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

J. DÍAZ, D. MITSCHE, G. PERARNAU, AND X. PÉREZ-GIMÉNEZ

Abstract. Given any two vertices \( u, v \) of a random geometric graph, denote by \( d_E(u, v) \) their Euclidean distance and by \( d_G(u, v) \) their graph distance. The problem of finding upper bounds on \( d_G(u, v) \) in terms of \( d_E(u, v) \) has received a lot of attention in the literature [1, 2, 6, 8]. In this paper, we improve these upper bounds for values of \( r = \omega(\sqrt{\log n}) \) (i.e. 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 \( d_G(u, v) \) in terms of \( d_E(u, v) \).

Keywords: Random geometric graphs, Graph distance, Euclidean distance, Diameter.

1. Introduction

Given a positive integer \( n \), and a non-negative real \( r \), we consider a random geometric graph \( G \in \mathcal{G}(n, r) \) defined as follows. The vertex set \( V \) of \( G \) is obtained by choosing \( n \) points independently and uniformly at random in the square \( S_n = [-\sqrt{n}/2, \sqrt{n}/2]^2 \) (Note that, with probability 1, no point in \( S_n \) is chosen more than once, and thus we assume \( |V| = n \)). For notational purposes, we identify each vertex \( v \in V \) with its corresponding geometric position \( v = (v_x, v_y) \in S_n \), where \( v_x \) and \( v_y \) denote the usual \( x \)- and \( y \)-coordinates in \( S_n \). Finally, the edge set of \( G \in \mathcal{G}(n, r) \) is constructed by connecting each pair of vertices \( u, v \) by an edge if and only if \( d_E(u, v) \leq r \), where \( d_E \) denotes the Euclidean distance in \( S_n \).

Random geometric graphs were first introduced in a slightly different setting by Gilbert [3] to model the communications between radio stations. Since then, several closely related variants on 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 the monograph by Penrose [10].

The properties of \( \mathcal{G}(n, r) \) are usually investigated from an asymptotic perspective, as \( n \) grows to infinity and \( r = r(n) \). Throughout the paper, we use the following standard notation for the asymptotic behavior of sequences of non-negative numbers \( a_n \) and \( b_n \): \( a_n = O(b_n) \) if \( \limsup_{n \to \infty} a_n/b_n \leq C < +\infty \); \( a_n = \Omega(b_n) \) 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 \( \lim_{n \to \infty} a_n/b_n = 0 \). Finally, a sequence of events \( H_n \) holds asymptotically almost surely (a.a.s.) if \( \lim_{n \to \infty} \Pr(H_n) = 1 \).

1991 Mathematics Subject Classification. Primary: 05C80.
Key words and phrases. random graphs.
Partially supported by the CYCIT: TIN2007-66523 (FORMALISM).
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 $d_G(u,v) \geq d_E(u,v)/r$ deterministically, 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.** The particular problem has risen quite a bit of interest in recent years. Given any two $v, u \in V$, most of the work related to this problem has been devoted to study upper bounds on $d_G(u,v)$ in terms of $d_E(u,v)$ and $r$, that hold a.a.s. Ellis, Martin and Yan [2] showed that there exists some large constant $K$ such that for every $r \geq r_c$, $G \in \mathcal{G}(n,r)$ satisfies a.a.s. $d_G(u,v) \leq Kd_E(u,v)/r$ for every $u$ and $v$. This result was extended by Bradonjic 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 $d_E(u,v) = \Omega((\log n/r)^2)$. Friedrich, Sauerwald and Stauffer [6] improved this last result by showing that the result holds a.a.s. for every $u$ and $v$ satisfying $d_E(u,v) = \Omega((\log n/r)^2)$. They also proved that if $r = o(r_c)$, a linear upper bound of $d_G(u,v)$ in terms of $d_E(u,v)$ is no longer possible. In particular, a.a.s. there exist vertices $u$ and $v$ with $d_E(u,v) \leq 3r$ 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 $\text{diam}(G)$, 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 to the problem of broadcasting information (see [1, 6]).

Figure 1. Graph distance vs. Euclidean distance between two points $u$ and $v$ in $V$.

**Theorem 1.1.** Let $G \in \mathcal{G}(n,r)$ be a random geometric graph with $r \geq r_c$. A.a.s., for every pair of vertices $u, v \in V(G)$ we have:

(i) if $d_E(u,v) \geq 20r \log n$, then $d_G(u,v) \geq \frac{d_E(u,v)}{r} \left(1 + \frac{1}{2(rd_E(u,v))^{2/3}}\right)$, and

---

The result is stated in the unit ball random geometric graph model, but can be adapted to our setting.
In order to prove (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. For the proof of (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 1.2. (1) We do not know if the condition $d_E(u,v) \geq 20r \log n$ in the lower bound in (i) can be improved. (2) The constant 70 in the condition $r \geq 70\sqrt{\log n}$ of (ii) (as well as those in the definition of $\gamma$) is not optimized, and could be made slightly smaller. However, our method as it is, cannot be extended all the way down to $r \geq \sqrt{\log n/\pi} = r_c$. (3) The error term in (ii) is $\gamma r^{-4/3} = \Theta\left(\max\left\{\frac{\log n}{r^2 + rd_E(u,v)}^{2/3}, \frac{\sqrt{\log n}}{r}^4, r^{-4/3}\right\}\right)$, which is $o(1)$ iff $r = \omega(\sqrt{\log n}) = \omega(r_c)$. Hence, for $r = \omega(r_c)$, statement (ii) implies that a.a.s.

$$d_G(u,v) \leq \left\lceil (1 + o(1)) \frac{d_E(u,v)}{r}\right\rceil,$$

thus improving the result in [2].

