



*Citation for published version:*

Friedrich, T, Sauerwald, T & Stauffer, A 2013, 'Diameter and broadcast time of random geometric graphs in arbitrary dimensions', *Algorithmica*, vol. 67, no. 1, pp. 65-88. <https://doi.org/10.1007/s00453-012-9710-y>

*DOI:*

[10.1007/s00453-012-9710-y](https://doi.org/10.1007/s00453-012-9710-y)

*Publication date:*

2013

*Document Version*

Peer reviewed version

[Link to publication](#)

The final publication is available at [link.springer.com](http://link.springer.com)

## University of Bath

### Alternative formats

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

<b>Noname manuscript No.</b> (will be inserted by the editor)
--

---

# Diameter and Broadcast Time of Random Geometric Graphs in Arbitrary Dimensions

Tobias Friedrich · Thomas Sauerwald ·  
Alexandre Stauffer

**Abstract** A random geometric graph (RGG) is defined by placing  $n$  points uniformly at random in  $[0, n^{1/d}]^d$ , and joining two points by an edge whenever their Euclidean distance is at most some fixed  $r$ . We assume that  $r$  is larger than the critical value for the emergence of a connected component with  $\Omega(n)$  nodes. We show that, with high probability (w.h.p.), for any two connected nodes with a Euclidean distance of  $\omega\left(\frac{\log n}{r^{d-1}}\right)$ , their graph distance is only a constant factor larger than their Euclidean distance. This implies that the diameter of the largest connected component is  $\Theta(n^{1/d}/r)$  w.h.p. We also prove that the condition on the Euclidean distance above is essentially tight.

We also analyze the following randomized broadcast algorithm on RGGs. At the beginning, only one node from the largest connected component of the RGG is informed. Then, in each round, each informed node chooses a neighbor independently and uniformly at random and informs it. We prove that w.h.p. this algorithm informs every node in the largest connected component of an RGG within  $\Theta(n^{1/d}/r + \log n)$  rounds.

## 1 Introduction

We study Random Geometric Graphs (RGGs) in  $d \geq 2$  dimensions. An RGG is a graph resulting by placing  $n$  nodes independently and uniformly at random on  $[0, n^{1/d}]^d$  and creating edges between pairs of nodes if and only if their Euclidean distance is at most  $r$ . These graphs have been studied intensively in relation to subjects such as cluster analysis, statistical physics, and wireless sensor networks [25]. Traditionally, most work on RGGs is restricted to two dimensions. However, wireless sensor networks also expand in three dimensions. Examples are sensors in water bodies [1] and sensor networks based on the use of flying anchors [19]. Another motivation for RGGs in arbitrary dimensions is multivariate statistics of high-dimensional data [23]. In this case the coordinates of the nodes of the RGG represent

---

A conference version with slightly weaker results appeared in the 22nd International Symposium on Algorithms and Computation. This work was done while the first two authors were at International Computer Science Institute Berkeley supported by the German Academic Exchange Service (DAAD) and the third author was at the Computer Science Division of the University of California Berkeley.

---

T. Friedrich  
Friedrich-Schiller-Universität Jena, Germany

T. Sauerwald  
Max-Planck-Institut für Informatik, Saarbrücken, Germany

A. Stauffer  
Microsoft Research, Redmond WA, USA

different attributes of the data. The metric imposed by the RGG then depicts the similarity between data elements in the high-dimensional space. Also in bioinformatics, RGGs in up to four dimensions have been observed to give an excellent fit for various global and local measures of protein-protein interaction networks [16].

Several algorithms and processes have been studied on RGGs. One prominent example is the cover time of random walks. Avin and Eralc [3] considered RGGs in two dimensions when the coverage radius is a constant large enough to assure that the RGG is connected with probability  $1 - o(1)$ . They proved that in this regime, the cover time of an RGG is  $\Theta(n \log n)$  with probability  $1 - o(1)$ , which is optimal up to constant factors. This has been improved by Cooper and Frieze [6] who gave a more precise estimate of the cover time on RGGs that also extends to larger dimensions. However, all these works are restricted to the case where the probability that the RGG is connected approaches 1 as  $n \rightarrow \infty$ .

We are interested in a wider range for  $r$ . All the following results hold for the regime where the RGG is likely to contain a connected component with  $\Omega(n)$  nodes. Bradonjić et al. [4] proved for RGGs in  $d = 2$  dimensions that, with probability  $1 - \mathcal{O}(n^{-1})$ , for any two connected nodes with a minimum Euclidean distance of  $\Omega(\log^{3.5} n / r^2)$ , their graph distance is only a constant factor larger than their Euclidean distance. We establish this result for all dimensions  $d \geq 2$  under the weaker condition that the minimum Euclidean distance is  $\omega\left(\frac{\log n}{r^{d-1}}\right)$ . For this, we have to employ a different proof technique since the approach of Bradonjić et al. [4] strongly depends on restrictions imposed by the geometry in two dimensions. Our result implies that the diameter of the largest connected component is  $\Theta(n^{1/d}/r)$  with high probability<sup>1</sup>; this was previously open for  $d \geq 3$  and matches the corresponding bound for  $d = 2$  [4, 8]. Our techniques are inspired by percolation theory and we believe them to be useful for other problems, such as estimating the cover time for the largest connected component of RGGs.

## Broadcasting information

We use the aforementioned structural result of RGGs to study the problem of broadcasting information in RGGs. We study the well known randomized rumor spreading algorithm which is also known as the push algorithm [11]. In this algorithm, in every round, each informed node chooses a neighbor independently and uniformly at random and informs it. We are interested in the runtime, i.e., how long it takes to spread a piece of information from an arbitrary node of the largest connected component to all other nodes in that component.

The obvious lower bound of this process on an arbitrary graph  $G$  is  $\Omega(\text{diam}(G) + \log n)$ , where  $\text{diam}(G)$  denotes the diameter of the largest connected component. A matching upper bound of  $\mathcal{O}(\text{diam}(G) + \log n)$  is known for complete graphs [14, 24], hypercubes [11], expander graphs [13, 26], several Cayley graphs [9], bounded-degree graphs [11], and RGGs in two dimensions [4]. In this paper we prove that RGGs in  $d \geq 3$  dimensions also allow an optimal broadcast time of  $\mathcal{O}(\text{diam}(G) + \log n) = \mathcal{O}(n^{1/d}/r + \log n)$  w.h.p. This generalizes the two-dimensional result of Bradonjić et al. [4] and significantly improves upon the general bound of  $\mathcal{O}(\Delta \cdot (\text{diam}(G) + \log n))$  [11], since for sparse RGGs (where  $r = \Theta(1)$ ) the maximum degree is  $\Delta = \Theta(\log n / \log \log n)$ . Note that our result implies that all nodes get informed after  $\mathcal{O}(n^{1/d}/r + \log n)$  rounds for connected RGGs as well.

The rest of this paper is organized as follows. In Section 2, we give a precise definition of the random broadcast algorithm and the random geometric graph, as well as introduce some notation and state our results. In Section 3, we derive an upper bound on the length

<sup>1</sup> By “with high probability” (short: w.h.p.), we denote an event that holds with probability at least  $1 - \mathcal{O}(n^{-1})$ .

of the shortest path between two nodes in an RGG given their Euclidean distance is large enough. We show in Section 4 that this condition on the Euclidean distance is tight up to constants. In Section 5, we perform the runtime analysis of the random broadcast algorithm. We close in Section 6 with some concluding remarks.

## 2 Precise Model and Results

We consider the following random broadcast algorithm also known as the push algorithm [11]. We are given an undirected graph  $G$ . At the beginning, called round 0, a node  $s$  of  $G$  owns a piece of information, i.e., it is informed. In each subsequent round  $1, 2, \dots$ , every informed node chooses a neighbor independently and uniformly at random and informs that neighbor. We are interested in the runtime of this algorithm, which is the time taken until every node in  $G$  gets informed; in the case of  $G$  being disconnected, we require every node in the same connected component as  $s$  to get informed. The runtime of this algorithm is a random variable denoted by  $\mathcal{R}(s, G)$ . Our aim is to prove bounds on  $\mathcal{R}(s, G)$  that hold with high probability, i.e., with probability  $1 - \mathcal{O}(n^{-1})$ .

We study  $\mathcal{R}(s, G)$  for the case of a random geometric graph  $G$  in arbitrary dimension  $d \geq 2$ . We define the random geometric graph in the space  $\Omega := [0, n^{1/d}]^d$  equipped with the Euclidean norm, which we denote by  $\|\cdot\|_2$ . The most natural definition of RGG is stated as follows.

**Definition 1 (cf. [23])** *Let  $\mathcal{X}_n = \{X_1, X_2, \dots, X_n\}$  be  $n$  points in  $\Omega$  chosen independently and uniformly at random. The random geometric graph  $\mathcal{G}(\mathcal{X}_n; r)$  has node set  $\mathcal{X}_n$  and edge set  $\{(x, y) : x, y \in \mathcal{X}_n, \|x - y\|_2 \leq r\}$ .*

In our analysis, it is more advantageous to use the following definition.

**Definition 2 (cf. [23])** *Let  $N_n$  be a Poisson random variable with mean  $n$  and let  $\mathcal{P}_n = \{X_1, X_2, \dots, X_{N_n}\}$  be points chosen independently and uniformly at random from  $\Omega$ ; equivalently,  $\mathcal{P}_n$  is a Poisson Point Process over  $\Omega$  with intensity 1. The random geometric graph  $\mathcal{G}(\mathcal{P}_n; r)$  has node set  $\mathcal{P}_n$  and edge set  $\{(x, y) : x, y \in \mathcal{P}_n, \|x - y\|_2 \leq r\}$ .*

The following basic lemma says that any result that holds in the setting of Definition 2 with sufficiently large probability holds with similar probability in the setting of Definition 1.

**Lemma 1** *Let  $\mathcal{A}$  be any event that holds with probability at least  $1 - \alpha$  in  $\mathcal{G}(\mathcal{P}_n; r)$ . Then,  $\mathcal{A}$  also holds in  $\mathcal{G}(\mathcal{X}_n; r)$  with probability  $1 - \mathcal{O}(\alpha\sqrt{n})$ .*

*Proof.* In this proof, we use subscripts to indicate the space over which the probabilities are calculated. Recall that  $N_n$  denotes the number of nodes in  $\mathcal{P}_n$ . Then it follows by Stirling's formula that  $\Pr_{\mathcal{G}(\mathcal{P}_n; r)}[N_n = n] = \Theta(1/\sqrt{n})$ . Note that conditioned on  $N_n = n$ ,  $\mathcal{G}(\mathcal{P}_n; r)$  is a realization of  $\mathcal{G}(\mathcal{X}_n; r)$ . Let  $\mathcal{A}^c$  denote the complement of  $\mathcal{A}$ . Therefore

$$\Pr_{\mathcal{G}(\mathcal{X}_n; r)}[\mathcal{A}^c] = \Pr_{\mathcal{G}(\mathcal{P}_n; r)}[\mathcal{A}^c \mid N_n = n] \leq \frac{\Pr_{\mathcal{G}(\mathcal{P}_n; r)}[\mathcal{A}^c]}{\Pr_{\mathcal{G}(\mathcal{P}_n; r)}[N_n = n]} = \mathcal{O}(\alpha\sqrt{n}). \quad \square$$

Henceforth, we consider an RGG given by  $G = \mathcal{G}(\mathcal{P}_n; r)$ , and refer to  $r$  as the coverage radius of  $G$ . It is known that, for  $d \geq 2$ , there exists a critical value  $r_c = r_c(d) = \Theta(1)$  such that if  $r > r_c$ , then with high probability the largest connected component of  $G$  has cardinality  $\Omega(n)$ . On the contrary, if  $r < r_c$ , each connected component of  $G$  has  $\mathcal{O}(\log n)$  nodes with probability  $1 - o(1)$  [23]. The exact value of  $r_c$  is not known, though some bounds have been derived in [18]. In addition, if  $r^d \geq \frac{\log n + \omega(1)}{b_d}$ , where  $b_d$  is the volume of the  $d$ -dimensional ball of radius 1, then  $G$  is connected with probability  $1 - o(1)$  [21, 22].

