Efficient Broadcast on Random Geometric Graphs

Milan Bradonjić* Robert Elsässer† Tobias Friedrich‡ Thomas Sauerwald§

Alexandre Stauffer¶

Abstract

A *Random Geometric Graph* (RGG) in two dimensions is constructed by distributing n nodes independently and uniformly at random in $[0, \sqrt{n}]^2$ and creating edges between every pair of nodes having Euclidean distance at most r, for some prescribed r. We 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 with probability $1 - \mathcal{O}(n^{-1})$ this algorithm informs every node in the largest connected component of an RGG within $\mathcal{O}(\sqrt{n}/r + \log n)$ rounds. This holds for any value of r larger than the critical value for the emergence of a connected component with $\Omega(n)$ nodes. In order to prove this result, we show that for any two nodes sufficiently distant from each other in $[0, \sqrt{n}]^2$, the length of the shortest path between them in the RGG, when such a path exists, is only a constant factor larger than the optimum. This result has independent interest and, in particular, gives that the diameter of the largest connected component of an RGG is $\Theta(\sqrt{n}/r)$, which surprisingly has been an open problem so far.

1 Introduction

A *Random Geometric Graph* (RGG) is a graph resulting from placing n nodes independently and uniformly at random on $[0, \sqrt{n}]^2$ and creating edges between pairs of nodes if and only if their Euclidean distance is at most some fixed r. These graphs have been studied intensively in relation to subjects such as cluster analysis, statistical physics, hypothesis testing [12], and wireless sensor networks [14]. One further application of RGGs is modeling data in a high-dimensional space, where the coordinates of the nodes of the RGG represent the attributes of the data. The metric imposed by the RGG then depicts the similarity between data elements in the high-dimensional space.

In this work, we are specifically interested in the problem of *broadcasting information* in random geometric graphs (RGG) in two dimensions. The study of information spreading in large networks has various applications in distributed computing. Typically, the broadcast algorithm should be simple, be resilient against failures, and work locally, i.e., the nodes cannot be assumed to have any prior knowledge about the global topology of the network. One simple algorithm of this kind is the random broadcast (a.k.a. the push algorithm) [8], which we study here. In this algorithm, in each round each informed node chooses a neighbor independently and uniformly at random and informs it.

The random broadcast algorithm has been first analyzed on complete graphs by Frieze and Grimmett [9], who proved that with probability $1 - o(1)$, the runtime is $\log_2 n + \ln n + o(\log n)$. This result was later further improved by Pittel [13]. Feige et al. [8] proved that on any graph, the runtime is at most $\mathcal{O}(n \log n)$ with probability $1 - \mathcal{O}(n^{-1})$, and that for any bounded-degree graph, $\mathcal{O}(\mathrm{diam}(G))$ rounds are sufficient, where $\mathrm{diam}(G)$ stands for the diameter of the graph. Furthermore, they established a runtime of $\mathcal{O}(\log n)$ on hypercubes and sufficiently dense random graphs with probability $1 - \mathcal{O}(n^{-1})$. In [6], two of the authors extended this result to other graphs by proving an upper bound of $\mathcal{O}(\mathrm{diam}(G) + \log n)$ for different Cayley graphs.

A different broadcasting model known as *Radio Broadcasting* has also been studied on RGGs [4, 10]. In this model, every transmission by a node is sent to all neighbors. However, if two (or more) transmissions are sent to the same node in one round, then this node cannot receive the message. In order to derive an efficient algorithm for radio broadcasting on these graphs, Lotker and Navarra solved this problem first on a grid [10]. Then, they emulated the corresponding grid protocol on RGGs, and obtained an asymptotically optimal algorithm for broadcasting in the case when

*Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

†Department of Computer Science, University of Paderborn, 33102 Paderborn, Germany

‡Department 1: Algorithms and Complexity, Max-Planck-Institut für Informatik, 66123 Saarbrücken, Germany

§School of Computing Science, Simon Fraser University, Burnaby B.C. V5A 1S6, Canada

¶Computer Science Division, University of California, Berkeley, CA 94720, USA

the graph is connected with high probability. However, the result of [10] only holds if each node is aware of its own position. Later, Czumaj and Wang considered various scenarios with respect to the local knowledge of each node in the graph, and showed that in many settings radio broadcasting[1] can be solved in time $\mathcal{O}(\operatorname{diam}(G))$ [4].

A problem related to broadcasting that has already been studied for RGGs is the cover time of random walks [1, 3]. In [1], Avin and Ercal considered random geometric graphs in two dimensions when the coverage radius is a constant larger than the minimum coverage radius that assures the RGG to be 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. Recently, Cooper and Frieze [3] gave a more precise estimate of the cover time on RGGs that extends also to larger dimensions. However, all of these works are restricted to the case where the probability that the RGG is connected goes to 1 as $n \to \infty$.

In this work, we analyze a wider range for r and we focus on the regime where the RGG is likely to contain a connected component with $\Omega(n)$ nodes. We prove that if one node from the largest connected component of an RGG uses the random broadcast algorithm to disseminate a piece of information, then with probability $1 - \mathcal{O}(n^{-1})$, all nodes in the same connected component receive the information within $\mathcal{O}(\sqrt{n}/r + \log n)$ rounds. In particular, if the RGG turns out to be connected, then all nodes get informed after $\mathcal{O}(\sqrt{n}/r + \log n)$ rounds.

In our proof for this result, we also show that for any two nodes having sufficiently large Euclidean distance, their distance in the RGG is just a constant factor larger than the optimum. In particular, this result shows that the diameter of the largest connected component of an RGG is $\Theta(\sqrt{n}/r)$ in the case where a connected component with $\Omega(n)$ nodes is likely to exist. This result has independent interest and, to the best of our knowledge, was only previously known for the case when the RGG is connected with probability $1 - o(1)$ [5]. Our techniques are inspired by percolation theory and we believe them to be useful for other problems, like estimating the cover time for the largest connected component of RGGs.

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 for the length of the shortest path between two nodes in an RGG provided their Euclidean distance. In Section 4, we perform the runtime analysis of the random broadcast algorithm. We close in Section 5 with some conclusive remarks.

2 Precise Model and Results

We consider the following random broadcast algorithm also known as the push algorithm (cf. [8]). 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 $1, 2, \ldots$, each informed node chooses a neighbor independently and uniformly at random and transmits a copy of the information to that neighbor, which thus becomes informed. We are interested in the runtime of this algorithm, which is the time until every node in G gets informed; in 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 probability $1 - \mathcal{O}(n^{-1})$.