Theorem 1.1 gives an upper bound on the diameter as a corollary. First, observe that $d_E(u,v) \leq \sqrt{2n}$. From Theorem 10 in [2] for the particular case $d = 2$, one can easily deduce that if $r \geq r_c$ a.a.s.

$$\text{diam}(G) \leq \sqrt{\frac{2n}{r}} \left(1 + O\left(\sqrt{\frac{\log \log n}{\log n}}\right)\right). \quad (1)$$

Moreover, note that a.a.s. there exist $u$ and $v$ at distance $d_E(u,v) \geq \sqrt{2n} - 2\sqrt{2} \log n$; the probability that the squares of side $\log n$ at the corners of $S_n$ contain no vertices is $o(1)$. Applying Theorem 1.1 to these vertices $u$ and $v$, we obtain the following result.

\textbf{Corollary 1.3.} Let $G \in \mathcal{G}(n,r)$ be a random geometric graph with $r \geq 70\sqrt{\log n}$. A.a.s. we have

$$\text{diam}(G) \leq \sqrt{\frac{2n}{r}} \left(1 + \gamma r^{-4/3}\right),$$

where $\gamma$ is defined as in Theorem 1.1.

Observe that our bound on the diameter stated in Corollary 1.3 improves the one in (1) derived from [2] whenever $r = \Omega\left(\frac{\log^{5/8} n}{(\log \log n)^{1/8}}\right)$. 

(ii) if $r \geq 70\sqrt{\log n}$, then $d_G(u,v) \leq \left\lceil \frac{d_E(u,v)}{r} \left(1 + \gamma r^{-4/3}\right)\right\rceil$
where $\phi_\delta > 0$. Then, for every $\delta > 0$ we have
\[
\Pr(X \geq (1 + \delta)\mathbb{E}(X)) \leq \left(\frac{1 + \delta}{e^\delta}\right)^N,
\]
and for any $0 < \delta < 1$ we have
\[
\Pr(X \leq (1 - \delta)\mathbb{E}(X)) \leq \left((1 - \delta)e^\delta\right)^N.
\]

**Proof.** By Markov’s inequality, we have for every $\beta > 0$
\[
\Pr(X \geq (1 + \delta)\mathbb{E}(X)) = \Pr(e^{\beta X} \geq e^{\beta(1 + \delta)\mathbb{E}(X)}) \leq \prod \frac{\mathbb{E}(e^{\beta X_i})}{e^{\beta(1 + \delta)\mathbb{E}(X)}} = (\varphi_{X_i}(\beta))^N e^{-\beta(1 + \delta)N/\mu},
\]
where $\varphi_{X_i}(\beta) = \mathbb{E}(e^{\beta X_i}) = \frac{\mu}{\mu - \beta}$ is the moment-generating function of an exponentially distributed random variable with parameter $\mu$. Thus,
\[
\Pr(X \geq (1 + \delta)\mathbb{E}(X)) \leq \left(\frac{\mu}{\mu - \beta}\right)^N e^{-\beta(1 + \delta)N/\mu}
\]
\[
= \exp \left(N \left(- \log \left(1 - \frac{\beta}{\mu}\right) - (1 + \delta)\frac{\beta}{\mu}\right)\right).
\]
Setting $\frac{\beta}{\mu} = \frac{\delta}{1 + \delta}$, we have
\[
\Pr(X \geq (1 + \delta)\mathbb{E}(X)) \leq \exp (N \log (1 + \delta - \delta)) = \left(\frac{1 + \delta}{e^\delta}\right)^N.
\]
The lower tail is proved similarly. $\square$

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 [10] for a detailed account of the subject). Here we sketch it.

Consider the following related model of a random geometric graph given vertices $u$ and $v$. Let $V = \{u, v\} \cup V'$, where $V'$ is a set obtained as a homogeneous Poisson point process of intensity 1 in the square $S_n$ of area $n$. In other words, $V'$ consists of $N$ points in the square $S_n$ chosen independently and uniformly at random, where $N$ is a Poisson random variable of mean $n$. We add two labelled vertices $u$ and $v$, whose position is also selected independently and uniformly at random in $S_n$. Exactly as we did for the model $\mathcal{G}(n, r)$, we connect by an edge $u$ and $v$ in $V$ if $d_G(u, v) \leq r$. We denote this new model by $\mathcal{G}_{u,v}(n, r)$.

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 $A \subseteq S_n$ of area $a$ has a Poisson distribution with mean $a$; and the number of points of $V'$ in disjoint regions of $S_n$ are independently distributed. Moreover, by conditioning $\mathcal{G}_{u,v}(n, r)$ upon the event $N = n - 2$, we recover the original distribution of $\mathcal{G}(n, r)$. Therefore, since $\Pr(N = n - 2) = \Theta(1/\sqrt{n})$, any event holding in $\mathcal{G}_{u,v}(n, r)$ with probability at least $1 - o(n)$ must hold in $\mathcal{G}(n, r)$ with probability at least $1 - o(\sqrt{n})$.