Our main result is stated in the next theorem. It shows that if  $r > r_c$ , with probability  $1 - \mathcal{O}(n^{-1})$ ,  $\mathcal{R}(s, G) = \mathcal{O}(n^{1/d}/r + \log n)$  for all  $s$  inside the largest connected component of  $G$ . Note that  $r_c$  does not depend on  $n$ , but if  $r$  is regarded as a function of  $n$ , then here and in what follows,  $r > r_c$  means that this strict inequality must hold in the limit as  $n \rightarrow \infty$ .

**Theorem 2** *For a random geometric graph  $G = \mathcal{G}(\mathcal{P}_n; r)$  in  $d \geq 2$  dimensions, if  $r > r_c$ , then  $\mathcal{R}(s, G) = \mathcal{O}(n^{1/d}/r + \log n)$  with probability  $1 - \mathcal{O}(n^{-1})$  for all nodes  $s$  of  $G$ .*

The proof of Theorem 2, which we provide in Section 5, requires an upper bound on the length of the shortest path between nodes of  $G$ . Our result on this matter, which is stated in the next theorem, provides that for any two nodes that are sufficiently distant in  $\Omega$ , the distance between them in the metric induced by  $G$  is only a constant factor larger than the optimum with probability  $1 - \mathcal{O}(n^{-1})$ . In particular, this result implies that the diameter of the largest connected component of  $G$  is  $\mathcal{O}(n^{1/d}/r)$ , a result previously known only for two dimensions or values of  $r$  that yield a connected  $G$  with probability  $1 - o(1)$ .

For all  $v_1, v_2 \in G$ , we say that  $v_1$  and  $v_2$  are connected if there exists a path in  $G$  from  $v_1$  to  $v_2$ , and define  $d_G(v_1, v_2)$  as the graph distance between  $v_1$  and  $v_2$  on  $G$ , that is,  $d_G(v_1, v_2)$  is the length of the shortest path from  $v_1$  to  $v_2$  in  $G$ . Also, we denote the Euclidean distance between the locations of  $v_1$  and  $v_2$  by  $\|v_1 - v_2\|_2$ . Clearly, the length of the shortest path between two nodes  $v_1$  and  $v_2$  in  $G$  satisfies  $d_G(v_1, v_2) \geq \frac{\|v_1 - v_2\|_2}{r}$ .

**Theorem 3** *If  $d \geq 2$  and  $r > r_c$ , for any two connected nodes  $v_1$  and  $v_2$  in  $G = \mathcal{G}(\mathcal{P}_n; r)$  such that  $\|v_1 - v_2\|_2 = \omega\left(\frac{\log n}{r^{d-1}}\right)$ , we obtain  $d_G(v_1, v_2) = \mathcal{O}(\|v_1 - v_2\|_2/r)$  with probability  $1 - \mathcal{O}(n^{-1})$ .*

**Remark 4** In the proof of Theorem 3, we focus on the case  $r = \mathcal{O}(\log^{1/d} n)$ . If  $r = \omega(\log^{1/d} n)$ , then the upper bound on  $d_G(v_1, v_2)$  above follows from [8, Theorem 8].

**Remark 5** It is not hard to see that our proof establishes that there exist large enough constants  $C = C(d)$  and  $n_0 = n_0(d)$  such that  $d_G(v_1, v_2) = \mathcal{O}(\|v_1 - v_2\|_2/r)$  for all connected pairs  $v_1, v_2$  for which  $\|v_1 - v_2\|_2 \geq C \frac{\log n}{r^{d-1}}$  for all  $n \geq n_0$ . We used the  $\omega$  notation above only to simplify the statement of Theorem 3.

**Corollary 6** *If  $r > r_c$ , the diameter of the largest connected component of  $G = \mathcal{G}(\mathcal{P}_n; r)$  is  $\mathcal{O}(n^{1/d}/r)$  with probability  $1 - \mathcal{O}(n^{-1})$ .*

The statement of Theorem 3 generalizes and improves upon Theorem 2.3 of [4], which holds only for  $d = 2$  and  $\|v_1 - v_2\|_2 = \Omega(\log^{3.5} n/r^2)$ . The current paper not only improves upon the previous results, but also employs different proof techniques which are necessary to tackle the geometrically more involved case where  $d \geq 3$ .

Our last result establishes that the condition for  $\|v_1 - v_2\|_2$  in Theorem 3 is necessary. We show that there exists a pair of nodes in the largest connected component of the RGG whose graph distance is much larger than the Euclidean distance. We prove this result in Section 4.

**Theorem 7** *Let  $r^d = o(\log n)$ . Then, with probability  $1 - \mathcal{O}(n^{-1})$ , there exist two connected nodes  $v_1$  and  $v_2$  in  $G = \mathcal{G}(\mathcal{P}_n; r)$  such that  $\|v_1 - v_2\|_2 \leq 3r$  but  $d_G(v_1, v_2) = \Omega\left(\frac{\log n}{r^d}\right)$ .*

**Remark 8** In other words, since  $r^d = o(\log n)$ , Theorem 7 above establishes that there exists a pair of connected nodes  $v_1, v_2$  such that  $\|v_1 - v_2\|_2 = o\left(\frac{\log n}{r^{d-1}}\right)$  for which  $d_G(v_1, v_2) = \omega\left(\frac{\|v_1 - v_2\|_2}{r}\right)$ . This means that the condition for  $\|v_1 - v_2\|_2$  in Theorem 3 (see also Remark 5) is essentially tight.

### 3 The Diameter of the Largest Connected Component

We devote this section to proving Theorem 3. We consider  $G = \mathcal{G}(\mathcal{P}_n; r)$ . Recall that we assume  $r > r_c$  and  $r^d = \mathcal{O}(\log n)$ . (When  $r^d = \omega(\log n)$ ,  $G$  is connected with probability  $1 - o(1)$  and Theorem 3 becomes a slightly different version of [8, Theorem 8].) Note also that  $r = \Omega(1)$  since  $r_c = \Theta(1)$ . We show that, for any two connected nodes  $v_1$  and  $v_2$  of  $G$  such that  $\|v_1 - v_2\|_2 = \omega\left(\frac{\log n}{r^{d-1}}\right)$ , we have  $d_G(v_1, v_2) = \mathcal{O}(\|v_1 - v_2\|_2/r)$  with probability  $1 - \mathcal{O}(n^{-1})$ .

The proof follows by a renormalization argument that allows us to relate the largest connected component of  $G$  with the so-called percolation cluster of the square lattice. We start in Section 3.1 discussing some results for the lattice that we will need later. Then, in Section 3.2, we discuss the renormalization argument, whose main ideas we present here. We tessellate  $\Omega$  into cubes that are large enough but whose side length is  $\Theta(r)$ . We say that a cube is open if the graph induced by the nodes of  $G$  inside the cube contains a unique large component and the large components of two adjacent open cubes intersect. We make the cubes large enough so that the probability that any given cube is open is sufficiently close to 1 and, consequently, the set of open cubes percolates. Then, the main idea is that, in order to bound the size of the shortest path between two vertices  $v_1$  and  $v_2$  of  $G$ , we consider a path of adjacent cubes  $j_1, j_2, \dots, j_\ell$  of the tessellation so that cube  $j_1$  contains  $v_1$  and cube  $j_\ell$  contains  $v_2$ . Note that we can choose  $j_1, j_2, \dots, j_\ell$  so that  $\ell = \Theta(\|v_1 - v_2\|_2/r)$ . If all the cubes in this path are open, then we use the largest components of these cubes to obtain a path from  $v_1$  to  $v_2$ . Note that, since the side length of the cubes is  $\Theta(r)$ , any shortest path within a given cube contains only a constant number of edges, so the shortest path on  $G$  between  $v_1$  to  $v_2$  within the cubes  $j_1, j_2, \dots, j_\ell$  has length  $\Theta(\ell) = \Theta(\|v_1 - v_2\|_2/r)$ . So, with this renormalization, we can reduce the problem of estimating the graph distance between  $v_1$  and  $v_2$  to the problem of finding the shortest path of open cubes from a cube that is close enough to  $v_1$  to a cube that is close enough to  $v_2$ . However, in general, not all the cubes in the path will be open. When this happens, we need to estimate how far away from the cubes  $j_1, j_2, \dots, j_\ell$  the path from  $v_1$  to  $v_2$  goes. The details of this analysis is given in Section 3.3. The main idea behind this step is that, if cube  $j_k$  is closed, then we consider all closed cubes that can be reached from  $j_k$  by a path of adjacent closed cubes. In other words, we consider the connected component of closed cubes that contains  $j_k$ . The boundary of this connected component is a collection of open cubes, so the nodes of  $G$  inside the largest components of these open cubes yield a detour around  $j_k$ ; i.e., they contain a path that goes around  $j_k$  and reaches some cube  $j_{k'}$  with  $k' > k$ . We show that, whenever the Euclidean distance between  $v_1$  and  $v_2$  is sufficiently large, these detours only increase the graph distance by a constant factor. Now we proceed to the detailed proof.

#### 3.1 Lattice

For  $m \geq 0$ , whose value we will set later, let  $\mathbb{S}_m$  be the elements of  $\mathbb{Z}^d$  contained in the cube of side length  $m$  centered at the origin (i.e.,  $\mathbb{S}_m = \{i \in \mathbb{Z}^d: \|i\|_\infty \leq m/2\}$ ). Let  $L$  be the graph with vertex set  $\mathbb{S}_m$  such that an edge between two vertices  $i, j \in \mathbb{S}_m$  exists if and only if  $\|i - j\|_\infty = 1$  (see Figure 1(a)). It is easy to see that the maximum degree  $\Delta$  of  $L$  is  $\Delta = 3^d - 1$ . The following is an adaptation of a well-known result regarding the number of lattice animals [15, Theorem 4.20] to our lattice  $L$ . We include the proof here for the sake of completeness.