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

Definition 1. (cf. [12]) *Let $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ be points in Ω 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 $\{(x, y) \colon x, y \in \mathcal{X}_n, \|x - y\|_2 \leqslant r\}$.*

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

Definition 2. (cf. [12]) *Let N_n be a Poisson random variable with parameter n and let $\mathcal{P}_n = \{X_1, X_2, \ldots, X_{N_n}\}$ be points chosen independently and uniformly at random from Ω; i.e., $\mathcal{P}_n$ is a Poisson Point Process over Ω with intensity 1. The random geometric graph $\mathcal{G}(\mathcal{P}_n; r)$ has node set $\mathcal{P}_n$ and edge set $\{(x, y) \colon x, y \in \mathcal{P}_n, \|x - y\|_2 \leqslant r\}$.*

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

Lemma 2.1. *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 $1 - \mathcal{O}(\alpha\sqrt{n})$.*

[1] In [4] the so-called gossiping problem has been considered, i.e., each node possesses a different message, and all these messages have to be disseminated efficiently to every node in the graph. However, by solving the gossiping problem they also solved the broadcasting problem.

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 there exists a critical value r_c for the coverage radius such that if $r > r_c$, then with high probability the largest connected component of G has cardinality $\Omega(n)$ and all the other connected components have cardinality $\mathcal{O}(\log^2 n)$. On the contrary, if $r < r_c$, each connected component of G has $\mathcal{O}(\log n)$ nodes with probability $1 - o(1)$ [12]. The exact value of r_c is not known, though some bounds have been derived in [11]. In addition, if $r = \sqrt{\frac{\log n + \omega(1)}{\pi}}$, then G is connected with probability $1 - o(1)$.[2]

Our main result is stated in the next theorem. It shows that if $r > r_c$, then for all s inside the largest connected component of G, $\mathcal{R}(s, G) = \mathcal{O}(\sqrt{n}/r + \log n)$ with probability $1 - \mathcal{O}(n^{-1})$. 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 $n \to \infty$.

Theorem 2.2. *For a random geometric graph $G = \mathcal{G}(\mathcal{P}_n; r)$, if $r > r_c$, then $\mathcal{R}(s, G) = \mathcal{O}(\sqrt{n}/r + \log n)$ with probability $1 - \mathcal{O}(n^{-1})$ for all node s inside the largest connected component of G.*

The proof of Theorem 2.2, which we provide in Section 4, requires an upper bound for the length of the shortest path between nodes of G. Our result on this matter, which is stated in the next theorem, gives that for any two nodes that are sufficiently distant from each other in Ω, 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 $\mathcal{O}(\sqrt{n}/r)$, a result previously known only for $r = \sqrt{\frac{\log n + \omega(1)}{\pi}}$.

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 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 smallest path between two nodes v_1 and v_2 in G must satisfy $d_G(v_1, v_2) \geqslant \|v_1 - v_2\|_2/r$.

Theorem 2.3. *If $r > r_c$, for any two connected nodes v_1 and v_2 in $G = \mathcal{G}(\mathcal{P}_n; r)$ such that $\|v_1 - v_2\|_2 = \Omega(\log^{3.5} n/r^2)$, we obtain $d_G(v_1, v_2) = \mathcal{O}(\|v_1 - v_2\|_2/r)$ with probability $1 - \mathcal{O}(n^{-1})$.*

[2] At this point it is important to remark that all logarithms in this paper refer to the natural logarithm.

Corollary 2.4. *If $r > r_c$, the diameter of the largest connected component of $G = \mathcal{G}(\mathcal{P}_n; r)$ is $\mathcal{O}(\sqrt{n}/r)$ with probability $1 - \mathcal{O}(n^{-1})$.*

3 The Diameter of the Largest Connected Component

We devote this section to prove Theorem 2.3. We consider $G = \mathcal{G}(\mathcal{P}_n; r)$ with $r > r_c$ and assume that $r = \mathcal{O}(\sqrt{\log n})$. (When $r = \omega(\sqrt{\log n})$, G is connected with probability $1 - o(1)$ and Theorem 2.3 becomes a slightly different version of [5, Theorem 8].) We show that for any two connected nodes v_1 and v_2 of G such that $\|v_1 - v_2\|_2 = \Omega(\log^{3.5} n/r^2)$, we obtain $d_G(v_1, v_2) = \mathcal{O}(\|v_1 - v_2\|_2/r)$ with probability $1 - \mathcal{O}(n^{-1})$.

We first take two fixed nodes v_1 and v_2 satisfying the conditions above and show that $d_G(v_1, v_2) = \mathcal{O}(\|v_1 - v_2\|_2/r)$ with probability $1 - \mathcal{O}(n^{-3})$. Then, we would like to take the union bound over all pairs of nodes v_1 and v_2 to conclude the proof for Theorem 2.3, however the number of nodes in G is a random variable and hence the union bound cannot be employed directly. We resort to the following lemma to extend the result to all pairs of nodes v_1 and v_2.

Lemma 3.1. *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, $\mathbf{Pr}[\mathcal{E}(w_1, w_2)] \geqslant 1 - p$, with $p > 0$. Then,*

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

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

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

□

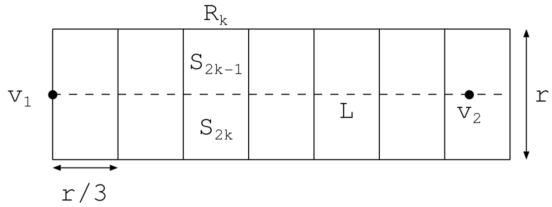

Figure 1: Illustration for the calculation of $d_G(v_1, v_2)$, with the large $r \times r/3$ rectangle R_k and the cells S_{2k-1} and S_{2k} contained in R_k.

We use Figure 1 as a reference to show how to find a path from v_1 to v_2. Take the line L that contains v_1 and v_2 and draw a sequence of adjacent rectangles starting from v_1 until we draw a rectangle that contains v_2. Each rectangle has two sides with length $r/3$ that are parallel to L and two other sides with length r that are perpendicular to L such that their middle point is contained in L. Let κ be the number of such rectangles and refer to them as $R_1, R_2, \ldots, R_\kappa$. For each $k \in [1, \kappa]$, L splits R_k into two identical, smaller rectangles which we denote by S_{2k-1} and S_{2k} and refer to as cells.

