LEARNING RANDOM POINTS FROM GEOMETRIC GRAPHS OR ORDERINGS

JOSEP DÍAZ, COLIN MCDIARMID, AND DIETER MITSCHE

ABSTRACT. Suppose that there is a family of n random points $\mathbf{X}_v$ for $v \in V$, independently and uniformly distributed in the square $\mathcal{S}_n = [-\sqrt{n}/2, \sqrt{n}/2]^2$ of area n. We do not see these points, but learn about them in one of the following two ways.

Suppose first that we are given the corresponding random geometric graph $G \in \mathscr{G}(n, r)$, where distinct vertices u and v are adjacent when the Euclidean distance $d_E(\mathbf{X}_u, \mathbf{X}_v)$ is at most r. If the threshold distance r satisfies $n^{3/14} \ll r \ll n^{1/2}$, then the following holds with high probability. Given the graph G (without any geometric information), in polynomial time we can approximately reconstruct the hidden embedding, in the sense that, 'up to symmetries', for each vertex v we find a point within distance about r of $\mathbf{X}_v$; that is, we find an embedding with 'displacement' at most about r.

Now suppose that, instead of being given the graph G, we are given, for each vertex v, the ordering of the other vertices by increasing Euclidean distance from v. Then, with high probability, in polynomial time we can find an embedding with the much smaller displacement error $O(\sqrt{\log n})$.

Keywords: Random geometric graphs, unit disk graphs, approximate embedding, vertex orders.

1. INTRODUCTION

In this section, we first introduce geometric graphs and random geometric graphs, the approximate realisation problem for such graphs, and families of vertex orderings; and we then present our main theorems, give a brief sketch of their proofs, and finally give an outline of the rest of the paper.

1.1. Random geometric graphs. Suppose that we are given a non-empty finite set V, and an embedding $\Psi : V \to \mathbb{R}^2$, or equivalently a family $(\mathbf{x}_v : v \in V)$ of points in $\mathbb{R}^2$, where $\Psi(v) = \mathbf{x}_v$. Given also a real *threshold distance* $r > 0$, we may define the *geometric graph* $G = G(\Psi, r)$ or $G = G((\mathbf{x}_v : v \in V), r)$ with vertex set V by, for each pair u, v of distinct elements of V, letting u and v be adjacent if and only if $d_E(\mathbf{x}_u, \mathbf{x}_v) \leq r$. Here d_E denotes Euclidean distance, $d_E(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|_2$. Note that the (abstract) graph G consists of

its vertex set V and its edge set (with no additional geometric information). A graph is called *geometric* if it may be written as $G(\Psi, r)$ as above, and then (Ψ, r) is called a *realisation* of the graph. Since we may rescale so that $r = 1$, a geometric graph may also be called a *unit disk graph* (UDG) [12].

Given a positive integer n, and a real $r > 0$, the *random geometric graph* $G \in \mathscr{G}(n, r)$ with vertex set $V = [n]$ is defined as follows. Start with n random points $\mathbf{X}_1, \ldots, \mathbf{X}_n$ independently and uniformly distributed in the square $\mathcal{S}_n = [-\sqrt{n}/2, \sqrt{n}/2]^2$ of area n; let $\Psi(v) = \mathbf{X}_v$ for each $v \in V$; and form the geometric graph $G = G(\Psi, r)$ or $G = G((\mathbf{X}_v \colon v \in V), r)$.

Random geometric graphs were first introduced by Gilbert [11] to model communications between radio stations. Since then, several related variants of these graphs have been widely used as models for wireless communication, and have also been extensively studied from a mathematical point of view. The main reference on random geometric graphs is the monograph by Penrose [19]; see also the survey of Walters [26]. The properties of $G \in \mathscr{G}(n, r)$ are usually investigated from an asymptotic perspective, as n grows to infinity and $r = r(n) = o(\sqrt{n})$.

A sequence A_n of events holds *with high probability* (whp) if $\mathbb{P}(A_n) \to 1$ as $n \to \infty$. For example, it is well known that $r_c = \sqrt{\log n/\pi}$ is a sharp threshold function for the connectivity of the random geometric graph $G \in \mathscr{G}(n, r)$. This means that, for every $\varepsilon > 0$, if $r \le (1 - \varepsilon)r_c$, then G is whp disconnected, whilst if $r \ge (1 + \varepsilon)r_c$, then G is whp connected (see [19] for a more precise result). We shall work with much larger r, so our random graphs will whp be (highly) connected.

Given a graph G, we define the *graph distance* $d_G(u, v)$ between two vertices u and v to be the least number of edges in a path between u and v if u and v are in the same component, and if not then we let the distance be ∞. Observe that in a geometric graph G with a given realisation (Ψ, r), each pair of vertices u and v must satisfy $d_G(u, v) \ge d_E(\Psi(u), \Psi(v))/r$, since each edge of the embedded geometric graph has length at most r. For a finite simple graph G with n vertices, let $A(G)$ denote its adjacency matrix, the $n \times n$ symmetric matrix with i, j entry 1 if ij is an edge and 0 otherwise. (We write ij for an edge rather than the longer form $\{i, j\}$.)

1.2. Approximate realisation for geometric graphs.

For a geometric graph G with vertex set V, the *realisation problem* for G has as input the adjacency matrix $A(G)$, and consists in finding some realisation (Ψ, r). It is known that for UD graphs, the realisation problem (also called the unit disk graph reconstruction problem) is NP-hard [4], and it remains NP-hard even if

we are given all the distances between pairs of vertices in some realisation [2], or if we are given all the angles between incident edges in some realisation [5]. Given that these results indicate the difficulty in finding exact polynomial time algorithms, researchers naturally turned their attention to finding good *approximate* realisations (for deterministic problems).

Previous work on approximate realisation

There are different possible measures of 'goodness' of an embedding. Motivated by the localisation problem for sensor networks, see for example [7], (essentially) the following scale-invariant measure of quality of embedding was introduced in [17]: given a geometric graph $G = (V, E)$, and an embedding $\Phi : V \to \mathbb{R}^2$ and threshold distance $r > 0$, if G is not a clique we let

$$Q_G(\Phi) = \frac{\max_{xy \in E} \|\Phi(x) - \Phi(y)\|_2}{\min_{xy \notin E} \|\Phi(x) - \Phi(y)\|_2}$$

(where we insist that $x \neq y$); and let $Q_G(\Phi) = (1/r) \max_{xy \in E} \|\Phi(x) - \Phi(y)\|_2$ if G is a clique. Observe that if (Φ, r) is a realisation of G then $Q_G(\Phi) \leq 1$. The aim is to find an embedding $\Phi : V \to \mathbb{R}^2$ with say $r = 1$ which minimises $Q_G(\Phi)$, or at least makes it small. The random projection method [24] was used in [17] to give an algorithm that, for an n-vertex UD graph G, outputs an embedding Φ with $Q_G(\Phi) = O(\log^{3.5} n \sqrt{\log \log n})$; that is, it approximates feasibility in terms of the measure Q_G up to a factor of $O(\log^{3.5} n \sqrt{\log \log n})$. On the other hand, regarding inapproximability, it was shown in [15] that, for any $\varepsilon > 0$, it is NP-hard for UD graphs to compute an embedding Φ with $Q_G(\Phi) \leq \sqrt{3/2} - \varepsilon$.

In this paper we do not aim to control a goodness measure like Q (though see the discussion following Theorem 1.3). Instead, we find whp a 'good' embedding Φ, which is 'close' to the hidden original random embedding Ψ. We investigate the approximate realisation problem for a random geometric graph, and for a family of vertex orderings (see later).

What we achieve for random geometric graphs is roughly as follows. We describe a polynomial time algorithm which, for a suitable range of values for r, whp finds an embedding Φ which 'up to symmetries' (see below for a detailed definition) maps each vertex v to within about distance r of the original random point $\Psi(v) = \mathbf{X}_v$. Observe that the mapping Φ must then satisfy the following properties whp: for each pair of vertices u, v with $d_E(\Psi(u), \Psi(v)) \leq r$ we have $d_E(\Phi(u), \Phi(v)) \leq (3 + \varepsilon)r$, and for each pair of vertices u, v with $d_E(\Psi(u), \Psi(v)) \geq (3 + \varepsilon)r$ we have $d_E(\Phi(u), \Phi(v)) > r$. Thus, adjacent pairs

of vertices remain quite close to being adjacent in Φ, and non-adjacent pairs of vertices that are sufficiently far apart remain non-adjacent in Φ.

For maps $\Phi_1, \Phi_2 : V \to \mathcal{S}_n$, the familiar *max* or *sup distance* is defined by

$$d_{\max}(\Phi_1, \Phi_2) = \max_{v \in V} d_E(\Phi_1(v), \Phi_2(v)).$$

Since there is no way for us to distinguish embeddings which are equivalent up to symmetries, we cannot hope to find an embedding Φ such that whp $d_{\max}(\Psi, \Phi)$ is small. There are 8 symmetries (rotations or reflections) of the square. We define the *symmetry-adjusted sup distance* d^* by

$$d^*(\Phi_1, \Phi_2) = \min_\sigma d_{\max}(\sigma \circ \Phi_1, \Phi_2) = \min_\sigma d_{\max}(\Phi_1, \sigma \circ \Phi_2),$$

where the minima are over the 8 symmetries σ of the square $\mathcal{S}_n$. This is the natural way of measuring distance 'up to symmetries'. If we let $\Phi_1 \sim \Phi_2$ when $\Phi_1 = \sigma \circ \Phi_2$ for some symmetry σ of $\mathcal{S}_n$, then it is easy to check that $\sim$ is an equivalence relation on the set of embeddings $\Phi : V \to \mathcal{S}_n$, and d^* is the natural sup metric on the set of equivalence classes.

Given $\alpha > 0$, we say that an embedding Φ has *displacement at most α* (from the hidden embedding Ψ) if $d^*(\Psi, \Phi) \le \alpha$. Consider the graph with three vertices u, v, w and exactly two edges uv and vw: if this is the geometric graph $G(\Psi, r)$, then $d_E(\Psi(u), \Psi(w))$ could be any value in $(r, 2r]$. Examples like this suggest that we should be happy to find an embedding Φ with displacement at most about r; and since our methods rely on graph distances, it is natural that we do not achieve displacement below r.

1.3. Vertex orderings. We also consider a related approximate realisation problem, with different information. As for a random geometric graph, we start with a family of n unseen points $\mathbf{X}_1, \ldots, \mathbf{X}_n$ independently and uniformly distributed in the square $\mathcal{S}_n$, forming the hidden embedding Ψ. (There is no radius r here, and there is no graph.) We are given, for each vertex v, the ordering τ_v of the vertices by increasing Euclidean distance from v. This is the *family of vertex orderings* corresponding to Ψ. Notice that with probability 1 no two distances between distinct points will be equal. Notice also that, if we had access to the complete ordering of the Euclidean distances between all pairs of distinct vertices in the hidden embedding Ψ, then we could read off the family of vertex orderings.

We shall see that, by using the family of vertex orderings, we can with high probability find an embedding with displacement error dramatically better than the bound we obtain for random geometric graphs.

1.4.   Main results. Suppose first that we are given a random geometric graph $G \in \mathscr{G}(n, r)$, with hidden original embedding Ψ, for example by being given the adjacency matrix $A(G)$. We of course know the number n of vertices and the number of edges, but we do not know the threshold distance r, and indeed we have no geometric information. Our goal is to find an embedding Φ such that whp it has displacement at most about r, for as wide as possible a range of values for r. However, first we need to consider how to estimate r. We determine the expected number of edges in $G \in \mathscr{G}(n, r)$ as a function of n and r (taking care over boundary effects), see Proposition 3.1; and use this value and the observed number of edges to calculate the estimator $\hat{r}$ of r.

Proposition 1.1. *Let $r = r(n) > 0$ satisfy $1/\sqrt{n} \ll r \ll \sqrt{n}$ as $n \to \infty$. Let $\rho = \sqrt{n}/r$ (so $\rho \to \infty$ as $n \to \infty$). Let $\omega = \omega(n)$ be a function tending to infinity with n arbitrarily slowly, and in particular satisfying $\omega \ll \sqrt{n}$. Then in $O(n^2)$ time we can compute an estimator $\hat{r}$ such that*

$$|\hat{r} - r| \ll \omega \cdot (n^{-1/2} + \rho^{-3/2}) \quad whp, \tag{1}$$

and thus $\hat{r}/r \to 1$ in probability as $n \to \infty$.

Our first theorem presents an algorithm to find an embedding Φ for a random geometric graph (given without any further information), which whp achieves displacement at most about r, for the range $n^{3/14} \ll r \ll \sqrt{n}$. Note that $3/14 \approx 0.21428$.

Theorem 1.2. *Let $r = r(n)$ satisfy $n^{3/14} \ll r \ll \sqrt{n}$, and consider the random geometric graph $G \in \mathscr{G}(n, r)$ (given say by the adjacency matrix $A(G)$) corresponding to the hidden embedding Ψ. Let $\varepsilon > 0$ be an arbitrarily small constant. There is an algorithm which in $O(n^2)$ time outputs an embedding Φ which whp has displacement at most $(1+\varepsilon)r$, that is, whp $d^*(\Psi, \Phi) \leq (1+\varepsilon)r$.*

In practice, after running the algorithm in this theorem, we would run a local improvement heuristic, even though this might not lead to a provable decrease in $d^*(\Psi, \Phi)$. For example, we might simulate a dynamical system where, for each vertex v, the corresponding point $\mathbf{x}_v$ tends to move towards the centre of gravity of the points $\mathbf{x}_w$ corresponding to the neighbours w of v (if $\mathbf{x}_v$ is not too close to the boundary of $\mathcal{S}_n$).

Our second theorem concerns the case when we learn not the random geometric graph but the family of vertex orderings; that is, for each vertex v, we learn the ordering τ_v of the vertices by increasing Euclidean distance from v, given as a linked list.

Theorem 1.3. *Suppose that we are given the family of vertex orderings corresponding to the hidden embedding* Ψ*. There is a linear time algorithm that outputs an embedding* Φ *which whp has displacement* $< 1.197\sqrt{\log n}$*; that is, whp* $d^*(\Psi, \Phi) < 1.197\sqrt{\log n}$*.*

Now let $\sqrt{\log n} \ll r \ll \sqrt{n}$, and let $G \in \mathscr{G}(n, r)$ be the random geometric graph corresponding to the hidden embedding Ψ. Then the constructed embedding Φ does well in terms of the measure Q_G introduced earlier: we have

$$Q_G(\Phi) < \frac{r + 1.197\sqrt{\log n}}{r - 1.197\sqrt{\log n}} < 1 + 2.4\sqrt{\log n}/r = 1 + o(1) \quad \text{whp.} \qquad (2)$$

Also, from the constructed embedding Φ we may form a second geometric graph $G' = G(\Phi, r)$. Then G' is close to G in the sense that 'we get only a small proportion of edges wrong'. We make this more precise in the inequality (3) below. It is easy to see that whp G has $(\frac{1}{2}+o(1))\pi r^2 n$ edges (see Proposition 3.1 for a more detailed result). We know from Theorem 1.3 that whp Φ has displacement $< 1.197\sqrt{\log n}$: assume that this event holds. If $d_E(\mathbf{X}_u, \mathbf{X}_v) \leq r - 2.394\sqrt{\log n}$ then $d_E(\Phi(u), \Phi(v)) \leq r$ so uv is an edge in G'; and similarly, if $d_E(\mathbf{X}_u, \mathbf{X}_v) \geq r + 2.394\sqrt{\log n}$ then $d_E(\Phi(u), \Phi(v)) > r$, so uv is not an edge in G'. Thus there could be a mistake with uv only if

$$r - 2.394\sqrt{\log n} < d_E(\mathbf{X}_u, \mathbf{X}_v) < r + 2.394\sqrt{\log n}.$$

But whp the number of unordered pairs $\{u, v\}$ of distinct vertices such that these inequalities hold is $(1+o(1))\, n \cdot 2\pi r \cdot 2.394\sqrt{\log n}$. Hence, whp the symmetric difference of the edge sets of G and G' satisfies

$$\left| E(G) \Delta E(G') \right| / \left| E(G) \right| < 9.6\sqrt{\log n}/r. \qquad (3)$$

Outline sketch of the proofs of Theorems 1.2 and 1.3

In order to prove these theorems, we first identify 4 'corner vertices' such that the corresponding points are close to the 4 corners of $\mathcal{S}_n$. To do this, for Theorem 1.2 we are guided by vertex degrees; and for Theorem 1.3 we look at the set of 'extreme' pairs $\{v, v'\}$ such that v' is farthest from v in the order τ_v, and v is farthest from v' in the order $\tau_{v'}$.

