European Journal of Combinatorics xxx (xxxx) xxx



Contents lists available at ScienceDirect

European Journal of Combinatorics



journal homepage: www.elsevier.com/locate/ejc

Reconstruction of random geometric graphs: Breaking the $\Omega(r)$ distortion barrier^{*} Varsha Dani^a, Josep Díaz^{b,1}, Thomas P. Hayes^c, Cristopher Moore^{d,2}

^a Department of Computer Science, Rochester Institute of Technology, Rochester, NY,

United States of America

^b Department of Comput. Science, UPC, Barcelona, Spain

^c Department of Computer Science, University at Buffalo, Buffalo, NY, United States of America

^d Santa Fe Institute, Santa Fe, NM, United States of America

ARTICLE INFO

Article history: Available online xxxx

ABSTRACT

Embedding graphs in a geographical or latent space, i.e. inferring locations for vertices in Euclidean space or on a smooth manifold or submanifold, is a common task in network analysis, statistical inference, and graph visualization. We consider the classic model of random geometric graphs where *n* points are scattered uniformly in a square of area *n*, and two points have an edge between them if and only if their Euclidean distance is less than r. The reconstruction problem then consists of inferring the vertex positions, up to the symmetries of the square, given only the adjacency matrix of the resulting graph. We give an algorithm that, if $r = n^{\alpha}$ for any $0 < \alpha < 1/2$, with high probability reconstructs the vertex positions with a maximum error of $O(n^{\beta})$ where $\beta = 1/2 - (4/3)\alpha$, until $\alpha \ge 3/8$ where $\beta =$ 0 and the error becomes $O(\sqrt{\log n})$. This improves over earlier results, which were unable to reconstruct with error less than r. Our method estimates Euclidean distances using a hybrid of graph distances and short-range estimates based on the number of common neighbors. We extend our results to the surface of the sphere in \mathbb{R}^3 and to hypercubes in any constant fixed dimension. © 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license

(http://creativecommons.org/licenses/by/4.0/).

An extended abstract of this paper was presented in ICALP-2022. *E-mail addresses:* varsha.dani@rit.edu (V. Dani), diaz@cs.upc.edu (J. Díaz), thayes@buffalo.edu (T.P. Hayes), moore@santafe.edu (C. Moore).

https://doi.org/10.1016/j.ejc.2023.103842

0195-6698/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article as: V. Dani, J. Díaz, T.P. Hayes et al., Reconstruction of random geometric graphs: Breaking the $\Omega(r)$ distortion barrier, European Journal of Combinatorics (2023) 103842, https://doi.org/10.1016/j.ejc.2023.103842.

¹ Partially supported by PID-2020-112581GB-C21 (MOTION).

² Partially supported by National Science Foundation grant IIS-1838251.

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

1. Introduction

Graph embedding is the art of assigning a position in some smooth space to each vertex, so that the graph's structure corresponds in some way to the metric structure of that space. If vertices with edges between them are geometrically close, this embedding can help us predict new or unobserved links, devise efficient routing strategies, and cluster vertices by similarity—not to mention (if the embedding is in two dimensions) give us a picture of the graph that we can look at and perhaps interpret. In social networks, this space might correspond literally to geography, or it might be a "latent space" whose coordinates measure ideologies, affinities between individuals, or other continuous demographic variables (e.g. [1]). In some applications the underlying space is known; in others we wish to infer it, including the number of dimensions, whether it is flat or hyperbolic, and so on.

The literature on graph embedding is vast, and we apologize to the many authors who we will fail to cite. However, despite the broad utility of graph embedding in practice (see [2] for a recent experimental review) many popular heuristics lack rigorous guarantees. Here we pursue algorithms that reconstruct the position of every vertex with high accuracy, up to a symmetry of the underlying space.

Many versions of the reconstruction problem, including recognizing whether a graph has a realization as a geometric graph, are NP-complete [3-5] in the worst case. Thus we turn to distributions of random instances, and design algorithms that succeed with high probability in the instance. For many inference problems, there is a natural generative model where a ground truth structure is "planted", and the instance is then chosen from a simple distribution conditioned on its planted structure. For community detection a.k.a. the planted partition problem, for instance, we can consider graphs produced by the stochastic block model, a generative model where each vertex has a ground-truth label, and each edge (u, v) exists with a probability that depends on the labels of u and v. Reconstructing these labels from the adjacency matrix then becomes a well-defined problem in statistical inference, which may or may not be solvable depending on the parameters of the model (e.g. [6-8]). In the same spirit, a series of papers has asked to what extent we can reconstruct vertex positions from the adjacency matrix in random geometric graphs, where vertex positions are chosen independently from a simple distribution.