We make use of this property throughout the article, and do all the analysis for a graph $G \in \mathcal{G}_{u,v}(n, r)$.

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 $X_1, \ldots, X_N$ be independent exponential random variables and let $X = X_1 + \cdots + X_N$. Then, for every $\delta > 0$ we have
\[
\Pr(X \geq (1 + \delta)\mathbb{E}(X)) \leq \left(\frac{1 + \delta}{e^\delta}\right)^N,
\]
and for any $0 < \delta < 1$ we have
\[
\Pr(X \leq (1 - \delta)\mathbb{E}(X)) \leq \left((1 - \delta)e^\delta\right)^N.
\]
2.1. Proof of statement (i). Our argument in this subsection depends only on the Euclidean
distance between \( u \) and \( v \), but not on their particular position in \( S_n \). Thus, let \( t = d_E(u, v) \) and
assume without loss of generality that \( u = (0,0) \) and \( v = (t,0) \).

The next lemma shows that short paths between vertices are contained in small strips. It is stated
in the more general context of a geometric graph \( G = (V, E) \) of radius \( r \), where the vertex set \( V \) is
a subset of points in the square \( S_n \) (not necessarily randomly placed), and edges connect (as usual)
every pair of vertices at Euclidean distance at most \( r \). For every \( \alpha > 0 \), consider the rectangle \( R = [0, t] \times \lbrack -\alpha, \alpha \rbrack \).

Lemma 2.2. Let \( G = (V, E) \) be a geometric graph with radius \( r \) in \( S_n \), and let \( u, v \in V \) such that
\( u = (0,0) \) and \( v = (t,0) \). Suppose that \( t = d_E(u, v) \geq kr - \frac{2\alpha^2}{kr} \), for some \( k \in \mathbb{Z}^+ \) and \( \alpha = o(kr) \). Then
all paths of length at most \( k \) from \( u \) to \( v \) are contained in \( R \).

Proof. Suppose that there exists a path from \( u \) to \( v \) in at most \( k \) steps. Let \( z = (a, b) \) the vertex with
largest \( y \)-coordinate in that path. Since \( \alpha \in [0, t] \), for any \( b \) we have,
\[
kr \geq \sqrt{a^2 + b^2} + \sqrt{(t-a)^2 + b^2} \geq 2\sqrt{t^2/4 + b^2}.
\]
Therefore,
\[
\frac{(kr)^2}{4} \geq \frac{t^2}{4} + b^2 \geq \frac{(kr - \frac{2\alpha^2}{kr})^2}{4} + b^2,
\]
where we used that \( t \geq kr - \frac{2\alpha^2}{kr} \). Using that \( \alpha = o(kr) \) we have
\[
b \leq \alpha \sqrt{1 - \alpha^2/(kr)^2} = (1 - o(1)) \alpha.
\]
Repeating the same argument for the vertex with smallest \( y \)-coordinate, we conclude that the path is
contained in \( R = [0, t] \times \lbrack -\alpha, \alpha \rbrack \).

Proposition 2.3. Let \( G \in \mathcal{G}_{u,v}(n, r) \) be a random geometric graph on \( S_n \), with \( u = (0,0) \) and \( v =
(t,0) \). Then, for every \( 0 < \delta < 2^{-1/3} \), we have that
\[
\Pr \left( d_G(u, v) \leq \frac{t}{r} \left( 1 + \frac{\delta}{(tr)^{2/3}} \right) \right) \leq \frac{t}{r} \exp \left( -\sqrt{\delta/2} (tr)^{2/3} \right) + \exp \left( -(1 - \sqrt{2\delta^3})^2 \frac{t}{2r} \right). \quad (2)
\]

Proof. Let \( k = d_G(u, v) \) and let \( \alpha = \sqrt{\delta/2} ((d_G(u, v))^{3}r^2/t)^{1/3} = o(d_G(u, v)r) \). Consider the event \( A_\alpha \)
that all the paths from \( u \) to \( v \) of length \( k \) are contained in the rectangle \( R = [0, t] \times \lbrack -\alpha, \alpha \rbrack \) and let \( B \)
the event defined by condition [2]. If \( B \) holds, then
\[
t \geq \frac{kr}{1 + \frac{\delta}{(tr)^{2/3}}} \geq kr \left( 1 - \frac{\delta}{(tr)^{2/3}} \right) = kr - \frac{2\alpha^2}{kr}.
\]
Since \( \alpha = o(kr) \), by Lemma 2.2 \( \Pr(B|A_\alpha) = 0 \).

Denote by \( v_1 \) the vertex with largest \( x \)-coordinate inside the rectangle \( R_1 = [0, r] \times [-\alpha, \alpha] \) (possibly \( v_1 = u \) if \( R_1 \) contains no other vertices of \( \mathcal{G}_{u,v}(n, r) \)). Note that \( v_1 \) might not be connected to \( u \), but observe that its \( x \)-coordinate is always greater or equal to the \( x \)-coordinate of any vertex \( u_1 \in R_1 \)
connected to \( u \) (see Figure 2). Let \( x_1 \) be the \( x \)-coordinate of \( v_1 \), and define the random variable \( a_1 = r - x_1 \). By definition, \( 0 \leq a_1 \leq r \). Since \( G \in \mathcal{G}_{u,v}(n, r) \), the number of vertices from \( V \) inside
a region of \( S_n \) is a Poisson random variable with mean equal to the area of that region. Hence, the
random variable \( a_1 \) satisfies

\[
\Pr(a_1 \geq \beta) = \begin{cases} 
    e^{-2\alpha\beta} & \text{if } 0 \leq \beta \leq r \\
    0 & \text{if } \beta > r.
\end{cases}
\] (3)

Thus, \( a_1 \) is stochastically dominated by an exponentially distributed random variable \( \tilde{a}_1 \) of parameter \( 2\alpha \). We assume that \( a_1 \) and \( \tilde{a}_1 \) are coupled together in the same probability space, so that \( a_1 = \min\{\tilde{a}_1, r\} \leq \tilde{a}_1 \).