To prove Theorem 1.2, we continue as follows. For a vertex v, we approximate the Euclidean distance between $\mathbf{X}_v$ and a corner by using the graph distance from v to the corresponding corner vertex, together with the estimate $\hat{r}$ of r; and then we place our estimate $\Phi(v)$ for $\mathbf{X}_v$ at the intersection of circles

of appropriate radius centred on a chosen pair of the corners. For each of the circles, whp $\mathbf{X}_v$ lies within a narrow annulus around it, so $\Phi(v)$ is close to $\mathbf{X}_v$.

In the proof of Theorem 1.3, we obtain a much better approximation to the Euclidean distance between $\mathbf{X}_v$ and a corner, by using the rank of v in the ordering from the corresponding corner vertex, and the fact that the number of points $\mathbf{X}_w$ at most a given distance from a given corner is concentrated around its mean. In this way, we obtain much narrower annuli, and a correspondingly much better estimate for $\mathbf{X}_v$.

1.5. Further related work. In this section we mention further related work.

Theorem 1 of [1] estimates Euclidean distances between points by r times the graph distance in the corresponding geometric graph. It is assumed that r is known, and the error is at most r plus a term involving the maximum radius of an empty ball. In the case of n points distributed uniformly and independently in $\mathcal{S}_n$, the authors of [1] need $r \geq n^{1/4}(\log n)^{1/4}$ in order to keep the error bound down to $(1 + o(1))r$ whp (so they need r a little larger than we do in Theorem 1.2).

In [18] the authors assume that they are given a slightly perturbed adjacency matrix (some edges were inserted, some were deleted) of n points in some metric space. Under fairly general conditions on insertion and deletion, the authors use the Jaccard index (the size of the intersection of the neighbourhood sets of the endpoints of an edge divided by the size of their union) to compute a 2-approximation to the graph distances.

The use of graph distances for predicting links in a dynamic social network such as a co-authorship network was experimentally analysed in [16]: it was shown that graph distances (and other approaches) can provide useful information to predict the evolution of such a network. In [21] the authors consider a deterministic and also a non-deterministic model, and show that using graph distances, and also using common neighbours, they are able to predict links in a social network. The use of shortest paths in graphs for embedding points was also experimentally analysed in [22]. See [14] for an approach using graph distances and multidimensional scaling to assign geometric locations to vertices (in such a way that the Euclidean distances between the points "best" match the graph distances), with various applications, in particular to graph drawing and graph realisation.

In [25] the authors consider a k-nearest neighbour graph on n points $\mathbf{X}_i$ that have been sampled iid from some unknown density in Euclidean space. They show how shortest paths in the graph can be used to estimate the unknown

density. In [23] the authors consider the following problem: given a set of indices (i, j, k, ℓ), together with constraints $d_E(\mathbf{X}_i, \mathbf{X}_j) < d_E(\mathbf{X}_k, \mathbf{X}_\ell)$ (without knowing the distances), construct a point configuration that preserves these constraints as well as possible. The authors propose a 'soft embedding' algorithm which not only counts the number of violated constraints, but takes into account also the amount of violation of each constraint. Furthermore, the authors also provide an algorithm for reconstructing points when only knowing the k nearest neighbours of each data point, and they show that the obtained embedding converges for $n \to \infty$ to the real embedding (w.r.t. to a metric defined by the authors), as long as $k \gg \sqrt{n \log n}$. This setup is similar to our Theorem 1.3 in the sense that we are given the ordinal ranking of all distances from a point (for each point), though note that we estimate points up to an error $O(\sqrt{\log n})$ rather than $o(\sqrt{n})$ (recall that our points are sampled from the $\sqrt{n} \times \sqrt{n}$ square $\mathcal{S}_n$). See [3] for a similar approach to detecting a d-dimensional manifold in an N-dimensional space with $d \ll N$ based on sufficiently dense point sampling, using approximations of geodesic distances by graph distances.

In a slightly different context, the algorithmic problem of computing the embedding of n points in Euclidean space given some or all pairwise distances was considered. If all $\binom{n}{2}$ pairwise distances are known, and we are given the positions of three points forming a triangle T, then we can easily find exact positions in $O(n)$ arithmetic operations: separately for each point $\mathbf{x}$ not in T, find its location with respect to T by intersecting the three circles corresponding to the distances from $\mathbf{x}$ to the points in T, using $O(1)$ arithmetic operations. In this way we use only the $O(n)$ distances involving at least one of the points in T. In [8, 9] the authors consider the problem of knowing only a subset of the distances (they know only small distances, as typical in sensor networks), and show that by patching together local embeddings of small subgraphs an approximate embedding of the points can be found quickly (in polynomial time).

The related problem of trying to detect latent information on communities in a geometric framework was studied by [20]. In this case, (visible) points of a Poisson process in the unit square are equipped with an additional hidden label indicating to which of two hidden communities they belong. The probability that two vertices are joined by an edge naturally depends on the distance between them, but also edges between vertices of the same label have a higher probability to be present than edges between vertices of different labels. The paper gives exact recovery results for a dense case, and also shows the impossibility of recovery in a sparse case.

1.6. Organisation of the paper. In Section 2 we recall or establish preliminaries; in Section 3 we investigate the distribution of the number of edges in $G \in \mathscr{G}(n, r)$, use the number of edges to give an estimator $\hat{r}$ for the threshold distance r, and estimate Euclidean distances using $\hat{r}$ and graph distances; in Section 4 we complete the proof of Theorem 1.2; in Section 5 we prove Theorem 1.3; and in Section 6 we conclude with some open questions.

2. Preliminaries

In this section we gather simple facts and lemmas that are used in the proofs of the main results. We start with a standard version of the Chernoff bounds for binomial random variables, see for example Theorem 2.1 and inequality (2.9) in [13].

Lemma 2.1. *(Chernoff bounds) Let X have the binomial distribution $\mathrm{Bin}(n, p)$ with mean $\mu = np$. For every $\delta > 0$ we have*

$$\mathbb{P}(X \leq (1 - \delta)\mu) \leq e^{-\delta^2 \mu/2}$$

and

$$\mathbb{P}(X \geq (1 + \delta)\mu) \leq e^{-\delta^2(1+\delta/3)^{-1}\mu/2};$$

and it follows that, for each $0 < \delta \leq 1$,

$$\mathbb{P}(|X - \mu| \geq \delta\mu) \leq 2e^{-\delta^2 \mu/3}.$$

For $\mathbf{x} \in \mathbb{R}^2$ and $r > 0$, let $B(\mathbf{x}, r)$ denote the closed ball of radius r around $\mathbf{x}$. We shall repeatedly use the following elementary fact.

Fact 2.2. *Let $G \in \mathscr{G}(n, r)$ be a random geometric graph. For each $\mathbf{x} \in \mathcal{S}_n$ let $\sigma_n(\mathbf{x})$ be the area of $B(\mathbf{x}, r) \cap \mathcal{S}_n$, and let $\rho_n(\mathbf{x}) = \sigma_n(\mathbf{x})/n$. Then for each vertex $v \in V = [n]$ and each point $\mathbf{x} \in \mathcal{S}_n$, $\deg_G(v)$ conditional on $\mathbf{X}_v = \mathbf{x}$ has distribution $\mathrm{Bin}(n-1, \rho_n(\mathbf{x}))$. More precisely, this gives a density function: for any Borel set $A \subseteq \mathcal{S}_n$,*

$$\mathbb{P}((\deg_G(v) = k) \wedge (\mathbf{X}_v \in A)) = \int_{\mathbf{x} \in A} \mathbb{P}(\mathrm{Bin}(n-1, \rho_n(\mathbf{x})) = k) \, d\mathbf{x}.$$

In particular, if $\rho^- \leq \rho_n(\mathbf{x}) \leq \rho^+$ for each $\mathbf{x} \in A$, then, conditional on $\mathbf{X}_v \in A$, $\deg_G(v)$ is stochastically at least $\mathrm{Bin}(n-1, \rho^-)$ and stochastically at most $\mathrm{Bin}(n-1, \rho^+)$.

The next lemma gives elementary bounds on the area $\sigma_n(\mathbf{z})$ for $\mathbf{z} \in \mathcal{S}_n$, in terms of the distance from $\mathbf{z}$ to a corner of $\mathcal{S}_n$ or to the boundary of $\mathcal{S}_n$.

Lemma 2.3. *Let $0 < s \leq r < \sqrt{n}/2$, and let $\mathbf{z} \in \mathcal{S}_n$.*
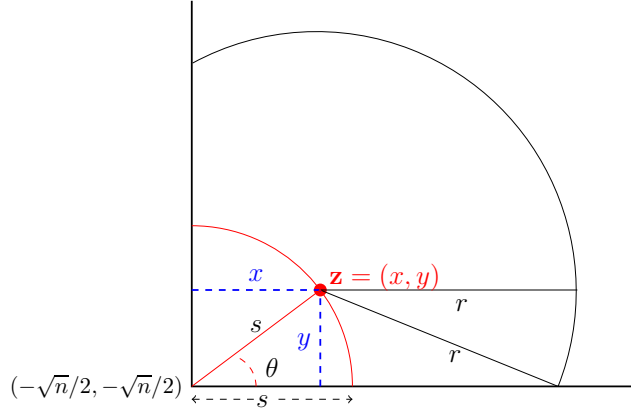
FIGURE 1. Figure for the proof of Lemma 2.3(ii)

(i) *If $\mathbf{z}$ is at distance at most s from some corner, then $\sigma_n(\mathbf{z}) \leq \frac{1}{4}\pi(r+s)^2$.*

(ii) *If $\mathbf{z}$ is at distance at least s from each corner, then $\sigma_n(\mathbf{z}) \geq \frac{1}{4}\pi r^2 + s(r - s/2)$.*

(iii) *If $\mathbf{z}$ is at distance at most s from the boundary, then $\sigma_n(\mathbf{z}) \leq \frac{1}{2}\pi r^2 + 2sr$.*

(iv) *If $\mathbf{z}$ is at distance at least s from the boundary and at distance at most r from at most one side of the boundary, then $\sigma_n(\mathbf{z}) \geq \frac{1}{2}\pi r^2 + 2s(r - s/2)$.*

Proof. Parts (i) and (iii) are easy. To prove parts (ii) and (iv), we observe first that, in the disk with centre $(0,0)$ and radius r, the set S of points (x,y) in the disk with $-s \leq x \leq 0$ and $y \geq 0$ has area at least $s(r - s/2)$. For if $(-s, y_1)$ is the point on the bounding circle with $y_1 > 0$, then $y_1 = \sqrt{r^2 - s^2} \geq r - s$, so the quadrilateral Q with corners $(0,0), (-s,0), (-s, y_1)$ and $(0,r)$ has area $\geq \frac{1}{2}s(r + r - s)$, and $Q \subseteq S$.

To prove part (ii) of the lemma (see Figure 1), it suffices to consider points $\mathbf{z} \in \mathcal{S}_n$ at distance equal to s from a corner, wlog from the bottom left corner $c_1 = (-\sqrt{n}/2, -\sqrt{n}/2)$. Suppose that $\mathbf{z} - c_1 = (x,y)$, so $x, y \geq 0$. Then, by the observation in the first paragraph,

$$\sigma_n(\mathbf{z}) - \tfrac{1}{4}\pi r^2 \geq xy + x(r - x/2) + y(r - y/2) \geq x(r - x/2) + y(r - y/2).$$

Now, set $x = s\cos\theta$, $y = s\sin\theta$ for some $0 \leq \theta \leq \pi/2$, and the right hand side of the previous inequality can be written as

$$
\begin{aligned}
f(\theta) &:= s\cos\theta(r - \tfrac{1}{2}s\cos\theta) + s\sin\theta(r - \tfrac{1}{2}s\sin\theta) \\
&= sr(\cos\theta + \sin\theta) - \tfrac{1}{2}s^2\cos^2\theta - \tfrac{1}{2}s^2\sin^2\theta \\
&= sr(\cos\theta + \sin\theta) - \tfrac{1}{2}s^2.
\end{aligned}
$$

We have $f''(\theta) = -sr(\cos\theta + \sin\theta) < 0$ for $0 < \theta < \pi/4$, so f is concave, and thus $f(\theta)$ is minimised over $0 \leq \theta \leq \pi/4$ at $\theta = 0$ or $\theta = \pi/2$. But

$f(0) = f(\pi/2) = sr - \frac{1}{2}s^2 = s(r - s/2)$, and part (ii) follows. Part (iv) follows similarly from the initial observation. $\square$

We shall depend heavily on the following result on the relation between graph distance and Euclidean distance for random geometric graphs (with slightly worse constants than the ones given in the original paper to make the expression cleaner).

Lemma 2.4. *[10][Theorem 1.1] Let $G \in \mathscr{G}(n, r)$ be a random geometric graph with $r \gg \sqrt{\log n}$. Then, whp, for every pair of vertices u, v we have:*

$$d_G(u, v) \leq \left\lceil \frac{d_E(\mathbf{X}_u, \mathbf{X}_v)}{r} \left(1 + \gamma\, r^{-4/3}\right) \right\rceil$$

where

$$\gamma = \max \left\{ 3000 \left(\frac{r \log n}{r + d_E(\mathbf{X}_u, \mathbf{X}_v)} \right)^{2/3}, \ \frac{4 \cdot 10^6 \log^2 n}{r^{8/3}}, \ 1000 \right\}.$$

We observed earlier that always $d_E(\mathbf{x}_u, \mathbf{x}_v) \leq r d_G(u, v)$; we next give a corollary of the last lemma which shows that whp this bound is quite tight.

Corollary 2.5. *There is a constant c ($\leq 6 \cdot 10^6$) such that, if $r \geq (\log n)^{3/4}$ for n sufficiently large, then whp, for every pair of vertices u, v we have:*

$$d_G(u, v) \leq d_E(\mathbf{X}_u, \mathbf{X}_v)/r + 1 + c \max\{n^{1/2} r^{-7/3}, n^{1/6} r^{-5/3}(\log n)^{2/3}\}.$$

Proof. By Lemma 2.4

$$r d_G(u, v) \leq d_E(\mathbf{X}_u, \mathbf{X}_v) + r + r \cdot d_E(\mathbf{X}_u, \mathbf{X}_v)\gamma\, r^{-7/3}. \tag{4}$$

But, for $r \geq (\log n)^{3/4}$, the second term in the maximum in the definition of γ is at most $4 \cdot 10^6$; and letting γ_1 denote the first term we have

$$
\begin{aligned}
d_E(\mathbf{X}_u, \mathbf{X}_v)\, \gamma_1\, r^{-7/3} &\leq\ 3000\, d_E(\mathbf{X}_u, \mathbf{X}_v)^{1/3}(r \log n)^{2/3} r^{-7/3} \\
&\leq\ 3000\, (2n)^{1/6} r^{-5/3}(\log n)^{2/3}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
&d_E(\mathbf{X}_u, \mathbf{X}_v)\, \gamma\, r^{-7/3} \\
&\leq\ \max\{3000\, (2n)^{1/6} r^{-5/3}(\log n)^{2/3}, (4 \cdot 10^6)(2n)^{1/2} r^{-7/3}\}
\end{aligned}
$$

and the lemma follows from (4). $\square$

In fact, all we shall need from the last two results is the following immediate consequence of the last one.

Corollary 2.6. *If $r \gg n^{3/14}$, then there exists $\varepsilon = \varepsilon(n) = o(1)$ such that whp, for every pair u, v of vertices, we have*

$$d_G(u, v) \leq d_E(\mathbf{X}_u, \mathbf{X}_v)/r + 1 + \varepsilon.$$

We consider the four corner points $c_i = c_i(n)$ of $\mathcal{S}_n$ in clockwise order from the bottom left: $c_1 = (-\sqrt{n}/2, -\sqrt{n}/2)$ (already defined), $c_2 = (-\sqrt{n}/2, \sqrt{n}/2)$, $c_3 = (\sqrt{n}/2, \sqrt{n}/2)$ and $c_4 = (\sqrt{n}/2, -\sqrt{n}/2)$. See Figure 2 for the points c_i and to illustrate the following lemma.

