Adv. Appl. Prob. **48,** 848–864 (2016)
*doi:*10.1017/apr.2016.31

ON THE RELATION BETWEEN GRAPH DISTANCE AND EUCLIDEAN DISTANCE IN RANDOM GEOMETRIC GRAPHS

J. DÍAZ,* *Universitat Politècnica de Catalunya and BGSMath*
D. MITSCHE,** *Université Nice Sophia Antipolis*
G. PERARNAU,*** *Universitat Politècnica de Catalunya*
X. PÉREZ-GIMÉNEZ,**** *University of Waterloo*

Abstract

Given any two vertices u, v of a random geometric graph $\mathcal{G}(n, r)$, denote by $d_{\mathrm{E}}(u, v)$ their Euclidean distance and by $d_{\mathrm{G}}(u, v)$ their graph distance. The problem of finding upper bounds on $d_{\mathrm{G}}(u, v)$ conditional on $d_{\mathrm{E}}(u, v)$ that hold asymptotically almost surely has received quite a bit of attention in the literature. In this paper we improve the known upper bounds for values of $r = \omega(\sqrt{\log n})$ (that is, for r above the connectivity threshold). Our result also improves the best known estimates on the diameter of random geometric graphs. We also provide a lower bound on $d_{\mathrm{G}}(u, v)$ conditional on $d_{\mathrm{E}}(u, v)$.

Keywords: Random geometric graph; graph distance; Euclidean distance; diameter

2010 Mathematics Subject Classification: Primary 05C80
Secondary 68R10

1. Introduction

Given a positive integer n and a nonnegative real function $r = r(n)$, a *random geometric graph* G on n vertices and *radius* r is defined as follows. The vertex set $V = V(G)$ is obtained by choosing n points independently and uniformly at random in the square $\mathcal{S}_n = [-\sqrt{n}/2, \sqrt{n}/2]^2$ (note that, with probability 1, no point in $\mathcal{S}_n$ is chosen more than once, and thus we assume that $|V| = n$). For notational purposes, we identify each vertex $v \in V$ with its corresponding geometric position $v = (x_v, y_v) \in \mathcal{S}_n$, where x_v and y_v denote the usual x- and y-coordinates in $\mathcal{S}_n$. For every two points $u, v \in \mathcal{S}_n$, we write $d_{\mathrm{E}}(u, v)$ for their Euclidean distance. Finally, the edge set $E = E(G)$ is constructed by connecting each pair of vertices $u, v \in V$ by an edge if and only if $d_{\mathrm{E}}(u, v) \leq r$. We denote this model of random geometric graphs by $\mathcal{G}(n, r)$, and use the notation $G \in \mathcal{G}(n, r)$ (or often simply $\mathcal{G}(n, r)$) to refer to a random outcome of this distribution. We will always assume that $r \leq \sqrt{2n}$, as for $r \geq \sqrt{2n}$ the graph obtained is always a clique.

Random geometric graphs were first introduced in a slightly different setting by Gilbert [3] in order to model the communications between radio stations. Since then, several closely related

Received 22 July 2014; revision received 5 September 2015.
* Postal address: Department of Computer Science, Universitat Politècnica de Catalunya, Jordi Girona Salgado 1, 08034 Barcelona, Spain. Email address: diaz@lsi.upc.edu
** Postal address: Laboratoire J. A. Dieudonné, Université Nice Sophia Antipolis, Parc Valrose, 06108 Nice cedex 02, France. Email address: dmitsche@unice.fr
*** Postal address: Department de Matemàtica Aplicada IV, Universitat Politècnica de Catalunya, Jordi Girona Salgado 1, 08034 Barcelona, Spain. Email address: guillem.perarnau@ma4.upc.edu
**** Postal address: Department of Combinatorics and Optimization, University of Waterloo, Waterloo ON, N2L 3G1, Canada. Email address: xperez@uwaterloo.ca

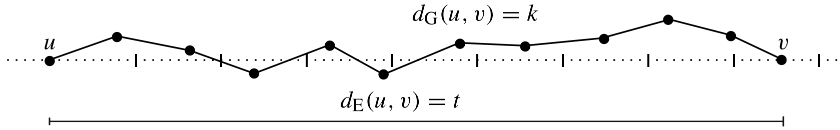

FIGURE 1: Graph distance versus Euclidean distance between two points u and v in V.

variants of these graphs have been widely used as a model for wireless communication, and have also been extensively studied from a mathematical point of view. The basic reference on random geometric graphs is Penrose [9] (see [10] for a more recent survey).

The properties of $\mathcal{G}(n, r)$ are usually investigated from an asymptotic perspective, as n grows to ∞ and $r = r(n)$. Throughout the paper we use the following standard notation for the asymptotic behavior of sequences of nonnegative numbers a_n and b_n: $a_n = O(b_n)$ if $\limsup_{n\to\infty} a_n/b_n \le C < +\infty$; $a_n = \Omega(b_n)$ if $b_n = O(a_n)$; $a_n = \Theta(b_n)$ if $a_n = O(b_n)$ and $a_n = \Omega(b_n)$; $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$; and $a_n = \omega(b_n)$ if $b_n = o(a_n)$. We also use $a_n \ll b_n$ and $b_n \gg a_n$ to denote $a_n = o(b_n)$. Finally, a sequence of events H_n holds *asymptotically almost surely* (a.a.s.) if $\lim_{n\to\infty} \mathbb{P}(H_n) = 1$.

It is well known that $r_c = \sqrt{\log n/\pi}$ is a sharp threshold function for the connectivity of a random geometric graph (see, e.g. [5] and [8]). This means that for every $\varepsilon > 0$, if $r \le (1-\varepsilon)r_c$ then $\mathcal{G}(n, r)$ is a.a.s. disconnected, whilst if $r \ge (1+\varepsilon)r_c$ then it is a.a.s. connected.

Given a connected graph G, we define the *graph distance* between two vertices u and v, denoted by $d_G(u, v)$, as the number of edges on a shortest path from u to v. Observe first that any pair of vertices u and v must satisfy $d_G(u, v) \ge d_E(u, v)/r$ deterministically by the triangle inequality, since each edge of a geometric graph has length at most r. The goal of this paper is to provide upper and lower bounds that hold a.a.s. for the graph distance of two vertices in terms of their Euclidean distance and in terms of r (see Figure 1).

Related work. This particular problem has given rise to quite a bit of interest in recent years. Given any two vertices $u, v \in V$, most of the work related to this problem has been devoted to studying the upper bounds on $d_G(u, v)$ in terms of $d_E(u, v)$ and r, that hold a.a.s. Ellis *et al.* [2] showed that there exists some large constant K such that for every $r \ge (1+\varepsilon)r_c$, $G \in \mathcal{G}(n, r)$ satisfies a.a.s. the following property: for every $u, v \in V$ such that $d_E(u, v) > r$,

$$d_G(u, v) \le K\frac{d_E(u, v)}{r}. \tag{1.1}$$

Their result is stated in the unit ball random geometric graph model, but it can be easily adapted into our setting. This result was extended by Bradonjić *et al.* [1] for the range of r for which $\mathcal{G}(n, r)$ has a giant component a.a.s., under the extra condition that $d_E(u, v) = \Omega(\log^{7/2} n/r^2)$. Friedrich *et al.* [4] improved this last result by showing that the result holds a.a.s. for every u and v satisfying $d_E(u, v) = \omega(\log n/r)$. They also proved that if $r = o(r_c)$, a linear upper bound of $d_G(u, v)$ in terms of $d_E(u, v)/r$ is no longer possible. In particular, a.a.s. there exist vertices u and v with $d_E(u, v) \le 3r$ and $d_G(u, v) = \Omega(\log n/r^2)$.

The motivation for the study of this problem stems from the fact that these results provide upper bounds for the diameter of $G \in \mathcal{G}(n, r)$, denoted by $\operatorname{diam}(G)$, that hold a.a.s., and the runtime complexity of many algorithms can often be bounded from above in terms of the diameter of G. For a concrete example, we refer the reader to the problem of broadcasting information (see [1] and [4]).

https://doi.org/10.1017/apr.2016.31 Published online by Cambridge University Press

One of the important achievements of our paper is to show that one can take the constant K for which (1.1) holds as $K = 1 + o(1)$ a.a.s., provided that $r = \omega(r_c)$. By the aforementioned result in [4], we know that the statement does not hold if $r = o(r_c)$.