Note that for any k and two points $x \in S_k$ and $x' \in S_{k+2}$, we obtain $\|x - x'\|_2 \leqslant \sqrt{(2r/3)^2 + (r/2)^2} \leqslant r$, that is, nodes in S_k and S_{k+2} are neighbors in G. For this reason, we say that the cell S_k is adjacent to the cells S_{k-2} and S_{k+2}. Note that v_1 belongs to both S_1 and S_2. We would like to find a path from v_1 to v_2 that starts at either S_1 or S_2 and moves along adjacent cells, but some S_k may contain no node.

Our choice for the length of the largest sides of the rectangle R_k is intended to achieve the following property. For any path in G that crosses the region $\bigcup_{i=1}^{\kappa} R_i$, in the sense that there exists an edge of the path that intersects $\bigcup_{i=1}^{\kappa} R_i$, it must be the case that the path contains a node inside $\bigcup_{i=1}^{\kappa} R_i$. This property is crucial in our analysis, since it guarantees that a path crossing two rectangles R_j and R_k provides a path from a node in R_j to a node in R_k in G and can be used to move around cells that contain no nodes.

We refer to a cell as empty if it contains no node. For any empty cell S_k with S_{k-2} being nonempty, we follow the shortest path from a node in S_{k-2} to some nonempty $S_{k'}$ for $k' \geqslant k+1$. Note that there is always such a k' since $R_\kappa = S_{2\kappa-1} \cup S_{2\kappa}$ contains v_2. Our aim is to give a bound for the length of the detour around empty cells. The path starts at $v_1 \in R_1$. For $3 \leqslant k \leqslant 2\kappa$, if S_k is empty and S_{k-2} is not empty, let D_k be the length of the shortest path from S_{k-2} to some $S_{k'}$ for $k' \geqslant k+1$. If S_k is not empty, we set $D_k = 0$. Also, if S_k and S_{k-2} are both empty, then we also set $D_k = 0$, since the detour around S_{k-2} will either go around S_k as well or lead to S_{k-1}, from which we can obtain an edge to S_{k+1} or a detour that goes around S_k. With these definitions we can write $d_G(v_1, v_2) \leqslant \kappa + \sum_{k=3}^{2\kappa} D_k$.

In order to calculate D_k, we exploit the idea of crossings from continuum percolation. For an odd number $k \geqslant 1$, we consider the cells $S_{k-2}, S_{k-1}, S_k, S_{k+1}$. Let $Q_k(1)$ be the rectangle containing all these cells, that is, $Q_k(1) = R_{(k-1)/2} \cup R_{(k+1)/2}$. Let $Q_k(\gamma)$ be a rectangle having the same center as $Q_k(1)$ and whose sides are parallel to those of $Q_k(1)$ and have length given by γ times the side lengths of $Q_k(1)$ (in other words, $Q_k(\gamma)$ is a stretched version of $Q_k(1)$). Then, for any odd number $k \geqslant 1$ and $\gamma > 1$, we define the annulus $A(S_k, \gamma) = A(S_{k+1}, \gamma) = Q_k(\gamma) \setminus Q_k(1)$ (see Figure 2(a)).

An annulus $A(S_k, \gamma)$ can be decomposed into two horizontal rectangles ($Z_1Z_4Z_5Z_{12}$ and $Z_{11}Z_6Z_7Z_{10}$ in Figure 2(b)) and two vertical rectangles ($Z_1Z_2Z_9Z_{10}$ and $Z_3Z_4Z_7Z_8$ in Figure 2(b)). For a horizontal rectangle, we define a horizontal crossing as a path in G completely contained in the rectangle and that connects the left to the right side of the rectangle, i.e., with the first node of the path being within distance r to the left side of the rectangle and the last node of the path being within distance r to the right side of the rectangle. Similarly, for a vertical rectangle, we define a vertical crossing as a path in G that is completely contained in the rectangle and that connects the top to the bottom side of the rectangle. For an annulus $A(S_k, \gamma)$, we define $\mathcal{F}(A(S_k, \gamma))$ as the event that both horizontal rectangles of $A(S_k, \gamma)$ have a horizontal crossing and that both vertical rectangles of $A(S_k, \gamma)$ have a vertical crossing. This event is illustrated in Figure 2(c). Note that when $\mathcal{F}(A(S_k, \gamma))$ happens, then the aforementioned crossings provide a cycle around S_k.

We now explain how to use the annuli to find detours around an empty cell S_k. Note that S_1 and S_2 contain v_1 and, consequently, are not empty. Now suppose that S_{k-2} is not empty and is connected to v_1, i.e., there is a path from v_1 to a node inside S_{k-2}. If S_k is also not empty, then the node inside S_k is a neighbor of the node in S_{k-2}, and we obtain a path from v_1 to S_k. Now, assume that S_k is empty. We want to use the path from v_1 to S_{k-2} to construct a path from v_1 to some $S_{k'}$ with $k' \geqslant k+1$. Clearly, for any $\gamma > 1$, the annulus $A(S_k, \gamma)$ intersects neither S_{k-2} nor S_k, but does intersect S_{k+2}. Take γ' such that $\mathcal{F}(A(S_k, \gamma'))$ happens and let $H \subset Q_k(\gamma')$ be the largest region delimited by the cycle surrounding S_k that is induced by the crossings of $A(S_k, \gamma')$. If $v_1 \notin H$, then the path from v_1 to S_{k-2} provides a path from S_{k-2} to the crossings of $A(S_k, \gamma')$. If the crossings intersect some nonempty $S_{k'}$, $k' \geqslant k+1$, then there is a path

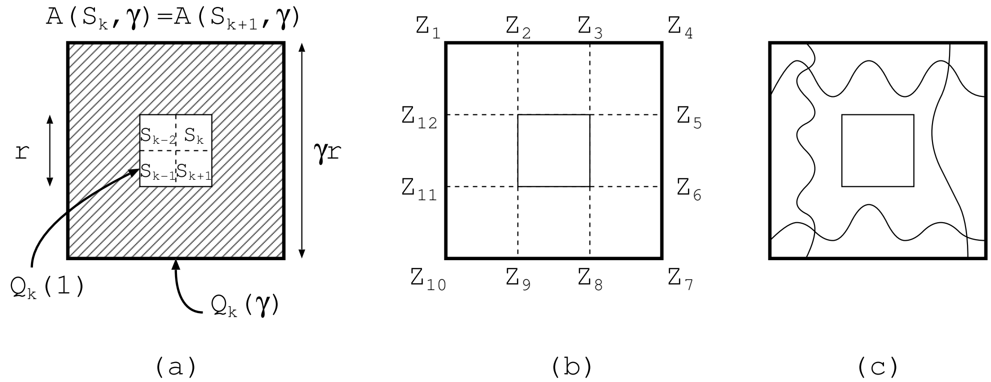

Figure 2: Illustration for the annulus $A(S_k, \gamma)$. Part (a) shows the annulus (highlighted region) and the cells S_{k-2}, S_{k-1}, S_k, and S_{k+1} in the middle. Part (b) shows the decomposition of $A(S_k, \gamma)$ into horizontal and vertical rectangles. And part (c) illustrates the event $\mathcal{F}(A(S_k, \gamma))$ for the left-to-right and top-to-bottom crossings (depicted as curvy lines) of $A(S_k, \gamma)$.

