To prove the lower bound on the speed, however, we need estimates on the time taken for the infection to cross the boundary of $m_3 r_n S$ for *every* $a(r_n^{-1}) \leq m_3 \leq r_n^{-1} - g(r_n^{-1})$ where $a(x) = o(x)$ and $g(x) = o(x)$ as $x \to \infty$ (see definition prior to Theorem 3.1). We proceed as follows. We set $a(r_n^{-1}) = n^{1/9}$ (which is $o(r_n^{-1})$ by (2.2)). For $m_3 \geq n^{1/9}$, let $m$ be the smallest integer that is a multiple of $MK_n$ and such that $S \supseteq m\frac{r_n}{\Delta}S \supseteq m_3 r_n S$. Since $\Delta \in [4,5]$ we have that

$$4n^{1/9} \leq 4m_3 \leq m \leq 5m_3 + 5MK_n \leq 6m_3.$$

Here we use $K_n = \frac{\log n}{nr_n^2} \leq \log n$. By the last sentence in the previous paragraph and the above equation, we have that if $V_m \cap \{T_{max} \leq (\log n)^8\} \cap \Gamma(x_0)$ occurs, then $\left(G(x_0) \setminus I\left(\frac{30D_1m_3}{nr_n^2}\right)\right) \bigcap m_3r_nS = \phi$. This conclusion holds for each $n^{1/9} \leq m_3 \leq r_n^{-1} - MK_n$. Since $MK_n = M\frac{\log n}{nr_n^2} \leq M \log n \leq (\log n)^2$ and $r_n^{-1} = O(\sqrt{n})$ (see (2.2)), we have from (3.18) that

$$\mathbb{P}\left(\bigcap_{n^{1/9} \leq m_3 \leq r_n^{-1} - (\log n)^2} V_m\right) \geq 1 - \frac{1}{n^7}.$$

From (3.19) and the estimate for $\Gamma(x_0)$ in (3.7), we have that

$$\mathbb{P}\left(\bigcap_{n^{1/9} \leq m_3 \leq r_n^{-1} - (\log n)^2} V_m \bigcap \{T_{max} \leq (\log n)^8\} \bigcap \Gamma(x_0)\right) \geq 1 - \frac{2}{n^7} - e^{-\theta_1 nr_n^2},$$

where $\theta_1$ is as in (3.7). Since $(\log n)^2 = o(r_n^{-1})$ by virtue of (2.2), this implies the lower bound on the speed in Theorem 3.1. $\square$

*Proof of Lemma 3.6*: For a constant $D_2 > 0$, we let $B = \left\{T(R_2) > \frac{2D_2m}{nr_n^2}\right\}$, $A = B \cap E_{n,tot}$ and use (3.6) to obtain that

$$\mathbb{P}(A) = \int \nu_p(A_\omega)\mu(d\omega) = \int_{E_{n,tot}} \nu_p(B_\omega)\mu(d\omega) \tag{3.20}$$

where as mentioned in Section 3.2, $\nu_p(B_\omega)$ denotes the probability that event $B$ occurs for a fixed configuration of points $\omega$. From the discussion in the paragraph preceding Lemma 3.8 we have that if $\omega \in E_{n,tot}$, then the passage time $T(R_2)$ of $R_2$ satisfies

$$T(R_2) = \sum_{i=1}^{q-1} t(h_i) \leq \sum_{i=1}^{q} X_i \leq \sum_{i=1}^{10Mm} X_i$$

where $\{X_i\}_i$ are i.i.d random variables with $X_i = \min_{1 \leq j \leq \beta_1 nr_n^2} t_{i,j}$ and $t_{i,j}$ are i.i.d exponential with unit mean. Here $\beta_1 > 0$ is as in (3.9). Thus

$$\nu_p(B_\omega) \leq \mathbb{P}\left(\sum_{i=1}^{10Mm} X_i > \frac{2D_2m}{nr_n^2}\right)$$

where the right hand side expression does not depend on $\omega$. Integrating over $\omega$, we have from (3.20) that

$$\mathbb{P}\left(\left\{T(R_2) > \frac{2D_2m}{nr_n^2}\right\} \bigcap E_{n,tot}\right) = \mathbb{P}\left(\sum_{i=1}^{10Mm} X_i > \frac{2D_2m}{nr_n^2}\right). \tag{3.21}$$

Since $\beta_1 n r_n^2 X_i$ is exponentially distributed with mean one, we use Chernoff bound and obtain for $D_2 > 0, s \in (0,1)$ that

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{10Mm} X_i > \frac{2D_2 m}{n r_n^2}\right) &\leq \left(\mathbb{E}\exp\left(sX_1\beta_1 n r_n^2\right)\right)^{10Mm} e^{-2s\beta_1 D_2 m} \\ &= \left(\frac{1}{1-s}\right)^{10Mm} e^{-2s\beta_1 D_2 m}.\end{aligned}$$

Therefore fixing $s = \frac{1}{2}$ and choosing the constant $D_2 > 0$ sufficiently large, we have for all $n \geq N_0$ sufficiently large and all $m \geq n^{1/9}$ that the last expression above is no more than $2^{10Mm} e^{-\beta_1 D_2 Mm} \leq e^{-\delta_1 m}$ for some positive constant $\delta_1$. $\square$

*Proof of Lemma 3.8*: Let $E_d(n)$ denote the event that every square in the set of $\frac{r_n}{\Delta} \times \frac{r_n}{\Delta}$ squares $\{S_i\}_i$, contains less than $K \log n$ nodes for some constant $K \geq 1$. Using $n r_n^2 \leq c_2 \log n$ (see (2.2)), we have

$$\mathbb{P}(E_d(n)) \geq 1 - \frac{1}{n^{10}} \tag{3.22}$$

if $K$ is sufficiently large. For a fixed $i$, let $\mathcal{E}_i$ denote the set of edges with at least one endvertex in $S_i'$. The square $S_i'$ contains $(MK_n)^2$ squares in $\{S_j\}_j$. Therefore if $E_d(n)$ occurs, the number of nodes in the $3MK_n\frac{r_n}{\Delta} \times 3MK_n\frac{r_n}{\Delta}$ square with the same centre as $S_i'$ is less than $(3MK_n)^2 K \log n$. Consequently the number of edges in $\mathcal{E}_i$ is less than $C_1(K_n^2 \log n)^2 \leq C_2(\log n)^6$ for some positive constants $C_1$ and $C_2$. Here we use $K_n = \frac{\log n}{n r_n^2}$. Arguing as in the derivation of (3.21) in proof of Lemma 3.6 above, we average over the configurations and get

$$\begin{aligned}\mathbb{P}\left(T_i > (\log n)^8\right) &\leq \mathbb{P}\left(\left\{T_i > (\log n)^8\right\} \cap E_d(n)\right) + \frac{1}{n^{10}} \\ &\leq \mathbb{P}\left(\sum_{i=1}^{C_1(\log n)^6} t_i > (\log n)^8\right) + \frac{1}{n^{10}},\end{aligned}$$

where $t_i$ are i.i.d exponential with unit mean. We have

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{C_1(\log n)^6} t_i > (\log n)^8\right) &\leq \mathbb{P}\left(\bigcup_{i=1}^{C_1(\log n)^6} \left\{t_i > C_1^{-1}(\log n)^2\right\}\right) \\ &\leq C_1(\log n)^6 e^{-C_1^{-1}(\log n)^2}.\end{aligned}$$

Thus
$$\mathbb{P}\left(T_i > (\log n)^8\right) \le C_1(\log n)^6 e^{-C_1^{-1}(\log n)^2} + \frac{1}{n^{10}} \le \frac{2}{n^{10}}$$
for all $n$ sufficiently large. Since the maximum possible number of squares in $\{S'_i\}_i$ is $\left(\frac{\Delta}{r_n}\right)^2 = O(n)$ by (2.2), we have that

$$\mathbb{P}\left(T_{max} > (\log n)^8\right) \le \sum_i \mathbb{P}(T_i > (\log n)^8) \le \frac{O(n)}{n^{10}}$$

proving (3.19). □

## 3.4 Proof of Theorem 3.1: Upper bound on speed

At time $t = 0$, the node $x_0$ of $G$ closest to the origin is infected. As before, we assume initially the occurrence of the event $\Gamma(x_0)$ that there exists a path of edges from $x_0$ to the boundary of $S$ (see (i)-(ii) prior to (3.7)). For a fixed $\log n \le m \le r_n^{-1} - 5$, we now look at the path $\pi_m$ through which the infection first reaches the boundary of $mr_nS$. More precisely, let $\pi = (h_0, ..., h_b)$ be a self-avoiding path of edges such that:
(iii) $h_0$ contains $x_0$ as one of its endvertex, exactly one endvertex of $h_b$ lies in $S \setminus mr_nS$ and
(iv) all other endvertices of the edges $\{h_i\}_i$ lie in $mr_nS$.
Such a path definitely exists because of the occurrence of the event $\Gamma(x_0)$. Define $T(\pi) = \sum_{i=0}^{b} t(h_i)$ to be the passage time of $\pi$ and let $\pi_m$ be that path whose passage time is $T(\pi_m) = \min_\pi T(\pi)$, where the minimum is taken over all paths satisfying (iii)-(iv) above. Such a unique path exists since the passage times are continuous random variables.