Lemma 2.7. *Let $r = r(n)$ satisfy $\sqrt{\log n} \ll r \ll \sqrt{n}$ and consider the random geometric graph $G \in \mathscr{G}(n, r)$. Let $\omega = \omega(n)$ tend to infinity with n arbitrarily slowly, and in particular assume that $\omega^2 \leq r/2$ and $\omega \ll r/\sqrt{\log n}$. Then whp the following holds: (a) for each $i = 1, \ldots, 4$, there exists $v_i \in V$ such that $\mathbf{X}_{v_i} \in B(c_i, \omega)$ and $\deg_G(v_i) < \frac{1}{4}\pi r^2 + \frac{1}{3}\omega r$; and (b) for each $v \in V$ such that $\mathbf{X}_v \notin \cup_{i=1}^{4} B(c_i, \omega)$ we have $\deg_G(v) > \frac{1}{4}\pi r^2 + \frac{1}{2}\omega r$.*

Proof. (a) Fix $i \in [4]$. Note first that

$$\mathbb{P}(\mathbf{X}_v \notin B(c_i, \tfrac{1}{7}\omega) \text{ for each } v \in V) = (1 - \tfrac{\pi}{4n}(\tfrac{\omega}{7})^2)^n \leq e^{-\frac{\pi}{196}\omega^2} = o(1);$$

so whp there exists $v_i \in V$ such that $\mathbf{X}_{v_i} \in B(c_i, \tfrac{1}{7}\omega)$. Let $Z_n^{(i)}$ be the number of vertices v such that $\mathbf{X}_v \in B(c_i, \tfrac{1}{7}\omega)$. Then $\mathbb{E}[Z_n^{(i)}] = \frac{\pi}{196}\omega^2$. For each $\mathbf{x} \in B(c_i, \tfrac{1}{7}\omega)$, by Lemma 2.3 (i),

$$\sigma_n(\mathbf{x}) \leq \tfrac{1}{4}\pi(r + \tfrac{1}{7}\omega)^2 \leq \tfrac{1}{4}\pi r^2 + \tfrac{1}{4}\omega r$$

for n sufficiently large; and then, by Lemma 2.1 and Fact 2.2, for each vertex v

$$\mathbb{P}(\deg_G(v) \geq \tfrac{1}{4}\pi r^2 + \tfrac{1}{3}\omega r \mid \mathbf{X}_v \in B(c_i, \tfrac{1}{7}\omega))$$
$$\leq \quad \mathbb{P}(\text{Bin}(n, (\tfrac{1}{4}\pi r^2 + \tfrac{1}{4}\omega r)/n) \geq (\tfrac{1}{4}\pi r^2 + \tfrac{1}{4}\omega r) + \tfrac{1}{12}\omega r) \leq e^{-\Theta(\omega^2)}.$$

Hence

$$\mathbb{P}(\text{for some } v \in V, \ (\mathbf{X}_v \in B(c_i, \tfrac{1}{7}\omega)) \wedge (\deg_G(v) \geq \tfrac{1}{4}\pi r^2 + \tfrac{1}{3}\omega r))$$
$$\leq \quad \sum_{v \in V} \mathbb{P}(\mathbf{X}_v \in B(c_i, \tfrac{1}{7}\omega)) \, \mathbb{P}(\deg_G(v) \geq \tfrac{1}{4}\pi r^2 + \tfrac{1}{3}\omega r \mid \mathbf{X}_v \in B(c_i, \tfrac{1}{7}\omega))$$
$$\leq \quad \mathbb{E}[Z_n^{(i)}] \, e^{-\Theta(\omega^2)} \ = \ o(1).$$

Thus whp there exists $v_i \in V$ such that $\mathbf{X}_{v_i} \in B(c_i, \tfrac{1}{7}\omega)$ and $\deg_G(v_i) < \frac{1}{4}\pi r^2 + \frac{1}{3}\omega r$. This gives part (a) of the lemma.

(b) Let $j_0 = \lfloor \omega \rfloor$ and $j_1 = \lceil r/\omega \rceil$. For all integers $i \in [4]$ and $j_0 \leq j \leq j_1$, let $B_i^j = B(c_i, j) \cap \mathcal{S}_n$. Consider first the central part of the square $\mathcal{S}_n$, omitting parts near the corners: let $C_n = \mathcal{S}_n \setminus \cup_i B_i^{j_1}$. By Lemma 2.3 (ii), for each

$\mathbf{x} \in C_n$ we have

$$\sigma_n(\mathbf{x}) \geq \tfrac{1}{4}\pi r^2 + j_1(r - j_1/2) \geq \tfrac{1}{4}\pi r^2 + \tfrac{1}{2}j_1 r$$

for n sufficiently large. Hence, by Lemma 2.1 and Fact 2.2,

$$\mathbb{P}(\deg_G(v) \leq \tfrac{1}{4}\pi r^2 + \tfrac{1}{2}\omega r \mid \mathbf{X}_v \in C_n) \leq e^{-\Theta(j_1^2)}.$$

Since $\omega \ll r/\sqrt{\log n}$, we have $j_1^2 \geq (r/\omega)^2 \gg \log n$. Thus $ne^{-\Theta(j_1^2)} = o(1)$, and so whp there is no vertex v such that $\mathbf{X}_v \in C_n$ and $\deg_G(v) \leq \tfrac{1}{4}\pi r^2 + \tfrac{1}{2}\omega r$.

We need a little more care near the corners. Let $i \in [4]$ and let j be an integer with $j_0 \leq j \leq j_1$. The area of $B_i^{j+1} \setminus B_i^j$ is $\tfrac{1}{4}\pi(2j + 1)$. For each point $\mathbf{x} \in B_i^{j+1} \setminus B_i^j$, $\mathbf{x}$ is at distance at least j from each corner of $\mathcal{S}_n$, so by Lemma 2.3 (ii) we have

$$\sigma_n(\mathbf{x}) \geq \sigma^{(j)} := \tfrac{1}{4}\pi r^2 + j(r - j/2) = \tfrac{1}{4}\pi r^2 + (1 + o(1))jr.$$

Also,

$$\tfrac{1}{4}\pi r^2 + \tfrac{1}{2}\omega r \leq \tfrac{n-1}{n}\sigma^{(j)} - (\tfrac{1}{2} + o(1))jr.$$

Thus, by Lemma 2.1 and Fact 2.2,

$$\mathbb{P}\big(\deg_G(v) \leq \tfrac{1}{4}\pi r^2 + \tfrac{1}{2}\omega r \mid \mathbf{X}_v \in B_i^{j+1} \setminus B_i^j\big) \leq e^{-\Theta(j^2)}.$$

Therefore, for each $i \in [4]$,

$$\mathbb{P}\big(\text{for some } v \in V, (\mathbf{X}_v \in B_i^{j_1} \setminus B_i^{j_0}) \wedge (\deg_G(v) \leq \tfrac{1}{4}\pi r^2 + \tfrac{1}{2}\omega r)\big)$$

$$\leq \sum_{v \in V} \sum_{j=j_0}^{j_1-1} \mathbb{P}(\mathbf{X}_v \in B_i^{j+1} \setminus B_i^j) \, \mathbb{P}\big(\deg_G(v) \leq \tfrac{1}{4}\pi r^2 + \tfrac{1}{2}\omega r \mid \mathbf{X}_v \in B_i^{j+1} \setminus B_i^j\big)$$

$$\leq n \sum_{j \geq j_0} \frac{\tfrac{1}{4}\pi(2j + 1)}{n} e^{-\Theta(j^2)} = o(1).$$

Hence whp $\deg_G(v) > \tfrac{1}{4}\pi r^2 + \tfrac{1}{2}\omega r$ for each vertex v such that $\mathbf{X}_v$ is not in one of the four corner regions $B_i^{j_0}$; and so we have completed the proof of part (b). $\square$

The above lemma shows us how to find 4 vertices v such that whp the corresponding points $\mathbf{X}_v$ are close to the four corner points c_i of $\mathcal{S}_n$.

Lemma 2.8. *Let $r = r(n)$ satisfy $\sqrt{\log n} \ll r \ll \sqrt{n}$, and consider the random geometric graph $G = \mathcal{G}(n, r)$. Let $\omega = \omega(n)$ be any function tending to infinity as $n \to \infty$. There is a polynomial-time (in n) algorithm which, on input $A(G)$, finds four vertices v_1, v_2, v_3, v_4 such that whp the following holds: for some (unknown) symmetry π of $\mathcal{S}_n$,*

$$\mathbf{X}_{v_i} \in B(\pi(c_i), \omega) \quad \text{for each } i \in [4].$$
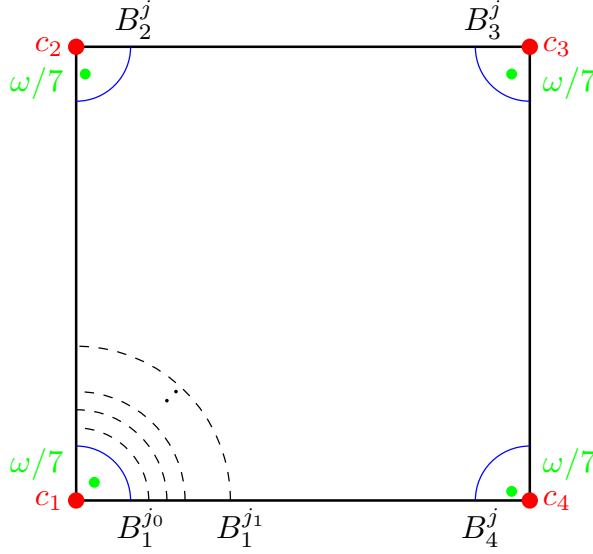
FIGURE 2. Choosing points in the 4 corners of the $\sqrt{n} \times \sqrt{n}$ square $\mathcal{S}_n$

Proof. Consider the following algorithm: pick a vertex of minimal degree, call it u_1, and mark u_1 and all its neighbours. Continue iteratively on the set of unmarked vertices, until we have found four vertices $u_1, \ldots, u_4$. (Whp each vertex has degree at most $1.1\,\pi r^2$; so after at most 3 steps, at most $3.3\,\pi r^2 + 3 = o(n)$ vertices are marked, and so whp we will find $u_1, \ldots, u_4$.) Let u_1' be a vertex amongst u_2, u_3, u_4 maximising the graph distance from u_1, and list the four vertices as v_1, v_2, v_3, v_4 where $v_1 = u_1$ and $v_3 = u_1'$ (and v_2 and v_4 are the other two of the vertices u_i listed in some order). We shall see that whp v_1, v_2, v_3, v_4 are as required.

By Lemma 2.7, whp the vertices $u_1, \ldots, u_4$ are each within distance ω of a corner of $\mathcal{S}_n$, and the marking procedure ensures that the four corners involved are distinct. If u_i and u_j are such that $\mathbf{X}_{u_i}$ and $\mathbf{X}_{u_j}$ are within distance ω of opposite corners of $\mathcal{S}_n$, then $d_E(\mathbf{X}_{u_i}, \mathbf{X}_{u_j}) \geq \sqrt{2n} - 2\omega$ and so $d_G(u_i, u_j) \geq (1 + o(1))\sqrt{2n}/r$. If $\mathbf{X}_{u_i}$ and $\mathbf{X}_{u_j}$ are within distance ω of adjacent corners, then $d_E(\mathbf{X}_{u_i}, \mathbf{X}_{u_j}) \leq \sqrt{n} + \omega$; and so, since we may assume wlog that $\omega \ll \sqrt{n}$, whp $d_G(u_i, u_j) \leq (1 + o(1))\sqrt{n}/r$ by Corollary 2.5. Hence, whp $u_1 = v_1$ and $u_1' = v_3$ are within distance ω of opposite corners, as are the other two of the chosen vertices. For each $i \in [4]$, denote the corner closest to $\mathbf{X}_{v_i}$ by $c_{\sigma(i)}$. Then whp σ is a permutation of $[4]$, and $c_{\sigma(1)}$ and $c_{\sigma(3)}$ are opposite corners; and so $c_{\sigma(1)}, \ldots, c_{\sigma(4)}$ lists the corners of $\mathcal{S}_n$ in either clockwise or anticlockwise order. Thus σ extends to a (unique) symmetry π of $\mathcal{S}_n$, and we are done. $\square$

Having found four vertices $v_1, \ldots, v_4$ such that the points $\mathbf{X}_{v_i}$ are close to the four corner vertices of $\mathcal{S}_n$, for each other vertex $v \in V(G)$ we will be able to use the graph distances from v to each of $v_1, \ldots, v_4$ to obtain an approximation to $\mathbf{X}_v$.

3. Estimating r and Euclidean distances

In this section, we use the preliminary results from the last section to see how to estimate the threshold distance r, and Euclidean distances between points, sufficiently accurately to be able to prove Theorem 1.2 in the next section. We start by considering the number of edges in $G \in \mathcal{G}(n, r)$ (with more precision than will be needed here).

3.1. Expectation and variance of the number of edges in $G \in \mathcal{G}(n, r)$.

Proposition 3.1. *Let n be a positive integer, let $0 < r \le \sqrt{n}$, and let M be the random number of edges in $G \in \mathcal{G}(n, r)$. Then*

$$\mathbb{E}[M] = \tfrac{1}{2}(n-1)\pi r^2 \big(1 - \tfrac{8}{3\pi}(r/\sqrt{n}) + \tfrac{1}{2\pi}(r^2/n)\big) \tag{5}$$

and

$$\mathbf{Var}(M) < \tfrac{1}{2}\pi n r^2 + 4\pi^2 \sqrt{n} r^5. \tag{6}$$

Observe that, by (5), for $0 < r = r(n) \ll \sqrt{n}$ in particular it follows that

$$\mathbb{E}[M] = (1 + o(1))\tfrac{1}{2}\pi n r^2. \tag{7}$$

We will use one lemma in the proof of this proposition. For each point z in the square $[0, 1]^2$, let $f(z)$ be the area of the ball $B(z, 1)$ to the right of the line $x = 1$, and let $g(z)$ be the area of $B(z, 1) \cap [1, 2]^2$.

Lemma 3.2. *Let the random variable Z be uniformly distributed over the square $[0, 1]^2$. Then*

- $\mathbb{E}[f(Z)] = \alpha = \tfrac{2}{3}$
- $\mathbb{E}[g(Z)] = \beta = \tfrac{1}{8}$.
- $\mathbb{E}[B(Z, 1) \cap [0, 1]^2] = \pi - 4\alpha + 4\beta \approx 0.974926$.

Proof. Let X_1 be uniformly distributed on $[0, 1]$. Then $\mathbb{E}[f(Z)]$ is the expected area of $B((1 - X_1, 0), 1)$ to the right of the line $x = 1$, which is

$$
\begin{aligned}
\int_0^1 \left(\int_x^1 2\sqrt{1 - t^2}\, dt \right) dx &= \int_0^1 \left(\int_0^t 2\sqrt{1 - t^2}\, dx \right) dt \\
&= \int_0^1 2t\sqrt{1 - t^2}\, dt \\
&= [-\tfrac{2}{3}(1 - t^2)^{3/2}]_0^1 = \tfrac{2}{3}.
\end{aligned}
$$

Now let X_1, X_2, Y_1, Y_2 be independent random variables, each uniformly distributed on $[0, 1]$. Let $X = X_1 + X_2$ and $Y = Y_1 + Y_2$. Then $\mathbb{P}(X \le t) = \frac{1}{2}t^2$ for $0 \le t \le 1$, and therefore we may take the density function f_0 of X to satisfy $f_0(t) = t$ for $0 \le t \le 1$. Observe that $Z_1 = (1 - X_1, 1 - Y_1)$ is uniformly distributed over $[0, 1]^2$, $Z_2 = (1 + X_2, 1 + Y_2)$ is uniformly distributed over $[1, 2]^2$, and $d_E(Z_1, Z_2)^2 = X^2 + Y^2$. Then

$$
\begin{aligned}
\mathbb{E}[g(Z)] &= \mathbb{P}(X^2 + Y^2 \le 1) \\
&= \int_0^1 \mathbb{P}(Y \le \sqrt{1 - x^2}) f_0(x)\, dx \\
&= \int_0^1 \tfrac{1}{2}(1 - x^2)\, x\, dx \\
&= \tfrac{1}{2}[x^2/2 - x^4/4]_0^1 \;=\; \tfrac{1}{8},
\end{aligned}
$$

as required. $\square$

Proof of Proposition 3.1. For $u \ne v$ in $[n]$, let $Y(u, v)$ be 1 if u and v are adjacent in G, and be 0 otherwise. Let $p = \mathbb{E}[Y(u, v)]$. Then $M = \sum_{u < v} Y(u, v)$, and $\mathbb{E}[M] = \binom{n}{2}p$.

Now let Z be uniformly distributed over the square $\mathcal{S}_n = [-\frac{1}{2}\sqrt{n}, \frac{1}{2}\sqrt{n}]^2$. Observe that Z is at distance at most r from a given side of the square $\mathcal{S}_n$ with probability $r/\sqrt{n}$, and Z is in the $r \times r$ subsquare touching a given corner point of $\mathcal{S}_n$ with probability r^2/n. Thus, by Lemma 3.2, the expected area of $B(Z, r) \cap \mathcal{S}_n$ is

$$
\pi r^2 - 4(r/\sqrt{n})\,\alpha r^2 + 4(r^2/n)\beta r^2 = \pi r^2 - \tfrac{8}{3}(r/\sqrt{n})r^2 + \tfrac{1}{2}(r^2/n)r^2.
$$

But this equals np, so

$$
\mathbb{E}[M] = \tfrac{1}{2}(n-1)np = \tfrac{1}{2}(n-1)\pi r^2\big(1 - \tfrac{8}{3\pi}(r/\sqrt{n}) + \tfrac{1}{2\pi}(r^2/n)\big),
$$

as required in (5).