entirely contained in H from S_{k-2} to a node inside $S_{k'}$. If such a $S_{k'}$ does not exist, it must be the case that $v_2 \in H$. Since v_1 and v_2 are connected, there is a path from v_2 to the crossing of $A(S_k, \gamma')$, and, consequently, there is a path from S_{k-2} to v_2 completely contained in H. Now, if $v_1 \in H$ and $S_{k'}$ as above exists, then the path from v_1 to v_2 intersects the crossings of $A(S_k, \gamma')$ and can be used to obtain a path completely contained in H from S_{k-2} to $S_{k'}$. Finally, if $v_1 \in H$ and $v_2 \in H$, then there is a path from v_1 to v_2 entirely contained in H.

This shows that whenever v_1 and v_2 are connected, we can use the annuli to move from S_k to $S_{k'}$, $k' \geqslant k+1$, or to move directly to v_2. Note that the construction of S_k and $A(S_k, \gamma)$ are independent from v_1 and v_2 being connected and are taken for two arbitrarily fixed nodes v_1 and v_2. This means that our calculations to follow are not conditioned on v_1 and v_2 being connected. However, when v_1 and v_2 turn out to be connected, then this construction provides a path from v_1 to v_2.

Once we know that $\mathcal{F}(A(S_k, \gamma))$ occurs for some γ, we can easily bound D_k by the following straightforward geometric lemma.

Lemma 3.2. *Let Q be a rectangle with side lengths s and αs. Let w_1 and w_2 be two nodes of G contained in Q. If there exists a path between w_1 and w_2 entirely contained in Q, then $d_G(w_1, w_2) \leqslant 11\alpha s^2/r^2$.*

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, their distance is larger than r. Otherwise, we can 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 m be the number of nodes in the path. There are $m/2$ non-overlapping balls of radius $r/2$. For each ball, at least $1/4$ of its area is contained inside Q. Therefore, it must hold that

$$m \leqslant 2\frac{\text{Area}(Q)}{\pi(r/2)^2/4} = \frac{32\alpha s^2}{\pi r^2}. \qquad \square$$

The lemma below gives an upper bound for the probability that $A(S_k, \gamma)$ does not have the crossings.

Lemma 3.3. *There exist constants c and $\gamma_0 > 1$ such that for all $\gamma > \gamma_0$ and $1 \leqslant k \leqslant 2\kappa$, it holds that*

$$\mathbf{Pr}\left[\mathcal{F}(A(S_k, \gamma))\right] \geqslant 1 - \exp\left(-c\gamma r\right).$$

Proof. We build upon ideas from the proof [12, Lemma 10.5]. Recall the decomposition of $A(S_k, \gamma)$ into rectangles (refer to Figure 2(b)) and take the top rectangle $Z_1Z_4Z_5Z_{12}$. Its sides have lengths $(\gamma-1)r/2$ and $2\gamma r/3$. Therefore, the aspect ratio of the rectangle is $3(\gamma-1)/(4\gamma) \leqslant 3/4$, which increases with γ. We want to calculate the probability that such a rectangle has a horizontal crossing as γ increases. This is slightly different from the calculation in [12, Lemma 10.5], since there the aspect ratio is fixed and the side of the rectangle is allowed to vary. But clearly, for any rectangle with side lengths $(\gamma-1)r/2$ and $2\gamma r/3$, we can stretch the largest sides (while keeping the smallest sides unchanged) to make the aspect ratio be $3(\gamma_0-1)/(4\gamma_0)$, which we can then fix. Also, if there is a horizontal crossing in the stretched rectangle, there must be a horizontal crossing in the original one. Following along the lines of the proof [12, Lemma 10.5], we can then conclude that there are constants γ_0 and c such that for all $\gamma \geqslant \gamma_0$ a rectangle of side lengths $(\gamma-1)r/2$ and $2\gamma r/3$ has a horizontal crossing with probability larger than $1-e^{-c\gamma r}/4$. Applying the union bound over the 4 rectangles composing $A(S_k,\gamma)$ concludes the proof. $\square$

Now we use this lemma to bound the length of a detour. For any k, let Γ_k be the smallest value of $\gamma > \gamma_0$ for which $\mathcal{F}(A(S_k,\gamma))$ occurs. Suppose that S_k is empty and S_{k-2} is not empty. We want to obtain an upper bound for Γ_k. Note that once we know the value of Γ_k, we can apply Lemma 3.2 to conclude that $D_k \leqslant (22/3)\Gamma_k^2$. Since for each v_1 and v_2 there are at most $2\kappa = \mathcal{O}(\sqrt{n})$ cells, Lemma 3.3 gives that there is a constant c_1 such that with probability $1-\mathcal{O}(n^{-4})$ we obtain $D_k \leqslant c_1 \log^2 n/r^2$ for all k. Let $\mathcal{E}(v_1,v_2)$ be the event that $D_k \leqslant c_1 \log^2 n/r^2$ for a fixed pair of nodes v_1 and v_2, and all k. Thus, $\mathbf{Pr}\left[\mathcal{E}(v_1,v_2)\right] \geqslant 1-\mathcal{O}(n^{-4})$.