**Lemma 9** *A subset of vertices  $V \subseteq \mathbb{S}_m$  is called a lattice animal if the subgraph of  $\mathbb{S}_m$  induced by  $V$  is connected. Then, the number of lattice animals that contain the origin and have exactly  $k$  vertices is at most  $3^{\Delta k}$ .*

*Proof.* We follow [15, Theorem 4.20]. Construct a random subgraph  $H$  of  $L$  by keeping each edge of  $L$  independently with probability  $1/2$ ; otherwise, we remove the edge. Let  $H_0$  be the subgraph of  $H$  induced by the vertices that are in the same connected component of the origin in  $H$ . Clearly,  $H_0$  is a lattice animal. Now, we refer to the edges of  $L$  that have exactly one endpoint in  $H_0$  as the *boundary edges* of  $H_0$ . The probability that  $H_0$  has  $a_v$  vertices,  $a_e$  edges, and  $a_b$  boundary edges is exactly  $w_{a_v, a_e, a_b} 2^{-a_e - a_b}$ , where  $w_{a_v, a_e, a_b}$  is the number of possible choices for the subgraph  $H_0$  with exactly  $a_v$  vertices,  $a_e$  edges and  $a_b$  boundary edges. Clearly, the probability that  $H_0$  has exactly  $a_v$  vertices is

$$\sum_{a_e, a_b} w_{a_v, a_e, a_b} 2^{-a_e - a_b} \leq 1.$$

Then, since the maximum degree of  $L$  is  $\Delta$ , we have that  $a_b \leq a_v \Delta$  and  $a_e \leq a_v \Delta/2$ . Plugging this into the equation above gives the total number of choices for  $H_0$  such that  $H_0$  has  $a_v$  vertices as

$$\sum_{a_e, a_b} w_{a_v, a_e, a_b} \leq 2^{3a_v \Delta/2} \leq 3^{a_v \Delta},$$

which concludes the proof of Lemma 9.  $\square$

Now, let  $X = (X_i)_{i \in \mathbb{S}_m}$  be a collection of binary random variables. For two vertices  $i, j \in \mathbb{S}_m$ , let  $d_L(i, j)$  be their graph distance in  $L$ . Also, for any  $i \in \mathbb{S}_m$  and  $k \geq 0$ , let  $\mathcal{F}_k(i)$  be the  $\sigma$ -field generated by all  $X_j$  with  $d_L(i, j) > k$ . Then, for  $k \geq 0$  and  $p \in (0, 1)$ , we say that  $X$  is a *k-dependent site percolation process on  $L$  with probability  $p$*  if, for any  $i \in \mathbb{S}_m$ , we have  $\Pr[X_i = 1] \geq p$  and  $\Pr[X_i = 1 \mid \mathcal{F}_k(i)] = \Pr[X_i = 1]$ ; i.e.,  $X_i$  is independent of any collection  $(X_j)_j$  for which the distance between  $i$  and  $j$  in  $L$  is larger than  $k$  for all  $j$  in the collection. Let  $L(X)$  be the subgraph of  $L$  induced by the vertices  $i$  with  $X_i = 1$ . The following lemma is a direct application of a result by Liggett, Schonmann and Stacey [17, Theorem 1.3] that provides that  $L(X)$  stochastically dominates an independent site percolation process. For two processes  $X = (X_i)_{i \in \mathbb{S}_m}$  and  $Y = (Y_i)_{i \in \mathbb{S}_m}$ , we say that  $L(X)$  stochastically dominates  $L(Y)$  if there exists a coupling between  $X$  and  $Y$  such that  $L(Y)$  is a subgraph of  $L(X)$  with probability 1.

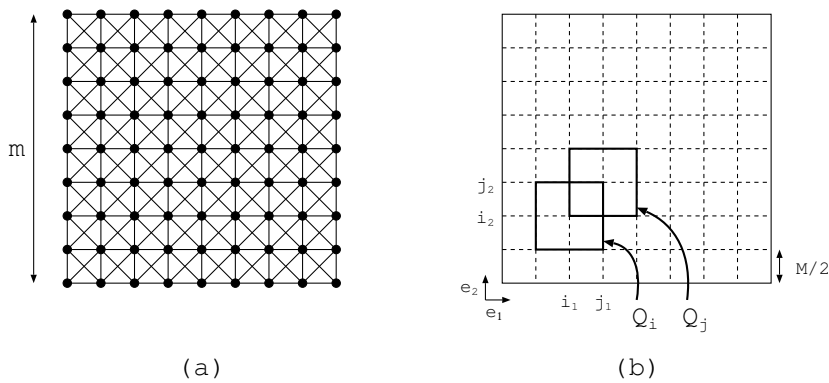
**Lemma 10 ([17, Theorem 1.3])** *For given constants  $\alpha > 0$  and  $k \geq 0$ , let  $L(X)$  be the subgraph of  $L$  obtained via a  $k$ -dependent site percolation process  $X$  with probability  $1 - e^{-\alpha}$ . If  $\alpha$  is large enough, then there exists a positive constant  $c < 1$  depending only on  $k$  and  $d$  so that  $L(X)$  stochastically dominates a collection of independent Bernoulli random variables with mean  $1 - e^{-c\alpha}$ .*

### 3.2 Renormalization argument

Fix a sufficiently large constant  $M > 0$ . For each  $i = (i_1, i_2, \dots, i_d) \in \mathbb{Z}^d$ , define the cube

$$Q_i = (i_1 Mr/2, i_2 Mr/2, \dots, i_d Mr/2) + [-Mr/2, Mr/2]^d,$$

which is centered at  $(i_1 Mr/2, i_2 Mr/2, \dots, i_d Mr/2)$  and has sides of length  $Mr$  (see Figure 1(b)). Let  $\mathcal{Q}$  be the set of cubes  $Q_i$  having center inside  $\Omega$  and set  $m$  so that  $\mathbb{S}_m = \{i: Q_i \in \mathcal{Q}\}$ . Note that  $m = \Theta(n^{1/d}/r)$  and the cubes in  $\mathcal{Q}$  cover the whole of  $\Omega$ . We call two cubes  $Q_i$  and  $Q_j$  neighbors if  $\|i - j\|_\infty \leq 1$ . Note that in this case  $i$  and  $j$  are also neighbors in  $L$ . Therefore each cube has at most  $\Delta = 3^d - 1$  neighbors, and there are at most  $K = \left\lceil \frac{n^{1/d}}{Mr/2} \right\rceil^d = \Theta(n/r^d)$  cubes in  $\mathcal{Q}$ .



**Figure 1** (a) The graph  $L$  over  $\mathbb{S}_m$ . (b) Illustration of the neighboring cubes  $Q_i$  and  $Q_j$ .

We say that a parallelepiped  $R$  in  $\mathbb{R}^d$  has a *crossing component* if there exists a connected component inside  $R$  such that, for each face of  $R$ , there exists at least one node of the component within Euclidean distance  $r$  of the face. Define the *region* of a component of  $G$  as the set of points of  $\Omega$  within Euclidean distance  $r$  of at least one node of the component. Also, we say that

a connected component has *spatial diameter*  $k$  if the region of  
the connected component has diameter  $k$ , (1)

where for any set  $A \subset \Omega$  we define the diameter of  $A$  as  $\sup_{x,y \in A} \|x - y\|_2$ . Then, for each  $i \in \mathbb{S}_m$ , let  $\mathcal{E}_i$  be defined as the event where both of the following happen:

- (i) For each neighbor  $Q_j$  of  $Q_i$ , the parallelepiped  $Q_i \cap Q_j$  contains a crossing component.
- (ii)  $Q_i$  contains only one connected component with spatial diameter larger than  $Mr/5$ .

Note that when  $\mathcal{E}_i$  happens for some  $Q_i$ , then (ii) above implies that the largest component of  $Q_i$  intersects the crossing components of all parallelograms  $Q_i \cap Q_j$ , where  $Q_j$  is a neighbor of  $Q_i$ . Moreover, for two  $i$  and  $j$  neighbors in  $L$ , we have that, if  $\mathcal{E}_i$  and  $\mathcal{E}_j$  both happen, then the crossing components of  $Q_i$  and  $Q_j$  intersect. The following lemma is a direct consequence of a result of Penrose and Pisztor [20, Theorem 2] when  $r = \Theta(1)$ . We extend it so that we can handle the case  $r = \omega(1)$ . For any set  $A \subset \mathbb{R}^d$  and positive number  $\gamma > 0$ , we denote  $A^\gamma = \{x \in \mathbb{R}^d : x/\gamma \in A\}$  as the  $\gamma$ -enlargement of  $A$ ; e.g., if  $A$  is the cube of side length 1, then  $A^\gamma$  is the cube of side length  $\gamma$ .

**Lemma 11** *Let  $R$  be a parallelepiped whose side lengths are independent of  $n$  and whose smallest side is 1. Then there exists a positive number  $\gamma_0$  and constant  $c = c(d)$  so that, for all  $\gamma > \gamma_0$  and  $r > r_c$ ,*

- (i)  $\Pr[R^{\gamma r}$  has a crossing component]  $\geq 1 - \exp(-c\gamma r^d)$ .
- (ii)  $\Pr[R^{\gamma r}$  has only one component with spatial diameter at least  $\gamma r/5$ ]  $\geq 1 - \exp(-c\gamma r^d)$ .

*Proof.* We first scale the graph by  $1/r$ ; i.e., we consider the space  $\Omega^{1/r}$ , which is a cube of side length  $n^{1/d}/r$ , we take the nodes of the graph to be  $\mathcal{P}_n^{1/r}$  and declare a pair of nodes to be adjacent iff their Euclidean distance is at most 1. The graph obtained in this way has the same topology as  $G$ , but the geometric properties are different. For example, since the expected number of nodes remains  $n$ , the density of nodes per unit volume becomes  $\lambda = \frac{n}{(n^{1/d}/r)^d} = r^d$ . This means that any result obtained from the model where nodes are given by a Poisson point process with intensity 1 and coverage radius  $r$  can be translated to a model where nodes are given by a Poisson point process with intensity  $\lambda = r^d$  and



coverage radius 1. For the remainder of the proof we will then let  $G'$  stand for the graph given by the latter model. Our goal is then to study the components of  $G'$  in  $R^\gamma$ .

Now we turn to prove the first inequality. Define  $\lambda_c = r_c^d$ . Using the superposition property of Poisson processes, we can see the set of nodes of  $G'$  as the union of  $k$  independent Poisson point processes with intensity  $\lambda/k$ . Let  $k$  have a value such that  $\lambda/k > \lambda_c$  and, since  $\lambda_c$  is constant, we can set  $\lambda/k$  constant. Note that if  $R^\gamma$  has a crossing component for any of these  $\lambda/k$  Poisson point processes, then  $R^\gamma$  has a crossing component for  $G'$ . It follows by [20, Theorem 2] that the probability that this happens for a Poisson point process with intensity larger than  $\lambda_c$  is at least  $1 - e^{-c'\gamma}$  for some constant  $c' > 0$ . Therefore, using the fact that the  $k$  Poisson point processes are mutually independent, it follows that

$$\Pr [R^\gamma \text{ has a crossing component for } G'] \geq 1 - \left(e^{-c'\gamma}\right)^k \geq 1 - \exp(-c\gamma r^d),$$

where the last inequality follows since  $\lambda/k$  is constant, which implies that  $k = \Theta(r^d)$ .

For the second inequality, we need a different approach since, unlike the event “ $R^\gamma$  has a crossing component,” the event that  $R^\gamma$  has only one component with spatial diameter larger than  $\gamma/5$  is not increasing. Here we assume that  $\lambda$  is larger than some arbitrarily large constant  $\lambda_0$ ; otherwise, we can simply use the result from [20, Theorem 2]. For large  $\lambda$ , we tessellate  $\Omega^{1/r}$  into cubes of side length  $\frac{1}{2d}$ , and note that nodes in adjacent cubes are neighbors in  $G'$ . For  $G'$ , we define the region of a component as the set of points of  $\Omega^{1/r}$  within Euclidean distance 1 of the nodes of the component. So, if a node  $v$  is in a given cube of the tessellation, then all the points inside this cube belong to the region of the component of  $v$ . Now note that there exists a positive constant  $c' = c'(d)$  such that, if there are two or more components in  $R^\gamma$  with spatial diameter larger than  $\gamma/5$ , then there must exist a path with at least  $\frac{\gamma}{5c'}$  cubes where any two consecutive cubes in the path intersect in at least one point and each cube in the path contains no node of  $G'$ . By Lemma 9, the total number of paths with exactly  $k$  cubes is at most

$$\frac{(b\gamma)^d}{\left(\frac{1}{2d}\right)^d} 3^{\Delta k},$$

where  $b$  is the size of the largest side of  $R$ . The probability that all cells in these paths have no node is exactly  $\exp\left(-\lambda \left(\frac{1}{2d}\right)^d k\right)$ . Therefore,

$$\begin{aligned} \Pr [R^\gamma \text{ has only one component with spatial diameter larger than } \gamma/5] \\ \geq 1 - \sum_{k=\frac{\gamma}{5c'}}^{\infty} (2b\gamma d)^d 3^{\Delta k} \exp\left(-\lambda \left(\frac{1}{2d}\right)^d k\right) \geq 1 - \exp(-c\gamma r^d), \end{aligned}$$

where the last inequality holds since  $\lambda > \lambda_0$  is large enough.  $\square$

Now it follows from Lemma 11 that

$$\Pr [\mathcal{E}_i] \geq 1 - \exp(-c_1 r^d M) \tag{2}$$

for all  $M$  large enough, where  $c_1$  is a positive constant depending only on  $d$ .