To bound $T(\pi_m)$ we recall the event $E_d(n)$ defined prior to (3.22). If $E_d(n)$ occurs, then each node has less than $K_1 \log n$ neighbours for some fixed constant $K_1 > 0$. Therefore, if $E_d(n)$ occurs, then the number of edges of $G$ is less than $K_1 n \log n$. If $e_1, ..., e_T$ denotes the set of edges, we then have that
$$t(e_i) \ge_{st} \min_{1 \le j \le K_1 n \log n} t_j =: X_0,$$
where $\{t_j\}_j$ are i.i.d. exponential with unit mean and $\ge_{st}$ denotes stochastic

domination. Since $\pi_m$ contains at least $\frac{m}{4}$ edges, we then have that

$$T(\pi_m) \geq_{st} \frac{m}{4} X_0.$$

Using the estimate

$$\mathbb{P}\left(X_0 \geq \frac{1}{n\sqrt{n}\log n}\right) = 1 - O\left(\frac{1}{\sqrt{n}}\right),$$

we have from the above discussion that

$$\mathbb{P}\left(\left\{T(\pi_m) \geq \frac{m}{4n\sqrt{n}\log n}\right\} \bigcap E_d(n) \bigcap \Gamma(x_0)\right) = 1 - O\left(\frac{1}{\sqrt{n}}\right).$$

Therefore,

$$\mathbb{P}\left(\bigcap_{\log n \leq m \leq r_n^{-1}-5} \left\{T(\pi_m) \geq \frac{m}{4n\sqrt{n}\log n}\right\} \bigcap E_d(n) \bigcap \Gamma(x_0)\right) = 1 - O\left(\frac{r_n^{-1}}{\sqrt{n}}\right)$$

and the final expression is $1 - o(1)$ as $n \to \infty$, since $nr_n^2 \longrightarrow \infty$. We note that if $T(\pi_m) \geq \frac{m}{4n\sqrt{n}\log n}$, then $I\left(\frac{m}{8n\sqrt{n}\log n}\right) \subseteq mr_nS$. From the estimates of the probabilities of the events $E_d(n)$ and $\Gamma(x_0)$ in (3.22) and (3.7), respectively, we therefore get the upper bound on the speed with $a(r_n^{-1}) = \log n = o(r_n^{-1})$ and $g(r_n^{-1}) = 5 = o(r_n^{-1})$, (by (2.2)). $\square$

## 3.5 Proof of Corollary 3.2

*Proof of (3.3)*: Let $m$ be a multiple of $MK_n$ (the constant $M$ as in Lemma 3.4) that satisfies

$$m\frac{r_n}{\Delta}S \subseteq S \subseteq (m + MK_n)\frac{r_n}{\Delta}S.$$

Using $\Delta \in [4,5]$ and $K_n = \frac{\log n}{nr_n^2} = o(r_n^{-1})$ by (2.2), we get $4r_n^{-1} \leq m \leq 6r_n^{-1}$ for all $n$ sufficiently large. The square $m\frac{r_n}{\Delta}S$ is the largest square contained in $S$ to which the tiling argument of the proof of Theorem 4.3 described in Section 4.3 can be applied. Consequently, there exists a backbone of low passage time connections as described in the paragraph preceding (3.18).

Suppose first that the event $V_m$ defined prior to (3.18) and the event $\{T_{max} \leq (\log n)^8\}$ defined prior to Lemma 3.8 occurs and let $U_m = V_m \cap$

$\{T_{max} \leq (\log n)^8\}$. Let $\Gamma(x_0)$ as defined prior to (3.7) and $E_d(n)$ be as defined prior to (3.22). We have from (3.7), (3.18), (3.22) and (3.19) that $U_m \cap \Gamma(x_0) \cap E_d(n)$ occurs with probability $1-o(1)$. By the proof of lower and upper bound on the speed in Theorem 4.3, we therefore have with probability $1-o(1)$ that the time elapsed $T_0$ before all nodes of $G(x_0) \cap m\frac{r_n}{\Delta}S$ are infected satisfies

$$\frac{4D_1 r_n^{-1}}{n\sqrt{n}\log n} \leq \frac{D_1 m}{n\sqrt{n}\log n} \leq T_0 \leq \frac{D_2 m}{nr_n^2} \leq \frac{6D_2 r_n^{-1}}{nr_n^2} \tag{3.23}$$

for some positive constants $D_1$ and $D_2$. The first and the last inequalities are true by our choice of $m$. We claim that by time $\frac{6D_2r_n^{-1}}{nr_n^2} + (\log n)^8 \leq \frac{7D_2r_n^{-1}}{nr_n^2}$, all nodes of $G(x_0)$ are infected. This is true since $\{T_{max} \leq (\log n)^8\}$ occurs. Here we use the fact that $(\log n)^8 = \frac{o(r_n^{-1})}{nr_n^2}$ by (2.2). This proves the lower and upper bound in (3.3) and the lower bound in (3.4).

*Proof of (3.4)*: The lower bound in (3.4) is proved above. To prove the upper bound in (3.4), we recall the event $E_d(n)$ defined prior to (3.22) that the number of nodes of each square in $\{S_k\}_k$ is less than $K\log n$ and the event $U_m$ defined in the proof of (3.3) above. Also, $x_0$ denotes the node of $G$ closest to the origin. Let $\Gamma_1(x_0)$ denote the event that $x_0 \in S(K_n)$ and the component $G(x_0)$ contains at least one node outside $S(3K_n)$. As before, $S(K_n)$ is the $MK_n\frac{r_n}{\Delta} \times MK_n\frac{r_n}{\Delta}$ square with centre at the origin, where $M$ is the constant in Proposition 3.4. We now write

$$\begin{aligned}\mathbb{E}T_{elap} &= \mathbb{E}T_{elap}\mathbf{1}(U_m \cap \Gamma_1(x_0)) + \mathbb{E}T_{elap}\mathbf{1}(U_m \cap \Gamma_1^c(x_0) \cap E_d(n)) \\ &\quad + \mathbb{E}T_{elap}\mathbf{1}(U_m \cap \Gamma_1^c(x_0) \cap E_d^c(n)) + \mathbb{E}T_{elap}\mathbf{1}(U_m^c) \\ &\leq \mathbb{E}T_{elap}\mathbf{1}(U_m \cap \Gamma_1(x_0)) + \mathbb{E}T_{elap}\mathbf{1}(\Gamma_1^c(x_0) \cap E_d(n)) \\ &\quad + \mathbb{E}T_{elap}\mathbf{1}(E_d^c(n)) + \mathbb{E}T_{elap}\mathbf{1}(U_m^c)\end{aligned} \tag{3.24}$$

and evaluate each term separately.

For the first term, we note that $\Gamma_1(x_0)$ occurs and therefore there is a path $\pi_1$ of edges from $x_0 \in S(K_n)$ that crosses $S(3K_n)$. By an analogous argument as in the two paragraphs following (3.17), the path $\pi_1$ intersects the backbone $\mathcal{P}$ (present due to the occurrence of $U_m$). Thus we have from the proof of upper bound of (3.3) above that

$$\mathbb{E}T_{elap}\mathbf{1}(U_m \cap \Gamma_1(x_0)) \leq \frac{7D_2r_n^{-1}}{nr_n^2}. \tag{3.25}$$

We now show that each of the remaining term in (3.24) is $\frac{o(r_n^{-1})}{nr_n^2}$.

To evaluate the second term, we write $\Gamma_1^c(x_0) = \Gamma_{1,1}(x_0) \cup \Gamma_{1,2}(x_0)$, where $\Gamma_{1,1}(x_0)$ is the event that $x_0 \notin S(K_n)$ and $\Gamma_{1,2}(x_0)$ is the event that $G(x_0)$ is contained in $S(3K_n)$. If $\Gamma_{1,2}(x_0) \cap E_d(n)$ occurs, then the component containing $x_0$ is completely contained in $S(3K_n)$. The time elapsed before no new nodes are infected is bounded above by the sum of the passage times of edges contained in the square $S(3K_n)$. Since $E_d(n)$ occurs, the square $S(3K_n)$ contains less than $(3MK_n)^2 \log n \leq (\log n)^4$ nodes and therefore less than $(\log n)^8$ edges if $n$ is sufficiently large. Here we use $K_n = \frac{\log n}{nr_n^2} \leq \log n$ for all $n$ sufficiently large since $nr_n^2 \longrightarrow \infty$. Since passage time of any edge has unit mean, this implies that

$$\mathbb{E}(T_{elap}\mathbf{1}(\Gamma_{1,2}(x_0) \cap E_d(n))) \leq \mathbb{E} \sum_{i=1}^{(\log n)^8} t_i = (\log n)^8$$

for all $n$ sufficiently large. In the above, $\{t_i\}_i$ are i.i.d Exponential with unit mean. Using (2.2) we have that the right hand side of the above is $\frac{o(r_n^{-1})}{nr_n^2}$.

We estimate $\mathbb{E}(T_{elap}\mathbf{1}(\Gamma_{1,1}(x_0) \cap E_d(n)))$ and the third and the fourth terms in (3.24) together. We note that if $\Gamma_{1,1}(x_0)$ occurs, then $S(K_n)$ is empty. Again using standard Binomial estimates (see e.g. Chapter 1 of Penrose (2003)), we have that

$$\mathbb{P}(\Gamma_{1,1}(x_0)) \leq e^{-\theta_1 (MK_n)^2 nr_n^2}$$

for some constant $\theta_1 > 0$ and for all $n$ sufficiently large. Choosing $M$ larger if necessary we have that

$$(MK_n)^2 nr_n^2 = \frac{M^2(\log n)^2}{nr_n^2} \geq \frac{10 \log n}{\theta_1},$$

so that $\mathbb{P}(\Gamma_{1,1}(x_0)) \leq \frac{1}{n^{10}}$. Here we use (2.2) and $K_n = \frac{\log n}{nr_n^2}$. Thus using Cauchy-Schwarz inequality, we have that

$$\begin{aligned}
\mathbb{E}T_{elap}\mathbf{1}(\Gamma_{1,1}(x_0) \cap E_d(n)) &\leq \mathbb{E}T_{elap}\mathbf{1}(\Gamma_{1,1}(x_0)) \\
&\leq \left(\mathbb{E}T_{elap}^2\right)^{1/2} \mathbb{P}(\Gamma_{1,1}(x_0))^{1/2} \\
&\leq \frac{1}{n^5} \left(\mathbb{E}T_{elap}^2\right)^{1/2}. \qquad (3.26)
\end{aligned}$$

Similarly, we bound the third term above as

$$\mathbb{E}T_{elap}\mathbf{1}(U_m^c) \leq \left(\mathbb{E}T_{elap}^2\right)^{1/2} \mathbb{P}(U_m^c)^{1/2}$$

where we use Cauchy Schwarz inequality in the final estimate. From (3.18) and Lemma 3.8 we have that

$$\mathbb{P}(U_m^c) \leq \mathbb{P}(V_m^c) + \mathbb{P}(T_{max} > (\log n)^8) \leq \frac{2}{n^8} + \frac{1}{n^8} \leq \frac{3}{n^8}$$

for all $n$ sufficiently large. Thus, the third term is bounded above by $C_1 \frac{\left(\mathbb{E}T_{elap}^2\right)^{1/2}}{n^4}$ for some positive constant $C_1$.