We want to apply Azuma's inequality to $\sum_{k=3}^{2\kappa} D_k$ under the condition that $\mathcal{E}(v_1,v_2)$ happens. Noting that $\mathbf{E}\left[D_k \mid \mathcal{E}(v_1,v_2)\right] \leqslant \mathbf{E}\left[D_k\right] / \mathbf{Pr}\left[\mathcal{E}(v_1,v_2)\right]$, we proceed to derive an upper bound for $\mathbf{E}\left[D_k\right]$. The probability that S_{k-2} is not empty and S_k is empty is $e^{-r^2/6}(1-e^{-r^2/6})$. Recall that $D_k \leqslant (22/3)\Gamma_k^2$. Therefore, $\mathbf{Pr}\left[D_k \geqslant \ell\right] \leqslant 1-\mathbf{Pr}\left[\mathcal{F}(A(S_k,\sqrt{(3/22)\ell}))\right] \leqslant \exp(-c\sqrt{(3/22)\ell}\, r)$. We can then write $\mathbf{E}\left[D_k\right] = e^{-r^2/6}(1-e^{-r^2/6})\sum_{\ell=1}^{\infty}\mathbf{Pr}\left[D_k \geqslant \ell\right] \leqslant e^{-r^2/6}\int_0^\infty \mathbf{Pr}\left[D_k \geqslant \ell\right] d\ell$, where the last inequality follows from $\mathbf{Pr}\left[D_k \geqslant \ell\right]$ being a non-increasing function of ℓ. Since we have an exponential upper bound for $\mathbf{Pr}\left[D_k \geqslant \ell\right]$ with $\ell \geqslant (22/3)\gamma_0^2$, we obtain

$$\begin{aligned}\mathbf{E}\left[D_k\right] \leqslant& e^{-r^2/6}(22/3)\gamma_0^2 \\ &+ e^{-r^2/6}\int_{\ell=(22/3)\gamma_0^2}^{\infty} e^{-c\sqrt{(3/22)\ell}\,r} d\ell \\ =&\mathcal{O}(1).\end{aligned}$$

Using the linearity property of expectations and $\mathbf{Pr}\left[\mathcal{E}(v_1,v_2)\right] \geqslant 1-\mathcal{O}(n^{-4})$, we obtain $\mathbf{E}\left[d_G(v_1,v_2) \mid \mathcal{E}(v_1,v_2)\right] \leqslant \mathbf{E}\left[d_G(v_1,v_2)\right] / \mathbf{Pr}\left[\mathcal{E}(v_1,v_2)\right] = \mathcal{O}(\kappa) = \mathcal{O}(\|v_1-v_2\|_2/r)$.

If the event $\mathcal{E}(v_1,v_2)$ holds, we have $\mathbf{E}\left[D_k \mid \mathcal{E}(v_1,v_2)\right] = \mathcal{O}(1)$ and $\Gamma_k \leqslant c_1' \log n/r$ for all k and some constant c_1', which yields $D_k \leqslant c_1 \log^2 n/r^2$. Letting $\lambda = 4c_1' \log n/r$, this implies that for two cells S_k and $S_{k'}$ such that $|k-k'| \geqslant \lambda$, the annuli $A(S_k,\lambda/4)$ and $A(S_{k'},\lambda/4)$ do not intersect, and consequently, the random variables D_k and $D_{k'}$ are independent. Now we split the random variables $D_1, D_2, \ldots, D_{2\kappa}$ into groups of independent random variables. Define the index set $I_j = \{k\colon 3 \leqslant k \leqslant 2\kappa, k \equiv j \pmod{\lambda}\}$. We can write $d_G(v_1,v_2) = \kappa + \sum_{j=0}^{\lambda-1}\sum_{k\in I_j} D_k$, where the second sum contains independent random variables. Note that $\mathbf{Pr}\left[\sum_{k\in I_j} D_k - \sum_{k\in I_j}\mathbf{E}\left[D_k\right] \geqslant c_2|I_j|\right]$ can be upper bounded by $1-\mathbf{Pr}\left[\mathcal{E}(v_1,v_2)\right] + \mathbf{Pr}\left[\sum_{k\in I_j} D_k - \sum_{k\in I_j}\mathbf{E}\left[D_k\right] \geqslant c_2|I_j| \mid \mathcal{E}(v_1,v_2)\right]$. In order to apply Azuma's inequality to the last term, we need to write $\sum_{k\in I_j}\mathbf{E}\left[D_k\right]$ in terms of $\sum_{k\in I_j}\mathbf{E}\left[D_k \mid \mathcal{E}(v_1,v_2)\right]$. Since $\mathbf{E}\left[D_k\right] \geqslant \mathbf{E}\left[D_k \mid \mathcal{E}(v_1,v_2)\right]\mathbf{Pr}\left[\mathcal{E}(v_1,v_2)\right] = \mathbf{E}\left[D_k \mid \mathcal{E}(v_1,v_2)\right] - \mathcal{O}(n^{-4})$, we derive that for each j,

$$\begin{aligned}&\mathbf{Pr}\left[\sum_{k\in I_j} D_k - \sum_{k\in I_j}\mathbf{E}\left[D_k\right] \geqslant c_2|I_j|\right] \\ &\leqslant 1-\mathbf{Pr}\left[\mathcal{E}(v_1,v_2)\right] \\ &\quad + 2\exp\left(-\frac{(c_2+\mathcal{O}(n^{-4}))^2|I_j|^2 r^4}{2c_1^2\log^4 n}\right).\end{aligned}$$

Since $|I_j| \geqslant \kappa/\lambda = \Omega(\|v_1-v_2\|_2/\log n)$, the probability above is smaller than $\mathcal{O}(n^{-4}) + \exp\left(-\frac{c_3\|v_1-v_2\|_2^2 r^4}{\log^6 n}\right)$, for some constant c_3. We solve the first sum by the union bound, obtaining

$$\begin{aligned}&\mathbf{Pr}\left[\sum_{k=1}^{2\kappa} D_k - \sum_{k=1}^{2\kappa}\mathbf{E}\left[D_k\right] \geqslant 2c_2\kappa\right] \\ &\leqslant \mathcal{O}(\lambda n^{-4}) + \lambda\exp\left(-\frac{c_3\|v_1-v_2\|_2^2 r^4}{\log^6 n}\right) \\ &= \mathcal{O}(n^{-3}),\end{aligned}$$

for any v_1 and v_2 such that $\|v_1-v_2\|_2 \geqslant c_4\log^{3.5} n/r^2$, for some constant c_4. Hence, by setting the constant c_2 properly, for a fixed pair of nodes v_1, v_2 such that $\|v_1-v_2\|_2 = \Omega(\log^{3.5} n/r^2)$, $d_G(v_1,v_2) = \mathcal{O}(\|v_1-v_2\|_2/r)$ with probability $1-\mathcal{O}(n^{-3})$. Applying Lemma 3.1 concludes the proof of Theorem 2.3.