Now we set  $X_i = \mathbf{1}(\mathcal{E}_i)$  for all  $i \in \mathbb{S}_m$ . By construction,  $\mathcal{E}_i$  does not depend on the events  $\mathcal{E}_j$  for which  $d_L(i, j) \geq 2$  since, in this case, the set of nodes in  $Q_i$  and the set of nodes in  $Q_j$  are disjoint. Therefore,  $(X_i)_{i \in \mathbb{S}_m}$  is a 1-dependent site percolation process with probability  $1 - \exp(-c_1 r^d M)$ . Since  $M$  can be made arbitrarily large, we can apply Lemma 10 to find a collection of independent Bernoulli random variables  $Y = (Y_i)_{i \in \mathbb{S}_m}$  with mean  $e^{-c_2 r^d M}$  so that  $L(Y)$  is a subset of  $L(X)$ . Moreover,  $M$  is large enough so that  $L(Y)$  has a giant component with probability  $1 - \exp(-\Theta((n^{1/d}/r)^{d-1}))$  [15].

### 3.3 Finding the path

In this section we will give the proof of Theorem 3.

*Proof of Theorem 3.* We take two fixed nodes  $v_1$  and  $v_2$  satisfying the conditions of Theorem 3 and show that the probability that  $v_1$  and  $v_2$  are connected by a path and  $d_G(v_1, v_2) = \omega(\|v_1 - v_2\|_2/r)$  is  $\mathcal{O}(n^{-3})$ . Then, we would like to take the union bound over all pairs of nodes  $v_1$  and  $v_2$  to conclude the proof of Theorem 3; however, the number of nodes in  $G$  is a random variable and hence the union bound cannot be employed directly. We employ the following lemma to extend the result to all pairs of nodes  $v_1$  and  $v_2$ .

**Lemma 12** *Let  $\mathcal{E}(w_1, w_2)$  be an event associated to a pair of nodes  $w_1, w_2 \in G = \mathcal{G}(\mathcal{P}_n, r)$ . Assume that, for all pairs of nodes,  $\Pr[\mathcal{E}(w_1, w_2)] \geq 1 - p$ , with  $p > 0$ . Then,*

$$\Pr \left[ \bigcap_{w_1, w_2 \in G} \mathcal{E}(w_1, w_2) \right] \geq 1 - 9n^2p - e^{-\Omega(n)}.$$

*Proof.* We condition on  $N_n \leq 3n$ . Using a Chernoff bound for Poisson random variables, it follows that  $\Pr[N_n > 3n] \leq e^{-\Omega(n)}$ . Let  $\mathcal{E}^c(w_1, w_2)$  denote the complement of  $\mathcal{E}(w_1, w_2)$ . Note that  $\Pr[\mathcal{E}^c(w_1, w_2) \mid N_n \leq 3n] \leq \frac{\Pr[\mathcal{E}^c(w_1, w_2)]}{\Pr[N_n \leq 3n]} \leq \frac{p}{1 - e^{-\Omega(n)}}$ , for all  $w_1, w_2 \in G$ . Therefore, using the definition of conditional probabilities and the union bound, we obtain

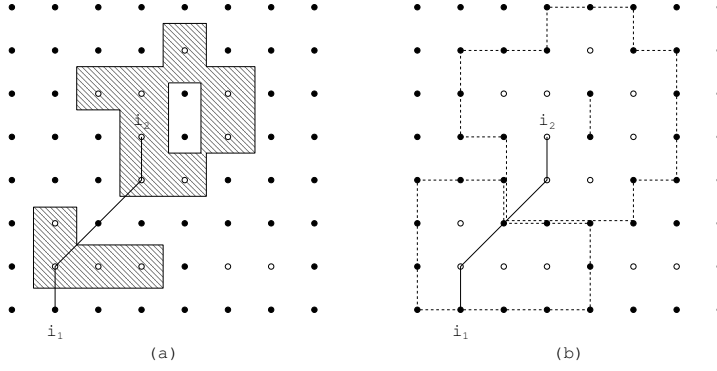
$$\begin{aligned} \Pr \left[ \bigcup_{w_1, w_2 \in G} \mathcal{E}^c(w_1, w_2) \right] &\leq \Pr \left[ \bigcup_{w_1, w_2 \in G} \mathcal{E}^c(w_1, w_2) \mid N_n \leq 3n \right] \Pr[N_n \leq 3n] + \Pr[N_n > 3n] \\ &\leq 9n^2 \cdot \max_{w_1, w_2 \in G} \Pr[\mathcal{E}^c(w_1, w_2) \mid N_n \leq 3n] + e^{-\Omega(n)} \\ &\leq 9n^2p + e^{-\Omega(n)}. \quad \square \end{aligned}$$

We now show that, for any fixed pair of nodes  $v_1, v_2$  of  $G$  such that  $\|v_1 - v_2\|_2 = \omega\left(\frac{\log n}{r^{d-1}}\right)$ , either  $v_1$  and  $v_2$  are in different connected components or  $d_G(v_1, v_2) = \mathcal{O}(\|v_1 - v_2\|_2/r)$ . Let  $i_1$  be the closest vertex of  $\mathbb{S}_m$  from  $v_1$  and  $i_2$  be the closest vertex of  $\mathbb{S}_m$  from  $v_2$ . Clearly,  $v_1 \in Q_{i_1}$  and  $v_2 \in Q_{i_2}$ . We use some ideas from Antal and Pisztorá [2]. For any connected subset  $H$  of  $\mathbb{S}_m$ , let  $\partial H$  be the set of vertices of  $\mathbb{S}_m \setminus H$  from which there exists an edge to a vertex in  $H$ ; that is,  $\partial H$  is the outer boundary of  $H$ . Note that  $|\partial H| \leq \Delta|H|$ . Recall that  $Y = (Y_i)_{i \in \mathbb{S}_m}$  is a collection of i.i.d. Bernoulli random variables as defined in the last paragraph of Section 3.2, and let  $L'(Y)$  be the graph induced by the *closed* vertices of  $L$  (which are the vertices  $j$  for which  $Y_j = 0$ ). For each  $j \in \mathbb{S}_m$ , if  $j$  is closed, let  $Z_j$  be the connected component of  $L'(Y)$  containing  $j$  and let  $\widehat{Z}_j = \partial Z_j$ . If  $j$  is open, then set  $Z_j = \emptyset$  and  $\widehat{Z}_j = \{j\}$ . Note that  $Z_j$  contains only closed vertices and  $\widehat{Z}_j$  contains only open vertices. Moreover,  $\widehat{Z}_j$  separates  $Z_j$  from  $\mathbb{S}_m \setminus (Z_j \cup \widehat{Z}_j)$  in the sense that any path in  $L$  from a vertex in  $\{j\} \cup Z_j$  to a vertex in  $\mathbb{S}_m \setminus (Z_j \cup \widehat{Z}_j)$  must contain a vertex of  $\widehat{Z}_j$ . Now, let  $A_j = \cup_{k: \|k-j\|_\infty \leq 1} Z_k$  and  $\widehat{A}_j = \cup_{k: \|k-j\|_\infty \leq 1} \widehat{Z}_k$ .

Now we give an upper bound for the tails of  $|Z_j|$  and  $|A_j|$ .

**Lemma 13** *Let  $j \in \mathbb{S}_m$ . Then, there exists a positive constant  $c$  such that, for all large enough  $z > 0$ ,*

$$\Pr[|Z_j| \geq z] \leq \exp(-cr^d Mz) \quad \text{and} \quad \Pr[|A_j| \geq z] \leq \exp(-cr^d Mz).$$



**Figure 2** Illustration for the sets  $Z_j$  and  $\widehat{Z}_j$ . Open vertices are shown as black balls and closed vertices as white balls. (a) A path from  $i_1$  to  $i_2$  and shaded areas illustrate  $Z_j$  for  $j$  in the path. (b) The dotted cycles show  $\widehat{Z}_j$  for  $j$  in the path from  $i_1$  to  $i_2$ .

*Proof.* We prove the lemma for  $Z_j$ . The result for  $A_j$  follows from the same argument. Note that  $Z_j$  is a lattice animal as defined in Lemma 9. Therefore, there are at most  $3^{k\Delta}$  choices for the vertices of  $Z_j$  such that  $|Z_j| = k$ . Since the probability that a given vertex is closed is  $1 - e^{-c_2 r^d M}$ , we have  $\Pr[|Z_i| \geq z] \leq \sum_{k=z}^{\infty} 3^{k\Delta} e^{-c_2 r^d M k}$ , which converges as  $z \rightarrow \infty$  as long as  $M$  is large enough.  $\square$

Therefore, for some sufficiently large constant  $c_3$ , we obtain  $\Pr[|Z_j| \geq c_3 \frac{\log m}{r^d}] = \mathcal{O}(m^{-3d})$  and  $\Pr[|A_j| \geq c_3 \frac{\log m}{r^d}] = \mathcal{O}(m^{-3d})$ , and using the union bound over the  $m^d$  choices for  $j$ , we conclude that, for all  $j \in \mathbb{S}_m$ , we have  $|Z_j| \leq c_3 \frac{\log m}{r^d}$  and  $|A_j| \leq c_3 \frac{\log m}{r^d}$  with probability  $1 - \mathcal{O}(m^{-2d})$ .

Now we take an arbitrary path  $j_1, j_2, \dots, j_\ell$  in  $L$  such that  $j_1 = i_1$ ,  $j_\ell = i_2$  and  $\ell \leq \|i_1 - i_2\|_1$ . For  $2 \leq k \leq \ell - 1$  we consider the set  $\widehat{Z}_{j_k}$ . Note that, since for every  $j \in L$  the set  $\widehat{Z}_j$  separates  $Z_j$  from  $\mathbb{S}_m \setminus (Z_j \cup \widehat{Z}_j)$ , we know that  $\bigcup_{k \in [2, \ell-1]} \widehat{Z}_{j_k}$  contains a connected component with at least one vertex from each  $\widehat{Z}_{j_k}$ ,  $2 \leq k \leq \ell - 1$ . We call this component the *bridging* component and denote it by  $B(i_1, i_2)$ . For  $i_1$  and  $i_2$  we consider the sets  $\widehat{A}_{i_1}$  and  $\widehat{A}_{i_2}$ .