1.1. Random geometric graph models

Let *n* be an integer and let r > 0 be real. Let $V = \{v_i\}_{i=1}^n$ be a set of points chosen uniformly at random in the square $[0, \sqrt{n}]^2$. Then the *random geometric graph* $G \in \mathcal{G}(n, r)$ has vertex set *V* and edge set $E = \{(u, v) : ||u - v|| < r\}$ where ||u - v|| denotes the Euclidean distance. (We will often abuse notation by identifying a vertex with its position.)

This is simply a rescaling of the *unit disk model* where $(u, v) \in E$ if the unit disks centered at u and v intersect, [9]. However, we follow previous authors in changing the density of the graph by varying r rather than varying the density of points in the plane. Since the square has area n, the density is always 1: that is, the expected number of points in any measurable subset is equal to its area.

It is also natural to consider a Poisson model, where the points are generated by a Poisson point process with intensity 1. In that case the number of vertices fluctuates but is concentrated around *n*, and the local properties of the two models are asymptotically the same. We will occasionally refer to the Poisson model below.

Note that the number of points in a region of area *A* is binomially distributed in the uniform model, and Poisson distributed with mean *A* in the Poisson model. In both cases, the probability that such a region of area is empty is at most e^{-A} ; this is exact in the Poisson model, and is an upper bound on the probability $(1 - A/n)^n$ in the uniform model.

Random geometric graphs (RGGs) were first introduced by Gilbert in the early 1960s to model communications between radio stations [10]. Since then, RGGs have been widely used as models for wireless communication, in particular for wireless sensor networks. Moreover, RGG have also been extensively studied as mathematical objects, and much is known about their asymptotic properties,

V. Dani, J. Díaz, T.P. Hayes et al.

see for example [11,12]. One well-known result is that $r_c = \sqrt{\log n/\pi}$ is a sharp threshold for connectivity for $G \in \mathcal{G}(n, r)$ in the square in both the uniform and Poisson models: that is, for any $\varepsilon > 0$, with high probability (i.e., with probability tending to 1 as $n \to \infty$) *G* is connected if $r > (1 + \varepsilon)r_c$ and disconnected if $r < (1 - \varepsilon)r_c$.

More generally, we can define RGGs on any compact Riemannian submanifold, by scattering n points uniformly according to the surface area or volume. We then define the edges as $E = \{(u, v) : \|u - v\|_g \le r\}$ where $\|\cdot\|_g$ is the geodesic distance, i.e. the arc length of the shortest geodesic between u and v. On the sphere in particular this includes the cosine distance, since $\|u - v\|_g$ is a monotonic function of the angle between u and v.

1.2. The reconstruction problem

Given the adjacency matrix *A* of a random geometric graph defined on a smooth submanifold *M*, we want to find an embedding $\phi : V \to M$ which is as close as possible to the true positions of the vertices. We focus on the max distance $\max_v \|\phi(v) - v\|$ where we identify each vertex *v* with its true position. (In the sequel we sometimes say "distortion" or "error" for subsets of vertices or single vertices.)

However, if we are only given *A*, the most we can ask is for ϕ to be accurate up to *M*'s symmetries. In the square, for instance, applying a rotation or reflection to the true positions results in exactly the same adjacency matrix. Thus we define the *distortion* $d^*(\phi)$ as the minimum of the maximum error achieved by composing ϕ with some element of the symmetry group Sym(*M*),

$$d^{*}(\phi) = \min_{\sigma \in \operatorname{Sym}(M)} \max_{v \in V} \left\| (\sigma \circ \phi)(v) - v \right\|.$$
(1)

As in previous work, our strategy is to estimate the distances between pairs of vertices, and then use geometry to find points with those pairwise distances. We focus on the case where $M = \begin{bmatrix} 0, \sqrt{n} \end{bmatrix}^2$ and $\| \cdot \|$ is the Euclidean distance. However, many of our results apply more generally, both in higher dimensions and on curved manifolds.

For those who enjoy group theory, if *M* is the square then Sym(M) is the dihedral group D_8 . For the sphere, Sym(M) is the continuous group O(3) of all rotations and reflections, i.e., all 3×3 orthogonal matrices.