Now consider the variance of M. We may expand $\mathbb{E}[M^2]$ as

$$
\mathbb{E}[M^2] = \sum_{\{u,v\}} \sum_{\{u',v'\}} \mathbb{E}[Y(u, v)Y(u', v')] \tag{8}
$$

where $\{u, v\}$ and $\{u', v'\}$ are both unordered pairs of distinct vertices. If the pairs $\{u, v\}$ and $\{u', v'\}$ are identical, then $\mathbb{E}[Y(u, v)Y(u', v')] = \mathbb{E}[Y(u, v)] = p$. If the vertices $u, v, u', v' \in [n]$ are all distinct then $\mathbb{E}[Y(u, v)Y(u', v')] = p^2$. Now suppose that $u, v, w \in [n]$ are distinct. Then

$$
\mathbb{E}[Y(u, v)Y(u, w)] = \mathbb{E}[Y(u, v)]\,\mathbb{E}[Y(u, w)|Y(u, v) = 1] \le p \cdot \pi r^2/n.
$$

Call a point in the square $\mathcal{S}_n$ *central* if it is not within distance r of a side of the square. Then

$$p > \mathbb{E}[Y(u,v)|\mathbf{X}_u \text{ central}]\mathbb{P}[\mathbf{X}_u \text{ central}] > (\pi r^2/n)(1 - 4r/\sqrt{n}) = \pi r^2/n - 4\pi(r/\sqrt{n})^3.$$

Thus

$$\mathbb{E}[Y(u,v)Y(u,w)] < p\left(p + 4\pi(r/\sqrt{n})^3\right) \leq p^2 + 4\pi^2(r/\sqrt{n})^5$$

(using $p \leq \pi r^2/n$). But there are at most n^3 such terms in the sum in (8), so

$$\mathbb{E}[M^2] < \mathbb{E}[M] + (\mathbb{E}[M])^2 + n^3 \cdot 4\pi^2(r/\sqrt{n})^5.$$

Hence, since $\mathbb{E}[M] < \frac{1}{2}\pi n r^2$,

$$\mathbf{Var}(M) < \tfrac{1}{2}\pi n r^2 + 4\pi^2\sqrt{n}r^5,$$

as required in (6). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

If $r \gg n^{1/6}$ then both $n^{1/2}r$ and $n^{1/4}r^{5/2}$ are $\ll (r^2/n)nr^2 = r^4$, and so by Chebyshev's inequality

$$M = \tfrac{1}{2}n\pi r^2\left(1 - \tfrac{8}{3\pi}(r/\sqrt{n}) + \tfrac{1+o(1)}{2\pi}(r^2/n)\right) \text{ whp.}$$

If $nr^2 \to \infty$ and $r \ll \sqrt{n}$ then similarly

$$M = \tfrac{1+o(1)}{2}n\pi r^2 \text{ whp.}$$

3.2. The estimator $\hat{r}$: proof of Proposition 1.1. Let $\mu(n,r)$ denote $\mathbb{E}[M]$, given in equation (5), and note that $\mu(n,r)$ is increasing in r. We can determine M from the adjacency matrix in $O(n^2)$ time; and then, by repeated bisection (for example) we can find $\hat{r}$ such that $|\mu(n,\hat{r}) - M| < 1$, in $O(\log n)$ arithmetic operations. Thus we can calculate the estimator $\hat{r}$ in $O(n^2)$ time.

Let $\sigma = \sigma(n,r)$ denote the standard deviation of M, and let $\beta = \beta(n,r) = \sqrt{n}r + n^{1/4}r^{5/2}$. By (6) we have $\sigma < \sqrt{\pi/2}\sqrt{n}r + 2\pi n^{1/4}r^{5/2} < 7\beta$. Let $\omega = \omega(n) \to \infty$ as $n \to \infty$, and satisfy $\omega^2 \ll \min\{r,1\}\sqrt{n}$. Let $\omega' = \omega'(n) \to \infty$ as $n \to \infty$, and satisfy $\omega' \ll \omega$. By Chebyshev's inequality, whp $|M - \mu(n,r)| \ll \omega'\sigma$, and so whp $|\mu(n,r) - \mu(n,\hat{r})| < \omega'\sigma + 1$. It follows that

$$|\mu(n,\hat{r}) - \mu(n,r)| \ll \omega\beta \quad \text{whp.} \tag{9}$$

We want to bound $x := |\hat{r} - r|$. Let $f(y) = |\mu(n,r+y) - \mu(n,r)|$, and observe that if $y' > y \geq 0$ or $y' < y \leq 0$ then $f(y') > f(y)$. By (5), for y such that $|y| = o(r)$ we have $\mu(n,r+y) = \mu(n,r)(1 + 2y/r + o(|y|/r))$. Thus, for such a y, by (7)

$$f(y) = (2 + o(1))\mu(n,r)|y|/r = (1 + o(1))\pi r n|y|. \tag{10}$$

We make two observations. (a) For $r \leq n^{1/6}$, we have $\sqrt{n}r \geq n^{1/4}r^{5/2}$ so $\beta \leq 2\sqrt{n}r$; and $r^2 n/\omega \gg \omega\sqrt{n}r$ (since $\omega^2 \ll \sqrt{n}r$)), so $r^2 n/\omega \gg \omega\beta$. (b) For $r \geq n^{1/6}$, we have $\sqrt{n}r \leq n^{1/4}r^{5/2}$ so $\beta \leq 2n^{1/4}r^{5/2}$; and $r^2 n/\omega \gg \omega n^{1/4}r^{5/2}$ (since $\omega^2 \ll \sqrt{n} \ll n^{3/4}r^{-1/2}$), so $r^2 n/\omega \gg \omega\beta$. This in both cases $r^2 n/\omega \gg \omega\beta$. Hence, by (10), setting $y = \pm r/\omega$, we have

$$f(y) = (1 + o(1))\pi r^2 n/\omega \gg \omega\beta.$$

But now, by (9) and the observed monotonicity of f, whp $|x| < |y| \ll r$. It follows that whp $f(x) = (1 + o(1))\pi rn|x|$, so $|x| = (1 + o(1))f(x)/(\pi rn)$; and thus by (9), whp

$$|x| \ll \omega\beta/(\pi rn) \leq \omega \cdot (n^{-1/2} + n^{-3/4}r^{3/2}).$$

Hence

$$|\hat{r} - r| = |x| \ll \omega \cdot (n^{-1/2} + \rho^{-3/2}) \quad \text{whp},$$

which completes the proof of Proposition 1.1.

3.3. The estimator $\hat{r}$ and Euclidean distance.
In this subsection we restrict r to be large enough so that we can use Corollary 2.6.

Lemma 3.3. *Let* $r = r(n)$ *satisfy* $n^{3/14} \ll r \ll \sqrt{n}$. *Then there exists* $\delta = \delta(n) \to 0$ *as* $n \to \infty$ *such that whp, for all pairs* u, v *of vertices,*

$$\hat{r}d_G(u, v) + \delta\hat{r} \geq d_E(\mathbf{X}_u, \mathbf{X}_v) \geq \hat{r}d_G(u, v) - (1 + \delta)\hat{r}. \tag{11}$$

Observe that by (11) and Proposition 1.1 we can estimate each value $d_E(\mathbf{X}_u, \mathbf{X}_v)$ up to an additive error of $(\frac{1}{2} + o(1))\hat{r} = (\frac{1}{2} + o(1))r$ whp.

Proof. By Corollary 2.6, there exists $\delta = \delta(n) \to 0$ sufficiently slowly that whp, for all pairs vertices u, v we have

$$rd_G(u, v) \geq d_E(\mathbf{X}_u, \mathbf{X}_v) \geq rd_G(u, v) - (1 + \delta/2)r. \tag{12}$$

Let $\rho = \sqrt{n}/r$, as in Proposition 1.1. By Corollary 2.6, whp

$$\max_{u,v} d_G(u, v) \leq \sqrt{2n}/r + 2 \leq 2\rho$$

for n sufficiently large (where the maximum is over all pairs u, v of vertices). Hence, by (12),

$$
\begin{aligned}
\hat{r}d_G(u, v) + 2\rho|\hat{r} - r| &\geq d_E(\mathbf{X}_u, \mathbf{X}_v) \\
&\geq \hat{r}d_G(u, v) - 2\rho|\hat{r} - r| - (1 + \delta/2)r. \tag{13}
\end{aligned}
$$

By Proposition 1.1, whp $|\hat{r} - r| \leq \omega\rho^{-3/2}$. Thus whp

$$2\rho|\hat{r} - r| = o(1),$$

by choosing ω to be growing sufficiently slowly (which we may, since $\rho \to \infty$). We may assume that $\delta \to 0$ sufficiently slowly, so that $\delta r \to \infty$ as $n \to \infty$, and in particular, whp $2\rho|\hat{r} - r| = o(\delta r)$, and thus, since by Proposition 1.1 we have $\hat{r}/r \to 1$ whp, we have $2\rho|\hat{r} - r| \le \delta\hat{r}$ whp. Also, by the same Proposition and again choosing δ tending to 0 with n sufficiently slowly, whp

$$(1 + \delta/2)r = (1 + \delta/2)(1 + o(1))\hat{r} \le (1 + \delta)\hat{r}.$$

Putting these bounds into (13) completes the proof of the lemma. $\square$

4.   PROOF OF THEOREM 1.2

In this section, we prove Theorem 1.2, on the reconstruction of random geometric graphs.

Throughout this section, let $\omega = \omega(n)$ be any function tending to infinity slowly as $n \to \infty$, and in particular such that $\omega \ll \sqrt{\log n}$. We shall assume at various places without further comment that n is sufficiently large. Let $\mathcal{B}_1$ be the event that we find vertices $v_1, \ldots, v_4$ such that $d_E(\mathbf{X}_{v_i}, \pi(c_i)) < \omega$ for each $i = 1, \ldots, 4$, for some (unknown) random symmetry $\pi = \pi(\Psi)$ of $\mathcal{S}_n$. By Lemma 2.8, $\mathcal{B}_1$ holds whp. Let σ_0 denote the identity symmetry. If $\mathcal{B}_1$ does not hold then let us set $\pi = \sigma_0$ (the choice of π as σ_0 will not be important). Now let σ be any given symmetry. Observe that $\mathcal{B}_1$ holds for Ψ if and only if it holds for $\sigma^{-1} \circ \Psi$; on $\mathcal{B}_1$, $\pi(\Psi) = \sigma$ if and only if $\pi(\sigma^{-1} \circ \Psi) = \sigma_0$, and Ψ and $\sigma^{-1} \circ \Psi$ have the same distribution. Thus for each symmetry σ

$$\mathbb{P}(\mathcal{B}_1 \wedge (\pi = \sigma)) = \mathbb{P}(\mathcal{B}_1 \wedge (\pi = \sigma_0)), \qquad (14)$$

and so $\mathbb{P}(\pi = \sigma) \to \frac{1}{8}$ as $n \to \infty$. Since we are using the symmetry-adjusted measure d^*, we may treat the random symmetry π as if it were the identity, as we shall check below. We set $\Phi(v_i) = c_i$, and still have to assign $\Phi(v)$ for all other vertices $v \in V(G)$. Let $\mathcal{B}_2$ be the event that $\mathcal{B}_1$ holds and π is the identity. Thus $\mathbb{P}(\mathcal{B}_2) \to \frac{1}{8}$ as $n \to \infty$. Recall that we are given $\varepsilon > 0$: we may assume wlog that $\varepsilon < \frac{1}{2}$ say. The main step in the proof will be to show that Φ can be defined for all vertices in such a way that

$$\text{conditional on } \mathcal{B}_2, \text{ we have } d_{\max}(\Psi, \Phi) < (1 + \varepsilon)r \text{ whp.} \qquad (15)$$

(Of course $d^*(\Psi, \Phi) \le d_{\max}(\Psi, \Phi)$.)

Let us prove the claim (15). By Proposition 1.1, whp $|\hat{r} - r| \ll \omega\rho^{-3/2}$. Hence, whp, for each pair u, v of distinct vertices

$$\hat{r}(d_G(u, v) + \varepsilon/4) \le r(d_G(u, v) + \varepsilon/4) + \omega\rho^{-3/2}(2d_G(u, v)).$$

But by Corollary 2.6, whp, for each pair u, v of vertices

$$rd_G(u,v) \leq d_E(\mathbf{X}_u, \mathbf{X}_v) + 2r \leq \sqrt{2n} + 2r \leq 2\sqrt{n},$$

so

$$\omega\rho^{-3/2}d_G(u,v) = \frac{\omega r^{1/2}\, rd_G(u,v)}{n^{3/4}} \leq \frac{2\omega r^{1/2}}{n^{1/4}} = 2\omega\rho^{-1/2} = o(r);$$

and thus

$$\hat{r}(d_G(u,v) + \varepsilon/4) \leq r(d_G(u,v) + \varepsilon/3).$$

By the same argument we obtain whp, for each pair u, v of vertices

$$\begin{aligned}
\hat{r}(d_G(u,v) - (1 + \varepsilon/4)) &\geq r(d_G(u,v) - (1 + \varepsilon/4)) - \omega\rho^{-3/2}d_G(u,v) \\
&\geq r(d_G(u,v) - (1 + \varepsilon/3)).
\end{aligned}$$

By Lemma 3.3 with $\delta = \varepsilon/5$, whp, for each pair u, v of vertices

$$\hat{r}(d_G(u,v) + \varepsilon/5) \geq d_E(\mathbf{X}_u, \mathbf{X}_v) \geq \hat{r}(d_G(u,v) - (1 + \varepsilon/5)).$$

But $\omega + \hat{r}\varepsilon/5 \leq \hat{r}\varepsilon/4$ whp. Hence, conditional on the event $\mathcal{B}_2$, whp, for each $i \in [4]$ and vertex $v \in V^- = V \setminus \{v_1, \ldots, v_4\}$, we have

$$\begin{aligned}
r(d_G(v,v_i) + \varepsilon/3) &\geq \hat{r}(d_G(v,v_i) + \varepsilon/4) \geq d_E(\mathbf{X}_v, \mathbf{X}_{v_i}) + \omega \\
&\geq d_E(\mathbf{X}_v, c_i) \\
&\geq d_E(\mathbf{X}_v, \mathbf{X}_{v_i}) - \omega \geq \hat{r}(d_G(v,v_i) - (1 + \varepsilon/4)) \geq r(d_G(v,v_i) - (1 + \varepsilon/3)).
\end{aligned}$$

Let $\mathcal{B}_3$ be the event that these last inequalities hold, so $\mathcal{B}_3$ holds whp.

Condition on the events $\mathcal{B}_2$ and $\mathcal{B}_3$, and let v_1, v_2, v_3, v_4 be the 'corner' vertices found. We shall show that $d_{\max}(\Psi, \Phi) < (1+\varepsilon)r$ (deterministically): this will establish (15), since then

$$\mathbb{P}(d_{\max}(\Psi, \Phi) \geq (1 + \varepsilon)r \mid \mathcal{B}_2) \leq \mathbb{P}(\overline{\mathcal{B}_3})/\mathbb{P}(\mathcal{B}_2) = o(1).$$

By symmetry, we may assume for convenience that $v_i = i$ for each $i \in [4]$.

For each $i \in [4]$, let $Q(n, i)$ denote the quarter of $\mathcal{S}_n$ containing the corner c_i. For each vertex $v \in V^-$ there is a 'nearest corner' in terms of graph distance to the vertices $v_1, \ldots, v_4$. We need a lemma concerning such a nearest corner.