We proceed to define in a similar way the points \( v_i \) and the values \( x_i \) and \( a_i \), for any \( 2 \leq i \leq k \). Let \( v_i \) be the vertex with largest \( x \)-coordinate inside the rectangle \( R_i = (x_{i-1} + a_{i-1}, x_{i-1} + r] \times [-\alpha, \alpha] \), and let \( x_i \) be the \( x \)-coordinate of \( v_i \). Define \( a_i = x_{i-1} + r - x_i \). If \( R_i \) contains no vertex of \( \tilde{G}_{u,v}(n, r) \), then add an extra vertex \( v_i = (x_{i-1} + a_{i-1}, 0) \) (so in that case \( x_i = x_{i-1} + a_{i-1} \) and \( a_i = r - a_{i-1} \)).

Observe that \( 0 \leq a_i \leq r - a_{i-1} \), so the rectangles \( R_1, R_2, \ldots, R_k \) are disjoint. Moreover,

\[
\Pr(a_i \geq \beta) = \begin{cases} 
    e^{-2\alpha\beta} & \text{if } 0 \leq \beta \leq r - a_{i-1} \\
    0 & \text{if } \beta > r - a_{i-1},
\end{cases}
\] (4)

for every \( 1 \leq i \leq k \) (by defining \( a_0 = 0 \)). Therefore, \( a_1, a_2, \ldots, a_k \) are stochastically dominated by a sequence \( \tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_k \) of i.i.d. exponentially distributed random variables of parameter \( 2\alpha \), such that \( a_i = \min\{\tilde{a}_i, r - a_{i-1}\} \leq \tilde{a}_i \) for all \( 1 \leq i \leq k \).

![Figure 2](image)

Figure 2. The points \( u_1, v_1 \) coincide but \( u_2 \) and \( v_2 \) do not

Note that the vertices \( u, u_1, v_2, \ldots, v_k \) may not induce a connected path in \( \tilde{G}_{u,v}(n, r) \), since the Euclidean distance between two consecutive ones may be greater than \( r \). However, the fact that \( v_i \) is the vertex with largest \( x \)-coordinate inside \([0, x_{i-1} + r] \times [-\alpha, \alpha]\) and together with a straightforward induction argument yield to the following claim: if \( u = u_0, u_1, u_2, \ldots, u_k \) is a path contained in \( R \), then for every \( 1 \leq i \leq k \) the \( x \)-coordinate of \( u_i \) is at most \( x_i \) (see again Figure 2). We will now show that \( x_k < t \) with the desired probability.

Define

\[
a = \sum_{i=1}^{k} a_i \quad \text{and} \quad \tilde{a} = \sum_{i=1}^{k} \tilde{a}_i.
\]

Expanding recursively from the relations \( x_i = x_{i-1} + r - a_i \) and \( x_1 = r - a_1 \), we get

\[
x_k = \sum_{i=1}^{k} (r - a_i) = kr - a.
\]
ON THE RELATION BETWEEN GRAPH DISTANCE AND EUCLIDEAN DISTANCE IN RANDOM GEOMETRIC GRAPHS

Let us consider the event that \( \tilde{a}_i \leq r/2 \) for all \( 1 \leq i \leq k \). In particular, this event implies that \( a_i = \tilde{a}_i \) for all \( i \), and therefore \( x_k = kr - \tilde{a} \). Since each \( \tilde{a}_i \) is exponentially distributed with parameter \( 2\alpha \),

\[
\Pr(\exists i : a_i > r/2) \leq k \Pr(a_1 > r/2) = ke^{-\alpha r},
\]

so \( x_k = kr - \tilde{a} \) with probability at least \( 1 - ke^{-\alpha r} \).

Moreover, notice that

\[
\mathbb{E}(\tilde{a}) = k\mathbb{E}(\tilde{a}_1) = \frac{k}{2\alpha}.
\]

By Lemma 2.1 for all \( 0 < \varepsilon < 1 \),

\[
\Pr\left(\tilde{a} \leq (1 - \varepsilon)\frac{k}{2\alpha}\right) \leq ((1 - \varepsilon)e)^k < e^{-\varepsilon^2 k/2}.
\]

where we have used that \( \log(1 - x) < -x - \frac{x^2}{2} \), for any \( 0 < x < 1 \).