A similar problem has been studied by Muthukrishnan and Pandurangan [7]. They proposed a new technique to study several problems on random geometric graphs—the so-called *bin-covering* technique—which tries to cover the endpoints of a path by bins. They considered, among others, the problem of determining $D_{\mathrm{G}}(u, v)$, which is the length of the shortest Euclidean path connecting u and v. Recently, Mehrabian and Wormald [6] studied a similar problem to the one in [7]. They deployed n points uniformly in $[0, 1]^2$, and connected any pair of points with probability $p = p(n)$, independently of their distance. In this model, they determined the ratio of $D_{\mathrm{G}}(u, v)$ and $d_{\mathrm{E}}(u, v)$ as a function of p.

The following theorem is the main result of our paper.

Theorem 1.1. *Let $G \in \mathcal{G}(n, r)$ be a random geometric graph on n vertices and radius* $0 < r \leq \sqrt{2n}$. *For every pair of vertices $u, v \in V(G)$ with $d_{\mathrm{E}}(u, v) > r$ (as otherwise the statement is trivial) a.a.s., we have*

(i) *if* $d_{\mathrm{E}}(u, v) \geq \max\{12(\log n)^{3/2}/r,\ 21 r \log n\}$ *then*

$$d_{\mathrm{G}}(u, v) \geq \left\lfloor \frac{d_{\mathrm{E}}(u, v)}{r}\left(1 + \frac{1}{2(r d_{\mathrm{E}}(u, v))^{2/3}}\right) \right\rfloor;$$

(ii) *if* $r \geq 224\sqrt{\log n}$ *then*

$$d_{\mathrm{G}}(u, v) \leq \left\lceil \frac{d_{\mathrm{E}}(u, v)}{r}(1 + \gamma r^{-4/3}) \right\rceil,$$

where

$$\gamma = \gamma(u, v) = \max\left\{1358\left(\frac{3r \log n}{r + d_{\mathrm{E}}(u, v)}\right)^{2/3}, \frac{4 \times 10^6 \log^2 n}{r^{8/3}}, 30\,000^{2/3}\right\}.$$

Proof. (i) We first observe that all the short paths between two points must lie in a certain rectangle. Then we show that, by restricting the construction of the path on that rectangle, no very short path exists.

(ii) We proceed similarly. We restrict our problem to finding a path contained in a narrow strip. In this case, we show that a relatively short path can be constructed. We believe that the ideas in the proof can be easily extended to show the analogous result for d-dimensional random geometric graphs for all fixed $d \geq 2$. □

Remark 1.1. (i) Note that the condition $d_{\mathrm{E}}(u, v) \geq \max\{12(\log n)^{3/2}/r, 21 r \log n\}$ in the lower bound of Theorem 1.1(i) can be replaced by $d_{\mathrm{E}}(u, v) \geq 21 r \log n$ if $r \geq \sqrt{4/7}(\log n)^{1/4}$, and by $d_{\mathrm{E}}(u, v) \geq 12(\log n)^{3/2}/r$ if $r \leq \sqrt{4/7}(\log n)^{1/4}$. We do not know whether this condition can be made less restrictive, besides improving the multiplicative constants involved (which we did not attempt to optimize).

(ii) Similarly, the constant 224 in the condition $r \geq 224\sqrt{\log n}$ of Theorem 1.1(ii) (as well as those in the definition of γ) is not optimized either, and could be made slightly smaller. However, our method as is cannot be extended all the way down to $r \geq \sqrt{\log n/\pi} = r_c$.

(iii) Moreover, the error term in Theorem 1.1(i) is $(2(r d_E(u, v))^{2/3})^{-1} = O(1/\log n) = o(1)$.

(iv) Finally, the error term in Theorem 1.1(ii) is

$$\gamma r^{-4/3} = \Theta\biggl(\max\biggl\{\biggl(\frac{\log n}{r^2 + r d_{\mathrm{E}}(u,v)}\biggr)^{2/3}, \biggl(\frac{\sqrt{\log n}}{r}\biggr)^4, r^{-4/3}\biggr\}\biggr),$$

which is $o(1)$ if and only if $r = \omega(\sqrt{\log n}) = \omega(r_c)$. Hence, for $r = \omega(r_c)$, Theorem 1.1(ii) implies that a.a.s.

$$d_{\mathrm{G}}(u,v) \le \biggl\lceil (1 + o(1))\frac{d_{\mathrm{E}}(u,v)}{r} \biggr\rceil,$$

thus improving the result in [2].

Theorem 1.1 gives an upper bound on the diameter as a corollary. First, observe that $d_{\mathrm{E}}(u,v) \le \sqrt{2n}$. From [2, Theorem 10] for the particular case $d = 2$, one can deduce that if $r \ge (1+\varepsilon) r_c$ a.a.s.

$$\mathrm{diam}(G) \le \frac{\sqrt{2n}}{r}\biggl(1 + O\biggl(\sqrt{\frac{\log\log n}{\log n}}\biggr)\biggr). \tag{1.2}$$

Directly from Theorem 1.1, we have, for $r \ge 224\sqrt{\log n}$,

$$\mathrm{diam}(G) \le \biggl\lceil \frac{\sqrt{2n}}{r}(1 + \widehat{\gamma} r^{-4/3}) \biggr\rceil, \tag{1.3}$$

where

$$\widehat{\gamma} = \Theta\biggl(\biggl(\frac{r \log n}{\sqrt{n}}\biggr)^{2/3} + \frac{\log^2 n}{r^{8/3}} + 1\biggr).$$

(In fact, (1.3) holds for all $r \ge (1+\varepsilon) r_c$ as a consequence of (1.2).) From straightforward computations, one can check that (1.3) improves (1.2) provided that $r = \Omega(\log^{5/8} n/(\log\log n)^{1/8})$.

On the other hand, for the lower bound on the diameter, observe that for any function ω growing arbitrarily slowly with n, we can a.a.s. find two vertices u and v, each at distance at most ω from one corner (opposite from each other) of the square $\mathcal{S}_n$. For two such vertices, we trivially (and deterministically) have

$$\mathrm{diam}(G) \ge d_{\mathrm{G}}(u,v) \ge \biggl\lceil \frac{\sqrt{2n}(1 - \Theta(\omega/\sqrt{n}))}{r} \biggr\rceil. \tag{1.4}$$

Assuming that $\sqrt{\log^3 n/n} \ll r \ll \sqrt{n}/\log n$, applying the bound from Theorem 1.1 to these vertices, we have a.a.s.

$$\begin{aligned} \mathrm{diam}(G) &\ge d_{\mathrm{G}}(u,v) &(1.5)\\ &\ge \biggl\lfloor \frac{\sqrt{2n}(1 - \Theta(\omega/\sqrt{n}))}{r}(1 + \Theta(r^{-2/3} n^{-1/3})) \biggr\rfloor \\ &\ge \biggl\lceil \frac{\sqrt{2n}(1 - \Theta(\omega/\sqrt{n}) + \Theta(r^{-2/3} n^{-1/3}))}{r} - 1 \biggr\rceil. &(1.6) \end{aligned}$$

Assuming the additional constraint $r \ll n^{1/10}$, we have $r^{-2/3} n^{-1/3} \gg r/\sqrt{n}$, and also $r^{-2/3} n^{-1/3} \gg \omega/\sqrt{n}$ (for ω tending to ∞ sufficiently slowly). In this case, our bound in (1.5) improves upon the trivial lower bound (1.4), and can be written as

$$\mathrm{diam}(G) \ge \biggl\lfloor \frac{\sqrt{2n}}{r}(1 + \Theta(r^{-2/3} n^{-1/3})) \biggr\rfloor. \tag{1.7}$$

Note that this is a.a.s. still valid if we drop the constraint $r \gg \sqrt{\log^3 n/n}$, since for $r = O(\sqrt{\log^3 n/n})$ the random geometric graph is a.a.s. disconnected (and has infinite diameter). Hence, by (1.3) and (1.7), we obtain the following corollary.

Corollary 1.1. *Let $G \in \mathscr{G}(n, r)$ be a random geometric graph on n vertices and radius $0 < r \le \sqrt{2n}$. We have a.a.s.*

(i) *if $r \ge (1+\varepsilon)r_c$ then*

$$\operatorname{diam}(G) \le \left\lceil \frac{\sqrt{2n}}{r}(1+\widehat{\gamma} r^{-4/3}) \right\rceil,$$

where

$$\widehat{\gamma} = \Theta\left(\left(\frac{r\log n}{\sqrt{n}}\right)^{2/3} + \frac{\log^2 n}{r^{8/3}} + 1\right);$$

(ii) *if $r \ll n^{1/10}$ then*

$$\operatorname{diam}(G) \ge \left\lfloor \frac{\sqrt{2n}}{r}(1+\Theta(r^{-2/3}n^{-1/3})) \right\rfloor.$$

2. Proof of Theorem 1.1

In order to simplify the proof of Theorem 1.1 we will make use of a technique known as de-Poissonization, which has many applications in geometric probability (see, e.g. [9] for a detailed account of the subject). Here we sketch it.