Lemma 4.1. *Fix $j \in [4]$. For each vertex $v \in V^-$ such that $d_G(v,v_j) = \min_{1\leq i \leq 4} d_G(v,v_i)$, the corresponding point $\mathbf{X}_v$ lies within distance at most r of the quarter $Q(n,j)$ of $\mathcal{S}_n$.*

Proof of Lemma 4.1. Suppose wlog that $j = 4$. Let $v \in V^-$, and suppose for a contradiction that $\mathbf{X}_v$ is not within distance r of $Q(n,4)$. Assume that $\mathbf{X}_v \in Q(n,1)$ (we shall consider other cases later). Let us first check that the minimum value of $d_E(\mathbf{x}, c_4) - d_E(\mathbf{x}, c_1)$ over all points $\mathbf{x}$ in $Q(n,1)$ at distance

$\geq r$ from $Q(n, 4)$ is attained at $\mathbf{x} = \mathbf{x}^*$ where $\mathbf{x}^* = (-r, 0)$. To see this, let us observe first that the minimum must be attained for some point $(-r, -\frac{\sqrt{n}}{2} + z)$ with $z \in [0, \frac{\sqrt{n}}{2}]$, as otherwise one could obtain a smaller solution by shifting horizontally to the right until hitting the line $y = -r$. Next, for a given point $\mathbf{x} = (-r, -\frac{\sqrt{n}}{2} + z)$ we have

$$d_E(\mathbf{x}, c_4) - d_E(\mathbf{x}, c_1) = \sqrt{z^2 + (\tfrac{\sqrt{n}}{2} + r)^2} - \sqrt{z^2 + (\tfrac{\sqrt{n}}{2} - r)^2}.$$

The derivative with respect to z of the previous expression is

$$\frac{z}{\sqrt{z^2 + (\tfrac{\sqrt{n}}{2} + r)^2}} - \frac{z}{\sqrt{z^2 + (\tfrac{\sqrt{n}}{2} - r)^2}},$$

which is clearly negative, since the denominator in the first term is bigger than in the second one. Hence, $d_E(\mathbf{x}, c_4) - d_E(\mathbf{x}, c_1)$ is decreasing in z, and so it is minimised at $\mathbf{x} = \mathbf{x}^*$, as we wished to show. Hence, all points $\mathbf{x}$ in $Q(n, 1)$ at distance $\geq r$ from $Q(n, 4)$ satisfy

$$
\begin{aligned}
d_E(\mathbf{x}, c_4) - d_E(\mathbf{x}, c_1) &\geq d_E(\mathbf{x}^*, c_4) - d_E(\mathbf{x}^*, c_1) \\
&= \sqrt{(\tfrac{\sqrt{n}}{2})^2 + (\tfrac{\sqrt{n}}{2} + r)^2} - \sqrt{(\tfrac{\sqrt{n}}{2})^2 + (\tfrac{\sqrt{n}}{2} - r)^2} \\
&= \sqrt{n/2 + \sqrt{n}r(1 + o(1))} - \sqrt{n/2 - \sqrt{n}r(1 + o(1))} \\
&= \sqrt{\tfrac{n}{2}\left(1 + \tfrac{r}{\sqrt{n}}(1 + o(1))\right)} - \sqrt{\tfrac{n}{2}\left(1 - \tfrac{r}{\sqrt{n}}(1 + o(1))\right)} \\
&= (\sqrt{2} + o(1))r. \qquad (16)
\end{aligned}
$$

But, since $\mathcal{B}_3$ holds,

$$d_G(v, v_1) \leq d_E(\mathbf{X}_v, c_1)/r + 1 + \varepsilon/3$$

and

$$d_G(v, v_4) \geq d_E(\mathbf{X}_v, c_4)/r - \varepsilon/3.$$

Hence, by (16)

$$
\begin{aligned}
d_G(v, v_4) &\geq d_E(\mathbf{X}_v, c_4)/r - \varepsilon/3 \\
&\geq d_E(\mathbf{X}_v, c_1)/r + \sqrt{2} + o(1) - \varepsilon/3 \\
&\geq d_G(v, v_1) + \sqrt{2} - 1 + o(1) - 2\varepsilon/3 \; > \; d_G(v, v_1)
\end{aligned}
$$

(for n sufficiently large), a contradiction. Thus we cannot have $\mathbf{X}_v \in Q(n, 1)$.

The case when $\mathbf{X}_v$ is in $Q(n, 3)$ is analogous. Finally consider the case when $\mathbf{X}_v$ is in $Q(n, 2)$. Now we shall check that the minimum value of $d_E(\mathbf{x}, c_4) - d_E(\mathbf{x}, c_2)$ over all points $\mathbf{x}$ in $Q(n, 2)$ at distance $\geq r$ from $Q(n, 4)$ is again attained at $\mathbf{x} = \mathbf{x}^* = (-r, 0)$ (or at $\mathbf{x} = (0, r)$). To see this, observe much as before that the minimum must be attained at distance exactly r from $(0, 0)$, as

otherwise one could obtain a smaller solution by shifting $\mathbf{x}$ along the straight line connecting $\mathbf{x}$ with $(0,0)$, until the distance from $(0,0)$ is exactly r. Next, for a given point $\mathbf{x} = (-r\cos\theta, r\sin\theta)$ with $\theta \in [0, \pi/2]$, $d_E(\mathbf{x}, c_4) - d_E(\mathbf{x}, c_2)$ is equal to

$$\sqrt{(\frac{\sqrt{n}}{2} + r\cos\theta)^2 + (\frac{\sqrt{n}}{2} + r\sin\theta)^2} - \sqrt{(\frac{\sqrt{n}}{2} - r\cos\theta)^2 + (\frac{\sqrt{n}}{2} - r\sin\theta)^2}.$$

The derivative with respect to θ of the above expression is

$$\frac{r\sqrt{n}(\cos\theta - \sin\theta)}{2\sqrt{(\frac{\sqrt{n}}{2} + r\cos\theta)^2 + (\frac{\sqrt{n}}{2} + r\sin\theta)^2}} - \frac{r\sqrt{n}(\sin\theta - \cos\theta)}{2\sqrt{(\frac{\sqrt{n}}{2} - r\cos\theta)^2 + (\frac{\sqrt{n}}{2} - r\sin\theta)^2}}.$$

For $\theta \in [0, \pi/4]$, $\cos\theta \geq \sin\theta$, the derivative is positive (or zero), whereas for $\theta \in [\pi/4, \pi/2]$ the derivative is negative (or zero). Hence, the minimum value of $d_E(\mathbf{x}, c_4) - d_E(\mathbf{x}, c_2)$ is attained at $\mathbf{x} = \mathbf{x}^* = (-r, 0)$ (or at $\mathbf{x} = (0, r)$), as we wished to show.

Therefore, by (16), all points $\mathbf{x}$ in $Q(n, 2)$ at distance $\geq r$ from $Q(n, 4)$ satisfy

$$\begin{aligned} d_E(\mathbf{x}, c_4) - d_E(\mathbf{x}, c_2) &\geq d_E(\mathbf{x}^*, c_4) - d_E(\mathbf{x}^*, c_2) \\ &= (\sqrt{2} + o(1))r. \end{aligned}$$

The remainder of the argument is as before, and so we have completed the proof of the lemma.   □

We now resume the proof of Theorem 1.2 (still assuming that the events $\mathcal{B}_2$ and $\mathcal{B}_3$ hold). Let $v \in V^-$ be such that $d_G(v, v_j) = \min_{1 \leq i \leq 4} d_G(v, v_i)$, as in the last lemma, and assume wlog that $j = 4$. Then by the last lemma, $\min_{i \in [3]} d_E(\mathbf{X}_v, c_i) \geq \sqrt{n}/2 - r$. Denote by α the angle $c_2 \mathbf{X}_v c_3$ at $\mathbf{X}_v$ between the segments $\mathbf{X}_v c_2$ and $\mathbf{X}_v c_3$, and by β the angle $c_1 \mathbf{X}_v c_2$ at $\mathbf{X}_v$ between the segments $\mathbf{X}_v c_1$ and $\mathbf{X}_v c_2$. We do not observe α or β directly, but clearly we have $\pi/4 \leq \alpha, \beta \leq \pi/2 + o(1)$ (see Figure 3, left picture), the bounds being attained if $\mathbf{X}_v$ is c_4 and if $\mathbf{X}_v$ is near $(0, 0)$, respectively.

For each $i = 1, 2, 3$, let $R_i(v) = \hat{r}(d_G(v, v_i) - \frac{1}{2})$ $\left(= \Theta(\sqrt{n})\right)$, and let $C_i(v)$ be the circle centred on the corner c_i with radius $R_i(v)$. Also, let $A_i(v)$ be the annulus centred on c_i formed by circles of radii $R_i(v) \pm \hat{r}(\frac{1}{2} + \frac{\varepsilon}{4})$. We can construct these three circles. (We can also construct the corresponding annuli – assuming as we may that we are given a rational ε – though we do not need to do so to prove the theorem.) Note that $\mathbf{X}_v$ must lie in each of the annuli; for, since $\mathcal{B}_3$ holds, $\hat{r}(d_G(v, v_i) + \varepsilon/4) \geq d_E(\mathbf{X}_v, c_i) \geq \hat{r}(d_G(v, v_i) - (1 + \varepsilon/4))$, and hence $R_i(v) + \hat{r}(\frac{1}{2} + \frac{\varepsilon}{4}) \geq d_E(\mathbf{X}_v, c_i) \geq R_i(v) - \hat{r}(\frac{1}{2} + \frac{\varepsilon}{4})$. It is convenient to consider the circles and annuli in pairs.
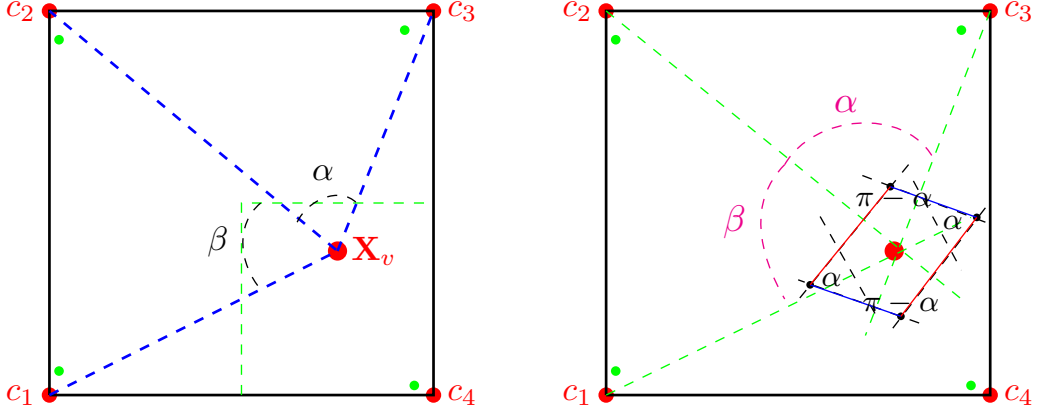
FIGURE 3. Illustration of the notation

Consider first the circles $C_2(v), C_3(v)$ and corresponding annuli $A_2(v), A_3(v)$. The circles intersect below the line c_2c_3 in a point $\mathbf{Y}_{23}(v)$, where the tangents are at angle $\alpha + o(1)$ (and $\pi - \alpha + o(1)$). The annuli intersect below the line c_2c_3 in a set $B_{23}(v)$ which is – up to lower order terms accounting for curvatures – a parallelogram $RH_{23}(v)$ with (interior) angles α and $\pi - \alpha$ (see Figure 3, right picture). (We chose to consider the circles and annuli with centres far from $\mathbf{X}_v$ so that curvatures would be negligible.) In fact, $RH_{23}(v)$ is a rhombus, as in each annulus the radii differ by the same value $(1 + \varepsilon/2)\hat{r}$; and since the heights are equal, the sides must be of equal length. Further, the point $\mathbf{Y}_{23}(v)$ is at the centre of the rhombus (up to lower order terms), where the diagonals cross. (It might happen that some part of the rhombus is actually outside $\mathcal{S}_n$, but since this makes the region which we know contains $\mathbf{X}_v$ smaller, it is only helpful for us.)

The circles $C_1(v), C_2(v)$ and corresponding annuli $A_1(v), A_2(v)$ behave in exactly the way described above for $C_2(v), C_3(v)$ and corresponding annuli. In particular, the annuli $A_1(v), A_2(v)$ intersect to the right of the line c_1c_2 in a set $B_{12}(v)$ which is close to a rhombus $RH_{12}(v)$ with angles β and $\pi - \beta$.

Now consider the circles $C_1(v), C_3(v)$ and corresponding annuli $A_1(v), A_3(v)$; and for convenience let us restrict our attention to the case when $\alpha, \beta \leq \pi/3 + o(1)$ (so $\mathbf{X}_v$ is not near the centre $(0,0)$ of $\mathcal{S}_n$; in the next paragraph we shall see why it suffices to have this assumption on α and β). The annuli $A_1(v)$, $A_3(v)$ intersect inside (or near) the bottom right quarter square $Q(n, 4)$ in a set $B_{13}(v)$ which is close to a rhombus $RH_{13}(v)$ with angles $\alpha + \beta$ and $\pi - \alpha - \beta$, where both these angles are in the interval between $\pi/2$ and $2\pi/3 + o(1)$.

Among these three pairs of circles and corresponding rhombi, we will consider one whose angles are closest to $\pi/2$. Let us check that there must be at least one with angles in the interval $[\pi/3, 2\pi/3]$ – we call the corresponding
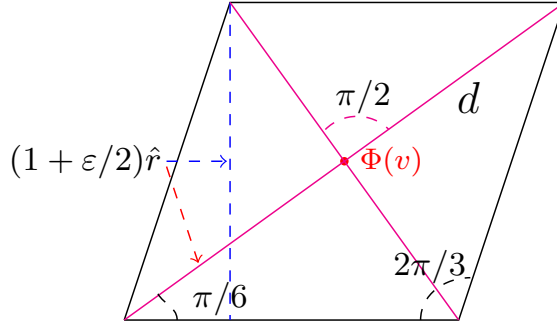
FIGURE 4. Angles in R_{23}, for the extreme case $\alpha = \pi/3$

rhombus *squarelike*. Indeed, suppose that this is not the case for $RH_{12}(v)$ or $RH_{23}(v)$. Then, since $\pi/4 \leq \alpha, \beta \leq \pi/2 + o(1)$, we must have $\alpha, \beta < \pi/3$. Then, however, $\pi/2 \leq \alpha + \beta < 2\pi/3$, and so $RH_{13}(v)$ is the desired squarelike rhombus. Further, the maximum distance from the centre $\mathbf{Y}_{13}(v)$ of the rhombus $RH_{13}(v)$ (the intersection of the diagonals) to a point in the set $B_{13}(v)$ is half the length d of the long diagonal (recall that we assume n sufficiently large, so that we can safely ignore curvature issues and we can approximate $B_{13}(v)$ arbitrarily well by a rhombus).

Pick a pair of circles and corresponding rhombus such that their angles are closest to $\pi/2$, and without loss of generality suppose that the rhombus is $RH_{23}(v)$. We set $\Phi(v)$ to be a point we calculate within distance 1 of $\mathbf{Y}_{23}(v)$ (or arbitrarily close to $\mathbf{Y}_{23}(v)$). Clearly, the further away the angles α and $\pi - \alpha$ are from $\pi/2$, the longer the long diagonal, and we may thus assume the worst case of $\alpha = \pi/3$ and $\pi - \alpha = 2\pi/3$. The shorter diagonal of such a rhombus splits it into two equilateral triangles, with height $(1 + \varepsilon/2)\hat{r}$; and thus half the length d of the longer diagonal is also $(1 + \varepsilon/2)\hat{r}$, see Figure 4. Thus in general

$$d/2 \leq (1 + \varepsilon/2)\hat{r} \leq (1 + \varepsilon)r.$$

Hence, the Euclidean distance from $\Phi(v)$ to any point inside $B_{23}(v)$ is at most $d/2 \leq (1 + \varepsilon)r$. But $\mathbf{X}_v \in B_{23}(v)$, so $d_E(\Phi(v), \mathbf{X}_v) \leq (1 + \varepsilon)r$. This holds for each $v \in V^-$, so we have found an embedding Φ with displacement at most $(1 + \varepsilon)r$. (If the point of the intersection of the two diagonals falls outside $\mathcal{S}_n$, then we project this point to the closest point on the boundary of $\mathcal{S}_n$, and clearly the distance to $\mathbf{X}_v$ can only decrease).

We have now established (15), and it remains only to justify treating the random symmetry π as the identity. We want to replace the conditioning on $\mathcal{B}_2$ in (15) by conditioning on $\mathcal{B}_1$.

Let $t > 0$ and let σ be a symmetry. Arguing as for (14), and noting also that $d^*(\Psi, \Phi) = d^*(\sigma^{-1} \circ \Psi, \Phi)$, we have

$$\mathbb{P}\left(\mathcal{B}_1 \wedge (d^*(\Psi, \Phi) \leq t) \wedge (\pi(\Psi) = \sigma)\right)$$
$$= \mathbb{P}\left(\mathcal{B}_1 \wedge (d^*(\sigma^{-1} \circ \Psi, \Phi) \leq t) \wedge (\pi(\sigma^{-1} \circ \Psi) = \sigma_0)\right)$$
$$= \mathbb{P}\left(\mathcal{B}_1 \wedge (d^*(\Psi, \Phi) \leq t) \wedge (\pi(\Psi) = \sigma_0)\right);$$

so, summing over σ we have

$$\mathbb{P}\left(\mathcal{B}_1 \wedge (d^*(\Psi, \Phi) \leq t)\right) = 8\,\mathbb{P}\left(\mathcal{B}_2 \wedge (d^*(\Psi, \Phi) \leq t)\right).$$

But, similarly to (14), we have $\mathbb{P}(\mathcal{B}_1) = 8\,\mathbb{P}(\mathcal{B}_2)$, so

$$\mathbb{P}(d^*(\Psi, \Phi) \leq t \mid \mathcal{B}_1) = \mathbb{P}(d^*(\Psi, \Phi) \leq t \mid \mathcal{B}_2).$$

Hence

$$\mathbb{P}(d^*(\Psi, \Phi) \leq (1 + \varepsilon)r \mid \mathcal{B}_1) = \mathbb{P}(d^*(\Psi, \Phi) \leq (1 + \varepsilon)r \mid \mathcal{B}_2) = 1 - o(1)$$

by (15). Since $\mathcal{B}_1$ holds whp, this completes the proof of Theorem 1.2.