Also, have from (3.22) that

$$\mathbb{E}T_{elap}\mathbf{1}(E_d^c(n)) \leq (\mathbb{E}T_{elap}^2)^{1/2}\mathbb{P}(E_d^c(n))^{1/2} \leq \frac{(\mathbb{E}T_{elap}^2)^{1/2}}{n^5}.$$

Thus from (3.26), the sum of $\mathbb{E}(T_{elap}\mathbf{1}(\Gamma_{1,1}(x_0) \cap E_d(n)))$ and the third and the fourth terms in (3.24) is bounded above by $C_2 \frac{\left(\mathbb{E}T_{elap}^2\right)^{1/2}}{n^4}$ for some positive constant $C_2$. Since the number of edges in $G$ is at most $n^2$, we have that $T_{elap} \leq \sum_{i=1}^{n^2} t_i$, where $t_i$ are i.i.d Exponential with unit mean. Hence by the AM-QM inequality we have that

$$\mathbb{E}T_{elap}^2 \leq \mathbb{E}n^2 \sum_{i=1}^{n^2} t^2(e_i) \leq C_2 n^4$$

for some positive constant $C_2$. Here we use the fact that $\mathbb{E}t(e)^2 < \infty$. Thus

$$\frac{\left(\mathbb{E}T_{elap}^2\right)^{1/2}}{n^4} \leq \frac{C_3}{n^2} = \frac{o(r_n^{-1})}{nr_n^2}$$

for some positive constant $C_3$ by (2.2).

*Proof of (3.5)*: To prove (3.5) we note from the proof of Theorem 4.3 that infection starting from the node $x_0$ closest to the origin crosses the boundary of $\frac{r_n^{-1}}{2}S$ with probability $1 - o(1)$. By the construction of giant component in the proof of (ii) in Theorem 4.3 of Ganesan (2012) we know that this path intersects the giant component with probability $1 - o(1)$. From the estimate on the size of the giant component in Theorem 4.3(ii) of Ganesan (2012), we know that the giant component contains at least $n - ne^{-\theta n r_n^2}$ nodes with probability $1 - o(1)$, for some constant $\theta > 0$. The equation (3.5) then follows. $\square$

# Chapter 4

# Convergence rate of locally determinable Poisson functionals

## 4.1 Introduction

Functionals of point processes arise naturally in computational geometry and Boolean models. The most common application (see e.g. Heinrich, Schmidt and Schmidt (2005), Møller (1994), Meester and Roy (1996)) is to estimate a certain parameter of the process from a single realization over a (possibly) large area. In such situations it is important to study how fast the proposed (consistent) estimator converges to the true value of the parameter in question.

Before we state the main result of this chapter, we present two examples, the Poisson Voronoi Tessellation and the Poisson Boolean Model, where our main result may be applied.

### 4.1.1 Poisson Voronoi Tessellation

Consider for example the Poisson Voronoi Tessellation defined on $\mathbb{R}^2$ as follows. For $\omega = \{y_1, y_2, ....\} \subset \mathbb{R}^2$ and $x \in \omega$ let

$$T(x, \omega) = \{z \in \mathbb{R}^2 : d(z, x) \leq d(z, y_j) \text{ for all } y_j \neq x\}$$

denote the *Voronoi tessellate* (Møller (1994)) containing the point $x$. Here and henceforth $d(a, b)$ represents the Euclidean distance between $a$ and $b$. Let $\mathcal{N}$ denote the realization of a Poisson process of unit intensity in $\mathbb{R}^2$ and let $\mathcal{J}$ denote the random Voronoi tessellation of $\mathbb{R}^2$ obtained from the points of $\mathcal{N}$. Such random tessellations are important in the study of many topics. See Bollobás and Riordan (2006b) for site percolation on Poisson Voronoi tessellation and Møller (1994) for more properties and applications. Let $f_V(x) = f_V(x, \mathcal{N})$ be the number of facets of the (random) tessellate of $\mathcal{J}$ containing the point $x \in \mathcal{N}$. For $n \geq 1$ define

$$X_V(nW) = \sum_{x \in nW \cap \mathcal{N}} f_V(x) \tag{4.1}$$

where $W = \left[-\frac{1}{2}, \frac{1}{2}\right]^2$. By ergodicity and stationarity, the scaled functional $\frac{X_V(nW)}{n^2}$ is an unbiased estimator of the mean intensity $\mu_V = \mathbb{E}\frac{X_V(mW)}{m^2}$ of the facets (see also Heinrich, Schmidt and Schmidt (2005), Møller (1994)). The following result determines the rate of convergence.

**Proposition 4.1.** *Fix $\gamma > 0, p > 1$ and $\delta > 0$. There exists positive constants $C_1 = C_1(\gamma, p, \delta)$ and $C_2 = C_2(\gamma, p, \delta)$ such that*

$$\mathbb{P}\left(\left|\frac{X_V(mW)}{m^2} - \mu_V\right| > \frac{1}{C_1 m^\delta \log m}\right) \leq \frac{C_2}{m^\gamma} \tag{4.2}$$

*and*

$$\mathbb{E}\left|\frac{X_V(mW)}{m^2} - \mu_V\right|^p \leq \frac{C_2}{m^\gamma} \tag{4.3}$$

*for all $m \geq 1$.*

The term $\gamma$ in the above equations is a lower bound on the rate of convergence of the bias in the estimator. Thus, the bias in the estimator converges at rate greater than $\gamma$ for *every* $\gamma > 0$. Here and henceforth, we use $\mathbb{P}$ and $\mathbb{E}$ to denote a generic probability measure and expectation operator, respectively.

### 4.1.2 Poisson Boolean Model

We consider the Poisson Boolean model consisting of a homogenous Poisson point process $\mathcal{N} = \{x_1, x_2, ...\}$ of intensity $\lambda$ in $\mathbb{R}^2$ and a sequence of non-negative independent and identically distributed (i.i.d.) random variables $\rho_1, \rho_2, \ldots$, independent of the Poisson process. Throughout we assume that $0 < \rho_1 \leq R$ *a.s.* for some $R > 0$. The point $x_i$ is associated with the mark $\rho_i$ and the resulting marked process $\mathcal{N}_M$ is Poisson and is called the Poisson Boolean model of continuum percolation on $\mathbb{R}^2$.

For $x \in \mathbb{R}^2$ and $r > 0$, let $S(x, r) = \{y : d(y, x) \leq r\}$ denote the ball of radius $r$ centred at $x$ and define

$$\lambda_c = \inf\{\lambda > 0 : \mathbb{P}(diam(\mathcal{C}(0)) = \infty) > 0\}, \tag{4.4}$$

where $\mathcal{C}(0)$ denotes the component of the occupied region $\cup_i S(x_i, \rho_i)$ containing the origin and $diam(A) = \sup\{d(x, y) : x, y \in A\}$ denotes the diameter of the set $A$. For $\lambda < \lambda_c$, we know by stationarity that a.s. the occupied region is a countable collection of disjoint bounded connected components $\{\mathcal{C}_j\}_j$; i.e., $\cup_i S(x_i, \rho_i) = \cup_j \mathcal{C}_j$ where $\mathcal{C}_j$'s are mutually disjoint and each $\mathcal{C}_j$ is a maximal connected component with finite diameter.

For $A \subset \mathbb{R}^2$, we let

$$X_B(A) = \sum_{x_i \in A \cap \mathcal{N}} f_B(x_i),$$

where $f_B(x_i) = f_B(x_i, \mathcal{N}_M)$ denotes the number of balls in the occupied component containing $x_i$. By stationarity, we know that $\mu_B = \mathbb{E}\frac{X_B(nW)}{n^2}$ represents the mean number of balls in the occupied component containing the origin. By ergodicity we know that

$$\frac{X_B(nW)}{n^2} - \mu_B \longrightarrow 0 \;\; a.s.$$

as $n \to \infty$. We have the following result regarding the rate of convergence of bias in the estimator.

**Proposition 4.2.** *For $\lambda < \lambda_c$, we have that $\mu_B < \infty$. Fix $\gamma > 0, p > 1$ and $\delta > 0$. There exists positive constants $C_1 = C_1(\gamma, p, \delta)$ and $C_2 = C_2(\gamma, p, \delta)$ such that*

$$\mathbb{P}\left(\left|\frac{X_B(mW)}{m^2} - \mu_B\right| > \frac{1}{C_1 m^\delta \log m}\right) \leq \frac{C_2}{m^\gamma} \tag{4.5}$$

*and*

$$\mathbb{E}\left|\frac{X_B(mW)}{m^2} - \mu_B\right|^p \leq \frac{C_2}{m^\gamma} \tag{4.6}$$

*for all* $m \geq 1$.

Proposition 4.1 and 4.2 are obtained as Corollaries of a more general result we prove below.

### 4.1.3 Convergence rate of Poisson functionals

Let $\mathcal{N} = \{x_1, x_2, ...\}$ be a Poisson point process on $\mathbb{R}^d$, $d \geq 2$, with intensity measure $\Lambda(.)$. On each point $x$ of $\mathcal{N}$ we place an independent and identically distributed mark $t_x$ defined on the probability space $(\mathcal{M}, \mathcal{F}_{\mathcal{M}}, \mu_M)$. The resulting marked point process $\mathcal{N}_M$ is also a Poisson process on $\mathbb{R}^d \times \mathcal{M}$ (see e.g. Daley and Jones (2008)). We denote $\mathbb{P}$ and $\mathbb{E}$ to be the probability measure and expectation operator, respectively, with respect to the marked process $\mathcal{N}_M$. For any set $A \subset \mathbb{R}^d$, we then have that

$$\mathbb{P}(\#\mathcal{N} \cap A = k) = e^{-\Lambda(A)}\frac{\Lambda(A)^k}{k!}.$$

For $x \in \mathbb{R}^d$, let $\mathbb{P}_x$ denote the probability measure of the process $\mathcal{N}_M$ conditioned to have a point at $x$.

For $x \in \mathbb{R}^d$ and $m > 0$, we define $B_m(x) = x + \left[-\frac{m}{2}, \frac{m}{2}\right]^d$ to be the cube of side length $m$ centred at $x$ and let $B_m^*(x) = B_m(x) \times \mathcal{M}$. Denote $B_m(0)$ simply as $B_m$.

Let $f(x, \omega)$ be any measurable real valued function defined for all pairs $(x, \omega)$ where $\omega \subset \mathbb{R}^d \times \mathcal{M}$ is countable and $x \in \mathbb{R}^d$. We wish to determine rate of convergence of functionals defined for $A \subset \mathbb{R}^d$ as

$$X(A) = \sum_{x \in \mathcal{N} \cap A} f(x, \mathcal{N}_M). \tag{4.7}$$

To state our main result regarding $X$, we assume that the functional $X$ and the intensity measure $\Lambda$ satisfy the following on all rectangles $A$ whose shortest edge has length at least one. (Here a rectangle is a set of the form $\prod_{j=1}^d [a_j, b_j]$ for some real numbers $a_j, b_j$.)