Consider the following related model of a random geometric graph G with two distinguished vertices u, v. The vertex set of G is $V = V(G) = \{u, v\} \cup V'$, where the position of u and v is selected independently and uniformly at random in $\mathcal{S}_n$, and where V' is a set obtained from a homogeneous Poisson point process of intensity 1 in the square $\mathcal{S}_n$ of area n. Observe that V' consists of N points in the square $\mathcal{S}_n$ chosen independently and uniformly at random, where N is a Poisson random variable of mean n. Exactly as we did for the model $\mathscr{G}(n, r)$, we connect $u_1, u_2 \in V$ by an edge if and only if $d_{\mathrm{E}}(u_1, u_2) \le r$. We denote this new model by $\widetilde{\mathscr{G}}_{u,v}(n, r)$.

The main advantage of defining $V' = V \setminus \{u, v\}$ as a Poisson point process is motivated by the following two properties: the number of points of V' that lie in any region $A \subseteq \mathcal{S}_n$ of area a has a Poisson distribution with mean a; and the number of points of V' in disjoint regions of $\mathcal{S}_n$ are independently distributed. Moreover, conditional on $N = n-2$, the distribution of $\widetilde{\mathscr{G}}_{u,v}(n, r)$ is the one of $\mathscr{G}(n, r)$. Therefore, since $\mathbb{P}(N = n-2) = \Theta(1/\sqrt{n})$, any event holding in $\widetilde{\mathscr{G}}_{u,v}(n, r)$ with probability at least $1 - o(f_n)$ must hold in $\mathscr{G}(n, r)$ with probability at least $1 - o(f_n\sqrt{n})$. We make use of this property throughout this paper, and perform all the analysis for a graph $G \in \widetilde{\mathscr{G}}_{u,v}(n, r)$ or related models of Poisson point processes.

We will need the following concentration inequality for the sum of independently and identically distributed exponential random variables. For the sake of completeness we provide the proof here.

Lemma 2.1. *Let $X_1, \ldots, X_m$ be independent exponential random variables of parameter $\lambda > 0$ (that is, expectation $1/\lambda$) and let $X = X_1 + \cdots + X_m$. Then, for every $\varepsilon > 0$, we have*

$$\mathbb{P}(X \ge (1+\varepsilon)\mathbb{E}(X)) \le \left(\frac{1+\varepsilon}{\mathrm{e}^{\varepsilon}}\right)^m,$$

and for any $0 < \varepsilon < 1$, we have

$$\mathbb{P}(X \le (1-\varepsilon)\mathbb{E}(X)) \le ((1-\varepsilon)\mathrm{e}^{\varepsilon})^m \le \mathrm{e}^{-\varepsilon^2 m/2}.$$

Proof. We will prove the bound for the upper tail. The bound for the lower tail is proved in a similar way and its proof is omitted. We have $\mathbb{E}X = m\mathbb{E}X_1 = m/\lambda$. By Markov's inequality, for every $0 < \beta < \lambda$ and every $\varepsilon > 0$, we have

$$\mathbb{P}(X \geq (1+\varepsilon)\mathbb{E}X) = \mathbb{P}(\mathrm{e}^{\beta X} \geq \mathrm{e}^{\beta(1+\varepsilon)m/\lambda}) \leq \frac{\prod \mathbb{E}(\mathrm{e}^{\beta X_i})}{\mathrm{e}^{\beta(1+\varepsilon)m/\lambda}} = (\varphi_{X_1}(\beta))^m \mathrm{e}^{-\beta(1+\varepsilon)m/\lambda},$$

where $\varphi_{X_1}(\beta) = \mathbb{E}(e^{\beta X_1}) = \lambda/(\lambda-\beta)$ is the moment generating function of an exponentially distributed random variable with parameter λ. Thus,

$$\mathbb{P}(X \geq (1+\varepsilon)\mathbb{E}X) \leq \left(\frac{\lambda}{\lambda-\beta}\right)^m \mathrm{e}^{-\beta(1+\varepsilon)m/\lambda}.$$

Now we set $\beta = (\varepsilon/(1+\varepsilon))\lambda$ to obtain

$$\mathbb{P}(X \geq (1+\varepsilon)\mathbb{E}X) \leq \left(\frac{1+\varepsilon}{\mathrm{e}^{\varepsilon}}\right)^m. \qquad \square$$

2.1. Proof of Theorem 1.1(i)

In this section we prove the lower bound in Theorem 1.1. For every $t \geq 0$, we introduce the following model of infinite random geometric graphs. The vertex set is constructed by adding two vertices $u = (0, 0)$ and $v = (t, 0)$ to a homogeneous Poisson point process of intensity 1 in the infinite plane $\mathbb{R}^2$. We denote this new model by $\widetilde{\mathcal{G}}^{\infty}_{u,v}(r, t)$.

The main task in the sequel is to show that, for any $t \geq \max\{12(\log n)^{3/2}/r,\ 21r\log n\}$, the lower bound of Theorem 1.1(i) holds with probability at least $1 - o(n^{-5/2})$ in $\widetilde{\mathcal{G}}^{\infty}_{u,v}(r, t)$ for the distinguished vertices $u = (0, 0)$ and $v = (t, 0)$. Combining this with an appropriate de-Poissonization argument will allow us to conclude the desired result for $\mathcal{G}(n, r)$.

In our next lemma we show that short paths connecting u and v are contained in small strips. The lemma is stated in the more general context of a deterministic geometric graph $G = (V, E)$ of radius r, where the vertex set V is an arbitrary subset of points in $\mathbb{R}^2$ (containing u and v) and edges connect (as usual) every pair of vertices at Euclidean distance at most r. For a given $r > 0$, for every $k \in \mathbb{N}$ and $\alpha > 0$, consider the rectangle

$$R(k, \alpha) = \left[\frac{-\alpha^2}{kr}, kr\right] \times [-\alpha, \alpha].$$

Lemma 2.2. *Given any $r, t, \alpha > 0$ and any $k \in \mathbb{N}$ satisfying $t \geq kr - 2\alpha^2/kr$, let G be a geometric graph of radius r in $\mathbb{R}^2$, and suppose that $u = (0, 0)$ and $v = (t, 0)$ are two vertices of G. Then all paths of length at most k from u to v are contained in $R(k, \alpha)$.*

Proof. If there is no path of length at most k from u to v, the statement of the lemma trivially holds. Thus, we suppose that $P = (u = z_0, z_1, \dots, z_\ell = v)$ is a path of length $\ell \leq k$, where $z_i = (x_i, y_i)$ for every $0 \leq i \leq \ell$. Also, note that it suffices to prove the lemma for α satisfying $t = kr - 2\alpha^2/kr$, since this trivially implies the statement for larger α. In particular, we have $\alpha^2 < (kr)^2/2$.

Write $x^+ = \max\{x_i : 0 \leq i \leq \ell\}$ and $x^- = \min\{x_i : 0 \leq i \leq \ell\}$. It is clear that $x^+ \leq \ell r \leq kr$ since every edge has length at most r. Let $0 \leq j \leq \ell$ be such that $x_j = x^-$ and observe that $x^- \leq x_0 = 0$. Then

$$kr \geq d_{\mathrm{E}}(u, z_j) + d_{\mathrm{E}}(z_j, v) \geq -x^- + (t - x^-) \geq kr - 2\left(x^- + \frac{\alpha^2}{kr}\right),$$

and we obtain $x^- \geq -\alpha^2/(kr)$.

https://doi.org/10.1017/apr.2016.31 Published online by Cambridge University Press

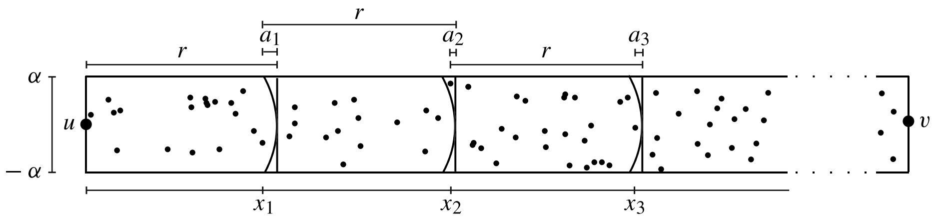

FIGURE 2: Example of some values of x_i and their corresponding a_i ($i \in \{1, 2, 3\}$).