Running time analysis

The $O(n^2)$ bound on the running time follows from two observations. On the one hand, by Proposition 1.1 at most $O(n^2)$ steps are needed for calculating $\hat{r}$. On the other hand, the four vertices close to the four corners can be found in time $O(n)$; and once they have been found, all graph distances from them can be computed in time $O(n^2)$ (for example, by breadth-first search). If ε is not rational, then the algorithm can be applied with some rational ε' such that $0 < \varepsilon' < \varepsilon$.

5.   Proof of Theorem 1.3

In this section, we prove Theorem 1.3, on approximate reconstruction from the family of vertex orderings. As in the algorithm in Theorem 1.2, the algorithm here has two main steps. In the first subsection we give a sketch of the method, in the second subsection we fill in details of step (a), in the third subsection we give several preliminary results needed to analyse step (b), and in the final subsection we complete the proof of the theorem.

5.1.   Sketch of the algorithm. The algorithm has two main steps.

- Step (a) We identify four vertices v_i such that the corresponding points are whp near the four corners of $\mathcal{S}_n$, and set $\Phi(v_i) = c_i$.
- Step (b) For each other vertex v we construct two circles, and two corresponding thin annuli both whp containing $\mathbf{X}_v$, centred on a chosen pair of corners, such that the circles meet at an angle between $\pi/3$

and $2\pi/3$; and we set $\Phi(v)$ to be the relevant point of intersection of the circles (which is essentially the centre of the rhombus formed by the intersection of the annuli, as before).

We obtain a much smaller displacement error than with random geometric graphs in Theorem 1.2 since our annuli are much thinner. We start with a sketch of the two steps (a) and (b) and of the proofs, before giving the full proofs. First, however, we introduce some useful notation.

Notation For each pair u, v of vertices, we let $k(u, v)$ be the rank of v in the vertex-ordering τ_u. Thus $k(u, u) = 1$; and if v is last in the order τ_u (farthest from u) then $k(u, v) = n$, and we let far(u) = v. We assume that, given vertices u and v, in constant time we can find the successor of v in the list specifying τ_u. Note that by reading through this list we can determine far(u) in $O(n)$ steps.

For $0 \leq s \leq \sqrt{2}$, let $\lambda(s)$ be the area of the set of points in the unit square $\mathcal{S}_1$, centred at $(0,0)$, within distance s of a fixed corner point, say $(-\frac{1}{2}, -\frac{1}{2})$. Observe that λ gives an increasing bijection from $[0, \sqrt{2}]$ to $[0, 1]$. For $0 \leq t \leq 1$ let $s(t) = \lambda^{-1}(t)$, the unique value s in $[0, \sqrt{2}]$ such that $\lambda(s) = t$. Also, define $\lambda_n(s) = \lambda(s/\sqrt{n})\, n$ for $0 \leq s \leq \sqrt{2n}$, so $\lambda_n(s)$ is the area of the set of points in the square $\mathcal{S}_n$ within distance s from a fixed corner point; and define $s_n(k) = s(k/n)\sqrt{n}$, and note that $\lambda_n(s_n(k)) = k$.

It will be convenient here to say that a sequence A_n of events holds *with very high probability* (wvhp) if $\mathbb{P}(A_n) = 1 - o(1/n)$ as $n \to \infty$. Finally, let $\omega = \omega(n) \to \infty$ slowly, and in particular assume as in the previous section that $\omega \ll \sqrt{\log n}$.

Sketch of step (a): finding points near the corners of $\mathcal{S}_n$

We call a pair of vertices $\{v, v'\}$ *extreme* if $v' = \text{far(v)}$ and $v = \text{far(v')}$. in the ordering $\tau_{v'}$. We show that for each pair of opposite corners c, c' whp there is an extreme pair $\{v, v'\}$ with $\mathbf{X}_v$ close to c and $\mathbf{X}_{v'}$ close to c', and we can find such extreme pairs quickly. Thus the following event $\mathcal{C}_1$ holds whp.

Let $\mathcal{C}_1$ be the event that this procedure yields vertices $v_1, \ldots, v_4$ such that, for some (unknown, random) symmetry π of $\mathcal{S}_n$, we have $\mathbf{X}_{v_i} \in B(\pi(c_i), \omega)$ for each $i \in [4]$ (so $\mathbf{X}_{v_i}$ is very close to the corner $\pi(c_i)$). Also, for given distinct vertices $v_1, \ldots, v_4$ let $\mathcal{C}_1(v_1, \ldots, v_4)$ be the event that $\mathcal{C}_1$ holds with this choice of the 'corner' vertices.

Sketch of step (b): constructing the circles and annuli

Suppose that the event $\mathcal{C}_1(v_1, \ldots, v_4)$ holds. We use the orders $\tau_{v_1}, \ldots, \tau_{v_4}$ to estimate the distances from the corners.

Let $V^- = V \setminus \{v_1, \ldots, v_4\}$. For each vertex $v \in V^-$, we define $i_0 = i_0(v)$ to be the least $j \in [4]$ such that $k(v_j, v) = \min_{i \in [4]} k(v_i, v)$ (picking the least j is just a tie-breaker). (Thus $\pi(c_{i_0})$ is likely to be the closest corner to $\mathbf{X}_v$.) Fix $v \in V^-$, and let $I^- = [4] \setminus \{i_0\}$. We consider the three orders τ_{v_i} for $i \in I^-$. (We do not use $\tau_{v_{i_0}}$, and do not consider distances from $\pi(c_{i_0})$, so that we work only with 'large' distances, and thus we do not need to worry about curvature, exactly as before.) We want to find a pair of thin annuli, centred on two of the three corners near the points $\mathbf{X}_{v_i}$ for $i \in I^-$, such that wvhp the 'near-rhombus' formed by the intersection of the two annuli is squarelike, and wvhp $\mathbf{X}_v$ is in this near-rhombus.

We shall see that, for each $i \in I^-$, we have $k(v_i, v) > 0.19\,n$ wvhp (so we will work only with 'large' distances); and for $i = i_0 \pm 1$, we have $k(v_i, v) < 0.91n$ wvhp. (Indices in $[4]$ are always taken mod 4.) Let $\alpha_0 = \frac{\pi}{9} + \frac{1}{\sqrt{3}} \approx 0.9264$: later we shall choose a rational constant α slightly bigger than α_0. When $k(v_i, v)$ is at least $0.19n$ and at most αn, we have a good estimate of $d_E(\pi(c_i), \mathbf{X}_v)$ (see Lemma 5.3). Also, as in the proof of Theorem 1.2, it suffices to consider the case when π is the identity map.

There are two cases depending on the rank $k(v_{i_0+2}, v)$ (note that v_{i_0} and v_{i_0+2} are at opposite corners of $\mathcal{S}_n$, and so $k(v_{i_0+2}, v)$ is likely to be the largest of the values $k(v_i, v)$ for $i \in I^-$): case (i) when $k(v_{i_0+2}, v) \leq \alpha n$, and case (ii) when $k(v_{i_0+2}, v) > \alpha n$. In case (i) we form three circles and three thin annuli, and then choose a best pair of them, as in the proof of Theorem 1.2. In case (ii), we just use the two circles and thin annuli centred on the corners c_{i_0-1} and c_{i_0+1} (see Lemma 5.4).

5.2. Filling in the details for step (a). We need to show that the method sketched above works. We first consider step (a), and show that indeed the event $\mathcal{C}_1$ holds whp. We need one deterministic preliminary lemma.

Lemma 5.1. *Let $\mathbf{x} \in \mathcal{S}_n$, and let*

$$t = \max\{d_E(\mathbf{x}, \mathbf{y}) : \mathbf{y} \in \mathcal{S}_n\} = \max_i d_E(\mathbf{x}, c_i)$$

(and note that $t \geq \sqrt{n/2}$). Then (assuming that n is sufficiently large)

$$\max\{d_E(\mathbf{x}, \mathbf{y}) : \mathbf{y} \in (\mathcal{S}_n \setminus \cup_i B^o(c_i, \omega))\} \leq t - \omega/3.$$

(Here B^o denotes an open ball.)

Proof of Lemma 5.1. Suppose wlog that $\mathbf{x}$ is in the bottom left quarter of $\mathcal{S}_n$ (containing c_1). It is easy to see that $d_E(\mathbf{x}, c_3) = t$, and $\max\{d_E(\mathbf{x}, \mathbf{y}) : \mathbf{y} \in (\mathcal{S}_n \setminus \cup_i B^o(c_i, \omega))\}$ is achieved at some point $\mathbf{y} \in \mathcal{S}_n$ with $d_E(\mathbf{y}, c_3) = \omega$.

Let $c_3 - \mathbf{x} = (a, b)$, so $\sqrt{n/2} \leq a, b \leq \sqrt{2n}$, and $t = \sqrt{a^2 + b^2}$. Suppose further wlog that $a \geq b$ (that is, $\mathbf{x}$ lies on or above the line $y = x$), and note that $a \leq 2b$.

Consider a point $\mathbf{y}$ with $d_E(c_3, \mathbf{y}) = \omega$. We claim that

$$d_E(\mathbf{x}, \mathbf{y}) \leq t - (1 + o(1))\omega/\sqrt{5}. \tag{17}$$

To see this, write $c_3 - \mathbf{y} = (p, q)$. Then $p, q \geq 0$ and $p^2 + q^2 = \omega^2$, so $p + q \geq \omega$; and we have

$$
\begin{aligned}
d_E(\mathbf{x}, \mathbf{y}) &= \left((a - p)^2 + (b - q)^2\right)^{1/2} \\
&= \left(t^2 - (1 + o(1))(2ap + 2bq)\right)^{1/2} \\
&= t\left(1 - (1 + o(1))(ap + bq)/t^2\right) \\
&= t - (1 + o(1))(ap + bq)/t.
\end{aligned}
$$

But $a \geq b$ and $p + q \geq \omega$, so

$$d_E(\mathbf{x}, \mathbf{y}) \leq t - (1 + o(1))b(p + q)/t \leq t - (1 + o(1))b\omega/t.$$

Also, $a \leq 2b$ so $b/t \geq b/\sqrt{4b^2 + b^2} = 1/\sqrt{5}$. The claim (17) now follows, and this completes the proof of the lemma.   □

Finding points near the corners of $\mathcal{S}_n$

Let $\mathcal{C}_2$ be the event that, for each $i \in [4]$, there is a vertex u_i such $\mathbf{X}_{u_i} \in B(c_i, \omega/4)$. Then $\mathcal{C}_2$ holds whp. To see this, note that, for a fixed $i \in [4]$

$$\mathbb{P}(\mathbf{X}_v \notin B(c_i, \omega/4) \text{ for each } v) = (1 - \tfrac{1}{4}\pi(\omega/4)^2/n)^n < e^{-\frac{\pi}{64}\omega^2} = o(1);$$

and use a union bound. From now on, assume that $\mathcal{C}_2$ holds. We claim that, for each vertex v, we have

$$\mathbf{X}_{\mathrm{far}(v)} \in \bigcup_{i=1}^{4} B(c_i, \omega). \tag{18}$$

To see this, suppose wlog that c_1 is the corner farthest from $\mathbf{X}_v$: then

$$d_E(\mathbf{X}_v, \mathbf{X}_{\mathrm{far}(v)}) \geq d_E(\mathbf{X}_v, c_1) - \omega/4,$$

since there is a point $\mathbf{X}_u$ in $B(c_1, \omega/4)$; and so, by Lemma 5.1, for each vertex v' such that $\mathbf{X}_{v'} \notin \bigcup_{i=1}^{4} B(c_i, \omega)$,

$$d_E(\mathbf{X}_v, \mathbf{X}_{v'}) \leq d_E(\mathbf{X}_v, c_1) - \omega/3 < d_E(\mathbf{X}_v, \mathbf{X}_{\mathrm{far}(v)}),$$

which establishes the claim (18).

Pick any vertex v, let $u = \mathrm{far}(v)$, and let $u' = \mathrm{far}(u)$. Then by (18) there is a pair of opposite corners c, c' such that $\mathbf{X}_u \in B(c, \omega)$ and $\mathbf{X}_{u'} \in B(c', \omega)$. We may take v_1 as u and v_3 as u'. Let us suppose wlog that c is c_1 and c' is c_3.

Now we must find two vertices such that the corresponding points are close to the opposite corners c_2 and c_4, and this needs more care. Let $r = 0.9\sqrt{n}$ (the exact value 0.9 is not important), and let $B_i = B(c_i, r)$. We claim that, if $\mathbf{x} \in \mathcal{S}_n \setminus (B_1 \cup B_3)$ then

$$\max\{d(c_2, \mathbf{x}), d(c_4, \mathbf{x})\} - \max\{d(c_1, \mathbf{x}), d(c_3, \mathbf{x})\} = \Omega(\sqrt{n}). \qquad (19)$$

Proof of claim (19). The circles of radius r centred on c_1 and c_3 meet twice, at points P and Q on the diagonal line between c_2 and c_4, both at distance t from the origin, where $t^2 = r^2 - n/2 = 0.31n$. The first maximum value is minimised at these points P and Q. Thus the first maximum value is at least $t + \sqrt{n/2} \approx 1.264\sqrt{n}$. The second maximum value is maximised at the four points on the sides of $\mathcal{S}_n$ at distance $0.1\sqrt{n}$ from c_2 and c_4. Thus the second maximum value is at most $\sqrt{1^2 + 0.1^2}\sqrt{n} \approx 1.005\sqrt{n}$. Hence the difference between the maxima is at least $0.25\sqrt{n}$, proving (19). $\qquad \square$

It follows now from (18) and (19) that if we can find a vertex v with $\mathbf{X}_v \in \mathcal{S}_n \setminus (B_1 \cup B_3)$, and we set $u = \mathrm{far}(v)$ and $u' = \mathrm{far}(u)$, then $\mathbf{X}_u$ is in one of the balls $B(c_2, \omega)$ and $B(c_4, \omega)$ and $\mathbf{X}_{u'}$ is in the other. It remains then to find such a vertex v.

For a measurable subset S of $\mathcal{S}_n$, let $N(S)$ be the random number of vertices v with $\mathbf{X}_v \in S$. Observe that $N(B(c_i, s_n(k))) \sim \mathrm{Bin}(n, k/n)$, with mean k. For $i = 1, 3$,

$$B_i \subseteq B(v_i, r + \omega) \subseteq B(c_i, r + 2\omega),$$

and whp (by Chebyshev's inequality),

$$N(B(c_i, r + 2\omega)) = (\pi/4)(r + 2\omega)^2 + O(\omega\sqrt{n}) = (\pi/4)r^2 + O(\omega\sqrt{n}),$$

so whp each vertex v with $\mathbf{X}_v \in B_i$ has rank $k(v_i, v) \leq (\pi/4)r^2 + \omega^2\sqrt{n}$. Also, for $i = 1, 3$,

$$B(c_i, r + n^{1/3}) \subseteq B(v_i, r + n^{1/3} + \omega) \subseteq B_i^+ := B(c_i, r + n^{1/3} + 2\omega),$$

and whp

$$N(B(c_i, r + n^{1/3})) = (\pi/4)(r + n^{1/3})^2 + O(\omega\sqrt{n}) = (\pi/4)r^2 + \Theta(n^{5/6}),$$

so whp each vertex v with $\mathbf{X}_v \notin B_i^+$ has rank $k(v_i, v) \geq (\pi/4)r^2 + n^{4/5}$. But the area of $\mathcal{S}_n \setminus (B_1^+ \cup B_3^+)$ is $\Omega(n)$, so whp there are (many) vertices v with $\mathbf{X}_v \notin (B_1^+ \cup B_3^+)$ and so with $\min\{k(v_1, v), k(v_3, v)\} \geq (\pi/4)r^2 + n^{4/5}$. Hence, by reading through τ_{v_1} and τ_{v_3} we can find a vertex v with $\min\{k(v_1, v), k(v_3, v)\} \geq (\pi/4)r^2 + n^{4/5}$; and then $\mathbf{X}_v \in \mathcal{S}_n \setminus (B_1 \cup B_3)$, and we are done. Note that we have looked at only four of the orders τ_v.