(i) There exists a positive constant $C_1$ independent of the choice of $A$ so that

$$C_1^{-1} \leq \frac{\Lambda(A)}{\ell(A)} \leq C_1, \tag{4.8}$$

where $\ell(A)$ refers to the Lebesgue measure of $A$.
(ii) There exists positive constants $p$ and $C_2$ independent of $A$ such that

$$\mathbb{E}\left|\frac{X(A)}{\Lambda(A)}\right|^p \leq C_2. \tag{4.9}$$

(iii) For every $v = (x,t) \in nW \times \mathcal{M}$, there exists $r_x = r_x(t, \mathcal{N}_M, n) < \infty$ *a.s.*, such that

$$f(x, \mathcal{N}_M \cup \{v\}) = f(x, \omega' \cup \{v\})$$

for all $\omega'$ satisfying $(\omega' \cup \{v\}) \cap B^*_{2r_x}(x) = (\mathcal{N}_M \cup \{v\}) \cap B^*_{2r_x}(x)$.
(iv) There exists a positive constant $\alpha$ and a constant $C(\alpha) > 0$ such that

$$\sup_{x\in\mathbb{R}^d} \mathbb{P}_x(r_x \geq m) \leq \frac{C}{m^\alpha} \tag{4.10}$$

for all $m \geq 1$.
Statement (i) essentially implies that the intensity measure is comparable to the Lebesgue measure. Integrability of the functional is described in (ii) and in (iii) we state the locally determinable property of $X$ : for every $v = (x,t) \in \mathbb{R}^d \times \mathcal{M}$, there exists a radius $r_x$, finite a.s., such that the value of $f(x, \mathcal{N}_M \cup \{v\})$ is determined by the restriction of $\mathcal{N}_M$ to the cube $B^*_{2r_x}(x)$. We call the smallest such $r_x$ to be the radius of determinability at $x$ for the realization $\mathcal{N}_M$. Finally, in (iv), we require mild tail conditions on $r_x$. Here and henceforth, we use the following notation: For $x \in \mathbb{R}^d$, we let $\mathbb{P}_x$ denote the probability measure of the process $\mathcal{N}_M$ conditioned to have a point at $x$.

Conditions (iii)-(iv) are analogous to but slightly different from the notion of stability discussed in Penrose and Yukich (2003), Baryshnikov and Yukich(2005), Penrose (2007).

The following is the main result of this chapter.

**Theorem 4.3.** *Suppose (i)-(iv) are satisfied for some positive constants $p > 1$ and $\alpha > d$. There exists positive constants $\gamma_1, \gamma_2$ and $C$ so that*

$$\mathbb{P}\left(|X(nW) - \mathbb{E}X(nW)| \geq \frac{C\Lambda(nW)}{n^{\gamma_1}\log n}\right) \leq \frac{1}{Cn^{\gamma_2}} \tag{4.11}$$

*for all $n \geq 1$. Also, if $0 < r < p$ and $0 < \gamma < \min\left(r\gamma_1, \left(1 - \frac{r}{p}\right)\gamma_2\right)$ are positive constants, then there exists a positive constant $C_1 = C_1(r,\gamma)$ such that*

$$\mathbb{E}\left|\frac{X(nW)}{\Lambda(nW)} - \mathbb{E}\frac{X(nW)}{\Lambda(nW)}\right|^r \leq \frac{C_1}{n^\gamma} \tag{4.12}$$

*for all* $n \geq 1$.

If $\frac{X(nW)}{\Lambda(nW)}$ is an estimator as in Sections 4.1.1 and 4.1.2, the quantity $\gamma$ in (4.12) is a lower bound on the rate of convergence of the bias.

While the above theorem guarantees the positivity of convergence rate, we are also interested to know how convergence rate varies with the decay rate of the radius of determinability. We have the following result.

**Theorem 4.4.** *Fix* $\delta \in \left[0, \frac{1}{2}\right)$ *and* $\eta > 0$. *Suppose that the functionals* $X$ *and* $\Lambda$ *satisfy (i)-(iv) for some constants* $p > \max\left(\frac{d+4\eta}{d-4\delta}, \frac{2\eta}{1-2\delta}\right)$ *and*

$$\alpha > \alpha_0 = d\left(\frac{2-a_0}{2-2a_0}\right) + \frac{p(\delta+\eta)}{(p-1)(1-a_0)}, \tag{4.13}$$

*where* $a_0 = 2\max\left(\frac{(2p\delta+2\eta)}{d(p-1)}, \frac{\eta}{p} + \delta\right)$. *We have that (4.11) holds for some* $\gamma_1 > \delta$ *and* $\gamma_2 > \eta$.

It is easy to check that $0 < \alpha_0 < \infty$. The term $\alpha_0$ tells us how the convergence rate is affected by the decay rate of the the radius of determinability. For $d \geq 3$, we set $\delta = 0$ and $\gamma = 1$ in the above result and use Borel-Cantelli Lemma to obtain that

$$\frac{X(nW)}{\Lambda(nW)} - \mathbb{E}\frac{X(nW)}{\Lambda(nW)} \longrightarrow 0 \ \text{ a.s. as } n \to \infty$$

provided (i)-(iv) are satisfied for some $p > 3$ and $\alpha > 2d + 5$.

Before we prove the above results, we state a sufficient condition that ensures that $X$ satisfies (4.9) for some constant $p \geq 1$. For $l \geq 1$, let $\mathcal{X} = \{x_1, ..., x_l\}$ denote any fixed set of $l$ points in $\mathbb{R}^d$ and let $t_i$ denote a random mark at $x_i$. Since the marked process $\mathcal{N}_M$ is Poisson, the term $\mathbb{E}f(x_1, \mathcal{N}_M \cup \cup_{i=1}^{l}(x_i, t_i))$ represents the expected value of $f$ at $x_1$ conditioned on the event that the marked process contains the points $\{(x_i, t_i)\}_{1 \leq i \leq l}$ (see e.g. Daley and Jones (2008)). Averaging over the marks, we then let

$$\mathbb{E}_{x_1, \mathcal{X}} f := \int_{\mathcal{M}} ... \int_{\mathcal{M}} \mathbb{E}f(x_1, \mathcal{N}_M \cup \cup_{i=1}^{l}(x_i, t_i))\mu_M(dt_1)...\mu_M(dt_l) \tag{4.14}$$

to denote the expected value of $f$ at the point $x_1$ conditioned on the event that $\mathcal{X} \subseteq \mathcal{N}$. We have the following result.

**Proposition 4.5.** *If for some integer $k \geq 1$ we have*

$$\sup_{\mathcal{X}} \mathbb{E}_{x_1,\mathcal{X}} |f|^k < \infty, \tag{4.15}$$

*where the supremum is taken over all sets $\mathcal{X} = \{x_1, ..., x_l\}$ having $l \leq k$ distinct points, then the functional $X$ satisfies (4.9) with $p = k$.*

In Section 4.3, we use the expression in (4.14) along with the Slivnyak-Mecke formula (Møller (1994)) to prove the above Proposition.

The chapter is organized as follows: In Section 4.2, we prove Propositions 4.1-4.2 assuming Theorems 4.3-4.4 and Proposition 4.5. In Section 4.3, we prove the Theorems and Proposition 4.5.

## 4.2 Proof of Propositions 4.1 and 4.2

We assume Theorems 4.3-4.4 and Proposition 4.5 in this section. In the next section, we prove Theorems 4.3-4.4 and Proposition 4.5.

### 4.2.1 Proof of Proposition 4.1

Let $f = f_V$ and $X = X_V$ be as defined in (4.1). We prove Proposition 4.1 using Theorem 4.4. To that end we prove that (i)-(iv) hold for *every* $p > 1$ and $\alpha > 0$. It is easy to check that (i) and (iii) holds: since $\Lambda(.)$ is the Lebesgue measure, (i) holds; see e.g. Penrose and Yukich (2003) for a proof that (iii) is satisfied.

It is well-known (see e.g. Baryshnikov and Yukich (2005)) that for $X_V$, the condition (iv) holds for every $\alpha > 0$. We give a brief proof for completeness. For $m \geq 2$, divide $B_{m+2(\log m)^2}$ into small squares each of whose side length is in the range $\left[\frac{(\log m)^2}{10}, \frac{(\log m)^2}{5}\right]$ and let $G_m$ denote the event that each square has a Poisson point. It is easy to check that

$$\mathbb{P}_0(G_m) \geq 1 - e^{-C(\log m)^4} \tag{4.16}$$

for some positive constant $C$. As before $\mathbb{P}_0$ denotes the probability measure of the Poisson process conditioned to have a point at the origin. If $G_m$ occurs, the following two statements hold: (a) for every point $y \in B_{m-2(\log m)^2} \cap (\mathcal{N} \cup \{0\})$, the corresponding tessellate $T(y, \mathcal{N} \cup \{0\}) \subseteq B_{m-(\log m)^2}$ and (b) for every point $z \in B_m^c \cap (\mathcal{N} \cup \{0\})$, we have $T(z, \mathcal{N} \cup \{0\}) \cap B_{m-(\log m)^2} = \phi$.

In particular, if $G_m$ occurs, the tessellate containing the origin is contained in $B_m$ no matter what the configuration is outside $B_m$ and the radius of determinability $r_0$ of the point at the origin (see assumption (iii) of Section 3.1) satisfies $r_0 \leq \frac{m}{2}$. By translation invariance and (4.16), this proves that (iv) holds for every $\alpha > 0$.