4 Broadcast Time

In this section we prove Theorem 2.2. 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 $\|v_1-v_2\|_2 = \Omega(\log^4 n/r^2)$, $\mathcal{R}(v_1,v_2) = \mathcal{O}(\|v_1-v_2\|_2/r)$. Initially, we assume that $r = \mathcal{O}(\sqrt{\log n})$. The case $r = \omega(\sqrt{\log n})$ is simpler, but since it uses different proof techniques, we deal with it in Section 4.1.

We start our treatment for the case $r = \mathcal{O}(\sqrt{\log n})$ with an easy lemma that shows that the time until a node informs a given neighbor is $\mathcal{O}(\log^2 n)$ with high probability.

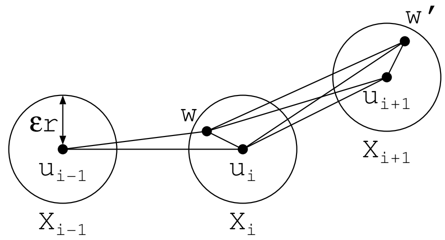

Figure 3: Illustration of the path considered to obtain $\mathcal{R}(v_1, v_2)$. The picture shows three consecutive nodes u_{i-1}, u_i, and u_{i+1} of the path from v_1 to v_2 and the balls X_{i-1}, X_i, and X_{i+1} around them. Two other nodes $w \in X_i$ and $w' \in X_{i+1}$ are depicted to illustrate the edges that arise from the construction of the X_i's.

LEMMA 4.1. *Let $r = \mathcal{O}(\sqrt{\log n})$. There exists a constant c such that for all pair of nodes w_1 and w_2 satisfying $\|w_1 - w_2\|_2 \leqslant r$, the following holds with probability $1 - \mathcal{O}(n^{-1})$,*

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

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_5 such that with probability $1 - \mathcal{O}(n^{-3})$ all nodes of a random geometric graph have degree smaller than $c_5 \log n$ [12] provided $r = \mathcal{O}(\sqrt{\log n})$. Therefore,

$$\begin{aligned}\mathbf{Pr}\left[\mathcal{R}(w_1, w_2) \geqslant t\right] &\leqslant \left(1 - \frac{1}{c_5 \log n}\right)^t \\ &\leqslant \exp\left(-\frac{t}{c_5 \log n}\right).\end{aligned}$$

If we set $t = 3c_5 \log^2 n$, we obtain that $\mathbf{Pr}\left[\mathcal{R}(w_1, w_2) \geqslant 3c_5 \log^2 n\right] \leqslant n^{-3}$ and, by Lemma 3.1 we conclude that $\mathcal{R}(w_1, w_2) \leqslant 3c_5 \log^2 n$ for all w_1, w_2 with probability $1 - \mathcal{O}(n^{-1})$. ☐

Before proceeding, note that the lemma above shows that $\mathcal{R}(v_1, v_2)$ can be upper bounded by $\mathcal{O}(d_G(v_1, v_2) \log^2 n)$. We 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 $r' = r(1 - 2\varepsilon)$. Since $r' > r_c$, $G' = \mathcal{G}(\mathcal{P}_n, r')$ contains a connected component of size $\Omega(n)$ with probability $1 - e^{-\Omega(\sqrt{n})}$. 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 random broadcast algorithm to transmit the information along this path, which gives a rather pessimistic upper bound, we enlarge the path using 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 $u_1, u_2, \ldots, u_m$ be a path from v_1 to v_2 in G', where $u_1 = v_1$ and $u_m = v_2$. 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_m to be the point where u_m is located; for $2 \leqslant i \leqslant m-1$, define X_i to be the ball with center at u_i and radius εr. Our goal is to get an upper bound for $\mathcal{R}(v_1, v_2)$ by following the path $X_1, X_2, \ldots, X_m$ (refer to Figure 3).

Define the random variable $T(X_i, X_{i+1})$, $1 \leqslant i \leqslant m-1$, 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. 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 $\|w - w'\|_2 \leqslant 2\varepsilon r + \|u_i - u_{i+1}\|_2 \leqslant r$. 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 can clearly obtain the following upper bound

$$\mathcal{R}(v_1, v_2) \leqslant \sum_{i=1}^{m-1} T(X_i, X_{i+1}).$$

Note that Lemma 4.1 gives $T(X_{m-1}, X_m) = \mathcal{O}(\log^2 n)$ 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 $1 \leqslant i \leqslant m-2$.

LEMMA 4.2. *For any $1 \leqslant i \leqslant m-2$, it holds that $\mathbf{E}\left[T(X_i, X_{i+1})\right] \leqslant 1/\varepsilon^2$.*

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 and let Y' be the number of nodes in X_{i+1}. Therefore, $\mathbf{E}\left[T(X_i, X_{i+1})\right] = \mathbf{E}\left[Y/Y'\right]$. We know that $Y \geqslant 1$ and $Y' \geqslant 1$, therefore, $Y-1$ and $Y'-1$ are Poisson random variables with mean πr^2 and $\pi\varepsilon^2 r^2$, respectively. Conditional on $Y-1 = k$, the value of $Y'-1$ is given by a Binomial distribution with mean

$k\frac{\pi\varepsilon^2 r^2}{\pi r^2} = k\varepsilon^2$. We obtain

$$\begin{aligned}
&\mathbf{E}\left[T(X_i, X_{i+1})\right] \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{k}\frac{k+1}{i+1}e^{-\pi r^2}\frac{(\pi r^2)^k}{k!}\binom{k}{i}(\varepsilon^2)^i(1-\varepsilon^2)^{k-i} \\
&= \frac{1}{\varepsilon^2}\sum_{k=0}^{\infty}e^{-\pi r^2}\frac{(\pi r^2)^k}{k!} \\
&\quad\cdot\sum_{i=0}^{k}\binom{k+1}{i+1}(\varepsilon^2)^{i+1}(1-\varepsilon^2)^{k-i} \\
&\leqslant \frac{1}{\varepsilon^2}\sum_{k=0}^{\infty}e^{-\pi r^2}\frac{(\pi r^2)^k}{k!} \\
&= \frac{1}{\varepsilon^2}.
\end{aligned}$$