Now write $y^+ = \max\{y_i : 0 \le i \le \ell\}$ and $y^- = \min\{y_i : 0 \le i \le \ell\}$. We will show that $y^+ \le \alpha$ and that $y^- \ge -\alpha$. For every $0 \le i \le \ell$, we have

$$kr \ge d_{\mathrm{E}}(u, z_i) + d_{\mathrm{E}}(z_i, v) = \sqrt{x_i^2 + y_i^2} + \sqrt{(t - x_i)^2 + y_i^2} \ge \sqrt{t^2 + 4y_i^2},$$

where we used the fact that the left-hand side of the last inequality is minimized at $x_i = t/2$. Using the fact that $t \ge kr - 2\alpha^2/kr$, we obtain

$$(kr)^2 \ge t^2 + 4y_i^2 \ge \left(kr - \frac{2\alpha^2}{kr}\right)^2 + 4y_i^2.$$

Thus, for every $0 \le i \le \ell$, we have $|y_i| \le \alpha\sqrt{1 - \alpha^2/(kr)^2} \le \alpha$, and so, in particular, $-\alpha \le y^- \le y^+ \le \alpha$. Using the bounds on x^+, x^-, y^+, and y^-, we conclude that P is contained in $R(k, \alpha)$. □

Proposition 2.1. *For every $t > r$, let $G \in \widetilde{\mathscr{G}}^{\infty}_{u,v}(r, t)$ be a random geometric graph on $\mathbb{R}^2$ with additional vertices $u = (0, 0)$ and $v = (t, 0)$. Then, for every $0 < \delta < 2^{-1/3}$, we have*

$$\mathbb{P}\left(d_{\mathrm{G}}(u, v) \le \left\lfloor \frac{t}{r}\left(1 + \frac{\delta}{(tr)^{2/3}}\right)\right\rfloor\right) \le \frac{(1 + o(1))t}{r} \exp\left(-\sqrt{\frac{\delta}{2}}(tr)^{2/3}\right) + \exp\left(-(1 - \sqrt{2\delta^3} - o(1))^2 \frac{t}{2r}\right).$$

Proof. We first set

$$k = \left\lfloor \frac{t}{r}\left(1 + \frac{\delta}{(tr)^{2/3}}\right)\right\rfloor \quad \text{and} \quad \alpha = \sqrt{\frac{\delta}{2}}\left(\frac{k^3 r^2}{t}\right)^{1/3}.$$

Observe that since $t > r$, we have $k \ge 1$. Let A_1 be the event that $d_{\mathrm{G}}(u, v) \le k$; that is, there exists a path P in $\widetilde{\mathscr{G}}^{\infty}_{u,v}(r, t)$ from u to v of length at most k. Our goal is to show that the probability of A_1 is small.

Let x_1 be the largest x-coordinate of the vertices inside the rectangle $R_1 = [0, r] \times [-\alpha, \alpha]$ (possibly $x_1 = 0$ if u is the only vertex in R_1). Define the random variable $a_1 = r - x_1$. We proceed similarly for every $2 \le i \le k$. We define x_i as follows. If $R_i = (x_{i-1} + a_{i-1}, x_{i-1} + r] \times [-\alpha, \alpha]$ is nonempty, let x_i be the largest x-coordinate of the vertices inside R_i; otherwise, set $x_i = x_{i-1} + a_{i-1}$ (see Figure 2). Then also define $a_i = x_{i-1} + r - x_i$.

Claim 2.1. If A_1 holds then $t \le x_k$.

Proof. Suppose that A_1 holds and let $P = (u = z_0, z_1, \dots, z_\ell = v)$ be one such path, and for every $0 \le i \le \ell$ let $\hat{x}_i$ be the x-coordinate of z_i. We will prove by induction on i that we have $\hat{x}_i \le x_i$. In particular, this implies that $t = \hat{x}_\ell \le x_\ell \le x_k$, and proves the claim.

Observe that

$$t \ge \frac{kr}{(1 + \delta/(tr)^{2/3})} \ge kr\left(1 - \frac{\delta}{(tr)^{2/3}}\right) = kr - \frac{2\alpha^2}{kr}. \tag{2.1}$$

Thus, we can use Lemma 2.2 to show that the path P is contained in the strip $\mathbb{R} \times [-\alpha, \alpha]$. Moreover, we must have $\hat{x}_1 - \hat{x}_0 \le r$ (since $u = z_0$ and z_1 are adjacent vertices). Therefore, our choice of x_1 and the fact that $z_1 \in \mathbb{R} \times [-\alpha, \alpha]$ imply that $\hat{x}_1 \le x_1$. So, the statement holds for $i = 1$. Now we inductively assume that $\hat{x}_{i-1} \le x_{i-1}$. We must have $\hat{x}_i \le \hat{x}_{i-1} + r$ (since z_{i-1} and z_i are adjacent vertices), and, therefore, $\hat{x}_i \le x_{i-1} + r$. Similarly as before, since $z_i \in \mathbb{R} \times [-\alpha, \alpha]$ and by the choice of x_i, we conclude that $\hat{x}_i \le x_i$, as desired. □

Thus, it suffices to show that $x_k \ge t$ with very small probability. We first study the random variables a_i. Define $a_0 = 0$. By the choice of x_i, for every $1 \le i \le k$, we have $0 \le a_i \le r - a_{i-1}$. Recall that $R_i = (x_{i-1} + a_{i-1}, x_{i-1} + r] \times [-\alpha, \alpha]$. Since $G \in \widetilde{\mathcal{G}}^{\infty}_{u,v}(r, t)$, the number of vertices from V inside a region of $\mathbb{R}^2$ is a Poisson random variable with mean equal to the area of that region. So, for every $2 \le i \le k$, we have

$$\mathbb{P}(a_i \ge \beta) = \begin{cases} \mathbb{P}((x_{i-1} + r - \beta, x_{i-1} + r] \\ \quad \times [-\alpha, \alpha] \text{ empty}) = \mathrm{e}^{-2\alpha\beta} & \text{if } 0 \le \beta \le r - a_{i-1}, \\ 0 & \text{if } \beta > r - a_{i-1}. \end{cases} \tag{2.2}$$

Thus, a_i is stochastically dominated by an exponentially distributed random variable $\widetilde{a}_i$ of parameter 2α. We assume that a_i and $\widetilde{a}_i$ are coupled together in the same probability space, so that $a_i = \min\{\widetilde{a}_i, r - a_{i-1}\} \le \widetilde{a}_i$.

Moreover, since the regions $R_1, R_2, \dots, R_k$ that define the random variables a_i are disjoint, the joint distribution of $a_1, a_2, \dots, a_k$ is stochastically dominated by the joint distribution of $\widetilde{a}_1, \widetilde{a}_2, \dots, \widetilde{a}_k$; that is, the distribution of k independent exponentially distributed random variables of parameter 2α.

Define

$$a = \sum_{i=1}^{k} a_i \quad \text{and} \quad \widetilde{a} = \sum_{i=1}^{k} \widetilde{a}_i.$$

Expanding recursively from the relations $x_i = x_{i-1} + r - a_i$ and $x_1 = r - a_1$, we have

$$x_k = \sum_{i=1}^{k} (r - a_i) = kr - a.$$

Consider the event A_2 defined by $\widetilde{a}_i \le r/2$ for all $1 \le i \le k$. Since we aim to bound the probability that x_k is large (or, equivalently, that a is small), we cannot use the fact that the joint distribution of the a_is is stochastically dominated by the ones of the $\widetilde{a}_i$s. Nevertheless, note that conditional on A_2, we have $a_i = \widetilde{a}_i$ for all $1 \le i \le k$; if $a_{i-1} \le r/2$ then $a_i \le r/2 \le r - a_{i-1}$, and from (2.2), $a_i = \widetilde{a}_i$. In other words, for every $\beta \ge 0$,

$$\mathbb{P}(a \le \beta, A_2) = \mathbb{P}(\widetilde{a} \le \beta, A_2).$$

https://doi.org/10.1017/apr.2016.31 Published online by Cambridge University Press

Since each $\widetilde{a}_i$ is exponentially distributed with parameter 2α and stochastically dominates a_i, we can bound the probability that A_2 does not occur; that is,

$$\mathbb{P}(\overline{A_2}) \le \sum_{i=1}^{k} \mathbb{P}\biggl(a_i > \frac{r}{2}\biggr) \le \sum_{i=1}^{k} \mathbb{P}\biggl(\widetilde{a}_i > \frac{r}{2}\biggr) = k\mathrm{e}^{-\alpha r}. \tag{2.3}$$

Therefore, using the bound on t given in (2.1), we have

$$\begin{aligned}\mathbb{P}(x_k \ge t) &\le \mathbb{P}(\overline{A_2}) + \mathbb{P}(x_k \ge t, A_2)\\ &\le k\mathrm{e}^{-\alpha r} + \mathbb{P}(kr - a > t, A_2)\\ &\le k\mathrm{e}^{-\alpha r} + \mathbb{P}\biggl(a \le \frac{2\alpha^2}{kr}, A_2\biggr)\\ &= k\mathrm{e}^{-\alpha r} + \mathbb{P}\biggl(\widetilde{a} \le \frac{2\alpha^2}{kr}, A_2\biggr)\\ &\le k\mathrm{e}^{-\alpha r} + \mathbb{P}\biggl(\widetilde{a} \le \frac{2\alpha^2}{kr}\biggr). \end{aligned}\tag{2.4}$$