If \( t < kr \left(1 - \frac{\delta}{(tr)^{2/3}}\right) = kr(1 + o(1)) \), by definition of \( B \), \( \Pr(B) = 0 \) and we are done. Thus, we may assume that \( t \geq kr(1 + o(1)) \). For \( \delta = \frac{1 - \sqrt{2}}{\sqrt{2}} \), from (6) and (5),

\[
\Pr(x_k \geq t) \leq \Pr\left(x_k \geq kr\left(1 - \frac{\delta}{(tr)^{2/3}}\right)\right)
\]

\[
\leq \Pr\left(x_k \geq kr - \frac{\sqrt{2}\delta^{3/2}kr}{t} \cdot \frac{k}{2\alpha}\right)
\]

\[
\leq \Pr\left(x_k \geq kr - (1 - \varepsilon + o(1))\frac{k}{2\alpha}\right)
\]

\[
\leq \Pr(\exists i : a_i > r/2) + \Pr\left(\forall i : a_i \leq r/2 \quad \text{and} \quad a \leq (1 - \varepsilon + o(1))\frac{k}{2\alpha}\right)
\]

\[
\leq \Pr(\exists i : a_i > r/2) + \Pr\left(\tilde{a} \leq (1 - \varepsilon + o(1))\frac{k}{2\alpha}\right)
\]

\[
\leq ke^{-\alpha r} + e^{\varepsilon^2 k/2}
\]

\[
\leq \frac{t}{r} \exp\left(-\sqrt{\delta/2(tr)^{2/3}}\right) + \exp\left(-(1 - \sqrt{2}\delta^{3/2})^2 \frac{t}{2r}\right).
\]

Hence, if \( A_a \) holds, then \( \Pr(B) \leq \Pr(x_k \geq t) \leq \frac{t}{r} \exp\left(-\sqrt{\delta/2(tr)^{2/3}}\right) + \exp\left(-(1 - \sqrt{2}\delta^{3/2})^2 \frac{t}{2r}\right) \)

and if \( \overline{A_a} \) holds, then \( \Pr(B) = 0 \). Thus, the proposition follows.

Proposition 2.4. Let \( \tilde{G}_{u,v}(n, r) \) be a random geometric graph in \( S_n \) with labelled vertices \( u \) and \( v \) such that \( d_E(u, v) \geq 20r \log n \). Then we have

\[
d_G(u, v) \leq \frac{d_E(u, v)}{r} \left(1 + \frac{1}{2(rd_E(u, v))^{2/3}}\right),
\]

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

Proof. As before, let \( t = d_E(u, v) \) and \( k = d_G(u, v) \). Also let \( \delta = 1/2 \).

Since \( t \geq 20r \log n \) and \( r \geq r_c = \Omega(\sqrt{\log n}) \), we have

\[
\sqrt{\delta/2(tr)^{2/3}} - \log(t/r) = \Omega(\log^{4/3} n),
\]
and
\[
\left(1 - \sqrt{2\delta^3}\right)^2 \frac{t}{2r} \geq \frac{5}{2} \log n.
\]

By Proposition 2.3, this implies that
\[
\Pr\left(k \leq \frac{t}{r} \left(1 + \frac{1}{2(\sqrt{r})^{2/3}}\right)\right) = o(n^{-5/2}.
\]
□

To finish the proof of statement (i) in Theorem 1.1, by de-Poissonizing \(\tilde{G}_{u,v}(n,r)\), we have that in \(\tilde{G}(n,r)\), statement (i) in Theorem 1.1 holds for our choice of \(u\) and \(v\), with probability at least \(1 - o(n^{-2})\). Note that this fact does not depend on the particular location of \(u\) and \(v\) in \(S_n\). The statement follows by taking a union bound over all at most \(n^2\) pairs of vertices.

2.2. Proof of statement (ii). As in Subsection 2.1, we pick two points in \(S_n\), and put \(t = d_E(u,v)\).

Let \(\gamma\) be as in the statement of Theorem 1.1. We assume first that \(u = (0,0)\) and \(v = (t,0)\), and consider a Poisson point process in the rectangle \(R = [0,t] \times [0,\alpha]\), for a certain \(\alpha \leq r\) that will be made precise later.

Let \(\tilde{G}_{R,u,v}(n,r)\) denote the random geometric graph on the rectangle \(R\), to which the points \(u\) and \(v\) are added. We will show that the probability of having \(d_G(u,v) \geq d_E(u,v) + \frac{d_E(u,v)}{r} (1 + \delta r^{-4/3})\) decays exponentially in \(\delta\). For each point \(z\) in \(R\) with \(x\)-coordinate \(s\), define the rectangle
\[
R_z = [s, s + \rho] \times [0,\alpha], \quad \text{where} \quad \rho = r - \frac{\alpha^2}{r}.
\]
We need the following auxiliary lemma.

Lemma 2.5. For any vertex \(z\) in \(R\), all vertices in \(R_z\) are connected to \(z\) (see Figure 3).

Proof. It is enough to show that the upper-left and the bottom-right corner of \(R_z\) are at distance at most \(r\). Then all vertices inside \(R_z\) are connected to one another, and in particular \(z\) is connected to every vertex in \(R_z\). A sufficient condition for that is
\[
\sqrt{\rho^2 + \alpha^2} \leq r,
\]
or equivalently
\[
\rho (1 - (\alpha/r)^2) = r \sqrt{r^2 - \alpha^2} = r \sqrt{1 - (\alpha/r)^2}.
\]
Since \(\sqrt{1-x} > 1-x\) for any \(0 < x < 1\), the lemma follows. □

![Figure 3. The rectangle \(R_z\)](image)