□

For any two connected nodes v_1 and v_2 such that $\|v_1 - v_2\|_2 = \Omega(\log^4 n/r^2)$ (we come back to the case $\|v_1 - v_2\|_2 = o(\log^4 n/r^2)$ at the end of this section), we know that there is a path like the ones derived in Section 3 for the proof of Theorem 2.3. In particular, we know that there is a path $v_1 = u_1, u_2, \ldots, u_{m-1}, u_m = v_2$ such that $m = \mathcal{O}(\sqrt{n}/r')$ and, provided $\mathcal{E}(v_1, v_2)$ holds, the annuli $A(S_k, \Gamma_k)$ and $A(S_{k'}, \Gamma_{k'})$ are disjoint if $|k - k'| \geqslant \lambda$. Recall that the cells $S_1, S_2, \ldots, S_{2\kappa}$ have side lengths $r/2$ and $r/3$, therefore, we need to take 6 adjacent cells together to obtain a rectangle with largest side length $2r$. Recall also that only every other cell is adjacent. Then, for k and k' such that $|k-k'| \geqslant \lambda+12$, the distance between any point in $A(S_k, \Gamma_k)$ and any point in $A(S_{k'}, \Gamma_{k'})$ is at least $2r$. Each annulus has at most $c_1 \log^2 n/r^2$ nodes in the path, therefore, letting $\lambda' = (c_1 \log^2 n/r^2)(\lambda + 12) = \mathcal{O}(\log^3 n/r^3)$, we obtain that for any two nodes u_i and u_j in the path such that $|i - j| \geqslant \lambda'$, $\|u_i - u_j\|_2 \geqslant 2r$ and, consequently, $T(X_i, X_{i+1})$ and $T(X_j, X_{j+1})$ are independent.

It is important to remark that the path has length $m = \mathcal{O}(\sqrt{n}/r')$, for all v_1 and v_2. Conditional on the existence of this particular path, the Poisson point process over $\Omega \setminus \bigcup_{i=1}^{m}\{u_i\}$, where the union is over the points where the nodes of the path are located, remains unchanged since $\bigcup_{i=1}^{m}\{u_i\}$ spans a set of measure 0 in Ω.

Let the index set $J_j = \{1 \leqslant i \leqslant m\colon i \equiv j \pmod{\lambda'}\}$. We can write $\mathcal{R}(v_1, v_2) = \mathcal{O}(\log^2 n) + \sum_{j=0}^{\lambda'-1}\sum_{i\in J_j} T(X_i, X_{i+1})$, where the first term comes from the time it takes for the random broadcast algorithm to inform v_2 once any neighbor of v_2 is informed. For all j, the term $\sum_{i\in J_j} T(X_i, X_{i+1})$ is given by the sum of independent geometric random variables. We apply the following Chernoff bound for Geometric random variables.

LEMMA 4.3. *Let $X_1, \ldots, X_n$ be independent geometric random variables, each having parameter p (and thus mean $1/p$), and let $X = \sum_{i=1}^{n} X_i$. Then, for any $\varepsilon > 0$,*

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

Using Lemma 4.3, we obtain that for each j,

$$\mathbf{Pr}\left[\sum_{i\in J_j} T(X_i, X_{i+1}) \geqslant (1+\varepsilon)\sum_{i\in J_j}\mathbf{E}\left[T(X_i, X_{i+1})\right]\right] \leqslant \exp\left(-\varepsilon^2\frac{|J_j|}{2(1+\varepsilon)}\right).$$

Note that $|J_j| = \Omega(d_G(v_1, v_2)r^3/\log^3 n) = \Omega(\log n)$, since $d_G(v_1, v_2) \geqslant \|v_1 - v_2\|_2/r = \Omega(\log^4 n/r^3)$. Using the fact that $\mathbf{E}\left[T(X_i, X_{i+1})\right] = \mathcal{O}(1)$ for all i and taking the union bound over all j allows us to conclude that for all pairs of connected nodes v_1 and v_2 such that $\|v_1 - v_2\|_2 = \Omega(\log^4 n/r^2)$, there is a constant c_6 for which

$$\mathbf{Pr}\left[\sum_{j=0}^{\lambda'-1}\sum_{i\in J_j} T(X_i, X_{i+1}) \geqslant c_6(m-2)\right] \leqslant n^{-3}.$$

Applying Lemma 3.1, we can conclude that for any two nodes v_1 and v_2 in the largest connected component of G for which $\|v_1 - v_2\|_2 = \Omega(\log^4 n/r^2)$, 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 $\|v_1 - v_2\|_2 = \Theta(\sqrt{n})$ and, consequently, $\mathcal{R}(v_1, v_2) = \Theta(\sqrt{n}/r)$.

Now we treat 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 a known fact from random geometric graphs [12, Theorem 10.18] that the second largest component of G' contains $\mathcal{O}(\log^2 n)$ nodes with probability $1 - \mathcal{O}(n^{-1})$. 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 $\mathcal{O}(\log^4 n)$, which is negligible in comparison to $\Theta(\sqrt{n}/r)$. The second case corresponds to the nodes that are within distance $o(\log^4 n/r^2)$ 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 $c_7 \log^4/r^2$, for some constant c_7 (the orientation of Q does not matter). Note that Q contains all nodes within distance $o(\log^4 n/r^2)$ 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 Lemma 3.3 can be used to show that $\mathcal{F}(Q' \setminus Q)$ holds with probability

$1 - e^{-\Omega(\log^4 n/r^2)}$. Thus, all nodes within distance $o(\log^4 n/r^2)$ are contained inside the crossings of $Q' \setminus Q$ and their distance to v_1 in G must be smaller than $44c_7^2 \log^8 n/r^6$ by Lemma 3.2. So using Lemma 4.1 we conclude that all nodes within distance $o(\log^4 n/r^2)$ to v_1 are informed after $\mathcal{O}(\log^{10} n/r^6)$ rounds, which is also negligible in comparison to $\Theta(\sqrt{n}/r)$.

4.1 Case $r = \omega(\sqrt{\log n})$. In this section we prove the following lemma, which deals with the case $r = \omega(\sqrt{\log n})$.