Thus, it remains to provide a good upper bound on the lower tail of $\widetilde{a}$. Note that $\mathbb{E}(\widetilde{a}) = \sum_{i=1}^{k}\mathbb{E}(\widetilde{a}_i) = k/2\alpha$. We use the definition of k and α, as well as Lemma 2.1 with $\varepsilon = (1-\sqrt{2}\delta^{3/2} - o(1))$, to show that

$$\mathbb{P}\biggl(\widetilde{a} \le \frac{2\alpha^2}{kr}\biggr) = \mathbb{P}\biggl(\widetilde{a} \le \frac{4\alpha^3}{k^2 r}\mathbb{E}(\widetilde{a})\biggr) \le \mathbb{P}(\widetilde{a} \le (\sqrt{2}\delta^{3/2} + o(1))\mathbb{E}(\widetilde{a})) \le \mathrm{e}^{-\varepsilon^2 k/2}. \tag{2.5}$$

Finally, we use (2.4), (2.5) and the definition of k, α, and ε to obtain

$$\mathbb{P}(x_k \ge t) \le \frac{(1+o(1))t}{r}\exp\biggl(-\sqrt{\frac{\delta}{2}}(tr)^{2/3}\biggr) + \exp\biggl(-(1-\sqrt{2\delta^3} - o(1))^2\frac{t}{2r}\biggr).$$

Since the event A_1 implies that $x_k \ge t$, $\mathbb{P}(A_1) \le \mathbb{P}(x_k \ge t)$ and the proposition follows. □

Proposition 2.2. *Let* $\widetilde{\mathcal{G}}^{\infty}_{u,v}(r,t)$ *be a random geometric graph in* $\mathbb{R}^2$ *with two additional distinguished vertices* $u = (0,0)$ *and* $v = (t,0)$ *such that*

$$t = d_{\mathrm{E}}(u,v) \ge \max\biggl\{\frac{12(\log n)^{3/2}}{r}, 21r\log n\biggr\}.$$

Then we have

$$d_{\mathrm{G}}(u,v) \le \biggl\lfloor \frac{t}{r}\biggl(1 + \frac{1}{2(rt)^{2/3}}\biggr)\biggr\rfloor$$

with probability at most $o(n^{-5/2})$.

Proof. Set $\delta = \frac{1}{2}$. Since $t \ge 12(\log n)^{3/2}/r$, we have

$$\sqrt{\frac{\delta}{2}}(tr)^{2/3} - \log\biggl(\frac{1-o(1)t}{r}\biggr) > \frac{5}{2}\log n,$$

and since $t \ge 21r\log n$,

$$(1-\sqrt{2\delta^3} - o(1))^2\frac{t}{2r} > \frac{5}{2}\log n.$$

https://doi.org/10.1017/apr.2016.31 Published online by Cambridge University Press

By Proposition 2.1, this implies that

$$\mathbb{P}\bigg(d_{\mathrm{G}}(u,v) \le \bigg\lfloor \frac{t}{r}\bigg(1+\frac{1}{2(rt)^{2/3}}\bigg)\bigg\rfloor\bigg) = o(n^{-5/2}). \qquad \square$$

The same conclusion in Proposition 2.1 must hold (for $t \le \sqrt{2n}$) if we restrict $\widetilde{\mathcal{G}}^{\infty}_{u,v}(r,t)$ to any arbitrary square $\widehat{\mathcal{S}}_n$ of area n containing $u=(0,0)$ and $v=(t,0)$ (that is, we consider the subgraph induced by the vertices lying inside of that square), since the graph distance between u and v can only increase when doing so. Moreover, by rotating and mapping an appropriate square $\widehat{\mathcal{S}}_n$ to $\mathcal{S}_n = [-\sqrt{n}/2, \sqrt{n}/2]^2$, we conclude that Theorem 1.1(i) holds in $\widetilde{\mathcal{G}}_{u,v}(n,r)$ with probability $1-o(n^{-5/2})$. Hence, in view of the de-Poissonization argument described at the beginning of Section 2, this same property holds in $\mathcal{G}(n,r)$ with probability $1-o(n^{-2})$, for a given pair of vertices u, v. The statement follows by taking a union bound over all at most n^2 pairs of vertices.

2.2. Proof of Theorem 1.1(ii)

In this section we complete the proof of Theorem 1.1. To derive the bound on the upper tail on the graph distance between $u, v \in V$, we first assume that $u=(0,0)$ and $v=(t,0)$ (for some $0 < t \le \sqrt{2n}$), and analyze $\widetilde{\mathcal{G}}^{\infty}_{u,v}(r,t)$ restricted to a suitable rectangle. Our goal is to find a path P from u to v inside of that rectangle that gives an appropriate upper bound on $d_{\mathrm{G}}(u,v)$. Then, we will use similar ideas to those at the end of Section 2.1 to derive the desired conclusion about $\mathcal{G}(n,r)$.

For every measurable set $S \subseteq \mathbb{R}^2$ containing u and v, let $\widetilde{\mathcal{G}}_{S,u,v}(r,t)$ denote the random geometric graph obtained as the subgraph of $\widetilde{\mathcal{G}}^{\infty}_{u,v}(r,t)$ induced by the vertices contained in S. Observe that $\widetilde{\mathcal{G}}_{S,u,v}(r,t)$ can also be constructed by taking as the vertex set a Poisson point process of intensity 1 in S, adding the vertices $u=(0,0)$ and $v=(t,0)$, and connecting any two vertices by an edge if they are at Euclidean distance at most r.

For every $0 < \alpha \le r$, we define the rectangle

$$S(t,\alpha) = [0,t] \times [0,\alpha].$$

(The precise value of α will be specified later; it will be different to the one given in the previous section.) Given α and r, we write $\rho = r - \alpha^2/r$. Then, for every point $z=(x_z,y_z) \in S$, we define the rectangle

$$S_z = S_z(\alpha) := [x_z, x_z+\rho] \times [0,\alpha].$$

We need the following auxiliary lemma.

Lemma 2.3. *Let $t > 0$ and $0 < \alpha \le r$. Then, for every pair of points $z \in S(t,\alpha)$ and $z' \in S_z(\alpha)$, we have $d_{\mathrm{E}}(z,z') \le r$ (see Figure 3).*

Proof. It is enough to show that the upper-left corner $z_1 = (x_z, \alpha)$ and the bottom-right corner $z_2 = (x_z+\rho, 0)$ of $S_z(\alpha)$ satisfy $d_{\mathrm{E}}(z_1,z_2) \le r$. Then all the points inside $S_z(\alpha)$ lie at distance at most r, and, in particular, $d_{\mathrm{E}}(z,z') \le r$.

We have

$$(d_{\mathrm{E}}(z_1,z_2))^2 = \rho^2 + \alpha^2 = r^2 - \alpha^2\bigg(1-\bigg(\frac{\alpha}{r}\bigg)^2\bigg) \le r^2,$$

and the lemma follows. $\square$

https://doi.org/10.1017/apr.2016.31 Published online by Cambridge University Press

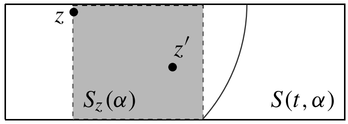

FIGURE 3: The rectangle S_z.

Our next task is to bound the graph distance between u and v in $\widetilde{\mathcal{G}}_{S(t,\alpha),u,v}(r,t)$ by finding a path of length at most $\lceil (t/r)(1+\delta r^{-4/3})\rceil$ from u to v, for some δ that will be made precise in the following proposition.

Proposition 2.3. *Let $F>0$ and $J>3(F+1)$ be constants and define $g(x)=x-\log(1+x)$. For every $J\le\delta\le Fr^{4/3}$, there exists an α such that the following holds. Fix $t\ge 0$ and consider $\widetilde{\mathcal{G}}_{S,u,v}(r,t)$ to be a random geometric graph with $u=(0,0)$, $v=(t,0)$ in the rectangle $S=S(t,\alpha)$. Then we have*

$$\mathbb{P}\biggl(d_{\mathrm{G}}(u,v)>\biggl\lceil\frac{t}{r}\biggl(1+\delta r^{-4/3}\biggr)\biggr\rceil\biggr)\le n\exp\biggl(-\frac{(F+1)\delta^{1/2}r^{4/3}}{2J^{3/2}}\biggr)+\exp\biggl(-g\biggl(\biggl(\frac{\delta}{J}\biggr)^{3/2}\biggr)\frac{t}{r}\biggr).$$

Proof. We will first define some parameters that will be useful in our analysis. Set $C=J^{-3/2}$ and let B be an arbitrary positive constant satisfying

$$B^2+\frac{2C}{B}<\frac{1}{F+1}. \tag{2.6}$$

Elementary analysis shows that such a B must exist. In fact, $B^2+2C/B=1/(F+1)$ has exactly two positive solutions B_1 and B_2 for any $0<C=J^{-3/2}<1/(3(F+1))^{3/2}$, and any $0<B_1<B<B_2<1/\sqrt{F+1}$ satisfies (2.6).