We will show how to find a path from  $v_1$  to  $v_2$  in  $G$  in three parts. We will bound the length of these parts by  $F_1$ ,  $F_2$ , and  $F_3$  so that this path from  $v_1$  to  $v_2$  in  $G$  contains  $F_1 + F_2 + F_3$  edges. Note that, since  $v_1$  and  $v_2$  are such that  $\|v_1 - v_2\|_2 = \omega\left(\frac{\log n}{r^{d-1}}\right)$  and  $|A_j| \leq c_3 \frac{\log m}{r^d} = \mathcal{O}(\log n)$  for all  $j$ ,  $A_{i_1}$  and  $A_{i_2}$  must be disjoint. Intuitively,  $\widehat{A}_{i_1}$  envelops the region  $Q_{i_1}$  so that, if there exists a path from  $v_1$  to  $v_2$  in  $G$ , this path must cross the region  $\bigcup_{k \in \widehat{A}_{i_1}} Q_k$ . Now, since  $\widehat{A}_{i_1}$  is a set of open vertices, it follows that, for each  $j \in \widehat{A}_{i_1}$ , the cube  $Q_j$  has a crossing component. For any connected set  $V \subseteq \mathbb{S}_m$  of open vertices, where connectivity is defined with respect to  $L$ , let  $\mathcal{C}(V)$  be the set of vertices of  $G$  that belong to the crossing component of at least one  $Q_j$  with  $j \in V$ . With this definition, the path from  $v_1$  to  $v_2$  must have a node in  $\mathcal{C}(\widehat{A}_{i_1})$ . Let  $F_1$  be the length of the shortest path between  $v_1$  and a node of  $\mathcal{C}(\widehat{A}_{i_1}) \cap \mathcal{C}(\widehat{Z}_{j_2})$ . Note that this node must exist since  $\widehat{A}_{i_1} \cap \widehat{Z}_{j_2} \neq \emptyset$  by construction. If we denote by  $\Psi$  the set  $A_{i_1} \cup \widehat{A}_{i_1}$ , then this path is completely contained inside  $\bigcup_{k \in \Psi} Q_k$ . Therefore, we can bound  $F_1$  using the following geometric lemma.

**Lemma 14** *Let  $I$  be a set of vertices of  $\mathbb{S}_m$  and  $Q = \bigcup_{i \in I} Q_i$ . Let  $w_1$  and  $w_2$  be two nodes of  $G$  inside  $Q$ . If there exists a path between  $w_1$  and  $w_2$  entirely contained in  $Q$ , then there*

exists a constant  $c > 0$  depending only on  $d$  such that

$$d_G(w_1, w_2) \leq c|I|M^d.$$

*Proof.* The shortest path between  $w_1$  and  $w_2$  that is contained inside  $Q$  has the property that, for any two non-consecutive nodes  $u$  and  $u'$  in the path, the Euclidean distance between  $u$  and  $u'$  is larger than  $r$ . Otherwise, we could take the edge  $(u, u')$  and make the path shorter. This means that, if we draw a ball of radius  $r/2$  around every other node of the path, then the balls will not overlap. Let  $\kappa$  be the number of nodes in the path. There are  $\kappa/2$  non-overlapping balls of radius  $r/2$ . For each ball, we know that at least  $1/2^d$  of its volume is contained inside  $Q$ . Therefore, it must hold that

$$\kappa \leq 2 \frac{\text{Area}(Q)}{b_d(r/2)^d/2^d} \leq 2^{2d+1} \frac{|I|(Mr)^d}{b_d r^d},$$

where  $b_d$  is the volume of the  $d$ -dimensional ball of radius 1.  $\square$

**Remark 15** The result in Lemma 14 also holds when  $I$  is replaced by any bounded subset of  $\mathbb{R}^d$  composed of the union of parallelograms with side length at least  $r$ ; in this case, we set  $|I|(Mr)^d$  to be the volume of this set.

Lemma 14 then establishes that there exists a constant  $c_4$  such that

$$F_1 \leq c_4 |A_{i_1} \cup \widehat{A}_{i_1}| M^d \leq c_4 (1 + \Delta) (|A_{i_1}| + 1) M^d = \mathcal{O}\left(\frac{\log m}{r^d}\right),$$

since  $|\widehat{A}_j| \leq \Delta |A_j| + \Delta = \mathcal{O}(\log m)$  for all  $j$ . Similarly, there is a path from  $v_2$  to a node inside  $\mathcal{C}(\widehat{A}_{i_2}) \cap \mathcal{C}(\widehat{A}_{j_{\ell-1}})$ , whose length we denote by  $F_2$ . An analogous derivation then gives  $F_2 = \mathcal{O}\left(\frac{\log m}{r^d}\right)$ . These paths must intersect  $\mathcal{C}(B(i_1, i_2))$  since they intersect  $\mathcal{C}(\widehat{A}_{j_2})$  and  $\mathcal{C}(\widehat{A}_{j_{\ell-1}})$ , respectively. Denote the length of the path in  $\mathcal{C}(B(i_1, i_2))$  that connects the two paths we found above by  $F_3$ . Using Lemma 14 we obtain a constant  $c_5$  such that

$$F_3 \leq c_5 |B(i_1, i_2)| M^d. \quad (3)$$

In order to bound  $|B(i_1, i_2)|$ , we use a coupling argument by Fontes and Newman [12] and a result of Deuschel and Pisztora [7, Lemma 2.3], which gives

$$\Pr \left[ \sum_{k=2}^{\ell-1} |\widehat{Z}_{j_k}| \geq \ell\alpha \right] \leq \Pr \left[ \sum_{k=2}^{\ell-1} |Z_{j_k}| \geq (\ell\alpha - 1)/\Delta \right] \leq \Pr \left[ \sum_{k=2}^{\ell-1} |\tilde{Z}_{j_k}| \geq (\ell\alpha - 1)/\Delta \right],$$

where the first inequality follows since  $|\widehat{Z}_j| \leq 1 + \Delta |Z_j|$  for all  $j$ , and the  $\tilde{Z}$ s are defined to be independent random variables such that  $\tilde{Z}_{j_k}$  has the same distribution as  $Z_{j_k}$ . From Lemma 13 we know that  $\tilde{Z}_{j_k}$  is stochastically dominated by an exponential random variable with parameter  $\mu = \Theta(r^d)$ . Note that the probability that the sum of  $\ell - 2$  i.i.d. exponential random variables with parameter  $\mu$  is larger than  $\alpha\ell$  is equal to the probability that a Poisson random variable with mean  $\alpha\ell\mu$  is at most  $\ell - 3$ . Then, it follows from standard properties of Poisson random variables that there exists a constant  $c_6$  such that, for any large enough  $\alpha$ , we have

$$\Pr \left[ \sum_{k=2}^{\ell-1} |\widehat{Z}_{j_k}| \geq \alpha\ell \right] \leq \exp(-c_6 \alpha \ell r^d).$$

Since  $\ell = \Theta(\|v_1 - v_2\|_2/r) = \omega\left(\frac{\log n}{r^d}\right)$ , it holds that  $\Pr\left[\sum_{k=2}^{\ell-1} |\widehat{Z}_{j_k}| \geq \alpha\ell\right] = \mathcal{O}(m^{-3d})$  for some large enough  $\alpha$ . Then, using (3), it follows that, with probability  $1 - \mathcal{O}(m^{-3d})$ ,

$$F_3 \leq c_5 \alpha \ell M^d = \mathcal{O}(\|i_1 - i_2\|_1) = \mathcal{O}\left(\frac{\|v_1 - v_2\|_2}{r}\right).$$

Putting everything together, with probability  $1 - \mathcal{O}(m^{-3d})$ , we obtain a path from  $v_1$  to  $v_2$  with length at most

$$F_1 + F_2 + F_3 = \mathcal{O}\left(\frac{\log n}{r^d} + \frac{\|v_1 - v_2\|_2}{r}\right). \quad (4)$$

By Lemma 12, the result above holds for all connected pairs of nodes  $v_1, v_2$  such that  $\|v_1 - v_2\|_2 = \omega\left(\frac{\log n}{r^{d-1}}\right)$ . Then using  $m = \Theta(n^{1/d})$  completes the proof.  $\square$

#### 4 Existence of Long Paths

In this section we give the proof of Theorem 7. We show that there exist two connected nodes  $v_1$  and  $v_2$  whose Euclidean distance is at most  $3r$  but whose graph distance is  $\Omega\left(\frac{\log n}{r^d}\right) = \omega(1)$ . The proof is based on a geometric construction, which we define here and use at the end of this section to complete the proof.

Fix  $\varepsilon > 0$  small enough with respect to  $d$ , and set

$$\alpha = \frac{\varepsilon \log n}{r^{d-1}}.$$

Note that the condition  $r^d = o(\log n)$  assures that  $\frac{\alpha}{r} = \omega(1)$ .

Tessellate  $\Omega$  into disjoint parallelograms  $A_1, A_2, \dots$  of the form  $[0, L']^{d-1} \times [0, L]$  where

$$L' = 2r \left(1 + \frac{1}{d}\right) \quad \text{and} \quad L = 2\alpha + r \left(1 + \frac{2}{d}\right).$$

Let  $\mathcal{A}$  denote the set of these cubes. This tessellation is illustrated in two dimensions in Figure 3(a) and in three dimensions in Figure 3(c).

Consider a fixed parallelogram  $A_i \in \mathcal{A}$  and let  $x_i$  be such that

$$A_i = x_i + [0, L']^{d-1} \times [0, L].$$

We will partition  $A_i$  into two regions; refer to Figure 3(b) for an illustration of this partition in two dimensions and Figure 3(c) for an illustration of the partition in three dimensions. For  $k = 1, 2, \dots, d$ , let  $e_k$  be the vector whose  $k$ th element is 1 and whose other elements are all 0. First, define the four points (see Figure 3(b,c))

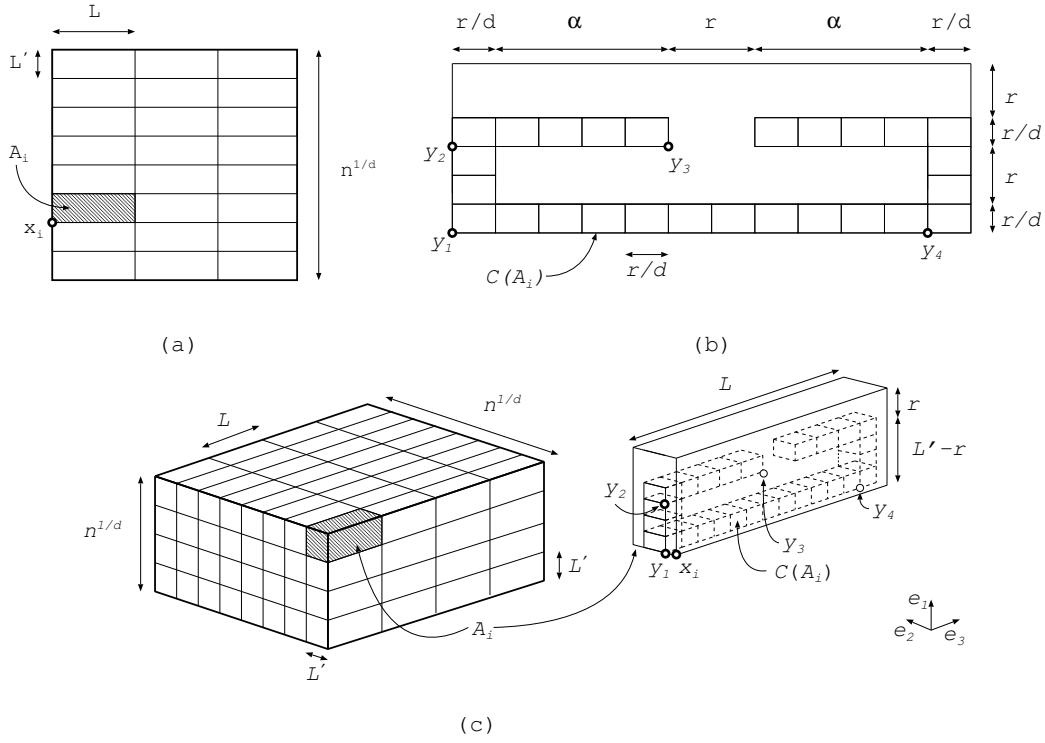
$$y_1 = x_i + \sum_{k=2}^{d-1} e_k \left(\frac{L' - r/d}{2}\right), \quad y_2 = y_1 + e_1 \left(L' - r - \frac{r}{d}\right),$$

$$y_3 = y_2 + e_d \left(\frac{L - r}{2}\right) \quad \text{and} \quad y_4 = y_1 + e_d \left(L - \frac{r}{d}\right).$$

(Note that, in two dimensions,  $y_1 = x_i$ .) We now define the parallelograms

$$Y_1 = y_1 + [0, r/d]^{d-1} \times [0, L], \quad Y_2 = y_2 + [0, r/d]^{d-1} \times [0, L],$$

$$Y_3 = y_3 + [0, r/d]^{d-1} \times [0, r]$$



**Figure 3** Geometric construction to show the existence of long paths.

and

$$Y_4 = (y_1 + [0, L' - r] \times [0, r/d]^{d-1}) \cup (y_4 + [0, L' - r] \times [0, r/d]^{d-1}).$$

Finally, we partition  $A_i$  into  $C(A_i)$  and  $A_i \setminus C(A_i)$ , where

$$C(A_i) = (Y_1 \cup Y_2 \cup Y_4) \setminus Y_3.$$

We see  $C(A_i)$  as the *core* of  $A_i$  and tessellate  $C(A_i)$  into disjoint cubes of side length  $r/d$ . In Figure 3(b,c),  $C(A_i)$  is the tessellated region that resembles a bracelet with an opening.