There is an (unknown) symmetry π of $\mathcal{S}_n$ such that $\mathbf{X}_{v_i} \in B(\pi(c_i), \omega)$ for each $i \in [4]$; and it follows that $\mathcal{C}_1$ holds whp, as required in step (a).

5.3. Preliminary results needed to analyse step (b).

In this subsection we give preliminary results concerning (i) the area function $\lambda(s)$ for $\mathcal{S}_1$, (ii) the angle $c_2\mathbf{x}c_4$ at a point $\mathbf{x} \in \mathcal{S}_1$ far from c_1, (iii) estimating values $s_n(k)$ from ranks k, (iv) estimating Euclidean distances from values $s_n(k)$, (v) bounds on the ranks $k(v_i, v)$ for $i \neq i_0(v)$, and (vi) how to deal with the case when $k(v_{i_0+2}, v)$ is large.

(i) On the area function $\lambda(s)$ for $\mathcal{S}_1$

If $0 \leq s \leq 1$ then clearly $\lambda(s) = \frac{1}{4}\pi s^2$. Let $1 < s < \sqrt{2}$. Let A be the point on the right side of $\mathcal{S}_1$ at distance s from the corner point c_1, so $A = (\frac{1}{2}, -\frac{1}{2} + \sqrt{s^2 - 1})$; and similarly $B = (-\frac{1}{2} + \sqrt{s^2 - 1}, \frac{1}{2})$ is the point on the top side of $\mathcal{S}_1$ at distance s from c_1. Let $\psi = \psi(s)$ be the angle Ac_1B, which is the angle subtended at c_1 by the curved part of the boundary of $\mathcal{S}_1 \cap B(c_1, s)$. We claim that

$$\psi(s) = \sin^{-1}(2s^{-2} - 1) \tag{20}$$

and

$$\lambda(s) = \tfrac{1}{2}s^2\psi(s) + \sqrt{s^2 - 1}. \tag{21}$$

To establish this claim, let θ be the angle Ac_1c_4. Then $\cos\theta = 1/s$ and so $\cos(2\theta) = 2\cos^2\theta - 1 = 2s^{-2} - 1$. But the angle Bc_1c_2 also equals θ, so $\psi + 2\theta = \pi/2$. Thus $\sin\psi = 2s^{-2} - 1$, giving the formula for ψ in (20). Also, the sum of the areas of the triangles c_1Ac_4 and c_1Bc_2 is $\sqrt{s^2 - 1}$, and the sector with straight sides c_1A and c_1B (and internal angle ψ) has area $\frac{1}{2}s^2\psi$; and these add up to $\lambda(s)$, establishing (21).

For $s = \frac{2}{\sqrt{3}}$ we have $\psi = \sin^{-1}\frac{1}{2} = \frac{\pi}{6}$; and so

$$\lambda(s) = \tfrac{1}{2} \cdot \tfrac{4}{3} \cdot \tfrac{\pi}{6} + \tfrac{1}{\sqrt{3}} = \tfrac{\pi}{9} + \tfrac{1}{\sqrt{3}} = \alpha_0 \approx 0.926416. \tag{22}$$

Also, for $1 \leq s \leq \frac{2}{\sqrt{3}}$ we have $\psi \geq \frac{\pi}{6}$.

(ii) On the angle $c_2\mathbf{x}c_4$ at a point $\mathbf{x} \in \mathcal{S}_1$ far from c_1

We need to consider values of s near to $\sqrt{2}$. We shall show that

for each $\mathbf{x} \in \mathcal{S}_1$ with $d_E(c_1, \mathbf{x}) \geq \frac{2}{\sqrt{3}}$, the angle $c_2\mathbf{x}c_4$ is at most $\frac{2\pi}{3}$. (23)

(We shall use this result in the proof of Lemma 5.4.) To prove (23), let F (for 'far' from c_1) be the set of points $(x, y) \in \mathcal{S}_1$ with $x + y \geq \frac{1}{\sqrt{3}}$. The line $x + y = \frac{1}{\sqrt{3}}$ meets the line $x = \frac{1}{2}$ at the point $P = (\frac{1}{2}, \frac{1}{\sqrt{3}} - \frac{1}{2})$, and meets the line $y = \frac{1}{2}$ at the point $Q = (\frac{1}{\sqrt{3}} - \frac{1}{2}, \frac{1}{2})$. Consider the midpoint $\mathbf{x}^* = (\frac{1}{2\sqrt{3}}, \frac{1}{2\sqrt{3}})$ of the segment of the line $x + y = \frac{1}{\sqrt{3}}$ between P and Q.

The point $\mathbf{x}^*$ is at distance $\frac{1}{\sqrt{6}}$ from the origin O. Thus the angle $c_2\mathbf{x}^*O$ is $\tan^{-1}\frac{1/\sqrt{2}}{1/\sqrt{6}} = \tan^{-1}\sqrt{3} = \pi/3$; and so the angle $c_2\mathbf{x}^*c_4$ is $2\pi/3$. We claim that,

$$\text{for each point } \mathbf{x} \in F, \text{ the angle } c_2\mathbf{x}c_4 \text{ is at most } \tfrac{2\pi}{3}. \tag{24}$$

This will follow from the above, once we check that the angle is maximised over $\mathbf{x} \in F$ at $\mathbf{x} = \mathbf{x}^*$. Clearly it is maximised at some point on the line $x + y = \frac{1}{\sqrt{3}}$. Note that the lines c_2c_4 and $x + y = \frac{1}{\sqrt{3}}$ are parallel. Let $a, b > 0$, and consider the parallel lines $y = 0$ and $y = b$. Consider the origin O and the points $C = (a, 0)$ on the line $y = 0$. For each point $\mathbf{z} = (z, b)$ on the line $y = b$, let $\theta(z)$ be the angle $O\mathbf{z}C$. It suffices now to show that $\theta(z)$ is maximised at $z = a/2$. Write $\theta(z)$ as $\tan^{-1}\frac{z}{b} + \tan^{-1}\frac{a-z}{b}$, and differentiate: we find

$$
\begin{aligned}
\theta'(z) &= \frac{1}{1 + (z/b)^2}\frac{1}{b} + \frac{1}{1 + ((a-z)/b)^2}\left(-\frac{1}{b}\right) \\
&= \frac{ab}{(b^2 + z^2)(b^2 + (a-z)^2)}(a - 2z)
\end{aligned}
$$

after some simplification. Thus indeed $\theta(z)$ is maximised at $z = a/2$; and we have established the claim (24).

Since P and Q lie on the line $x + y = \frac{1}{\sqrt{3}}$ and

$$d_E(c_1, P) = d_E(c_1, Q) = (1 + \tfrac{1}{3})^{1/2} = \tfrac{2}{\sqrt{3}} \ (\approx 1.1547),$$

it follows that $\mathcal{S}_1 \setminus B(c_1, \frac{2}{\sqrt{3}}) \subseteq F$. This completes the proof of (23).

(iii) Estimating values $s_n(k)$ from ranks k

We now check that we can quickly estimate the values $s_n(k)$ needed to determine radii of circles and annuli.

Lemma 5.2. *For fixed $0 < c_0 < c_1 < 1$, in linear time we can estimate all the values $s_n(k)$ for $c_0n \leq k \leq c_1n$ up to an additive error ± 1.*

Proof. Let I be the set of integers k with $c_0n \leq k \leq c_1n$. Let $t = \lceil n^{1/3} \rceil$ (the exact value is not critical), let J be the set of integers k with $c_0n - t \leq k \leq c_1n$ such that $t \mid k$, and note that $|J| = \Theta(n^{2/3})$. For any given value k, by repeated bisection (or another method), we can calculate an estimate $\hat{s}_n(k)$ of $s_n(k)$ up to accuracy $\pm\frac{1}{2}$ in time $O(\log n)$. Thus we can calculate all the estimates $\hat{s}_n(k)$ for $k \in J$ in time $o(n)$. For $k \in I \setminus J$, set the estimate $\hat{s}_n(k)$ to be $\hat{s}_n(j)$ where j is the largest element of J which is at most k. We can calculate all these estimates in time $O(n)$.

It remains to check that these latter estimates for $i \in I \setminus J$ are accurate up to an additive error ± 1. Recall that, for $0 \leq s \leq \sqrt{2n}$, we let $\lambda_n(s) = \lambda(s/\sqrt{n})\, n$, the area of $B(c_1, s) \cap \mathcal{S}_n$. Thus $\lambda_n(s_n(k)) = k$. For each $k \in I$, the curved part of the boundary of $B(c_1, s_n(k)) \cap \mathcal{S}_n$ has length $\Theta(\sqrt{n})$. But if the curved part

of the boundary of $B(c_1, s) \cap \mathcal{S}_n$ has length ℓ then

$$\lambda_n(s + \tfrac{1}{2}) - \lambda_n(s) > \tfrac{1}{2}\ell. \tag{25}$$

Let us see why this is true. For $r > 0$ and $0 \le \alpha < 2\pi$, let $\mu(r, \alpha)$ be the area of a sector of radius r and central angle α, so $\mu(r, \alpha) = \tfrac{1}{2}\alpha r^2$. Note that the curved part of the boundary of such a sector has length αr. Let P and Q be the points where the curved part of the boundary of $B(c_1, s + \tfrac{1}{2}) \cap \mathcal{S}_n$ meets the sides of the square $\mathcal{S}_n$, and let β be the angle Pc_1Q. Then

$$\lambda_n(s + \tfrac{1}{2}) - \lambda_n(s) \ge \mu(s + \tfrac{1}{2}, \beta) - \mu(s, \beta) = \tfrac{1}{2}\beta((s + \tfrac{1}{2})^2 - s^2) = \tfrac{1}{2}\beta(s + \tfrac{1}{4}) > \tfrac{1}{2}\ell,$$

proving (25).

By (25) we have

$$\lambda_n(s_n(k) + \tfrac{1}{2}) = k + \Theta(\sqrt{n}) > k + t = \lambda_n(s_n(k + t)),$$

and so

$$s_n(k) < s_n(k + t) < s_n(k) + \tfrac{1}{2}.$$

Now let $i \in I \setminus J$ and let j be the largest element of J which is at most i (so $\hat{s}_n(i) = \hat{s}_n(j)$). Then

$$|\hat{s}_n(i) - s_n(i)| \le |\hat{s}_n(j) - s_n(j)| + |s_n(j) - s_n(i)| < 1.$$

Thus the estimates for $i \in I \setminus J$ are accurate up to an additive error ± 1, as required. □

(iv) Estimating Euclidean distances from values $s_n(k)$ We have seen that in linear time we can estimate the values $s_n(k)$ up to an additive error ± 1. We can use these values to estimate Euclidean distances from corners. Recall that $\alpha_0 = \lambda(\tfrac{2}{\sqrt{3}}) \approx 0.9264$, see (22).

Lemma 5.3. *Assume that $\mathcal{C}_1(v_1, \dots, v_4)$ holds. There exists $\varepsilon > 0$ such that if $\alpha = \alpha_0 + \varepsilon$ then the following holds whp. For each $i \in [4]$ and each $v \in V^-$, if $k = k(v_i, v)$ satisfies $k = \Omega(n)$ and $k \le \alpha n$, then*

$$\left| d_E(\pi(c_i), \mathbf{X}_v) - s_n(k) \right| \le 1.19695\sqrt{\log n}. \tag{26}$$

Proof. Let $i \in [4]$. Let k be an integer with $k = \Omega(n)$ and $k \le \alpha n$. Observe that $N(B(c_i, s_n(k))) \sim \text{Bin}(n, k/n)$, with mean k. Let $0 < \eta < 1$: later we shall insist that η is sufficiently small that a certain inequality holds. By (22), by taking ε sufficiently small, we may ensure that $s(k/n) \le (1 + \eta)\tfrac{2}{\sqrt{3}}$ and the angle $\psi = \psi(s(k/n))$ satisfies $\psi \ge \psi_0$, where $\psi_0 = (1 - \eta)\tfrac{\pi}{6}$. Now, for a given

constant $c > 0$,

$$\lambda_n(s_n(k) + c\sqrt{\log n}) - \lambda_n(s_n(k))$$
$$\geq \quad (1 + o(1))\, \psi_0 \left((s_n(k) + c\sqrt{\log n})^2 - s_n(k)^2 \right)$$
$$= \quad (1 + o(1))\, 2c\, \psi_0\, s_n(k)\sqrt{\log n}.$$

Also, since $\psi \leq \frac{\pi}{2}$,

$$\lambda_n(s_n(k) + c\sqrt{\log n}) - \lambda_n(s_n(k))$$
$$\leq \quad (1 + o(1))\, \tfrac{\pi}{2} \left((s_n(k) + c\sqrt{\log n})^2 - s_n(k)^2 \right)$$
$$\leq \quad (1 + o(1))\pi\, c\, s_n(k)\sqrt{\log n},$$

so

$$1 \leq \lambda_n(s_n(k) + c\sqrt{\log n})/\lambda_n(s_n(k)) \leq 1 + O\left(\sqrt{\tfrac{\log n}{n}}\right) = 1 + o(1).$$

Let $X^+ = N(B(c_i, s_n(k) + c\sqrt{\log n}))$. By Lemma 2.1, since $s_n(k)\sqrt{\log n}/\mathbb{E}[X^+] = o(1)$,

$$\mathbb{P}(X^+ \leq k) = \mathbb{P}(X^+ \leq \mathbb{E}[X^+](1 - (1 + o(1))\, 2c\, \psi_0\, s_n(k)\sqrt{\log n}/\mathbb{E}[X^+]))$$
$$\leq \exp\{-(1 + o(1))\tfrac{1}{2} \left(2c\, \psi_0\, s_n(k)\sqrt{\log n}/\mathbb{E}[X^+]\right)^2)\mathbb{E}[X^+]\}$$
$$\leq \exp\{-(1 + o(1))\, 2\, c^2\psi_0^2\, s_n(k)^2 \log n/\mathbb{E}[X^+]\}.$$

But

$$\mathbb{E}[X^+] = (1 + o(1))\, k = \lambda_n(s_n(k)) \leq \tfrac{1}{4}\pi\, s_n(k)^2,$$

so

$$\mathbb{P}(X^+ \leq k) \leq \exp\left(-(1 + o(1))\frac{\frac{1}{4}\pi s_n(k)^2}{\mathbb{E}[X^+]}\frac{8}{\pi}\, c^2\psi_0^2 \log n \right)$$
$$\leq \exp\left(-(1 + o(1))\tfrac{2\pi}{9}c^2(1 - \eta)^2 \log n \right).$$

Note that $\sqrt{9/(2\pi)} \approx 1.196827$. Set $c = 1.1969$, and choose $\eta > 0$ sufficiently small that $\frac{2\pi}{9}c^2(1 - \eta)^2 > 1$. Now we have $\mathbb{P}(X^+ \leq k) = o(1/n)$.

Similarly, let $X^- = N(B(c_i, s_n(k) - c\sqrt{\log n}))$: then, with the same value of c,

$$\mathbb{P}(X^- \geq k) = o(1/n).$$

Thus we have seen that whp the following holds. For each $i \in [4]$ and each $v \in V^-$, if $k = k(v_i, v)$ satisfies $k = \Omega(n)$ and $k \leq \alpha n$, then

$$N(B(c_i, s_n(k) - 1.1969\sqrt{\log n})) < k \quad \text{and} \quad N(B(c_i, s_n(k) + 1.1969\sqrt{\log n})) > k.$$

Now, for each $i \in [4]$, $d_E(\pi(c_i), v_i) < \omega \ll \sqrt{\log n}$. Also

$$
\begin{aligned}
B(\pi(c_i), s_n(k) + 1.1969\sqrt{\log n}) \ & \subseteq \ B(v_i, s_n(k) + 1.1969\sqrt{\log n} + \omega) \\
& \subseteq \ B(\pi(c_i), s_n(k) + 1.1969\sqrt{\log n} + 2\omega) \\
& \subseteq \ B(\pi(c_i), s_n(k) + 1.19695\sqrt{\log n})
\end{aligned}
$$

(for n sufficiently large). But the first of these four balls contains more than k points $\mathbf{X}_u$, so $\mathbf{X}_v$ must be in the second ball, and so it is in the last one; that is $d_E(\pi(c_i), \mathbf{X}_v) \leq s_n(k) + 1.19695\sqrt{\log n}$. Similarly, $\mathbf{X}_v \notin B(\pi(c_i), s_n(k) - 1.19695\sqrt{\log n})$, and the lemma follows. $\square$

(v) Bounding the ranks $k(v_i, v)$ for $i \neq i_0(v)$

Condition throughout on the event $\mathcal{C}_1$, and on a particular choice of $v_1, \ldots, v_4$; that is, condition on the event $\mathcal{C}_1(v_1, \ldots, v_4)$. Let $V^- = V \setminus \{v_1, \ldots, v_4\}$. Recall that, for each $i \in [4]$ and $v \in V^-$, $k(v_i, v)$ is the rank of v in the order τ_{v_i}. Since v_i is very close to $\pi(c_i)$ whp, we may think of $k(v_i, v)$ as roughly the number of points $\mathbf{X}_u$ for $u \in V$ which are as close to $\pi(c_i)$ as $\mathbf{X}_v$ is. Let $\mathcal{C}_5$ be the event that, for each $j \in [4]$,

$$
\left| \{ u \in V^- : d_E(c_j, \mathbf{X}_u) < \tfrac{1}{2}\sqrt{n} - 2\omega \} \right| \geq \tfrac{\pi}{16}n - n^{2/3}.
$$

Then $\mathcal{C}_5$ holds whp, by Chebyshev's inequality. Let $i \in [4]$ and let $v \in V^-$. If $d_E(\pi(c_i), \mathbf{X}_v) \geq \tfrac{1}{2}\sqrt{n}$ then $d_E(\mathbf{X}_{v_i}, \mathbf{X}_v) \geq \tfrac{1}{2}\sqrt{n} - \omega$, and so each vertex u such that $d_E(\pi(c_i), \mathbf{X}_u) < \tfrac{1}{2}\sqrt{n} - 2\omega$ satisfies $d_E(\mathbf{X}_{v_i}, \mathbf{X}_u) < d_E(\mathbf{X}_{v_i}, \mathbf{X}_v)$; hence, if $d_E(\pi(c_i), \mathbf{X}_v) \geq \tfrac{1}{2}\sqrt{n}$ and $\mathcal{C}_5$ holds, then $k(v_i, v) > \tfrac{\pi}{16}n - n^{2/3}$.