Fix some δ with $J\le\delta\le Fr^{4/3}$, and set

$$\alpha=B\delta^{1/2}r^{1/3}. \tag{2.7}$$

In order to use Lemma 2.3, let us first show that $\alpha\le r$. Since $\delta\le Fr^{4/3}$ by hypothesis of the proposition, we have

$$\alpha\le(B\sqrt{F})r, \tag{2.8}$$

and $B\sqrt{F}<1$, since $B<1/\sqrt{F+1}$. Moreover, we have

$$\rho=r-\frac{\alpha^2}{r}\ge(1-B^2F)r. \tag{2.9}$$

We will consider the integer $k=\lceil (t/r)(1+\delta r^{-4/3})\rceil$ and let A_1 be the event $d_{\mathrm{G}}(u,v)\le k$; that is, there exists a path $P=(u=z_0,z_1,\dots,z_\ell,v)$ in $\widetilde{\mathcal{G}}_{S,u,v}(r,t)$ from u to v of length at most k. Such a path will only use vertices inside $S=S(t,\alpha)$, but due to some technical considerations in the argument, we extend the Poisson point process of our probability space to the semiinfinite strip $S(\alpha)=[0,\infty)\times[0,\alpha]$. Our goal is to show that the probability of A_1 is large.

As we did in the proof of Proposition 2.1, now we define random variables x_i and a_i for every $i\ge 1$. Set $x_0=0$ and $a_0=0$. For each $i\ge 1$, consider the rectangle $R_i=R_i(\alpha):=(x_{i-1}+\rho/2,x_{i-1}+\rho]\times[0,\alpha]$. If R_i contains at least a vertex, let z_i be the vertex with largest x-coordinate inside R_i. In such a case, define x_i to be the x-coordinate of z_i and $a_i=x_{i-1}+\rho-x_i$. Otherwise, we stop the process.

Let $\tau=\min\{i\ge 1\colon R_i \text{ contains no points}\}$ be the stopping time of the process.

https://doi.org/10.1017/apr.2016.31 Published online by Cambridge University Press

Claim 2.2. Conditional on $\tau \geq k$, if $x_{k-1} + \rho \geq t$ then A_1 holds.

Proof. Assume that $\tau \geq k$ and that $x_{k-1} + \rho \geq t$. Observe that for every $i < k$, we have $0 \leq a_i \leq \rho/2$. Moreover, by construction of the process, for every $1 \leq i < k$, we have $z_i \in R_i \subseteq S_{z_{i-1}}$ and, since $\alpha \leq r$, Lemma 2.3 implies that z_i is adjacent to z_{i-1}. Thus, the vertices $z_0, z_1, \dots, z_{k-1}$ form a path. In particular,

$$x_1 \geq \frac{\rho}{2}. \tag{2.10}$$

Since $x_{k-1} + \rho \geq t$, we know that there exists a value $\ell \leq k-1$ such that $x_{\ell-1} + \rho \geq t$ and also $x_{\ell-1} \leq t$, and, thus, by Lemma 2.3, z_ℓ and v are connected by an edge. The path $P = (u = z_0, z_1 \dots, z_\ell, v)$ has length $\ell + 1 \leq k$, connects u and v, and is *fully contained* in S. Therefore, A_1 is satisfied, which completes the proof of the claim. □

It suffices to show that we have with high probability $\tau \geq k$, and that conditional on it, with high probability $x_{k-1} + \rho \geq t$.

For every $0 \leq i < \tau$, let $A_2^{(i)}$ be the event that $a_j \leq \rho/2$ for every $1 \leq j \leq i$ and let $A_2 = A_2^{(k-1)}$ be the event that $\tau \geq k$. Conditional on A_2, it follows that the regions $R_1, \dots, R_{k-1}$ are disjoint. Hence, we deduce that conditional on A_2, the joint distribution of $a_1, \dots, a_{k-1}$ is the same as the joint distribution of $\widetilde{a}_1, \dots, \widetilde{a}_{k-1}$, with $\widetilde{a}_1, \dots, \widetilde{a}_{k-1}$ being $k-1$ independent exponentially distributed random variables with parameter α. In particular, conditional only on $A_2^{(i-1)}$, we also have that a_i is stochastically dominated by $\widetilde{a}_i$, and, hence,

$$\mathbb{P}(\overline{A_2}) = \sum_{i=1}^{k-1} \mathbb{P}\biggl(a_i \geq \frac{\rho}{2} \biggm| A_2^{(i-1)}\biggr) \leq \sum_{i=1}^{k-1} \mathbb{P}\biggl(\widetilde{a}_i \geq \frac{\rho}{2}\biggr).$$

Since $\alpha\rho/2 \geq (1 - B^2F)\alpha r/2 = (1 - B^2F)B\delta^{1/2}r^{4/3}/2$, we have

$$\mathbb{P}\biggl(\widetilde{a}_i \geq \frac{\rho}{2}\biggr) = \mathrm{e}^{-\alpha\rho/2} \leq \mathrm{e}^{-(1-B^2F)B\delta^{1/2}r^{4/3}/2},$$

and that

$$\mathbb{P}(\tau < k) = \mathbb{P}(\overline{A_2}) \leq n\mathrm{e}^{-(1-B^2F)B\delta^{1/2}r^{4/3}/2}. \tag{2.11}$$

Also, if we let $a = \sum_{i=1}^{k-1} a_i$ and $\widetilde{a} = \sum_{i=1}^{k-1} \widetilde{a}_i$, conditional on A_2 (or in other words, on $\tau \geq k$), by the same argument, for every $\beta \geq 0$, we have

$$\mathbb{P}(a \geq \beta, A_2) = \mathbb{P}(\widetilde{a} \geq \beta, A_2). \tag{2.12}$$

Observe that now $\mathbb{E}(\widetilde{a}) = (k-1)/\alpha$. Let A_3 be the event that $\widetilde{a} \leq (1 + C\delta^{3/2})((k-1)/\alpha)$. We first show that A_3 implies the event $\{k\rho - \widetilde{a} > t\}$. Conditional on A_3, using the definition of α, the fact that $\delta^{-3/2} \leq C$, and that $\delta \leq Fr^{4/3}$, we have

$$\begin{aligned}
k\rho - \widetilde{a} &> k\rho - \frac{(1 + C\delta^{3/2})(k-1)}{\alpha} \\
&\geq kr\biggl(1 - \frac{\alpha^2}{r^2} - \frac{(1 + C\delta^{3/2})}{\alpha r}\biggr) \\
&\geq t(1 + \delta r^{-4/3})\biggl(1 - \delta r^{-4/3}\biggl(B^2 + \frac{(\delta^{-3/2} + C)}{B}\biggr)\biggr)
\end{aligned}$$

https://doi.org/10.1017/apr.2016.31 Published online by Cambridge University Press

$$
\begin{aligned}
&\geq t(1+\delta r^{-4/3})\Big(1-\delta r^{-4/3}\Big(B^2+\frac{2C}{B}\Big)\Big)\\
&= t\Big[1+\delta r^{-4/3}\Big(1-\Big(\delta r^{-4/3}+1\Big)\Big(B^2+\frac{2C}{B}\Big)\Big)\Big]\\
&\geq t\Big[1+\delta r^{-4/3}\Big(1-(F+1)\Big(B^2+\frac{2C}{B}\Big)\Big)\Big]\\
&> t. \qquad (2.13)
\end{aligned}
$$

Now, we can use (2.12) and the upper-tail bound in Lemma 2.1 to prove

$$\mathbb{P}(\overline{A_3}) = \mathbb{P}\Big(\widetilde{a} \geq (1+C\delta^{3/2})\frac{k-1}{\alpha}\Big) \leq \mathrm{e}^{-g((\delta/J)^{3/2})(k-1)} \leq \mathrm{e}^{-g((\delta/J)^{3/2})(\lceil t/r\rceil-1)}. \qquad (2.14)$$