To prove that (ii) holds for every $p > 1$, we use Proposition 4.5. Fix integer $k \geq 1$ and let $\mathbb{P}_{\mathcal{X}}$ denote the probability measure of the process $\mathcal{N}_M$ conditioned to have points in a finite set $\mathcal{X}$. Assume that the origin is in $\mathcal{X}$. For $l \geq 2\max\{x : x \in \mathcal{X}\}$, we have that

$$\begin{aligned}\mathbb{P}_{\mathcal{X}}(f_V(0) = l) &= \mathbb{P}_{\mathcal{X}}(\{f_V(0) = l\} \cap G_{2l^{1/3}}) + \mathbb{P}_{\mathcal{X}}(G^c_{2l^{1/3}}) \\ &\leq \mathbb{P}_{\mathcal{X}}(\{f_V(0) = l\} \cap G_{2l^{1/3}}) + e^{-C_1(\log l)^4} \qquad (4.17)\end{aligned}$$

for some positive constant $C_1$, where the last estimate is analogous to (4.16). If $G_{2l^{1/3}}$ occurs, then by the discussion above, the tessellate $\mathcal{T}_0$ containing origin is contained in $B_{l^{1/3}}$ for all $l$ sufficiently large. Moreover, each tessellate intersecting $\mathcal{T}_0$ is also contained in $B_{l^{1/3}}$. Thus if $\#\mathcal{X} = k$, we have that

$$\mathbb{P}_{\mathcal{X}}(\{f_V(0) = l\} \cap G_{2l^{1/3}}) \leq \mathbb{P}(\#(\mathcal{N} \cap B_{l^{1/3}}) \geq l - k) \leq e^{-C_2 l^{2/3}}$$

for some positive constant $C_3$ depending only $k$ and not on the choice of $\mathcal{X}$. Thus from (4.17) and the above estimate, we have that $\mathbb{E}_{0,\mathcal{X}} f_V^k \leq C_k$ for some positive constant $C_k$ independent of the choice of $\mathcal{X}$. By translation invariance and Proposition 4.5, we have that (iii) holds for $p = k$. Since $k$ is arbitrary, we are done. □

### 4.2.2 Proof of Proposition 4.2

The first part of Proposition 4.2 follows from Chapter 3 of Meester and Roy (1996).

If we prove that assumptions (i)-(iv) in Section 4.1.3 are satisfied then the second part of Proposition 4.2 follows from Theorem 4.4. Clearly (i) is satisfied since $\Lambda(.)$ is the Lebesgue measure. To prove (iii) we place a ball of (random) radius $t$ at $x \in \mathbb{R}^2$. Let $\mathcal{C}_x$ denote the component containing the ball intersecting $x$ in the Poisson Boolean model. Since $\lambda < \lambda_c$, we know that $\mathcal{C}_x$ is bounded almost surely and therefore there exists $T = T(x,t) < \infty$ $a.s.$ such that $\mathcal{C}_0 \subseteq B_T(x)$. As before, $B_m(x)$ is the square of side length $m$ centred at $x$. We have that (iii) is satisfied by setting $r_x = T + 2R$.

To prove (ii) and (iv), we let $\mathbb{E}_{x_1,\mathcal{X}}f$ be the expectation as defined in (4.14) for a fixed finite set $\mathcal{X} \subset \mathbb{R}^d$ and $x_1 \in \mathcal{X}$.

**Proposition 4.6.** *Fix $\lambda < \lambda_c$. For every $k \geq 1$, we have that*

$$\sup_{\mathcal{X}} \mathbb{E}_{x_1,\mathcal{X}} f_B^k < \infty \tag{4.18}$$

*where supremum is over all sets $\mathcal{X} = \{x_1, ..., x_k\}$ containing $k$ vertices. Also,*

$$\sup_{x \in \mathbb{R}^d} \mathbb{P}_x(r_x \geq m) \leq e^{-C_2 m} \tag{4.19}$$

*for all $m \geq 1$ and for some positive constant $C_2$.*

We prove Proposition 4.6 at the end of this proof. From (4.18) and Proposition 4.5, we have that (ii) is satisfied for every $p > 1$. From (4.19), we have that (iv) is satisfied for every $\alpha > 0$. Thus (i)-(iv) are satisfied and Proposition 4.2 follows from Theorem 4.4. $\square$

*Proof of Proposition 4.6*: We first prove (4.18). Consider fixed points $y_1, ..., y_k$ and place a ball of radius $\rho_i'$ at $y_i$. Each $\rho_i'$ has the same distribution as the radius of a ball in the Poisson Boolean model. By stationarity, we let $y_1 = 0$. For $1 \leq i \leq k$, let $Z_i = S(y_i, 2R)$ denote the $2R$ ball centred at $y_i$ and let $\mathcal{C}(Z_i)$ denote the union of all occupied components in $\mathcal{N}_M$ intersecting $Z_i$. Let $\mathcal{C}_M(0)$ be the occupied cluster of the process $\mathcal{N}_M \bigcup \bigcup_{i=1}^k \{(y_i, \rho_i')\}$ intersecting the ball $Z_1$ centred at the origin. Clearly, $\mathcal{C}_M(0) \subseteq \bigcup_{i=1}^k \mathcal{C}(Z_i)$, the union of all the components. And therefore if diameter of $\mathcal{C}_M(0)$ is at least $m$, at least one of $\mathcal{C}(Z_i)$ must have diameter at least $\frac{m}{2k}$. Thus, given $\rho_1', .., \rho_k'$, we have that

$$\begin{aligned}
\mathbb{P}(diam(\mathcal{C}_M(0)) \geq m | (y_1, \rho_1'), ..., (y_k, \rho_k')) &\leq \mathbb{P}\left(\bigcup_{i=1}^k diam(\mathcal{C}(Z_i)) \geq \frac{m}{2k}\right) \\
&\leq k\mathbb{P}\left(diam(\mathcal{C}(Z_1)) \geq \frac{m}{2k}\right) \\
&\leq C_1 \mathbb{P}\left(diam(\mathcal{C}(0)) \geq \frac{m}{2k}\right)
\end{aligned}$$

for some constant $C_1 > 0$, where as before, $\mathcal{C}(0)$ denotes the component of the occupied region intersecting the origin in the process $\mathcal{N}_M$. Here the second inequality follows by translation invariance, and the last inequality follows from Example 2.1 of Meester and Roy (1996). Since all critical intensities are equal (Theorems 3.5, 4.3 and 4.4 of Meester and Roy (1996)), we have from Lemma 3.3 of Meester and Roy (1996) that

$$\mathbb{P}\left(diam(\mathcal{C}(0)) \geq \frac{m}{2k}\right) \leq e^{-C_2 m}, \tag{4.20}$$

for some constant $C_2 > 0$.

Let $E_d(m)$ denote the event that $B_m$ contains less than $4\lambda m^2$ points. It is easy to check that $\mathbb{P}(E_d(m)) \geq 1 - e^{-C_3 m^2}$ for some positive constant $C_3$. Suppose $E_d(m)$ occurs. If for a fixed $(y_1, \rho'_1), ..., (y_k, \rho'_k)$, we have $diam(\mathcal{C}_M(0)) \leq m$, then $\mathcal{C}_M(0)$ is contained in $B_m$ and consequently must contain less than $4\lambda m^2 + k \leq 5\lambda m^2$ balls. Thus if $N_0$ denotes the number of balls of the occupied cluster containing the origin in the process $\mathcal{N}_M \bigcup \bigcup_{i=1}^k \{(y_i, \rho'_i)\}$, we have that

$$\begin{aligned} &\mathbb{P}(E_d(m) \cap \{N_0 \geq 5\lambda m^2\} | (y_1, \rho'_1), ..., (y_k, \rho'_k)) \\ &\leq \ \mathbb{P}(diam(\mathcal{C}_M(0)) \geq m | (y_1, \rho'_1), ..., (y_k, \rho'_k)) \end{aligned}$$

and since $E_d(m)$ does not depend on $\{(y_i, \rho'_i)\}_i$, we get from (4.20) and the estimate on the probability of the event $E_d(m)$ above, that

$$\mathbb{P}(N_0 \geq 5\lambda m^2 | (y_1, \rho'_1), ..., (y_k, \rho'_k)) \leq e^{-C_2 m} + e^{-C_3 m^2} \leq e^{-C_4 m}$$

for some positive constant $C_4$. Thus

$$\mathbb{E}(N_0^k | (y_1, \rho'_1), ..., (y_k, \rho'_k)) \leq C_5$$

for some constant $C_5$ that depends on $k$ but not on the specific choice of $\{(y_i, \rho'_i)\}_i$ and integrating over $\rho'_1, ..., \rho'_k$, we have that $\mathbb{E}_{0,\mathcal{X}} f_B^k \leq C_5$.

To prove (4.19), we briefly define the notion of vacant circuits. A vacant circuit is a piecewise linear curve that has the same starting and ending point and is completely contained in the vacant region of the Poisson Boolean model. For $m \geq 1$, we say that a vacant circuit occurs in $B_{3m} \setminus B_m$ if there is a vacant circuit $\pi$ that surrounds $B_m$ and is contained in $B_{3m}$. Let $G_{1,m}$ denote the event that a vacant circuit occurs in $B_{3m} \setminus B_m$, $B_{5m} \setminus B_{3m}$ and in $B_{7m} \setminus B_{5m}$. We claim that

$$\mathbb{P}(G_{1,m}) \geq 1 - e^{-C_1 m} \tag{4.21}$$

for some positive constant $C_1$. Since the balls are bounded in radius by $R$ *a.s.*, we have that if $G_{1,m}$ occurs then changing the configuration inside $B_m$ will

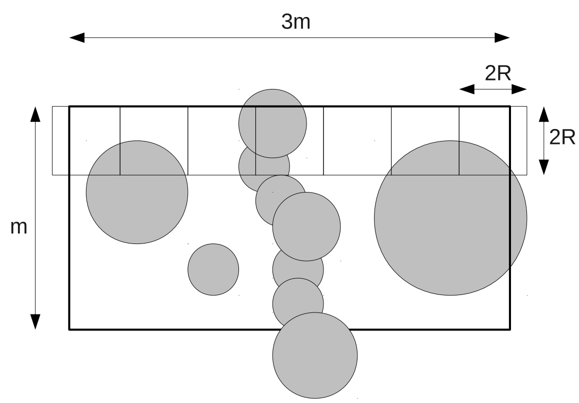



Figure 4.1: Occupied top-bottom crossing.

not affect the configuration outside $B_{7m}$. Thus the radius of determinability at the origin, $r_0$, as defined in (iv) of Section 3.1 can be bounded as

$$\mathbb{P}_0(r_0 \geq 4m) \leq \mathbb{P}(G^c_{1,m}) \leq e^{-C_1 m},$$

proving (4.19).