Now we define two events based on the regions described above. Consider the graph induced by the nodes inside  $A_i$ . Let  $\mathcal{E}_i$  be the event that there is no node inside  $A_i \setminus C(A_i)$  and let  $\mathcal{E}'_i$  be the event that each cube of the tessellation of  $C(A_i)$  contains at least one node. The usefulness of these two events is given by the following lemma.

**Lemma 16** *Suppose that both  $\mathcal{E}_i$  and  $\mathcal{E}'_i$  hold for some  $A_i \in \mathcal{A}$ . Then there exist two connected nodes  $v_1, v_2$  in  $G$  such that  $\|v_1 - v_2\|_2 \leq 3r$  and  $d_G(v_1, v_2) \geq 4\alpha/r$ .*

*Proof.* Since the cubes in the tessellation of  $C(A_i)$  have side length  $r/d$ , two nodes in adjacent cubes are neighbors in  $G$ . Then, when  $\mathcal{E}'_i$  happens, there is a path of nodes crossing through the bracelet represented by region  $C(A_i)$ . Consider the two cubes that are adjacent to the opening of the bracelet; more formally, these are the cubes  $y_3 - e_d(r/d) + [0, r/d]^d$  and  $y_3 + e_d r + [0, r/d]^d$ . Since  $\mathcal{E}'_i$  happens, there is one node in each such cube, call them  $v_1$  and  $v_2$ . Moreover,  $v_1$  and  $v_2$  are connected and

$$\|v_1 - v_2\|_2 \leq 2\frac{r}{d}\sqrt{d} + r \leq 3r.$$

Since  $\mathcal{E}_i$  also happens, the path from  $v_1$  to  $v_2$  must go through the bracelet, which gives that

$$d_G(v_1, v_2) \geq \frac{4\alpha}{r}. \quad \square$$

Now we have all the results we need to give the proof of Theorem 7.

*Proof of Theorem 7.* For any given  $A_i \in \mathcal{A}$ , we have

$$\Pr[\mathcal{E}_i] \geq \exp(-\text{vol}(A_i)) = \exp(-c_1 r^{d-1} \alpha),$$

for some constant  $c_1 = c_1(d) > 0$  and all large enough  $n$ . Using the fact that  $\text{vol}(Y_1) = \text{vol}(Y_2)$ , the number of cubes in the tessellation of  $C(A_i)$  is exactly

$$\frac{\text{vol}(C(A_i))}{(r/d)^d} \leq \frac{2\text{vol}(Y_1) + \text{vol}(Y_4)}{(r/d)^d} \leq \frac{1}{(r/d)^d} (2(r/d)^{d-1}L + 2(r/d)^{d-1}(L' - r)) = \Theta\left(\frac{\alpha}{r}\right).$$

Using this we obtain

$$\Pr[\mathcal{E}'_i] \geq (1 - \exp(-(r/d)^d))^{\frac{\text{vol}(C(A_i))}{(r/d)^d}} \geq \exp\left(-c_2 \frac{\alpha}{r}\right),$$

for some constant  $c_2 = c_2(d) > 0$  and all large enough  $n$ . Note that  $\mathcal{E}_i$  and  $\mathcal{E}'_i$  are independent, since they are defined in disjoint regions. Also, since  $r > r_c = \Theta(1)$ , there exists a constant  $c_3 = c_3(d) > 0$  for which

$$\frac{\alpha}{r} \leq c_3 r^{d-1} \alpha,$$

which gives

$$\Pr[\mathcal{E}_i \cap \mathcal{E}'_i] \geq \exp(-2c_3 r^{d-1} \alpha).$$

Now set  $\varepsilon = \frac{1}{100c_3}$ , which gives

$$\Pr[\mathcal{E}_i \cap \mathcal{E}'_i] \geq n^{-1/50}.$$

Using Lemma 16, we need only to show that there exists an  $A_i \in \mathcal{A}$  for which both  $\mathcal{E}_i$  and  $\mathcal{E}'_i$  occur. But these events are mutually independent over  $i$  since the  $A_i$  are disjoint. Noting that

$$|\mathcal{A}| = \frac{n}{L^{d-1}L} = \Theta\left(\frac{n}{\alpha r^{d-1}}\right),$$

it follows that the probability that there exists an  $i$  for which  $\mathcal{E}_i$  and  $\mathcal{E}'_i$  both hold is at least

$$1 - \left(1 - n^{-1/50}\right)^{|\mathcal{A}|} \geq 1 - \exp\left(-\Omega\left(\frac{n^{49/50}}{\log n}\right)\right).$$

This concludes the proof of Theorem 7.  $\square$

## 5 Broadcast Time

In this section we prove Theorem 2. Given two nodes  $v_1$  and  $v_2$ , let  $\mathcal{R}(v_1, v_2)$  be the time it takes for the random broadcast algorithm started at  $v_1$  to inform  $v_2$  for the first time. We assume in the sequel that  $v_1$  and  $v_2$  belong to the largest connected component of  $G$  and show that, provided  $\|v_1 - v_2\|_2 = \omega(\log^2 n)$ ,  $\mathcal{R}(v_1, v_2) = \mathcal{O}(\|v_1 - v_2\|_2/r)$ . (We deal with the easier case  $\|v_1 - v_2\|_2 = \mathcal{O}(\log^2 n)$  later.)

We assume that  $r = \mathcal{O}(\log^{1/d} n)$ . The case  $r = \omega(\log^{1/d} n)$  is simpler; since it uses different proof techniques, we handle it separately with the lemma below.

**Lemma 17** *If  $r = \omega(\log^{1/d} n)$ , then for all nodes  $s \in G$  we obtain  $\mathcal{R}(s, G) = \mathcal{O}(n^{1/d}/r + \log n)$  with probability  $1 - \mathcal{O}(n^{-1})$ .*

*Proof.* In order to prove Lemma 17, we consider a tessellation of  $\Omega$  into cubes of side length  $\min\left\{\frac{r}{2\sqrt{d}}, \frac{n^{1/d}}{2}\right\}$ , which we refer to as cells. (If  $\frac{n^{1/d}}{2}$  is not a multiple of  $\frac{r}{2\sqrt{d}}$ , then we make the cells in the last row or column of the tessellation be smaller than the others.) It is easy to verify that nodes in the same cell are neighbors in  $G$  and that a node in a given cell can only have neighbors in  $K = (1 + 4\sqrt{d})^d$  different cells. Let  $a_{\min}$  be the number of nodes inside the cell that contains the smallest number of nodes, and let  $a_{\max}$  be the number of nodes inside the cell that contains the largest number of nodes. Since  $r = \omega(\log^{1/d} n)$ , a standard Chernoff bound for Poisson random variables implies that there are constants  $c_1 < c_2$  such that a fixed cell contains at least  $c_1 r^d$  nodes and at most  $c_2 r^d$  nodes with probability larger than  $1 - n^{-2}$ . Using the union bound over all cells of the tessellation, we obtain the result that  $a_{\min}$  and  $a_{\max}$  are  $\Theta(r^d)$  with probability  $1 - \mathcal{O}(n^{-1})$ .

We are now in position to start our proof for Lemma 17. We index the cells by  $i \in \mathbb{Z}^d$  and let  $Z_i$  be the event that the cell  $i$  contains at least one informed node. We say that cells  $i$  and  $j$  are adjacent if and only if they intersect in a  $(d-1)$ -dimensional face. Therefore, each cell has exactly  $2d$  adjacent cells. Let  $C$  be the graph induced by this adjacency relation. Note that nodes in adjacent cells are neighbors in  $G$ .

Given two adjacent cells  $i$  and  $j$ , at any round of the random broadcast algorithm, an informed node in cell  $i$  chooses a node from cell  $j$  with probability larger than  $a_{\min}/(K a_{\max}) = \Theta(1)$ . We want to derive the time until  $Z_i = 1$  for all  $i$ . Given a path between two cells  $j_1, j_2 \in C$ , the number of rounds the information takes to be transmitted along this path can be bounded above by the sum of independent geometric random variables with mean  $\Theta(1)$ . Applying a Chernoff bound for geometric random variables (cf. Lemma 21), we infer that the number of rounds required to transmit the information from  $j_1$  to  $j_2$  is smaller than  $\mathcal{O}(\text{diam}(C) + \log n)$  with probability  $1 - e^{-\Omega(\text{diam}(C) + \log n)}$ . Since there are  $\mathcal{O}(n/r^d)$  cells and  $\text{diam}(C) = \mathcal{O}(n^{1/d}/r)$ , it follows that with probability  $1 - \mathcal{O}(n^{-1})$ ,  $Z_i = 1$  for all  $i$  after  $\mathcal{O}(n^{1/d}/r + \log n)$  rounds.

To complete the spreading within each cell, we consider a faulty version of the random broadcast algorithm, which proceeds as explained in Section 2, but when an informed node is about to transmit the information to a neighbor chosen independently and uniformly at random, this transmission fails with probability  $p \in [0, 1)$  independently of all other transmissions. Moreover, a node that was not informed at the beginning of the algorithm can only get informed if it receives the information from a transmission that did not fail. We denote by  $\mathcal{R}_p(s, G)$  the runtime of the faulty version of the random broadcast algorithm initiated at node  $s \in G$ . We use the following relation between  $\mathcal{R}(s, G)$  and  $\mathcal{R}_p(s, G)$ .