Recall that, given $v \in V^-$, the index $i_0 = i_0(v) \in [4]$ satisfies $k(v_{i_0}, v) = \min_{i \in [4]} k(v_i, v)$ (breaking ties by choosing the least such value i). Condition on $\mathcal{C}_5$ holding. Then, for each $v \in V^-$ and each $i \in [4] \setminus \{i_0\}$, we have $k(v_i, v) > \tfrac{\pi}{16}n - n^{2/3}$. (For, if not, then both $d_E(\pi(c_{i_0}), \mathbf{X}_v) < \tfrac{1}{2}\sqrt{n}$ and $d_E(\pi(c_i), \mathbf{X}_v) < \tfrac{1}{2}\sqrt{n}$, which is not possible since the distance between distinct corners is at least $\sqrt{n}$.) Note that $\pi/16 \approx 0.1963 > 0.19$. Hence, for each $i \in [4] \setminus \{i_0\}$ we have $k(v_i, v) > 0.19\,n$.

Next, we show that for $i = i_0 \pm 1$ (indices are taken modulo 4), we have $k(v_i, v) \leq \alpha n$. Assume wlog that $i_0 = 1$, and consider $i = 2$. We saw earlier that $\mathbf{X}_v$ is within distance ω of the quarter square containing $\pi(c_1)$. Recall that v_2 is close to the corner $\pi(c_2)$. The maximum distance from $\pi(c_2)$ to $\mathbf{X}_v$ is $(1 + o(1))\sqrt{5n}/2$. But $\lambda(\sqrt{5}/2) = \tfrac{5}{8}\sin^{-1}\tfrac{3}{5} + \tfrac{1}{2} \approx 0.902188$. Thus the area of $B(\pi(c_2), (1 + o(1))\sqrt{5n}/2)$ is $< 0.905n$. Hence, by Lemma 2.1, wvhp the number of vertices w with $\mathbf{X}_w \in B(\pi(c_2), (1 + o(1))\sqrt{5n}/2)$ is less than $0.91n$, so $k(v_2, v) < 0.91n < \alpha n$, as required.

(vi) The case when $k(v_{i_0+2}, v)$ is large

For $i = i_0 + 2$, we have $k(v_i, v) > 0.19n$, but the upper bound $k \leq \alpha n$ might or might not hold. In order to deal with both cases, we need another auxiliary lemma.

Lemma 5.4. *For $v \in V^-$, let $i_0 = i_0(v)$. Then the following holds whp. For each $v \in V^-$ with $k(v_{i_0+2}, v) > \alpha n$, the near-rhombus formed from the intersection of the two annuli centred on the corners $\pi(c_{i_0-1})$ and $\pi(c_{i_0+1})$ is squarelike, i.e., the angles in the near-rhombus are between $\pi/3$ and $2\pi/3$.*

Proof. Let A_1 be the event that $v_i \in B(\pi(c_i), \omega)$ for each $i \in [4]$. Then A_1 holds whp. Let A_2 be the event that $N(B(c_i, \frac{2}{\sqrt{3}}\sqrt{n}+\omega)) < \alpha n$ for each $i \in [4]$. Then A_2 holds wvhp by Lemma 2.1, since by (22) the area of $B(\pi(c_{i_0+2}), \frac{2}{\sqrt{3}}\sqrt{n}+\omega)$ is $(1 + o(1))\alpha_0 n$, and $\alpha > \alpha_0$.

For $v \in V^-$, let $A_3(v)$ be the event that $k(v_{i_0+2}, v) > \alpha n$, and let $A_4(v)$ be the event that $d_E(\pi(c_{i_0+2}), \mathbf{X}_v) \geq \frac{2}{\sqrt{3}}\sqrt{n}$. If $A_4(v)$ holds, then, by (23), the angle $c_{i_0-1}\mathbf{X}_v c_{i_0+1}$ is at most $2\pi/3$ (and clearly at least $\pi/2$). Hence the intersection of the two annuli centred on the corners $\pi(c_{i_0-1})$ and $\pi(c_{i_0+1})$ forms a near-rhombus such that the angles are between $\pi/3$ and $2\pi/3$, that is, it is squarelike. Thus we want to show that whp, for each $v \in V^-$, if $A_3(v)$ holds then $A_4(v)$ holds.

But on $A_1 \wedge A_2$, for each $v \in V^-$, if $A_4(v)$ fails then

$$
\begin{aligned}
k(v_{i_0+2}, v) &= \left|\{u \in V : d_E(v_{i_0+2}, \mathbf{X}_u) \leq d_E(v_{i_0+2}, \mathbf{X}_v)\}\right| \\
&\leq N(B(v_{i_0+2}, \tfrac{2}{\sqrt{3}}\sqrt{n})) \\
&\leq N(B(\pi(c_{i_0+2}), \tfrac{2}{\sqrt{3}}\sqrt{n} + \omega)) < \alpha n
\end{aligned}
$$

so $A_3(v)$ fails. In other words, on $A_1 \wedge A_2$, for each $v \in V^-$, if $A_3(v)$ holds then $A_4(v)$ holds; and since $A_1 \wedge A_2$ holds whp, this completes the proof. $\square$

5.4. Completing the proof of Theorem 1.3. In order to finish the proof of Theorem 1.3, we may assume wlog that, in step (a), in $O(n)$ time we have found 'corner' vertices $v_1, \ldots, v_4$ such that $\mathcal{C}_1(v_1, \ldots, v_4)$ holds; and we may assume wlog that the random permutation π is the identity map (as in the proof of Theorem 1.2). For each vertex $v \in V^-$, we form the rank list $R(v) = (k(v_i, v) : i = 1, \ldots, 4)$. We may do this in linear time, by reading through the four orders τ_{v_i}.

Recall that whp, for each vertex $v \in V^-$, $k(v_i, v) > 0.19\, n$ for each $i \neq i_0$, and $k(v_i, v) < 0.91\, n$ for $i = i_0 \pm 1$. Assume wlog that these inequalities

hold, and consider a vertex $v \in V^-$. We distinguish the two cases, whether $k(v_{i_0+2}, v) \leq \alpha n$ or not.

Case 1: $k(v_{i_0+2}, v) \leq \alpha n$.

In this case, the ideas of Theorem 1.2 can be applied. Let $I^- = [4] \setminus \{i_0\}$. By Lemmas 5.2 and 5.3, whp, for each vertex v and each $i \in I^-$, we can calculate the value $d_E(c_i, \mathbf{X}_v)$ up to an additive error of $1.19695\sqrt{\log n} + 1$. Now, exactly as in the proof of Theorem 1.2, we consider three circles $C_i(v)$ (with corresponding annuli $A_i(v)$) for $i \in I^-$, and pick a pair of circles meeting at an angle between $\pi/3$ and $2\pi/3$. We set $\Phi(v)$ to be a point we compute which is within distance 1 of (or arbitrarily close to) the relevant point where these circles meet, and then $d_E(\Phi(v), \mathbf{X}_v) < 1.197\sqrt{\log n}$ (for n sufficiently large).

Case 2: $k(v_{i_0+2}, v) > \alpha n$.

As in the last case, by Lemmas 5.2 and 5.3, we can calculate the values $d_E(c_{i_0-1}, \mathbf{X}_v)$ and $d_E(c_{i_0+1}, \mathbf{X}_v)$ up to an additive error of $1.19695\sqrt{\log n}+1$. In this case, by Lemma 5.4, the two circles (with corresponding annuli) centred on the corners c_{i_0-1} and c_{i_0+1} meet at an angle between $\pi/3$ and $2\pi/3$. As before, we set $\Phi(v)$ to be a point within distance 1 of the relevant point where these circles meet, and we find that $d_E(\Phi(v), \mathbf{X}_v) < 1.197\sqrt{\log n}$.

In both cases, all the calculations can be completed in linear time.

6. Concluding remarks

Recall that there is a family of n random points $\mathbf{X}_v$ for $v \in V$, independently and uniformly distributed in the square $\mathcal{S}_n = [-\sqrt{n}/2, \sqrt{n}/2]^2$ of area n. We do not see these points, but learn about them in one of the following two ways: (a) when we are given just the corresponding random geometric graph (for a suitable threshold distance r), and (b) when we have some geometric information. In case (a), we obtained an embedding Φ with displacement at most about r, but we required the threshold distance r to satisfy $r \gg n^{3/14}$, which yields rather a dense random geometric graph. In case (b), for each vertex v, we are given a list of all the vertices w ordered by increasing Euclidean distance from $\mathbf{X}_v$ of the corresponding points $\mathbf{X}_w$. In this case, we obtain an embedding Φ with dramatically less error.

Can we obtain lower displacement for these approximate reconstruction problems? Can we obtain similar low displacement for smaller values of r (yielding sparser random graphs)?

Another open issue is whether there is a different choice of non-trivial natural geometrical information that would help to extend the range of values of r

we can handle. Notice that exposing the real length of all the edges would trivialise the problem, as we saw in Subsection 1.5. Another natural line of research is to consider a region in the plane different from the square $\mathcal{S}_n$, for instance a disk of area n, still with n iid uniformly distributed random points $\mathbf{X}_v$ yielding a random geometric graph G. Here we cannot of course start from the corners, but we do have a boundary and we can identify vertices v of G with $\mathbf{X}_v$ near the boundary by looking at vertex degrees. Similarly, it would be interesting to investigate such problems with distributions different from the uniform distribution.

Also it would be interesting to generalise the problem to higher dimensions, to $\mathbb{R}^d$ for $d \geq 3$. We believe that for bounded dimension d, or indeed for sufficiently slowly growing dimension, similar results to those obtained in this paper could be obtained for n iid points uniformly distributed in the d-cube $\left[-n^{1/d}/2, n^{1/d}/2\right]^d$ of volume n.

Finally, let us mention the model where the underlying space is the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$ (with n iid uniformly distributed random points $\mathbf{X}_v$). See [6] for recent work on this model in high dimensions, where the main interest is to test whether we are looking at a random geometric graph from this model or at a corresponding Erdős-Rényi random graph. See also the references in [6] for other work on this model. For the approximate realization or reconstruction problem, there is now not even a boundary to start from!

Acknowledgement. We thank the referees for careful reading and helpful suggestions.

References

[1] E. Arias-Castro, A. Channarond, B. Pelletier, and N. Verzelen. On the estimation of latent distances using graph distances, Preprint available on arxiv: https://arxiv.org/pdf/1804.10611.pdf.

[2] J. Aspnes, D.K. Goldenberg and Y.R. Yang. On the computational complexity of sensor network localization. In: *ALGOSENSORS-04.*, 32–44. Lecture Notes in Computer Science 1312, Springer, 2004

[3] M. Bernstein, V. De Silva, V., J.C. Langford and J.B. Tenenbaum, J. B. Graph approximations to geodesics on embedded manifolds. Technical report, Department of Psychology, Stanford University, December 2000.

[4] H. Breu and D.G. Kirkpatrick. Unit disk graph recognition is NP-hard. *Compt. Geometry Theory Appl.*, 9(1-2):3–24, 1998.

[5] J. Bruck, J. Gao and A. Jiang. Localization and routing in sensor networks by local angle information. In: *ACM MOBIHOC-05*, 181–192. ACM Press. 2005.

[6] S. Bubeck, J. Ding, R. Eldan and M. Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures and Algorithms* 49(3): 503 – 532, 2016.

[7] N. Bulusu, J. Heidemann, D. Estrin and T. Tran. Self-configuring localization systems: Design and experimental evaluation. In: *ACM Trans. in Embedding Cmpt. Syst.*, 3:24–60. 2004.

[8] M. Cucuringu. Asap: An eigenvector synchronization algorithm for the graph realization problem. In: Distance Geometry, pages 177–195. Springer, 2013.

[9] M. Cucuringu, Y. Lipman, and A. Singer. Sensor network localization by eigenvector synchronization over the euclidean group. ACM Transactions on Sensor Networks, 8(3):19, 2012.

[10] J. Díaz, D. Mitsche, G. Perarnau and X. Pérez-Giménez. On the relation between graph distance and Euclidean distance in random geometric graphs. *Advances in Applied Probability*, 48(3): 848 – 864, 2016.

[11] E.N. Gilbert. Random plane networks. *J. Soc. Industrial Applied Mathematics*, 9(5), 533–543, 1961.

[12] W.K. Hale. Frequency assignment: theory and applications. *Proceedings of the IEEE*, 68(12): 1497–1514, 1980.

[13] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*. Wiley, New York, 2000.

[14] J.B. Kruskal, J.B. Seery, Designing network diagrams. In: Proceedings of the First General Conference on Social Graphics (pp. 22-50), 1980.

[15] F. Kuhn, T. Moscibroda, and R. Wattenhofer. Unit disk graph approximation. In: *ACM DIALM-POMC-04*, 17–23.

[16] D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. Journal of the Association for Information Science and Technology, 58(7):1019–1031, 2007.

[17] T. Moscibroda, R. O'Dell, M. Wattenhofer and R. Wattenhofer. Virtual coordinates for ad-hoc and sensor networks. In: *ACM DIALM-POMC-04*, 8–16. ACM Press. 2004.

[18] S. Parthasarathy, D. Sivakoff, M. Tian, and Y. Wang. A quest to unravel the metric structure behind perturbed networks. In: *SOCG 2017*, 53:1–53:16.

[19] M. Penrose, *Random Geometric Graphs*, Oxford Studies in Probability. Oxford U.P., 2003.

[20] A. Sankarararaman and F. Baccelli. Community detection on Euclidean random graphs. In: SODA 2018, 2181–2200.

[21] P. Sarkar, D. Chakrabarti, and A. W. Moore. Theoretical justification of popular link prediction heuristics. In: International Conference on Learning Theory, 2010.

[22] Y. Shang, W. Ruml, Y. Zhang, and M. P. Fromherz. Localization from mere connectivity. In: Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pages 201–212, 2003.

[23] Y. Terada and U. von Luxburg. Local ordinal embedding. In: Proceedings of the 31st International Conference on Machine Learning, pages 847–855, 2014.

[24] S. Vempala. Random projection: A new approach to VLSI layout. In: *FOCS-98*, 389–395. 1998.

[25] U. von Luxburg and M. Alamgir. Density estimation from unweighted k-nearest neighbor graphs: a roadmap. In: Advances in Neural Information Processing Systems, pages 225–233, 2013.

[26] M. Walters. Random geometric graphs, *Surveys in Combinatorics 2011*, Ed. Robin Chapman, Cambridge U.P., 365–402, 2011.

Universitat Politècnica de Catalunya, Dept. de Computer Science, 08034 Barcelona

E-mail address: diaz@cs.upc.edu. Partially supported by TIN2017-86727-C2-1-R.

Department of Statistics, Oxford University, 24 - 29 St Giles, Oxford OX1 3LB, UK

E-mail address: cmcd@stats.ox.ac.uk

Institut Camille Jordan (UMR 5208), Univ. de Lyon, Univ. Jean Monnet, 42023 Saint-Etienne, France

E-mail address: dieter.mitsche@univ-st-etienne.fr. Partially supported by IDEXLYON of Université de Lyon (Programme Investissements d'Avenir ANR16-IDEX-0005).