To prove (4.21), we use the ideas of occupied and vacant left-right crossings. As in Section 4.1.2, let $\mathcal{N} = \{x_1, x_2, ...\}$ denote a realization of the Poisson process and let $\rho_i$ denote the random radius at $x_i$. Fix $m \geq 1$ and consider the rectangle

$$Q(3m, m) := \left[-\frac{3m}{2}, \frac{3m}{2}\right] \times \left[-\frac{m}{2}, \frac{m}{2}\right].$$

A piecewise linear path $\pi$ is said to be a left-right crossing of $Q(3m, m)$ if $\pi$ is contained in $Q(3m, m)$ and $\pi$ intersects the left and right faces of $Q(3m, m)$. We say that $\pi$ is an *occupied* left-right crossing if $\pi$ is contained in the occupied region; i.e.,

$$\pi \subseteq (\cup_i S(x_i, \rho_i)) \bigcap Q(3m, m).$$

We say that $\pi$ is a *vacant* left-right crossing if it is contained in the vacant region; i.e.,

$$\pi \subseteq (\cup_i S(x_i, \rho_i))^c \bigcap Q(3m, m).$$

Let $LR^*(3m, m)$ denote the event that there exists a vacant left-right crossing of $Q(3m, m)$. If $\lambda < \lambda_c$, we claim that

$$\mathbb{P}(LR^*(3m, m)) \geq 1 - e^{-C_1 m}$$

for some positive constant $C_1$. Indeed, if a vacant left-right crossing does not occur, then an occupied top-bottom crossing occurs. Consider the $2R \times 2R$ squares intersecting the top edge of $Q(3m, m)$ as shown in Figure 4.1. The number of such squares is at least $\frac{3m}{2R}$ and at most $\frac{3m}{R}$. Enumerate them as $\{H_i\}_i$. If there exists a top-bottom crossing of $Q(3m, m)$, necessarily $diam(\mathcal{C}(H_i)) \geq m/2$ for some $i$. Thus,

$$\begin{aligned}
\mathbb{P}(LR(m, 3m)) &\leq \mathbb{P}\left(\bigcup_i \left\{diam(\mathcal{C}(H_i)) \geq \frac{m}{2}\right\}\right) \\
&\leq \frac{3m}{R}\mathbb{P}\left(diam(\mathcal{C}(H_1)) \geq \frac{m}{2}\right) \\
&\leq C_1 m \mathbb{P}\left(diam(\mathcal{C}(0)) \geq \frac{m}{2}\right) \\
&\leq C_1 m e^{-C_2 m}
\end{aligned}$$

for some positive constants $C_1$ and $C_2$ where the third inequality follows from Example 2.1 of Meester and Roy (1996) and the last inequality follows from Lemma 3.3 of Meester and Roy (1996).

Thus $\mathbb{P}(LR^*(3m, m)) \geq 1 - C_1 m e^{-C_2 m}$ and by FKG inequality

$$\mathbb{P}(G_m) \geq 1 - 4C_1 m e^{-C_2 m}$$

where $G_m$ is the event that there exists a vacant circuit contained in $B_{3m} \backslash B_m$. This implies (4.21). □

## 4.3 Proof of Theorems and Proposition 4.5

We prove Theorem 4.4 and obtain Theorem 4.3 as a Corollary.

### 4.3.1 Proof of Theorem 4.4

Without loss of generality, we assume that $f \geq 0$. Otherwise, we prove for $f^+ = f\mathbf{1}(f \geq 0)$ and $f^- = f\mathbf{1}(f < 0)$ separately. Fixing $\delta$ and $\eta > 0$, we first prove (4.11). The main idea in the proof is to divide the set $nW$ into cubes whose sides are of length $n^{1-\beta}$ each for an appropriately chosen $\beta \in (0,1)$ and decompose the functional $X$ into sums of independent random variables and then use a concentration inequality to estimate the sum.

Tile $B_n = nW$ into small cubes $\{S_i^{out}\}_i$ each having side length in the range $\left[n^{1-\beta}, 2n^{1-\beta}\right]$ for some $\beta \in (0,1)$ to be fixed later. Let $\{S_i^{out}\}_{1\le i\le m_n}$ denote the set of cubes that are completely contained inside $nW$, where $m_n$ is an integer that satisfies

$$m_n(n^{1-\beta})^d \le \ell(nW) = n^d \tag{4.22}$$

where as before $\ell(.)$ denotes the Lebesgue measure. For each $i$, $1 \le i \le m_n$, let $S_i$ denote the cube with the same centre as $S_i^{out}$, such that $S_i \subset S_i^{out}$ and $d(\partial S_i^{out}, \partial S_i) = 4n^{1-2\beta}$.

Define the events

$$T_{inf}(S_i) = \bigcap_{x\in\mathcal{N}\cap S_i} \{r_x \le n^{1-2\beta}\} \text{ and } T_i = T_{inf}(S_i) \cap \{X(S_i) \le (\Lambda(S_i^{out}))^{1+\epsilon}\} \tag{4.23}$$

where $\epsilon$ is some positive constant to be chosen later and the term $r_x$ is the radius of determinability defined in the statement (iv) following (4.9). Write

$$X(nW) = \tilde{X}_1 + \tilde{X}_2 + X(nW \setminus (nW)_{in}) \tag{4.24}$$

where $(nW)_{in} = \bigcup_{i=1}^{m_n} S_i$,

$$\tilde{X}_1 = \sum_{i=1}^{m_n} X(S_i)\mathbf{1}(T_i) \quad \text{and} \quad \tilde{X}_2 = \sum_{i=1}^{m_n} X(S_i)\mathbf{1}(T_i^c),$$

The following result explains the rationale behind splitting $X(nW)$ as in (4.24).

**Lemma 4.7.** *For any $i$, $1 \le i \le m_n$, the event $T_{inf}(S_i)$ and the random variable $X(S_i)\mathbf{1}(T_i)$ are both determined by the restriction of the marked Poisson process to $S_i^{out} \times \mathcal{M}$; i.e., if $\omega \cap (S_i^{out} \times \mathcal{M}) = \omega' \cap (S_i^{out} \times \mathcal{M})$ for $\omega, \omega' \subset \mathbb{R}^d \times \mathcal{M}$, then $\omega \in T_{inf}(S_i)$ if and only if $\omega' \in T_{inf}(S_i)$ and $X(S_i)\mathbf{1}(T_i)(\omega) = X(S_i)\mathbf{1}(T_i)(\omega')$. Moreover,*

$$\mathbb{P}(T_{inf}(S_i)) \ge 1 - \frac{Cn^{(1-\beta)d}}{n^{(1-2\beta)\alpha}} \tag{4.25}$$

*for some constant $C > 0$.*

The estimate (4.25) is a consequence of tail condition (4.10) and we prove the above result at the end of this proof.

The approach in evaluating the terms in (4.24) is as follows. From Lemma 4.7 we know that for $i \neq j$, the random variables $X(S_i)\mathbf{1}(T_i)$ and $X(S_j)\mathbf{1}(T_j)$ are independent of each other. The first term $\tilde{X}_1$ in (4.24) is therefore a sum of independent random variables and can be estimated using a concentration inequality. We then use the comparability and integrability conditions (i) and (iii) to show that the remaining two terms in (4.24) are negligible provided the constants $\beta$ and $\epsilon$ in (4.23) are appropriately chosen.

The first and the third terms are estimated from the following result.

**Lemma 4.8.** *We have that*

$$\ell(nW \setminus (nW)_{in}) \leq Cn^{d-\beta} \tag{4.26}$$

*for some positive constant $C$. Also, for a fixed $\epsilon > 0$, there are positive constants $C_1$ and $C_2$ such that*

$$\mathbb{P}\left(|\tilde{X}_1 - \mathbb{E}\tilde{X}_1| \geq \frac{\Lambda(nW)}{n^\delta \log n}\right) \leq \exp\left(-C_1 \frac{n^{\delta_0}}{(\log n)^2}\right) \tag{4.27}$$

*and*

$$\mathbb{P}\left(X(nW \setminus (nW)_{in}) \geq \frac{\Lambda(nW)}{n^\delta \log n}\right) \leq C_2 \frac{(\log n)^p}{n^{p\beta - p\delta}}, \tag{4.28}$$

*where $\delta_0 = (d\beta - 2\delta) - (2d(1-\beta)\epsilon)$.*

The following result estimates for the second term.

**Lemma 4.9.** *There exists positive constants $C_1$ and $C_2$ so that*

$$\mathbb{E}\tilde{X}_2 \leq C_1 \frac{\Lambda(nW)}{n^{\delta_1}} \text{ and } \mathbb{P}\left(\tilde{X}_2 \geq \frac{\Lambda(nW)}{n^\delta \log n}\right) \leq C_2 \frac{\log n}{n^{\delta_1 - \delta}}, \tag{4.29}$$

*where $\delta_1 = p^{-1}(p-1)\min\left(dp\epsilon(1-\beta), (1-2\beta)\alpha - (1-\beta)d\right)$.*

We prove the above lemmas at the end of the proof.

Before we choose the parameters $\beta$ and $\epsilon$, we collect together the estimates. From (4.27), (4.28) and (4.29), we have for some constants $C_1, C_2 > 0$ that

$$\mathbb{P}\left(|X(nW) - \mathbb{E}\tilde{X}_1| \geq C_1 \frac{\Lambda(nW)}{n^\delta \log n}\right) \leq \frac{C_2}{n^{\delta_2}} + \exp\left(-C_1 \frac{n^{\delta_0}}{(\log n)^2}\right)$$

where

$$\delta_2 = \min(p\beta - p\delta, \delta_1 - \delta). \tag{4.30}$$

Since we desire $\mathbb{E}X(nW) - \mathbb{E}X(nW)$ in the left-hand side, we estimate $\mathbb{E}X(nW) - \mathbb{E}\tilde{X}_1$. From (4.24), we first write

$$0 \le \mathbb{E}X(nW) - \mathbb{E}\tilde{X}_1 = \mathbb{E}\tilde{X}_2 + \mathbb{E}X(nW \setminus (nW)_{in}).$$

Since $nW \setminus (nW)_{in}$ is a finite union of rectangles with disjoint interiors, each rectangle having diameter at least one, by (4.9) we have that

$$\mathbb{E}X(nW \setminus (nW)_{in}) \le (\mathbb{E}X^p(nW \setminus (nW)_{in}))^{1/p} \le C\Lambda(nW \setminus (nW)_{in}) \tag{4.31}$$

for some constant $C > 0$. Here the first and the second estimates follows from Holders inequality and Minkowski's inequality, respectively.