**Lemma 18** ([10, Theorem 6]) *For any graph  $G$ , any node  $s \in G$ , and any  $p \in [0, 1)$ , there exists a coupling between  $\mathcal{R}_p(s, G)$  and  $\mathcal{R}(s, G)$  such that*

$$\mathcal{R}_p(s, G) = \mathcal{O}\left(\frac{\mathcal{R}(s, G)}{1-p}\right).$$

Assume that each cell contains at least one informed node. We want to determine how many additional rounds are required until all nodes in  $G$  become informed. Note that each cell constitutes a clique with  $\Theta(r^d)$  nodes. According to the random broadcast algorithm, at any round, a node chooses a neighbor inside its own cell with probability larger than  $a_{\min}/(Ka_{\max}) = \Theta(1)$ . Therefore, a standard coupling argument can be used to show that the time taken until all nodes from a given cell get informed can be bounded above by the time the faulty version of the random broadcast algorithm with failure probability  $\Theta(1)$  takes to inform all nodes of a complete graph with  $\Theta(r^d)$  nodes. Thus, by [11, Theorem 4.1] and Lemma 18, all nodes of a given cell get informed within  $\mathcal{O}(\log r^d + \log n)$  steps with probability  $1 - \mathcal{O}(n^{-2})$ . Then, taking the union bound over all cells it follows that, with probability  $1 - \mathcal{O}(n^{-1})$ , all nodes of  $G$  get informed after  $\mathcal{O}\left(\frac{n^{1/d}}{r} + \log n\right)$  steps, which concludes the proof of Lemma 17.  $\square$

Now we give the proof for the more involved case  $r = \mathcal{O}(\log^{1/d} n)$ .

*Proof of Theorem 2 when  $r = \mathcal{O}(\log^{1/d} n)$ .* We start with a basic lemma that shows that the time until a node informs a given neighbor is  $\mathcal{O}(\log^2 n)$  with high probability.

**Lemma 19** *If  $r = \mathcal{O}(\log^{1/d} n)$ , then there exists a constant  $c$  such that, for all pairs of nodes  $w_1$  and  $w_2$  satisfying  $\|w_1 - w_2\|_2 \leq r$ , the following holds with probability  $1 - \mathcal{O}(n^{-1})$ ,*

$$\mathcal{R}(w_1, w_2) \leq c \log^2 n.$$

*Proof.* Note that, if the degree of  $w_1$  in  $G$  is  $k$ , then the number of rounds until  $w_1$  sends the information to  $w_2$  is given by a geometric random variable with mean  $k$ . It is easy to check that there is a constant  $c$  such that, with probability  $1 - \mathcal{O}(n^{-3})$ , all nodes of a random geometric graph have degree smaller than  $c \log n$  [23] provided  $r = \mathcal{O}(\log^{1/d} n)$ . Therefore,

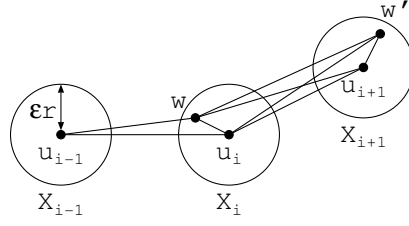
$$\Pr[\mathcal{R}(w_1, w_2) \geq t] \leq \left(1 - \frac{1}{c \log n}\right)^t \leq \exp\left(-\frac{t}{c \log n}\right).$$

If we set  $t = 3c \log^2 n$ , we obtain that  $\Pr[\mathcal{R}(w_1, w_2) \geq 3c \log^2 n] \leq \mathcal{O}(n^{-3})$  and, by Lemma 12, we conclude that  $\mathcal{R}(w_1, w_2) \leq 3c \log^2 n$  for all  $w_1, w_2$  with probability  $1 - \mathcal{O}(n^{-1})$ .  $\square$

Before proceeding, note that the lemma above shows that  $\mathcal{R}(v_1, v_2)$  can be bounded above by  $\mathcal{O}(d_G(v_1, v_2) \log^2 n)$ . We will derive a much better bound in the sequel.

Let  $r'$  be defined such that  $r_c < r' < r$ . Note that such an  $r'$  exists since  $r > r_c$ . For convenience, write  $r' = r(1 - 2\varepsilon)$ . Now let  $\delta > 0$  be sufficiently small and, using the thinning property of Poisson point processes, split  $\mathcal{P}_n$  into two Poisson point processes  $\mathcal{P}'_n$  and  $\mathcal{P}''_n$  with intensities  $1 - \delta$  and  $\delta$ , respectively. Since  $r' > r_c$ , we can set  $\delta$  sufficiently small so that  $G' = \mathcal{G}(\mathcal{P}'_n, r')$  contains a connected component of size  $\Omega(n)$  with probability  $1 - e^{-\Omega(n^{1-1/d})}$ . Note also that  $G'$  is a subgraph of  $G$ .

Our strategy to obtain an upper bound for  $\mathcal{R}(v_1, v_2)$  is the following. First, we assume that  $v_1$  and  $v_2$  belong to the largest connected component of  $G'$ . (We address the case where they do not belong to the largest connected component of  $G'$  at the end of this section.) Then, we take a path in  $G'$  from  $v_1$  to  $v_2$ . Instead of calculating the time it takes for the



**Figure 4** Illustration of the path considered to obtain  $\mathcal{R}(v_1, v_2)$ . The picture shows three consecutive nodes  $u_{i-1}$ ,  $u_i$ , and  $u_{i+1}$  of the path from  $v_1$  to  $v_2$  and the balls  $X_{i-1}$ ,  $X_i$ , and  $X_{i+1}$  around them. Two other nodes  $w \in X_i$  and  $w' \in X_{i+1}$  are depicted to illustrate the edges that arise from the construction of the  $X_i$ s.

random broadcast algorithm to transmit the information along this path, which gives a rather pessimistic upper bound, we enlarge the path using the fact that  $G'$  is a subgraph of  $G$  and calculate the time it takes for the random broadcast algorithm to transmit the information along this enlarged path.

Let  $\mathcal{E}_\kappa(v_1, v_2)$  be the event that there exists a path between  $v_1$  and  $v_2$  in  $G'$  with length at most  $\kappa$ . Let  $u_1, u_2, \dots, u_k$  be a *fixed* minimal path from  $v_1$  to  $v_2$  in  $G'$ , where  $u_1 = v_1$ ,  $u_k = v_2$  and  $k \leq \kappa$ ; we use the nomenclature *minimal* to refer to the property that, for all  $i$  and  $j$  with  $|i - j| > 1$ , the Euclidean distance between  $u_i$  and  $u_j$  is larger than  $r'$ . Note that, when  $\mathcal{E}_\kappa(v_1, v_2)$  occurs, we can find such a path. Now, for each  $i$ , we define the region  $X_i \subseteq \Omega$  in the following way. Set  $X_1$  to be the point where  $u_1$  is located and  $X_k$  to be the point where  $u_k$  is located; for  $2 \leq i \leq k - 1$ , define  $X_i$  to be the ball with center at  $u_i$  and radius  $\varepsilon r$ . Our goal is to get an upper bound for  $\mathcal{R}(v_1, v_2)$  by following the path  $X_1, X_2, \dots, X_k$  (refer to Figure 4).

Define the random variable  $T(X_i, X_{i+1})$ ,  $1 \leq i \leq k - 1$ , as the time the random broadcast algorithm takes to first inform a node in  $X_{i+1}$  given that it started in a node chosen uniformly at random from  $X_i$ . For convenience, we set  $T(X_i, X_{i+1}) = 0$  for all  $i \geq k$ . Note that, for any two nodes  $w \in X_i$  and  $w' \in X_{i+1}$ , the triangle inequality and the definition of  $X_i$  give  $\|w - w'\|_2 \leq 2\varepsilon r + \|u_i - u_{i+1}\|_2 \leq r$ . Therefore,  $w$  and  $w'$  are neighbors in  $G$ . Moreover, for any  $i$ , once the random broadcast algorithm informs a node inside  $X_i$ , then the node that receives the information is a uniformly random node from  $X_i$ . Thus, we set  $\kappa = C \frac{\|v_1 - v_2\|_2}{r}$  for some large constant  $C$  and have the following lower bound for all positive constants  $c$

$$\Pr[\mathcal{R}(v_1, v_2) \leq ck] \geq \Pr[\{\mathcal{R}(v_1, v_2) \leq ck\} \cap \mathcal{E}_\kappa(v_1, v_2)].$$

Once we know that  $\mathcal{E}_\kappa(v_1, v_2)$  occurs, we can fix a minimal path  $u_1, u_2, \dots, u_k$  according to any arbitrary order of the nodes of  $G$  and write

$$\begin{aligned} \Pr[\{\mathcal{R}(v_1, v_2) \leq ck\} \cap \mathcal{E}_\kappa(v_1, v_2)] &\geq \Pr\left[\left\{\sum_{i=1}^k T(X_i, X_{i+1}) \leq ck\right\} \cap \mathcal{E}_\kappa(v_1, v_2)\right] \\ &\geq \Pr\left[\sum_{i=1}^k T(X_i, X_{i+1}) \leq ck\right] - (1 - \Pr[\mathcal{E}_\kappa(v_1, v_2)]). \end{aligned}$$

It is important to remark that, once we have fixed the path  $u_1, u_2, \dots, u_k$ , the term  $\Pr\left[\sum_{i=1}^k T(X_i, X_{i+1}) \leq ck\right]$  can be estimated without any additional information on  $G$ . To see this, let  $\Psi$  be the subset of  $\Omega$  consisting of the location of the nodes  $u_1, u_2, \dots, u_k$  and the points of  $\Omega$  at which adding a node would violate the condition that  $u_1, u_2, \dots, u_k$  is a minimal path. Then, conditional on the existence of this particular path, the Poisson point process over  $\Omega \setminus \Psi$  remains unchanged. Also, for this choice of  $\kappa$ , it follows from Theorem 3

(more precisely Equation (4)) that  $1 - \mathbf{Pr}[\mathcal{E}_\kappa(v_1, v_2)] = \mathcal{O}(n^{-3})$  for all  $v_1$  and  $v_2$  such that  $\|v_1 - v_2\|_2 = \omega\left(\frac{\log n}{r^{d-1}}\right)$ . So we only need to derive a bound for  $\mathbf{Pr}\left[\sum_{i=1}^k T(X_i, X_{i+1}) \leq ck\right]$  for any fixed choice of  $u_1, u_2, \dots, u_k$  that forms a minimal path.

Note that Lemma 19 gives  $T(X_{k-1}, X_k) = \mathcal{O}(\log^2 n)$  with probability  $1 - \mathcal{O}(n^{-1})$ , for each choice of  $v_1$  and  $v_2$ . The next lemma gives the expectation of  $T(X_i, X_{i+1})$  for each  $1 \leq i \leq k-2$ .