Proposition 2.6. Let \(\tilde{G}_{R,u,v}(n,r)\) be a random geometric graph on \(R\), with \(u = (0,0)\) and \(v = (t,0)\). Let \(F > 0\) and \(J > \frac{3(F+1)}{2^{2/3}}\) be constants and define \(g(x) = x - \log(1+x)\). Then, for every \(J \leq \delta \leq Fr^{4/3},\)
we have that
\[
\Pr \left( d_G(u, v) > \left[ \frac{t}{r} \left( 1 + \delta r^{-4/3} \right) \right] \right) \leq n \exp \left( -\frac{(F + 1)\delta^{1/2}r^{4/3}}{2J^{3/2}} \right) + \exp \left( -g \left( \delta/J \right)^{3/2} \right) \frac{t}{r}.
\]

**Proof.** Set \( C = 1/J^{3/2} \), and let \( B \) be any positive constant satisfying
\[
B^2 + C/B \leq 1/(F + 1).
\]

Some elementary analysis shows that such \( B \) must exist. In fact, the equation \( B^2 + C/B = 1/(F + 1) \) has exactly two positive solutions \( B_1 \) and \( B_2 \) for any \( 0 < C < \frac{2}{(3(F + 1))^{3/2}} \), and any \( 0 < B_1 \leq B \leq B_2 < 1/\sqrt{F + 1} \) satisfies (7).

Let us consider the integer \( k = \left\lceil \frac{t}{r} \left( 1 + \delta r^{-4/3} \right) \right\rceil \). We will show that with very high probability there exists a path length at most \( k \) between \( u \) and \( v \). Such a path will only use vertices inside \( R \), but for technical reasons (the last of the rectangles \( R_i \) defined below might be further to the right than the point \((t, 0)\) or possibly be outside of the square) of the argument we extend the Poisson point process of our probability space to the semi-infinite strip \( R_\infty = [0, \infty) \times [0, \alpha] \).

We construct a sequence of vertices in a similar way as in the proof of Proposition 2.3. Set \( v_0 = u, x_0 = 0 \) and \( a_0 = 0 \). We make the choice of \( \alpha \) for this subsection now more precise. We set
\[
\alpha = B\delta^{1/2} r^{1/3},
\]
for some constant \( 0 < B < 1/\sqrt{F + 1} \) satisfying (7). Observe that the restriction \( \delta \leq Fr^{4/3} \) implies that
\[
\alpha \leq (B\sqrt{F})r < r,
\]
so our choice of \( \alpha \) is feasible, and moreover
\[
\rho = r - \alpha^2/r \geq (1 - B^2F)r. \tag{9}
\]

For each \( 1 \leq i \leq k - 1 \), define \( R'_i = (x_{i-1} + a_{i-1}, x_{i-1} + \rho] \times [0, \alpha] \), and let \( v_i \) be the vertex with largest \( x \)-coordinate inside \( R'_i \) (if \( R'_i \) is empty, then add an extra vertex \( v_i = (x_{i-1} + a_{i-1}, 0) \)). Define \( x_i \) to be the \( x \)-coordinate of \( v_i \) and \( a_i = x_{i-1} + \rho - x_i \). By the same considerations as in the proof of Proposition 2.3 (but replacing \( R_i \) by \( R'_i \), \( r \) by \( \rho \), and \( k \) by \( k - 1 \)), we deduce that \( a_1, a_2, \ldots, a_{k-1} \) are stochastically dominated by a sequence \( \tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_{k-1} \) of i.i.d. exponentially distributed random variables of parameter \( \alpha \), such that \( a_i = \min \{ \tilde{a}_i, \rho - a_{i-1} \} \leq \tilde{a}_i \) for all \( 1 \leq i \leq k - 1 \). Moreover, since \( \alpha \rho/2 \geq (1 - B^2F)\alpha r/2 = (1 - B^2F)B\delta^{1/2}r^{4/3}/2, \) we have
\[
\Pr(\tilde{a}_i > \rho/2) = e^{-\alpha \rho/2} \leq e^{-(1 - B^2F)B\delta^{1/2}r^{4/3}/2}.
\]

Thus, with probability at least \( 1 - ke^{-((1 - B^2F)B\delta^{1/2}r^{4/3})/2} \), for every \( 1 \leq i \leq k - 1 \) we have \( \tilde{a}_i \leq \rho/2 \), and therefore \( a_i = \tilde{a}_i \). This event implies that
\[
x_{k-1} = (k - 1)\rho - \tilde{a}, \tag{10}
\]
where \( \tilde{a} = \sum_{i=1}^{k-1} \tilde{a}_i \), and also that we did not add any extra vertices (i.e. all \( v_1, \ldots, v_{k-1} \) belong to the Poisson point process in \([0, \infty) \times [0, \alpha] \)).

By construction, each \( x_i \) belongs to the rectangle \( R'_i \subset R_{v_{i-1}} \) for every \( 1 \leq i \leq k - 1 \). Hence, by Lemma 2.5, the vertices \( v_0, v_1, \ldots, v_{k-1} \) form a connected path.
In view of all that, it suffices to show that \( x_{k-1} + \rho \geq t \) with sufficiently large probability. Note that, if this event holds, then \( v \) must belong to \( R_j \) for some \( 0 \leq j \leq k-1 \), and therefore \( u = v_0, v_1, v_2, \ldots, v_j, v \) is a connected path of length \( j + 1 \leq k \). (Observe that such a path is contained in \( R \), so our extension of the Poisson point process to \( R_\infty \) turned out to be harmless.)