Thus for some constant $C_1 > 0$ we have

$$0 \le \mathbb{E}X(nW) - \mathbb{E}\tilde{X}_1 \le \mathbb{E}\tilde{X}_2 + C\Lambda(nW \setminus (nW)_{in}) \le \quad C_1\tfrac{\Lambda(nW)}{n^{\delta_1}} + C_1\tfrac{\Lambda(nW)}{n^{\beta}}$$

where the last estimates follow from (4.29),(4.26) and (4.8). This implies that for some constants $C_1, C_2 > 0$ we have

$$\mathbb{P}\left(|X(nW) - \mathbb{E}X(nW)| \ge C_1\frac{\Lambda(nW)}{n^{\delta_3}\log n}\right) \le \frac{C_2}{n^{\delta_2}} + \exp\left(-C_1\frac{n^{\delta_0}}{(\log n)^2}\right) \tag{4.32}$$

where $\delta_3 = \min(\beta, \delta_1, \delta)$.

We now choose positive $\beta$ and $\epsilon$ such that $\delta_0 > 0$, $\delta_2 > \eta$ and $\delta_3 > \delta$. To get $\delta_0 > 0$, we need to choose $\epsilon$ so that

$$\epsilon < \frac{d\beta - 2\delta}{2d(1-\beta)}. \tag{4.33}$$

To get $\delta_2 > \eta$, we need $\delta_1 - \delta > \eta$ and $p\beta - p\delta > \eta$ (see (4.30)). The latter holds if

$$2\beta > \frac{2\eta}{p} + 2\delta. \tag{4.34}$$

(Since $\delta < \frac{1}{2}$ and $p > \frac{2\eta}{1-2\delta}$, the right hand side above inequality is strictly less than one.) The former holds if

$$\epsilon > \frac{\delta + \eta}{d(1-\beta)(p-1)} \tag{4.35}$$

and $(\alpha(1-2\beta)-d(1-\beta))\left(1-\frac{1}{p}\right)>\delta+\eta$ or equivalently if

$$2\beta<\frac{2\alpha-2d}{2\alpha-d}-\frac{2(\delta+\eta)p}{(p-1)(2\alpha-d)}. \tag{4.36}$$

Finally, for (4.33) and (4.35) to be true simultaneously, we need

$$\frac{\delta+\eta}{d(1-\beta)(p-1)}<\frac{d\beta-2\delta}{2d(1-\beta)}, \tag{4.37}$$

which is satisfied if

$$2\beta>2\frac{(2p\delta+2\eta)}{d(p-1)}. \tag{4.38}$$

The right hand side above inequality is less than one since $p>\frac{d+4\eta}{d-4\delta}$.

We choose $\beta$ satisfying (4.34), (4.36) and (4.38) (this is possible since $\alpha>\alpha_0$). Fixing such a $\beta$ ensures (4.37) and allows us to choose $\epsilon$ satisfying (4.33) and (4.35). This implies that $\delta_0>0$ and $\delta_2>\eta$. Since (4.34) is satisfied, we have that $\delta_3=\min(\delta_1,\delta)$ and since $\delta_2>\eta$, it follows from (4.30) that $\delta_3>\delta$. From (4.32), this proves that (4.11) holds for some $\gamma_1>\delta$ and $\gamma_2>\eta$.

*Proof of Lemma 4.7*: The first part follows from definition of radius of determinability $r_x$ in condition (iv). To prove (4.25), we first write

$$\mathbb{P}(T^c_{inf}(S_i)) \;=\; \mathbb{E}\mathbf{1}\left(\bigcup_{x\in\mathcal{N}\cap S_i} r_x\geq n^{1-2\beta}\right)\leq\mathbb{E}\sum_{x\in\mathcal{N}\cap S_i}\mathbf{1}\left(r_x\geq n^{1-2\beta}\right).$$

Using the Slivnyak-Mecke formula (Møller (1994)), we get that the last term equals

$$\int_{S_i}\mathbb{P}_x\left(r_x\geq n^{1-2\beta}\right)\Lambda(dx)\leq\int_{S_i}\frac{C_1}{n^{(1-2\beta)\alpha}}\Lambda(dx)\leq C_2\frac{n^{(1-\beta)d}}{n^{(1-2\beta)\alpha}}.$$

We have used (4.10) and (4.8), respectively, in obtaining the first and the second inequalities. ☐

*Proof of Lemma 4.8*: To prove the bound in (4.26), we first write

$$\ell(nW\setminus(nW)_{in})=\sum_{i=1}^{m_n}\ell(S_i^{out}\setminus S_i).$$

We have that $\ell(S_i^{out} \setminus S_i) \leq \frac{C_1 n^{(1-\beta)d}}{n^\beta}$ and from the estimate (4.22) we have that $m_n \leq C_2 n^{d\beta}$, for some positive constants $C_1$ and $C_2$, independent of $i$. This implies that

$$\ell(nW \setminus (nW)_{in}) \leq C_2 n^{d\beta} \frac{C_1 n^{(1-\beta)d}}{n^\beta} \leq C_3 \frac{\ell(nW)}{n^\beta},$$

for some positive constant $C_3$.

To prove (4.27), we write $\tilde{X}_1 - \mathbb{E}\tilde{X}_1 = \sum_{i=1}^{m_n} X_i$ where $X_i = X(S_i)\mathbf{1}(T_i) - \mathbb{E}X(S_i)\mathbf{1}(T_i)$. We have that $\mathbb{E}X_i = 0$ for every $i$. Also, since $0 \leq X(S_i)\mathbf{1}(T_i) \leq (\Lambda(S_i^{out}))^{1+\epsilon}$, we have that

$$|X_i| \leq 2(\Lambda(S_i^{out}))^{1+\epsilon} \triangleq c_i.$$

By Azuma-Hoeffding Inequality (Azuma (1967)) we have for $t \geq 0$ that

$$\mathbb{P}\left(|\tilde{X}_1 - \mathbb{E}\tilde{X}_1| \geq t\right) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^{m_n} c_i^2}\right)$$

and setting $t = \frac{\Lambda(nW)}{n^\delta \log n}$ we obtain

$$\mathbb{P}\left(|\tilde{X}_1 - \mathbb{E}\tilde{X}_1| \geq \frac{\Lambda(nW)}{n^\delta \log n}\right) \leq \exp\left(-\frac{1}{2(\log n)^2}\left(\frac{\Lambda(nW)}{n^\delta}\right)^2 \frac{1}{\sum_{i=1}^{m_n} c_i^2}\right).$$

By (4.8) and (4.22), we have for some constant $C > 0$ that

$$\sum_{i=1}^{m_n} c_i^2 = \sum_{i=1}^{m_n} \Lambda(S_i^{out})^{2+2\epsilon} \leq C m_n \ell(S_1^{out})^{2+2\epsilon} = C n^{d\beta}(n^{(1-\beta)d})^{2+2\epsilon}$$

and that $\Lambda(nW)^2 \geq C n^{2d}$. Thus we get for constants $C_1, C_2 > 0$ that

$$\left(\frac{\Lambda(nW)}{n^\delta}\right)^2 \frac{1}{\sum_{i=1}^{m_n} c_i^2} \geq C_1 \frac{n^{2d}}{n^{2\delta} n^{d\beta}(n^{(1-\beta)d})^{2+2\epsilon}} = C_1 n^{\delta_0},$$

where $\delta_0$ is as in the statement of this Lemma. Substituting in the above equation, we get (4.27).

Finally, to prove (4.28), we use Markov's inequality to get that

$$\begin{aligned}\mathbb{P}\left(X(nW \setminus (nW)_{in}) \geq \frac{\Lambda(nW)}{n^\delta \log n}\right) &\leq C_1 \frac{\mathbb{E}X^p(nW \setminus (nW)_{in})}{(\Lambda(nW))^p} n^{p\delta}(\log n)^p \\ &\leq C_2 \frac{\mathbb{E}X^p(nW \setminus (nW)_{in})}{n^{dp}} n^{p\delta}(\log n)^p\end{aligned}$$

for some positive constants $C_1$ and $C_2$, where the last equation follows from (4.31). By (4.26) and (4.8), we get (4.28). $\square$

*Proof of Lemma 4.9*: We first need an estimate of the probability of the event $T_i$ for every $1 \leq i \leq m_n$. We recall from (4.23) that $T_i$ is the intersection of two events. Using Markov's inequality, (4.9), (4.8) and the fact that $n^{d(1-\beta)} \leq \ell(S_i^{out}) \leq C_1 n^{d(1-\beta)}$ for some constant $C_1 > 0$ we estimate the probability for the latter event as

$$\mathbb{P}(X(S_i) \geq (\Lambda(S_i^{out}))^{1+\epsilon}) \leq \frac{\mathbb{E}X(S_i)^p}{(\Lambda(S_i^{out}))^{p+p\epsilon}} \leq \frac{\mathbb{E}X(S_i^{out})^p}{(\Lambda(S_i^{out}))^{p+p\epsilon}} \leq \frac{C_2}{n^{d(1-\beta)p\epsilon}}$$

for some constant $C_2 > 0$.

Using the estimate for $\mathbb{P}(T_{inf}(S_j^{out})^c)$ from Lemma 4.7, we then get

$$\mathbb{P}(T_i^c) \leq \mathbb{P}(X(S_i) \geq (\Lambda(S_i^{out}))^{1+\epsilon}) + \mathbb{P}(T_{inf}(S_i)^c) \leq \frac{C}{n^{\delta_1 p(p-1)^{-1}}} \tag{4.39}$$

where $\delta_1$ is as in the statement of this Lemma. By Holder's inequality, (4.9) and (4.8), we then have

$$\mathbb{E}X(S_i)\mathbf{1}(T_i^c) \leq (\mathbb{E}X^p(S_i))^{\frac{1}{p}} (\mathbb{P}(T_i^c))^{1-\frac{1}{p}} \leq \frac{C_1\Lambda(S_i)}{n^{\delta_1}}$$

for some constant $C_1 > 0$. Since $\mathbb{E}\tilde{X}_2 = \sum_{i=1}^{m_n} \mathbb{E}X(S_i)\mathbf{1}(T_i^c)$, we therefore have for constants $C_1, C_2 > 0$ that

$$\mathbb{E}\tilde{X}_2 \leq C_1 \sum_{i=1}^{m_n} \frac{\Lambda(S_i)}{n^{\delta_1}} \leq C_2 \frac{\Lambda(nW)}{n^{\delta_1}},$$

where the second inequality follows from $\cup_{i=1}^{m_n} S_i \subseteq nW$ and the final inequality follows from (4.8). We then apply Markov's inequality to obtain (4.29). $\square$

### 4.3.2 Proof of Theorem 4.3

To prove that (4.11) holds for *some* $\gamma_1 > 0$ and $\gamma_2 > 0$, we observe that the quantity $a_0 \to 0$ as $\delta \to 0$ and $\eta \to 0$. Here $a_0$ is as defined in Theorem 4.4. Moreover, for $\delta$ and $\eta$ positive, $\alpha_0$ is also positive. Thus, given $\alpha > d$ and $p > 1$, we can choose $\delta$ and $\eta$ appropriately so that $\alpha > \alpha_0 > d$ and hence (4.11) holds for $\gamma_1 > \delta$ and $\gamma_2 > \eta$.