By expanding the definition of x_{k-1}, we can write $x_{k-1} = (k-1)\rho - a$. Thus, using (2.11)–(2.14), we obtain

$$
\begin{aligned}
\mathbb{P}(\{\tau < k\}\cup\{x_{k-1}+\rho\leq t\}) &= \mathbb{P}(\overline{A_2}) + \mathbb{P}(x_{k-1}+\rho\leq t, A_2)\\
&\leq n\mathrm{e}^{-(1-B^2F)B\delta^{1/2}r^{4/3}/2} + \mathbb{P}(k\rho - a\leq t, A_2)\\
&\leq n\mathrm{e}^{-(1-B^2F)B\delta^{1/2}r^{4/3}/2} + \mathbb{P}(k\rho - \widetilde{a}\leq t)\\
&\leq n\mathrm{e}^{-(1-B^2F)B\delta^{1/2}r^{4/3}/2} + \mathbb{P}(\overline{A_3}) + \mathbb{P}(k\rho - \widetilde{a}\leq t, A_3)\\
&\leq n\mathrm{e}^{-(1-B^2F)B\delta^{1/2}r^{4/3}/2} + \mathrm{e}^{-g((\delta/J)^{3/2})(\lceil t/r\rceil-1)}. \qquad (2.15)
\end{aligned}
$$

Moreover, by the properties of B and the definition of C, we have

$$(1-B^2F)B > (1-B^2(F+1))B > 2C(F+1) = 2(F+1)J^{-3/2}.$$

Thus,

$$\mathbb{P}(x_{k-1}+\rho\leq t) \leq n\exp\Big(-\frac{(F+1)\delta^{1/2}r^{4/3}}{J^{3/2}}\Big) + \exp\Big(-g\Big(\Big(\frac{\delta}{J}\Big)^{3/2}\Big)\Big(\Big\lceil\frac{t}{r}\Big\rceil-1\Big)\Big),$$

concluding the proof of the proposition. □

Remark 2.1. Observe the trade-off between δ and the success probability in the proof of Proposition 2.3: for a given value of δ, we set $\alpha = \Theta(\sqrt{\delta}r^{1/3})$. That is, for a given radius r, the smaller δ, the smaller α. Proposition 2.3 computes the probability that a path using vertices only within a strip of width α can be found. Clearly, the smaller δ, the straighter a path has to be, and the smaller the rectangle in which we have to find a path has to be, therefore making also α smaller. On the other hand, for smaller α, the probability of indeed finding a path in such a small strip also gets smaller.

Proposition 2.4. *Given $t > 0$ and the vertices $u = (0,0)$ and $v = (t,0)$, let $\gamma = \gamma(t)$ be defined as in Theorem 1.1(ii). Let $\widetilde{\mathcal{G}}_{S,u,v}(r,t)$ be a random geometric graph in the rectangle $S = S(t,\alpha)$, with additional vertices u and v. Suppose that $r \geq 224\sqrt{\log n}$. Then, we have*

$$d_{\mathrm{G}}(u,v) > \Big\lceil\frac{t}{r}\Big(1+\gamma r^{-4/3}\Big)\Big\rceil$$

with probability at most $o(n^{-5/2})$.

https://doi.org/10.1017/apr.2016.31 Published online by Cambridge University Press

Proof. First, observe that, if $t \le r$ then $d_{\mathrm{G}}(u, v) = 1$ with probability 1 and the statement of the proposition holds trivially. Thus, we assume henceforth that $t > r$. Set $B = 0.01/(2.02\sqrt{2})$, $C = 10^{-4}$, $F = 1$, $D = 4 \times 10^6$, $E = 1358$, and $J = 10^{8/3}$. Set

$$\gamma' = \max\left\{E\left(\frac{\log n}{\lceil t/r\rceil - 1}\right)^{2/3}, D\frac{\log^2 n}{r^{8/3}}, 3^{2/3}J\right\}.$$

Note that $\gamma' \le \gamma$ for γ as given in Theorem 1.1: indeed, the second and the third term are equal, and for the first term, for $t > r$, it holds that$(3/(1 + t/r)) > (1/(\lceil t/r\rceil - 1))$. Therefore, it suffices to apply Proposition 2.3 with $\delta = \gamma'$. It is straightforward to check that the restrictions in (2.6) and $J > 3(F + 1)$, required in Proposition 2.3, hold. We also need to show that $J \le \gamma' \le Fr^{4/3}$. Note that $D\log^2/nr^{8/3} \le Fr^{4/3}$, since

$$r \ge 224\sqrt{\log n} \ge D^{1/4}\sqrt{\log n};$$

also

$$E\left(\frac{\log n}{\lceil t/r\rceil - 1}\right)^{2/3} \le Fr^{4/3},$$

since $\lceil t/r\rceil - 1 \ge 1$, and since $r^2 \ge E^{3/2}\log n$, which follows from our assumption of $r \ge 224\sqrt{\log n}$; and finally $3^{2/3}J \le Fr^{4/3}$ since $r = \Omega(\sqrt{\log n})$. Moreover, $\gamma' \ge 3^{2/3}J \ge J$.

Note that this choice of constants combined with (2.8) and (2.9) implies that

$$\alpha \le \frac{0.01}{2.02\sqrt{2}}r \le \frac{r}{3} \quad \text{and} \quad \rho \ge \frac{8r}{9} \ge \frac{8\alpha}{3}. \tag{2.16}$$

We will now apply (2.15) in the proof of Proposition 2.3 with this given δ, in order to show that $\mathbb{P}(d_{\mathrm{G}}(u, v) > k) = o(n^{-5/2})$. On the one hand, $\delta \ge D\log^2/nr^{8/3}$ implies that

$$\begin{aligned}\frac{(1 - B^2F)B\delta^{1/2}r^{4/3}}{2} - \log n &\ge \frac{(1 - B^2F)BD^{1/2}\log n}{2} - \log n \\ &> \frac{7.0009}{2}\log n - \log n \\ &= \frac{5.0009}{2}\log n.\end{aligned}$$

On the other hand, $\delta \ge 3^{2/3}J$ and $\delta \ge E(\log n/(\lceil t/r\rceil - 1))^{2/3}$ imply that

$$g\left(\left(\frac{\delta}{J}\right)^{3/2}\right)\left(\left\lceil\frac{t}{r}\right\rceil - 1\right) > \frac{(\delta/J)^{3/2}}{2}\left(\left\lceil\frac{t}{r}\right\rceil - 1\right) \ge \frac{1}{2}CE^{3/2}\log n > \frac{5.004}{2}\log n,$$

where we have used the fact that $g(x) \ge x/2$ if $x \ge 3$, and that $C = J^{-3/2}$.

Therefore, $\mathbb{P}(d_{\mathrm{G}}(u, v) > k) \le n^{-5.0009/2} + n^{-5.004/2} = o(n^{-5/2})$. □

Corollary 2.1. *The statement in Theorem 1.1(ii) holds.*

Proof. We will use an argument similar to that at the end of Section 2.1 to relate the models $\widetilde{\mathcal{G}}_{S(t,\alpha),u,v}(r, t)$ and $\widetilde{\mathcal{G}}_{u,v}(n, r)$. However, such an endeavor entails extra difficulties. Given two vertices $u, v \in \mathcal{S}_n = [-\sqrt{n}/2, \sqrt{n}/2]^2$ at Euclidean distance $t > 0$, there are exactly two isometries that map them to $(0, 0)$ and $(t, 0)$, denoted by π^+ and π^-. Unfortunately, the preimage of the rectangle $S(t, \alpha)$ under such isometries may not be entirely contained in the

https://doi.org/10.1017/apr.2016.31 Published online by Cambridge University Press

square $\mathcal{S}_n$. In order to overcome this obstacle, we just need to show that the internal vertices of the path from $(0, 0)$ to $(t, 0)$ that we constructed in the proof of Proposition 2.3 are contained in a smaller rectangle whose preimage under either π^+ or π^- is contained in $\mathcal{S}_n$. This will be enough for us to conclude the existence (with sufficiently high probability) of a path in $\widetilde{\mathcal{G}}_{u,v}(n, r)$ between u and v of the desired length.