Recall that \( C = 1/J^{3/2} \). Using the upper-tail bound in Lemma 2.1, we obtain

\[
\Pr \left( \tilde{a} \geq (1 + C\delta^{3/2}, k \frac{2}{2\alpha}) \right) < \left( \frac{1 + C\delta^{3/2}}{\alpha} \right)^k \leq e^{-kg((\delta/J)^{3/2})}.
\]

Combining this together with (10), we infer that, with probability at least

\[
1 - ke^{-\left(1-B^2F\right)B\delta^{3/2+r/3}} - e^{-kg((\delta/J)^{3/2})},
\]

we have

\[
x_{k-1} + \rho = k\rho - \tilde{a} > k\rho - \frac{(1 + C\delta^{3/2}) k}{2\alpha} = kr \left( 1 - \frac{\alpha \rho}{r^2} - \frac{(1 + C\delta^{3/2})}{2\alpha r} \right). \tag{11}
\]

From the definition of \( k \), the range of \( \delta \) and since \( \alpha = B\delta^{3/2+r/3} \), the event above implies

\[
x_{k-1} + \rho > t(1 + \delta r^{-4/3}) \left( 1 - \delta r^{-4/3} \left( B^2 + \frac{(\delta -3/2 + C)}{2B} \right) \right)
\]

\[
\geq t \left( 1 + \delta r^{-4/3} \left( 1 - (\delta r^{-4/3} + 1) \left( B^2 + \frac{C}{B} \right) \right) \right)
\]

\[
\geq t \left( 1 + \delta r^{-4/3} \left( 1 - (F + 1) \left( B^2 + \frac{C}{B} \right) \right) \right) > t,
\]

so

\[
\Pr(d_G(u, v) > k) \leq ke^{-\left(1-B^2F\right)B\delta^{3/2+r/3}} + e^{-kg((\delta/J)^{3/2})}
\]

\[
\leq ne^{-\left(1-B^2F\right)B\delta^{3/2+r/3}} + e^{-g((\delta/J)^{3/2})t/r} \tag{12}
\]

\[
\leq ne^{-\left(1-B^2(F+1)\right)B\delta^{3/2+r/3}} + e^{-g((\delta/J)^{3/2})t/r}
\]

\[
\leq ne^{-C(F+1)\delta^{3/2+r/3}} + e^{-g((\delta/J)^{3/2})t/r}, \tag{13}
\]

as desired. On the last step we used the fact that \( (1 - B^2(F + 1))B \geq C(F + 1) \), which easily follows from (7). This completes the proof of the proposition. Note that (12) may be stronger than (13) if we choose a constant \( B \) which satisfies (7) and maximises \( (1 - B^2F)B \).

**Proposition 2.7.** Let \( \gamma \) as in the statement of Theorem 1.1 and let \( \mathcal{G}_{R,u,v}(n, r) \) be a random geometric graph on \( R \), with \( u = (0, 0) \) and \( v = (t, 0) \). Suppose that \( r \geq 70\sqrt{\log n} \). Then we have

\[
d_G(u, v) > \frac{t}{r} \left( 1 + 70\gamma(rt)^{-2/3} \right),
\]

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

**Proof.** First, observe that, if \( t \leq r \), then \( d_G(u, v) = 1 \), and the statement holds trivially. Thus, we assume henceforth that \( t > r \).
Set $B = 47/50$, $C = 10^{-2}$, $F = 23/200$, $D = 70$, $E = 31$ and $J = 10^{-4/3}$. Recall that

$$
\gamma = \max \left\{ E \left( \frac{2r \log n}{r + t} \right)^{2/3}, D \frac{\log^2 n}{r^{8/3}}, 3^{2/3} J \right\}.
$$

We want to apply Proposition 2.6 with $\delta = \gamma$. It is straightforward to check that the restrictions (7) and $J > \frac{3(F + 1)}{277/3}$, required in Proposition 2.6, hold. We also need to show that $J \leq \gamma \leq Fr^{4/3}$. Notice that $\frac{D \log^2 n}{r^{8/3}} \leq Fr^{4/3}$, since $r \geq 70 \sqrt{\log n} \geq (D/F)^{1/4} \sqrt{\log n}$; also $E \left( \frac{2r \log n}{r + t} \right)^{2/3} \leq Fr^{4/3}$, since $r(r + t)/(2 \log n) > r^2/\log n \geq 4900 \geq (E/F)^{3/2}$; and finally $3^{2/3} J \leq Fr^{4/3}$ since $r = O(\sqrt{n})$.

Moreover, $\delta \geq 3^{2/3} J \geq J$.

Note that this choice of constants combined with (8) and (9) implies

$$
\alpha \leq r/3 \quad \text{and} \quad \rho \geq 8r/9 \geq 8\alpha/3.
$$

The proof concludes by applying (12) in the proof of Proposition 2.6 with this given $\delta$, showing that the upper bound on $\Pr(d_G(u, v) > k) = o(n^{5/2})$. On the one hand, $\delta \geq \frac{D \log^2 n}{r^{8/3}}$ implies

$$
\frac{(1 - B^2 F) B \delta^{1/2} r^{4/3}}{2} - \log n \geq \frac{(1 - B^2 F) B D^{1/2} \log n}{2} - \log n > \frac{7.01}{2} \log n - \log n = \frac{5.01}{2} \log n.
$$

On the other hand, $\delta \geq E(r \log n/t)^{2/3}$ and $\delta \geq 3^{2/3} J$ imply

$$
g \left( \frac{(\delta/J)^{3/2}}{r} \right) > \frac{(\delta/J)^{3/2}}{2r} \geq \frac{3}{2} C E^{3/2} \log n > \frac{5.17}{2} \log n,
$$

where we have used that $g(x) \geq x/2$ if $x \geq 3$.