To prove that (4.12) holds, we let $Z_n = \left|\frac{X(nW)}{\Lambda(nW)} - \mathbb{E}\frac{X(nW)}{\Lambda(nW)}\right|$ and $A_n = \left\{Z_n \leq n^{-\delta}\right\}$. For $r < p$, we have

$$\mathbb{E}Z_n^r = \mathbb{E}Z_n^r \mathbf{1}(A_n) + \mathbb{E}Z_n^r \mathbf{1}(A_n^c) \leq \frac{1}{n^{r\delta}} + \mathbb{E}Z_n^r \mathbf{1}(A_n^c).$$

To bound the second term above, we let $\theta_1 = \frac{r}{p} < 1$ and use Holder's inequality to obtain

$$\mathbb{E}Z_n^r \mathbf{1}(A_n^c) \leq (\mathbb{E}Z_n^p)^{\theta_1} (\mathbb{P}(A_n^c))^{1-\theta_1} \leq C_1 (\mathbb{E}Z_n^p)^{\theta_1} \left(\frac{1}{n^\eta}\right)^{1-\theta_1}$$

for some constant $C_1 > 0$, by our choice of $\eta$. Since $X$ satisfies (4.9) we have that

$$\mathbb{E}|Z_n|^p \quad \leq \quad C_1 \mathbb{E}\left|\frac{X(nW)}{\Lambda(nW)}\right|^p + C_1 \left|\mathbb{E}\frac{X(nW)}{\Lambda(nW)}\right|^p \leq C_2$$

for some constants $C_1, C_2 > 0$. Combining the estimates we have (4.12). □

### 4.3.3 Proof of Proposition 4.5

Fix $n \geq 1$ and let $A \subseteq nW$ be any rectangle whose shortest edge has length at least one. We have by the Slivnyak-Mecke formula (Møller (1994)) that

$$\begin{aligned}
\mathbb{E}X(A)^k &= \mathbb{E} \sum_{x_1 \in A \cap \mathcal{N}} \dots \sum_{x_k \in A \cap \mathcal{N}} f(x_1, \mathcal{N}_M) \dots f(x_k, \mathcal{N}_M) \\
&= \sum_{l=1}^{k} \binom{k}{l} \int_A \dots \int_A \int_{\mathcal{M}} \dots \int_{\mathcal{M}} \sum_{\mathcal{D}_l} \mathbb{E} f_1^{i_1} \dots f_l^{i_l} \\
&\qquad \mu_M(dt_1) \dots \mu_M(dt_n) \Lambda(dx_1) \dots \Lambda(dx_l)
\end{aligned}$$

where $f_j = f(x_j, \mathcal{N}'_M)$, the process $\mathcal{N}'_M = \mathcal{N}_M \bigcup \bigcup_{i=1}^{l} (x_{k_i}, t_{k_i})$ and the innermost summation is over the set $\mathcal{D}_l = \{(i_1, \dots, i_l) : i_1 + i_2 + \dots + i_l = k\}$.

Using the AM-GM inequality, the innermost term can be bounded above as

$$\mathbb{E} f_1^{i_1} \dots f_l^{i_l} \leq \frac{1}{k} \sum_{j=1}^{l} i_j \mathbb{E} f_j^k$$

and hence

$$\begin{aligned}\mathbb{E}X(A)^k &\leq \sum_{l=1}^{k}\binom{k}{l}\int_A\cdots\int_A\int_{\mathcal{M}}\cdots\int_{\mathcal{M}}\sum_{\mathcal{D}_l}\frac{1}{k}\sum_{j=1}^{l}i_j\mathbb{E}f_j^k \\ &\qquad \mu_M(dt_1)...\mu_M(dt_n)\Lambda(dx_1)...\Lambda(dx_l) \\ &= \sum_{l=1}^{k}\binom{k}{l}\sum_{\mathcal{D}_l}\frac{1}{k}\sum_{j=1}^{l}i_j\int_A\cdots\int_A\mathbb{E}_{x_j,\mathcal{X}}f_j^k\Lambda(dx_1)...\Lambda(dx_l).\end{aligned}$$

where $\mathbb{E}_{x_j,\mathcal{X}}f$ represents the expected value as in (4.14). By (4.15) we have that

$$\mathbb{E}X(A)^k \leq C_1\sum_{l=1}^{k}\binom{k}{l}\sum_{\mathcal{D}_l}\frac{1}{k}\sum_{j=1}^{l}i_j\int_A\cdots\int_A\Lambda(dx_1)...\Lambda(dx_l) \leq C_2(\Lambda(A))^k$$

for some positive constants $C_1$ and $C_2$. This proves that (iii) holds. □

# Bibliography

[1] K. Alam. (1972). Unimodality of the Distribution of an Order Statistic. *Ann. Math. Stat.*, **43**, 2041–2044.

[2] D. J. Aldous and W. S. Kendall. (2008). Short-length routes in low-cost networks via Poisson line patterns. *Adv. Appl. Probab.*, **40**, 1–21.

[3] D. J. Aldous, C. Bordenave and M. Lelarge. (2008). Near-minimal spanning trees: A scaling exponent in probability models. *Ann. Inst. Henri Poincare*, **44**, 962–976.

[4] K. Azuma. (1967). Weighted Sums of Certain Dependent Random Variables. *Tohoku Math. Journ.*, **19**, 357–367.

[5] F. Baccelli and C. Bordenave. (2005). The Radial Spanning Tree of a Poisson Point Process. *Ann. App. Prob.* **17** 305–359.

[6] Y. Baryshnikov and J. Yukich. (2005). Gaussian limits for random measures in geometric probability. *Ann. App. Probab.* **15** 213–253.

[7] B. Bollobás. (1985). *Random graphs.* Academic Press.

[8] B. Bollobás and Thomason. (1987). Threshold functions. *Combinatorica*, **7** 35–38.

[9] B. Bollobás. (1988). The chromatic number of random graphs. *Combinatorica*, **8** 49–56.

[10] B. Bollobás and O. Riordan. (2006). *Percolation.* Academic Press.

[11] B. Bollobás and O. Riordan. (2006b). The critical probability for random Voronoi percolation in the plane is 1/2. *Prob. Th. and Rel. Fields*, **136** 417–468.

[12] D. J. Daley and D. V-Jones. (2008). *An Introduction to the Theory of Point Processes II: General theory and structure.* Springer Verlag, Berlin.

[13] R. Durrett and X.-F. Liu. (1988). The contact process on a finite set. *Ann. Probab.*, **16**, 1158–1173.

[14] R. Durrett. (2006). *Random Graph Dynamics.* Cambridge University Press.

[15] P. Erdös. (1947). Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.*, **53**, 292–294.

[16] P. Erdös and A. Rényi. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hungarian Academy of Sciences*, **5**, 17–61.

[17] M. Franceschetti, O. Dousse, D. N. C. Tse and P. Thiran. (2007). Closing Gap in the Capacity of Wireless Networks via Percolation Theory. *IEEE Trans. Inform. Theory*, **53**, 1009–1018.

[18] G. Ganesan. (2012). Size of the giant component in a random geometric graph. *Accepted for publ. Ann. Inst. Henri Poincare.*

[19] E. N. Gilbert. (1961). Random Plane Networks. *J. Soc. Indust. Appl.*, **9**, 533–543.

[20] A. Gopalan, S. Banerjee, A. K. Das and S. Shakottai. (2011). Random Mobility and the Spread of Infection. *Proc. IEEE Infocomm*, pp. 999–1007.

[21] P. Gupta and P. R. Kumar. (1998). Critical Power for Asymptotic Connectivity in Wireless Networks. *Stoch. Anal., Control, Opt. and Appl.*, 547–566.

[22] G. Grimmett. (1999). *Percolation.* Springer Verlag.

[23] L. Heinrich, H. Schmidt and V. Schmidt. (2005). Limit Theorems for Stationary Tessellations with Random Inner Cell Structure. *Adv. App. Probab.* **37** 25–47.

[24] S. Janson, T. Luczak and A. Rucinski. (2000). *Random graphs.* John Wiley.

[25] T. M. Liggett, R. H. Schonmann and A. M. Stacey. (1997). Domination by Product Measures. *Ann. Probab.*, **25**, 71–95.

[26] J. Martin. (2002). Linear growth for greedy lattice animals.. *Stoch. Proc. and Appl.*, **98**, 43–66.

[27] R. Meester and R. Roy. (1996). *Continuum Percolation.* Cambridge Tracts in Mathematics.

[28] J. Møller. (1994). Lectures on Random Voronoi Tessellations. Springer Verlag, Berlin.

[29] S. Muthukrishnan and G. Pandurangan. The Bin-Covering Technique for Thresholding Random Geometric Graph Properties. *Proc. SODA,* 2005, 989–998.

[30] M. D. Penrose. (2003). *Random Geometric Graphs.* Oxford.

[31] M. D. Penrose and J. E. Yukich. (2003). Weak Law of Large Numbers In Geometric Probability. *Ann. App. Probab.* **13** 277–303.

[32] M. D. Penrose. (2007). Laws of large numbers in stochastic geometry with statistical applications. *Bernoulli.* **13** 1124–1150.

[33] A. Sarkar. (1995). *Some problems in Continuum percolation.* PhD Thesis, ISI Delhi.

[34] E. Shamir and J. Spencer. (1987). Sharp Concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica*, **7** 124–129.

[35] R. T. Smythe and J. C. Wierman. (2008). *First-Passage Percolation on the Square Lattice.* Springer-Verlag.

[36] J. M. Steele (1987). *Probability Theory and Combinatorial Optimization.* SIAM.