Recall the definition of α given in (2.7). Observe that from (2.10) together with (2.16), $x_1 \ge \rho/2 > 4\alpha/3$ with probability at least $1 - o(n^{-5/2})$. In particular, this event implies that z_1 is outside of the square $[0, 1.01\alpha] \times [0, \alpha]$. If z_ℓ (the last internal vertex of the path P found) is outside $[t - 1.01\alpha, t] \times [0, \alpha]$, we obtain a path connecting u and v of length $\ell + 1 \le k$ with all its internal vertices in $R := [1.01\alpha, t - 1.01\alpha] \times [0, \alpha]$. Otherwise, suppose that z_ℓ lies in $[t - 1.01\alpha, t] \times [0, \alpha]$. Then, also with probability $1 - o(n^{-5/2})$, we can find some point $\hat{z}_\ell$ in $[t - 1.01\alpha - r/2, t - 1.01\alpha) \times [0, \alpha]$: indeed, since $\rho \le r$, the region in which we want $\hat{z}_\ell$ is bigger than the regions S_i in the proofs of Proposition 2.3 and Proposition 2.4. We can now use Lemma 2.3 to show that z_ℓ can be replaced by $\hat{z}_\ell$ in P. Observe that $\hat{z}_\ell$ is connected to v, since $1.01\alpha + r/2 \le \rho$, and also $\hat{z}_\ell$ is connected to $z_{\ell-1}$, since its x-coordinate $\hat{x}_\ell$ satisfies

$$|\hat{x}_\ell - x_{\ell-1}| \le \max\left\{\rho, \frac{r}{2} - \frac{\rho}{2}\right\} \le \rho.$$

Thus, we can replace z_ℓ with $\hat{z}_\ell$, and obtain a new path connecting u and v of length $\ell + 1 \le k$ with all its internal vertices in R. We will show that either $\pi^+(R)$ or $\pi^-(R)$ is always contained in $\mathcal{S}_n$. We first introduce some definitions.

Consider two points $u = (x_u, y_u)$ and $v = (x_v, y_v)$ in $\mathcal{S}_n$. By symmetry we may assume that $x_u < x_v$ and $y_u \le y_v$. Let β be the angle of the vector $\boldsymbol{uv}$ with respect to the horizontal axis. Again by symmetry, we may consider $\beta \in [0, \pi/4]$.

We consider now two rectangles of dimensions $\alpha \times t$ placed on each side of the segment uv. Let R^+ be the rectangle to the left of $\boldsymbol{uv}$ (that is, $R^+ = \pi^+(R)$), and let R^- be the rectangle to the right of $\boldsymbol{uv}$ (also, $R^- = \pi^-(R)$). We will show that at least one of these rectangles contains a copy of R fully contained in $\mathcal{S}_n$. This choice will determine which of the isometries, π^+ or π^-, map R inside $\mathcal{S}_n$.

Note that the intersection of R^+ and R^- with each of the half-planes $x \le x_u$, $x \ge x_v$, $y \le y_u$, and $y \ge y_v$ gives four triangles. We call them T_u^+, T_v^-, T_u^-, and T_v^+, respectively. All these triangles are right-angled, and denote by t_u^+, t_v^-, t_u^-, and t_v^+ the side of the corresponding triangle that is parallel to the segment uv. Note that $|t_u^+| = |t_v^-|$ and $|t_u^-| = |t_v^+|$. Call a triangle T_w^*, with $w \in \{u, v\}$ and $* \in \{+, -\}$, *safe* if $|t_w^*| \le 1.01\alpha$. Note that if T_u^+ and T_v^+ are safe or fully contained in the square, then R^+ contains the desired rectangle R, and analogously for R^-.

Since we assumed that $\beta \le \pi/4$, we have $|t_u^+| = |t_v^-| = \alpha|\tan\beta| \le 1.01\alpha$. Thus, T_u^+ and T_v^- are safe. If $y_u = y_v$; that is, $\beta = 0$, it is clear that either R^+ or R^- contain the desired copy of R. Thus, we may assume that $\beta > 0$.

We can also assume that both u and v are on the boundary of $\mathcal{S}_n$, as otherwise we extend the line segment uv to the boundary of the square, and the original rectangles are contained in the new ones.

Recall that T_u^+ and T_v^- are safe. If $y_v < \sqrt{n}/2 - \alpha$ then T_v^+ is completely contained in the square, and, hence, R^+ satisfies the conditions. Similarly, if $y_u > -\sqrt{n}/2 + \alpha$, R^- satisfies the conditions. Thus, assume that $y_v \ge \sqrt{n}/2 - \alpha$ and $y_u \le -\sqrt{n}/2 + \alpha$. We will show that R^- contains the desired copy of R. As before, T_v^- is safe, so it remains to consider T_u^-. We have $|t_u^-| = \alpha \tan(\pi/2 - \beta)$. For $0 < \beta \le \pi/4$, $\tan(\pi/2 - \beta)$ is decreasing in β, and it therefore suffices to show that T_u^- is safe for the smallest possible value of β. Note that the minimal

https://doi.org/10.1017/apr.2016.31 Published online by Cambridge University Press

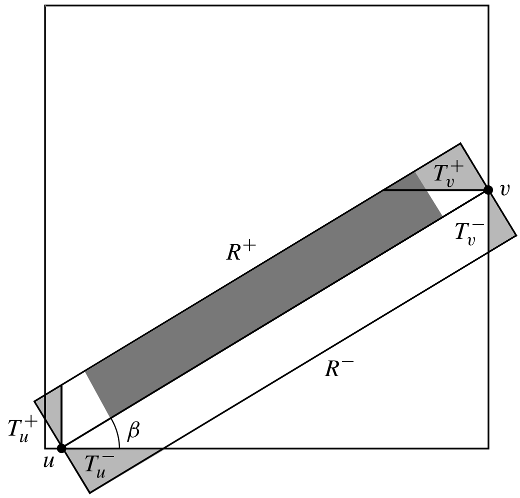

FIGURE 4: The dark shaded area represents the copy of R contained in $\mathcal{S}_n$.

angle β under our assumptions on y_u and y_v is obtained for $u = (-\sqrt{n}/2, -\sqrt{n}/2 + \alpha)$ and $v = (\sqrt{n}/2, \sqrt{n}/2 - \alpha)$, and, thus, $\beta \geq \arctan((\sqrt{n} - 2\alpha)/\sqrt{n})$, or, equivalently,

$$\tan\left(\frac{\pi}{2} - \beta\right) \leq \frac{\sqrt{n}}{\sqrt{n} - 2\alpha}.$$

In this case,

$$|t_u^-| \leq \alpha\left(\frac{\sqrt{n}}{\sqrt{n} - 2\alpha}\right) = \alpha\left(1 + \frac{2\alpha}{\sqrt{n} - 2\alpha}\right) \leq 1.01\alpha,$$

where the last inequality follows from the fact that

$$\alpha \leq \left(\frac{0.01}{2.02\sqrt{2}}\right) r \leq \left(\frac{0.01}{2.02}\right)\sqrt{n}$$

since we assumed that $r \leq \sqrt{2n}$ (see also (2.16)), and, therefore, $2\alpha/(\sqrt{n} - 2\alpha) \leq 0.01$.

Again, by de-Poissonizing $\widetilde{\mathcal{G}}_{u,v}(n, r)$, we can use Proposition 2.4 to show that for a given u and v in $G \in \mathcal{G}(n, r)$, Theorem 1.1(ii) holds with probability at least $1 - o(n^{-2})$. By taking a union bound over all at most n^2 possible pairs of vertices, Theorem 1.1(ii) follows. □

Acknowledgements

The authors are grateful to the anonymous referees for their constructive comments. This work was partially supported by the CYCIT (grant number TIN2013-66523 (CoMMas)).

References

[1] BRADONJIĆ, M. *et al.* (2010). Efficient broadcast on random geometric graphs. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, SIAM, Philadelphia, PA, pp. 1412–1421.

[2] ELLIS, R. B., MARTIN, J. L. AND YAN, C. (2007). Random geometric graph diameter in the unit ball. *Algorithmica*, **47,** 421–438.

[3] GILBERT, E. N. (1961) Random plane networks. *J. Soc. Indust. Appl. Math.* **9,** 533–543.

https://doi.org/10.1017/apr.2016.31 Published online by Cambridge University Press

[4] FRIEDRICH, T., SAUERWALD, T. AND STAUFFER, A. (2011). Diameter and broadcast time of random geometric graphs in arbitrary dimensions. In *Algorithms and Computation*, Springer, Heidelberg, pp. 190–199.

[5] GOEL, A. RAI, S. AND KRISHNAMACHARI, B. (2004). Sharp thresholds for monotone properties in random geometric graphs. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing,* ACM, New York, pp. 580–586.

[6] MEHRABIAN, A. AND WORMALD, N. (2013). On the Stretch Factor of Randomly Embedded Random Graphs. *Discrete Comput. Geom.* **49,** 647–658.

[7] MUTHUKRISHNAN, S. AND PANDURANGAN, G. (2005). The bin-covering technique for thresholding random geometric graph properties. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, ACM, New York, pp. 989–998.

[8] PENROSE, D. M. (1997). The longest edge of the random minimal spanning tree. *Ann. Appl. Prob.*, **7,** 340–361.

[9] PENROSE, M. (2003). *Random Geometric Graphs*. Oxford University Press.

[10] WALTERS, M. (2011). Random geometric graphs. In *Surveys in Combinatorics, 2011*, Cambridge University Press, pp. 365–401.