Therefore, $\Pr(d_G(u, v) > k) \leq n^{-5.01/2} + n^{-5.17/2} = o(n^{5/2})$. \hfill \square

**Corollary 2.8.** Statement (ii) in Theorem 1.1 is true.

**Proof.** Observe that from the proof of Proposition 2.6 together with (14), $x_1 \geq \rho/2 > 4\alpha/3$ with probability at least $1 - o(n^{-5/2})$. In particular, this event implies that $v_1$ is outside of the square $[0, 1.01 \alpha] \times [0, \alpha]$. Moreover, also with probability $1 - o(n^{-5/2})$, we can find some point $\hat{v}_j$ in $[t - 1.01 \alpha - r/2, t - 1.01 \alpha] \times [0, \alpha]$. It may happen that $v_j$ lies in $[t - 1.01 \alpha, t] \times [0, \alpha]$. However, in that case, we can replace $v_j$ with $\hat{v}_j$, and therefore we found a $u$-$v$ path of length $j + 1 \leq k$ with all internal vertices in $[1.01 \alpha, t - 1.01 \alpha] \times [0, \alpha]$. Indeed, we will show now that we can always fit such a rectangle $R' = [1.01 \alpha, t - 1.01 \alpha] \times [0, \alpha]$, suitably rotated and translated, into the square. We need first a few definitions.

Consider two points $u = (x_u, y_u)$ and $v = (x_v, y_v)$ in $\mathbb{R}^2$. By symmetry we may assume that $x_u < x_v$ and $y_u \leq y_v$. Let $\beta$ be the angle of the vector $u \hat{v}$ 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 $u \hat{v}$, and let $R^-$ be the rectangle to the right of $u \hat{v}$. We will show that at least one of these rectangles contains a copy of $R'$ fully contained in $S_n$.

Notice that the intersection of $R^+$ and $R^-$ with each of the halfplanes $x \leq x_u$, $x \geq x_v$, $y \leq y_u$ and $y \geq y_v$ gives 4 triangles. We call them $T^+_u$, $T^-_u$, $T^+_v$ and $T^-_v$ respectively. All these triangles are right-angled, and denote by $t^+_u$, $t^-_u$, $t^+_v$ and $t^-_v$ the side of the corresponding triangle that it is parallel to the segment $uv$. Notice that $|t^+_u| = |t^-_v|$ and $|t^-_u| = |t^+_v|$. Call a triangle $T^*_w$, with $w \in \{u, v\}$ and
we may assume that $\beta > 0$. Since we assumed that $\beta \leq \pi/4$, we have $|t^+_u| = |t^-_v| = |\alpha| \tan \beta | \leq 1.01\alpha$. 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 $S_n$, 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 $y_u \leq \sqrt{n}/2 - \alpha$, then $T^+_u$ is completely contained in the square, and hence $R^+$ satisfies the conditions. Similarly, if $y_u \geq -\sqrt{n}/2 + \alpha$, $R^-$ satisfies the conditions. Otherwise, $|y_u|, |y_v| \geq \sqrt{n}/2 - \alpha$, and the angle $\beta$ is at least $\arctan \left( \frac{\sqrt{n} - 2\alpha}{n} \right) > \pi/4$, which contradicts our assumption on $\beta$.

Again, by de-Poissonizing $G_{n,r}(n, r)$, we can use Proposition 2.7 to show that for given $u$ and $v$ in $G \in G(n, r)$, statement (ii) in Theorem 1.1 holds with probability at least $1 - o(n^{-2})$. By taking a union bound over all at most $n^2$ possible pairs of vertices, statement (ii) in Theorem 1.1 follows. \(\square\)

### 3. Open Problems

Theorem 1.1 establishes a relation between the graph distance and the Euclidean distance of two vertices $u$ and $v$ in $G(n, r)$ that holds a.a.s. simultaneously for all pairs of vertices. It would be interesting to find better concentration bounds on the values that $d_G(u, v)$ can take with high probability. Also, we would like to characterize the probability distributions of $\mathbb{E}(d_G(u, v) \mid d_E(u, v))$ and $\text{Var}(d_G(u, v) \mid d_E(u, v))$ (i.e. the expectation and variance of $d_G(u, v)$ given $d_E(u, v)$). What can we say about these distributions?

In the proof of statement (ii) in Theorem 1.1 we define a new random variable that stochastically dominates $d_G(u, v)$ and we give an upper bound for the probability that this random variable is too large. This argument can be easily adapted in the case $r = \omega(r_c)$, and provide the upper bound $\mathbb{E}(d_G(u, v) \mid d_E(u, v)) - d_E(u, v)/r = O (1 + \gamma d_E(u, v)r^{-7/3})$. Similarly, the proof of statement (i) in Theorem 1.1 can be adapted to give a lower bound on $\mathbb{E}(d_G(u, v) \mid d_E(u, v))$, but we need the further conditioning upon the event that $d_E(u, v)$ is large enough.

### References