**Lemma 20** *Given  $u_1, u_2, \dots, u_k$  is a minimal path, for any  $1 \leq i \leq k-2$ , it holds that  $T(X_i, X_{i+1})$  is a geometric random variable with  $\mathbf{E}[T(X_i, X_{i+1})] \leq c$ , for some constant  $c$  depending only on  $d, \varepsilon$  and  $\delta$ .*

*Proof.* Let  $w$  be a node chosen uniformly at random from  $X_i$ . Assume  $w \notin X_{i+1}$  (otherwise, the broadcast time from  $w$  to  $X_{i+1}$  is zero). Let  $Y$  be the number of neighbors of  $w$  in  $G$  and let  $Y'$  be 1 plus the number of nodes of  $\mathcal{P}_n''$  in  $X_{i+1}$ , where the term 1 is to account for  $u_{i+1}$ . Therefore,  $\mathbf{E}[T(X_i, X_{i+1})] \leq \mathbf{E}[Y/Y']$ . We know that  $Y' \geq 1$  and  $Y' - 1$  is a Poisson random variables with mean  $\mu' = \delta b_d \varepsilon^d r^d$ , where  $b_d$  is the volume of the  $d$ -dimensional ball of radius 1. We need to account for the fact that the path is minimal according to  $G'$ , but non-adjacent nodes in the path may still be neighbors in  $G$ . To solve this, note that there exists a constant  $K$  depending only on  $d$  and  $\varepsilon$  such that the number of nodes in the path that are neighbors of a given  $u_j$  in  $G$  is at most  $K$ , which follows since the path  $u_1, u_2, \dots, u_k$  is minimal with respect to  $G'$  and  $r' = (1 - 2\varepsilon)r$ . Also, note that  $Y \geq Y'$  and we can use a coupling argument to show that  $Y \leq Y' + K + Y''$ , where  $Y''$  is a Poisson random variable with mean  $\mu'' = b_d r^d - \delta b_d \varepsilon^d r^d$ . We then obtain

$$\begin{aligned} \mathbf{E}[T(X_i, X_{i+1})] &\leq \sum_{y'=0}^{\infty} \sum_{y''=0}^{\infty} \frac{1 + y' + K + y''}{1 + y'} \mathbf{Pr}[Y' = 1 + y'] \mathbf{Pr}[Y'' = y''] \\ &= \sum_{y'=0}^{\infty} \frac{1 + y' + K + \mu''}{1 + y'} \mathbf{Pr}[Y' = 1 + y'] \\ &= 1 + \frac{K + \mu''}{\mu'} \sum_{y'=0}^{\infty} \frac{(\mu')^{y'+1}}{(y' + 1)!} e^{-\mu'} \\ &\leq 1 + \frac{K + \mu''}{\mu'}. \quad \square \end{aligned}$$

Using a Chernoff bound for geometric random variables (cf. Lemma 21), we obtain

$$\mathbf{Pr}\left[\sum_{i=1}^k T(X_i, X_{i+1}) \geq (1+x)ck\right] \leq \exp\left(-x^2 \frac{k}{2(1+x)}\right),$$

where  $c$  is the constant in Lemma 20. Note that  $k = \Theta(\|v_1 - v_2\|_2/r) = \Omega(\log n)$  by construction, since  $\|v_1 - v_2\|_2 = \omega(\log^2 n)$ ,  $r = \mathcal{O}(\log^{1/d} n)$  and the path  $u_1, u_2, \dots, u_k$  is minimal.

Applying Lemma 12, we can conclude that, for any two nodes  $v_1$  and  $v_2$  in the largest connected component of  $G$  for which  $\|v_1 - v_2\|_2 = \omega(\log^2 n)$ , we obtain  $\mathcal{R}(v_1, v_2) = \Theta(\|v_1 - v_2\|_2/r)$ . Note that there exist  $v_1, v_2 \in G$  for which  $\|v_1 - v_2\|_2 = \Theta(n^{1/d})$  and, consequently,  $\mathcal{R}(v_1, v_2) = \Theta(n^{1/d}/r)$ .

Now we treat the two remaining cases. First, since  $G'$  is a subgraph of  $G$ , there may exist some nodes in the largest connected component of  $G$  that do not belong to the largest connected component of  $G'$ . Nevertheless, it is known that the second largest component of a random geometric graph  $G'$  contains  $\mathcal{O}(\log^{\frac{d}{d-1}} n)$  nodes with probability  $1 - \mathcal{O}(n^{-1})$  [23,

Theorem 10.18]. Therefore, since  $\mathcal{R}(w_1, w_2) = \mathcal{O}(\log^2 n)$  for every pair of neighbors  $w_1$  and  $w_2$ , we conclude that the time it takes to inform all the remaining nodes is  $\mathcal{O}(\log^{\frac{d}{d-1}+2} n)$ , which is negligible in comparison to  $\Theta(n^{1/d}/r)$ .

The second case corresponds to the nodes that are within Euclidean distance  $\mathcal{O}(\log^2 n)$  to the initially informed node, which is denoted here as  $v_1$ . Take  $Q$  to be a square centered at  $v_1$  with side length  $c \log^3 n$ , for some constant  $c$  (the orientation of  $Q$  does not matter). Note that  $Q$  contains all nodes within Euclidean distance  $\mathcal{O}(\log^2 n)$  of  $v_1$ . Now, take  $Q'$  to be a square centered at  $v_1$ , with the same orientation as  $Q$ , but with sides having twice the length of the sides of  $Q$ . Clearly,  $Q' \setminus Q$  is an annulus centered at  $v_1$  and, by standard results on random geometric graphs [23, Lemma 10.5 and Proposition 10.6], the probability that the intersection of the giant component of  $G$  with the largest connected component of  $Q'$  contains all the nodes of  $Q$  that are connected to  $v_1$  is at least  $1 - e^{-\Omega(\log^3 n)}$ . This happens because, with probability  $1 - e^{-\Omega(\log^3 n)}$ ,  $Q'$  has exactly one connected component with diameter larger than  $c \log^3 n/5$  and this component intersects the giant component of  $G$ . For all nodes within Euclidean distance  $\mathcal{O}(\log^2 n)$  of  $v_1$  that are connected to  $v_1$ , there exists a path from them to  $v_1$  completely contained in  $Q'$ ; hence, this path has  $\mathcal{O}\left(\frac{\log^{3d} n}{r^d}\right)$  edges by Lemma 14 and Remark 15. So using Lemma 19 we conclude that all nodes within Euclidean distance  $\mathcal{O}(\log^2 n)$  to  $v_1$  are informed after  $\mathcal{O}\left(\frac{\log^{3d+2} n}{r^d}\right)$  rounds, which is also negligible in comparison to  $\Theta(n^{1/d}/r)$ . This completes the proof of Theorem 2.  $\square$

## 6 Conclusion

We have analyzed the performance of the random broadcast algorithm in random geometric graphs in  $d$  dimensions. We proved that with probability  $1 - \mathcal{O}(n^{-1})$  the algorithm finishes within  $\mathcal{O}(n^{1/d}/r + \log n)$  steps, where  $r$  can be an arbitrary value above the critical coverage radius for the emergence of a connected component with  $\Omega(n)$  nodes. We also showed that for any two nodes  $v_1$  and  $v_2$  such that  $\|v_1 - v_2\|_2 = \omega\left(\frac{\log n}{r^{d-1}}\right)$ , the length of the shortest path between them in the random geometric graph is  $\mathcal{O}(\|v_1 - v_2\|_2/r)$ . In particular, this implies that the diameter of the largest connected component is  $\mathcal{O}(n^{1/d}/r)$ .

Our techniques may be useful to analyze other problems like the cover time of the largest connected component of RGGs. This would nicely complement results by Cooper and Frieze for connected RGGs [6] and for the largest connected component of Erdős-Rényi random graphs [5].

## References

1. I. Akyildiz, D. Pompili, and T. Melodia. Underwater acoustic sensor networks: research challenges. *Ad Hoc Networks*, 3:257–279, 2005.
2. P. Antal and A. Pisztor. On the chemical distance for supercritical Bernoulli percolation. *The Annals of Probability*, 24:1036–1048, 1996.
3. C. Avin and G. Ercal. On the cover time and mixing time of random geometric graphs. *Theoretical Computer Science*, 380(1-2):2–22, 2007.
4. M. Bradonjić, R. Elsässer, T. Friedrich, T. Sauerwald, and A. Stauffer. Efficient broadcast on random geometric graphs. In *21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '10)*, pages 1412–1421, 2010.
5. C. Cooper and A. Frieze. The cover time of the giant component of a random graph. *Random Structures & Algorithms*, 32(4):401–439, 2008.

- 1 6. C. Cooper and A. Frieze. The cover time of random geometric graphs. In *19th Annual*
- 2 *ACM-SIAM Symposium on Discrete Algorithms (SODA'09)*, pages 48–57, 2009.
- 3 7. J.-D. Deuschel and A. Pisztora. Surface order large deviations for high-density perco-
- 4 lation. *Probability Theory and Related Fields*, 104:467–482, 1996.
- 5 8. R. B. Ellis, J. L. Martin, and C. Yan. Random geometric graph diameter in the unit
- 6 ball. *Algorithmica*, 47(4):421–438, 2007.
- 7 9. R. Elsässer and T. Sauerwald. Broadcasting vs. mixing and information dissemination
- 8 on Cayley graphs. In *24th International Symposium on Theoretical Aspects of Computer*
- 9 *Science (STACS'07)*, pages 163–174, 2007.
- 10 10. R. Elsässer and T. Sauerwald. On the runtime and robustness of randomized broad-
- 11 casting. *Theoretical Computer Science*, 410(36):3414–3427, 2009.
- 12 11. U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized broadcast in networks.
- 13 *Random Structures and Algorithms*, 1(4):447–460, 1990.
- 14 12. L. Fontes and C. Newman. First passage percolation for random colorings of  $\mathbb{Z}^d$ . *The*
- 15 *Annals of Applied Probability*, 3(3):746–762, 1993.
- 16 13. N. Fountoulakis and K. Panagiotou. Rumor spreading on random regular graphs and ex-
- 17 panders. In *14th Inter. Workshop on Randomization and Comput. (RANDOM)*, volume
- 18 6302 of *LNCS*, pages 560–573, 2010.
- 19 14. A. Frieze and G. Grimmett. The shortest-path problem for graphs with random arc-
- 20 lengths. *Discrete Applied Mathematics*, 10:57–77, 1985.
- 21 15. G. Grimmett. *Percolation*. Springer Verlag, 2nd edition, 1999.
- 22 16. D. Higham, M. Rašajski, and N. Pržulj. Fitting a geometric graph to a protein–protein
- 23 interaction network. *Bioinformatics*, 24(8):1093, 2008.
- 24 17. T. Liggett, R. Schonmann, and A. Stacey. Domination by product measures. *The*
- 25 *Annals of Probability*, 25(1):71–95, 1997.
- 26 18. R. Meester and R. Roy. *Continuum percolation*. Cambridge University Press, 1996.
- 27 19. C.-H. Ou and K.-F. Ssu. Sensor position determination with flying anchors in three-
- 28 dimensional wireless sensor networks. *IEEE Transactions on Mobile Computing*, 7:
- 29 1084–1097, 2008.
- 30 20. M. Penrose and A. Pisztora. Large deviations for discrete and continuous percolation.
- 31 *Advances in Applied Probability*, 28:29–52, 1996.
- 32 21. M. D. Penrose. The longest edge of the random minimal spanning tree. *The Annals of*
- 33 *Applied Probability*, 7(2):340–361, 1997.
- 34 22. M. D. Penrose. On  $k$ -connectivity for a geometric random graph. *Random Struct.*
- 35 *Algorithms*, 15(2):145–164, 1999.
- 36 23. M. D. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.
- 37 24. B. Pittel. On spreading a rumor. *SIAM Journal on Applied Mathematics*, 47(1):213–223,
- 38 1987.
- 39 25. G. J. Pottie and W. J. Kaiser. Wireless integrated network sensors. *Commun. ACM*,
- 40 43(5):51–58, 2000.
- 41 26. T. Sauerwald. On mixing and edge expansion properties in randomized broadcasting.
- 42 *Algorithmica*, 56:51–88, 2010.
- 43
- 44
- 45
- 46
- 47
- 48
- 49

## 50 A Standard large deviation results

51 **Lemma 21 (Chernoff Bound for Sums of Geometric Variables)** *Let  $X_1, \dots, X_n$  be independent*

52 *geometric random variables, each having parameter  $p$  (and thus mean  $1/p$ ), and let  $X = \sum_{i=1}^n X_i$ . Then,*

53 *for any  $\varepsilon > 0$ ,*

$$54 \Pr \left[ X \geq (1 + \varepsilon) \frac{n}{p} \right] \leq \exp \left( -\frac{\varepsilon^2}{2(1 + \varepsilon)} n \right).$$

55

56

57

58

59

60

61

62

63

64

65