LEMMA 4.4. *If $r = \omega(\sqrt{\log n})$, then for all node $s \in G$, we obtain $\mathcal{R}(s, G) = \mathcal{O}(\sqrt{n}/r + \log n)$ with probability $1 - \mathcal{O}(n^{-1})$.*

REMARK 1. *We point out that Lemma 4.4 can be generalized to RGGs in higher dimensions. For dimension $d \geqslant 2$, the lemma holds with $\Omega = [0, n^{1/d}]^d$ and $r = \omega(\log^{1/d} n)$ as long as d is a constant independent from n.*

In order to prove Lemma 4.4, we consider a tessellation of Ω into squares of side-length $\min\{r/3, \sqrt{n}/2\}$, which we refer to as cells. (If $\sqrt{n}$ is not a multiple of $r/3$, then we make the cells in the last row or column of the tessellation be smaller than the others.) It is very 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 49 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 $r = \omega(\sqrt{\log n})$, a standard Chernoff bound for Poisson random variables can be used to show that there are constants $c_1 < c_2$ such that a fixed cell contains at least $c_1 r^2$ nodes and at most $c_2 r^2$ nodes with probability larger than $1 - n^{-2}$. Using the union bound over the cells of the tessellation, we obtain that $a_{\min}$ and $a_{\max}$ are $\Theta(r^2)$ with probability $1 - \mathcal{O}(n^{-1})$.

We are now in position to start our proof for Lemma 4.4. We index the cells by $i \in \mathbb{Z}^2$ 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 share an edge. Therefore, each cell has exactly 4 adjacent cells and this adjacency relation induces a 4-regular graph C over the cells.

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 $a_{\min}/(49a_{\max}) = \Theta(1)$. 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 upper bounded by the sum of independent Geometric random variables with mean $\Theta(1)$. Applying Lemma 4.3, 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 $\mathcal{O}(n/r^2)$ cells and $\text{diam}(C) = \mathcal{O}(\sqrt{n}/r)$, we obtain that with probability $1 - \mathcal{O}(n^{-1})$, $Z_i = 1$ for all i after $\mathcal{O}(\sqrt{n}/r + \log n)$ rounds.

Now, 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 from 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 4.5. ([7, 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 obtain how many additional rounds are required until all nodes in G become informed. Note that each cell constitutes a clique with $\Theta(r^2)$ nodes. According to the random broadcast algorithm, at any round, a node chooses a neighbor inside its own cell with probability larger than $a_{\min}/(49a_{\max}) = \Theta(1)$. Therefore, a standard coupling argument can be used to show that the time until all nodes from a given cell get informed can be upper bounded 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 $\Theta(r^2)$ nodes. Thus, from [8, Theorem 4.1] and Lemma 4.5, we obtain that all nodes of a given cell get informed within $\mathcal{O}(\log r^2)$ steps with probability $1 - \mathcal{O}(r^{-2})$.

Now we need to extend this result to all cells. For each cell i, let W_i be an independent Geometric random variable with parameter ρ (and thus mean $1/\rho$), where we assume $\rho = 1 - \mathcal{O}(r^{-2})$. Therefore, once $Z_i = 1$ for all cell i, then the time it takes until all nodes get informed can be upper bounded by $\mathcal{O}(\log r^2) \max_i W_i$, where the maximum is taken over all cells. Since we have $\Theta(n/r^2)$ cells, we obtain that all W_i's are smaller than $c \log(n/r^2)$ for some constant c with probability

$(1-(1-\rho)^{c\log(n/r^2)})^{\Theta(n/r^2)} \geqslant 1-\mathcal{O}(n^{-1})$ for a proper choice of c. Therefore, we obtain that $\mathcal{R}(s,G) \leqslant \mathcal{O}(\operatorname{diam}(C)+\log n+\log(r^2)\log(n/r^2)) = \mathcal{O}(\sqrt{n}/r+\log n)$, which concludes the proof of Lemma 4.4.

5 Conclusion

We have analyzed the performance of the random broadcast algorithm in random geometric graphs. We proved that with probability $1-\mathcal{O}(n^{-1})$ the algorithm finishes within $\mathcal{O}(\sqrt{n}/r)$ 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 $\|v_1-v_2\|_2 = \Omega(\log^3 n/r^2)$, 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 $\mathcal{O}(\sqrt{n}/r)$.

A challenging open problem is to extend our results to random geometric graphs in higher dimensions, since our proof takes advantage of some restrictions imposed by the geometry in two dimensions. In another direction, our techniques may be useful to analyze other problems like the cover time of the largest connected component of RGGs. This would nicely complement recent results by Cooper and Frieze [2, 3] for connected RGGs and for the largest connected component of Erdős-Rényi random graphs.

Acknowledgments

Milan Bradonjić would thank Los Alamos National Laboratory. Part of this work was funded through the Laboratory Directed Research and Development Program, and Center for Nonlinear Studies. Part of this work was done while Tobias Friedrich and Thomas Sauerwald were postdoctoral fellows at ICSI Berkeley supported by the German Academic Exchange Service (DAAD). Alexandre Stauffer acknowledges partial support from CAPES/Fulbright.

References

[1] 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.

[2] C. Cooper and A. Frieze. The cover time of the giant component of a random graph. *Random Structures & Algorithms*, 32(4):401–439, 2008.

[3] C. Cooper and A. Frieze. The cover time of random geometric graphs. In *19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'09)*, pages 48–57, 2009.

[4] A. Czumaj and X. Wang. Fast message dissemination in random geometric ad-hoc radio networks. In *18th International Symposium on Algorithms and Computation (ISAAC'07)*, pages 220–231, 2007.

[5] R. B. Ellis, J. L. Martin, and C. Yan. Random geometric graph diameter in the unit ball. *Algorithmica*, 47(4):421–438, 2007.

[6] R. Elsässer and T. Sauerwald. Broadcasting vs. Mixing and Information Dissemination on Cayley Graphs. In *24th International Symposium on Theoretical Aspects of Computer Science (STACS'07)*, pages 163–174, 2007.

[7] R. Elsässer and T. Sauerwald. On the runtime and robustness of randomized broadcasting. *Theoretical Computer Science*, 410(36):3414–3427, 2009.

[8] U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Randomized Broadcast in Networks. *Random Structures and Algorithms*, 1(4):447–460, 1990.

[9] A. Frieze and G. Grimmett. The shortest-path problem for graphs with random-arc-lengths. *Discrete Applied Mathematics*, 10:57–77, 1985.

[10] Z. Lotker and A. Navarra. Managing random sensor networks by means of grid emulation. In *Networking*, pages 856–867, 2006.

[11] R. Meester and R. Roy. *Continuum percolation.* Cambridge University Press, 1996.

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

[13] B. Pittel. On spreading rumor. *SIAM Journal on Applied Mathematics*, 47(1):213–223, 1987.

[14] G. J. Pottie and W. J. Kaiser. Wireless integrated network sensors. *Commun. ACM*, 43(5):51–58, 2000.