1.3. Our contribution and previous work

An intuitive way to estimate the Euclidean distance ||u - v|| in a random geometric graph is to relate it to the graph distance $d_G(u, v)$, i.e., the number of edges in a shortest path in the graph from u to v. The upper bound $||u - v|| \le rd_G(u, v)$ is obvious. Moreover, if the graph is dense enough, then shortest paths are fairly straight geometrically and most of their edges have Euclidean length almost r, and this upper bound is not too far from the truth, e.g. [13–16].

As far as we know, the best upper and lower bounds relating Euclidean distances to graph distances in RGGs are given in [17]. In [18] these bounds were used to reconstruct with distortion (1 + o(1))r when r is sufficiently large, namely if $r = n^{\alpha}$ for some $\alpha > 3/14$.

However, since the graph distance d_G is an integer, so the bound $||u - v|| \le rd_G(u, v)$ cannot distinguish Euclidean distances that are between two multiples of r. Thus, as discussed after the statement of Theorem 3.9 below, the methods of [18] cannot avoid a distortion that grows as $\Omega(r)$. Intuitively, the opposite should hold: as r grows the graph gets denser, neighborhoods get smoother, and more precise reconstructions should be possible.

We break this $\Omega(r)$ barrier by using a hybrid distance estimate. First we note that $rd_G(u, v)$ is a rather good estimate of ||u - v|| if ||u - v|| is just below a multiple of r, and we improve the bounds of [17] using a greedy routing analysis. Then, we combine rd_G with a more precise short-range estimate based on the number of neighbors that u and v have in common. In essence, we use a quantitative version of the popular heuristic that two vertices are close if they have a large Jaccard coefficient (see e.g. [19] for link prediction, and [20] for a related approach to small-world graphs).

V. Dani, J. Díaz, T.P. Hayes et al.



Fig. 1. Our results (solid) compared to those of [18] (dotted). If $r = n^{\alpha}$, our reconstruction has distortion $O(n^{\beta})$ where $\beta = 1/2 - (4/3)\alpha$, except for $\alpha > 3/8$ where the distortion is $O(\sqrt{\log n})$. The algorithm of [18] applies when $\alpha > 3/14$ and gives $\beta = \alpha$, i.e. distortion O(r). Our results apply for any constant $0 < \alpha < 1/2$ and give lower distortion when $\alpha > 3/14$.

As a result, we obtain distortion d^* that decreases with r: namely, if $r = n^{\alpha}$ for $\alpha > 0$, then $d^* = O(n^{\beta})$ where $\beta = \frac{1}{2} - \frac{4}{3}\alpha$, until for $\alpha \ge 3/8$ where $d^* = O(\sqrt{\log n})$. (Note that any $\alpha > 0$ puts us well above the connectivity threshold.) Since it uses graph distances, the running time of our algorithm is essentially the same as that of All-Pairs Shortest Paths. To our knowledge, this is the smallest distortion achieved by any known polynomial-time algorithm. We compare our results with those of [18] in Fig. 1.

We show that our results extend to higher dimensions and some curved manifolds as well. With small modifications, our algorithm works in the *m*-dimensional hypercube for any fixed *m* (the distortion depends on *m*, but the running time does not). We also sketch a proof that it works on the surface of the sphere, using spherical geometry rather than Euclidean geometry, solving an open problem posed in [18]. Our techniques are designed to be easy to apply on a variety of curved manifolds and submanifolds, although we leave the fullest generalizations to future work.

We also show that our results can be extended to "soft" random geometric graphs, *i.e.*, random geometric graphs in which a fraction of the edges are independently deleted from the adjacency list representation before it is shown to us. Perhaps surprisingly, our techniques work very well in this setting, with only minor modifications.

We use $N(u) = \{w : (u, w) \in E\}$ to denote the graph neighborhood of a vertex u, and B(u, r) to denote the geometrical ball of radius r around it. Our results, as well as many of the cited results, hold with high probability (w.h.p.) in the random instance $G \in \mathcal{G}(n, r)$, i.e. with probability tending to 1 as $n \to \infty$. When we consider randomized algorithms, the probability is over both $\mathcal{G}(n, r)$ and the randomness of the algorithm.

1.4. Other related work

In the statistics community there are a number of consistency results for maximum-likelihood methods (e.g. [21]) but it is not clear how the accuracy of these methods scales with the size or density of the graph, or how to find the maximum-likelihood estimator efficiently. There are also results on the convergence of spectral methods, using relationships between the graph Laplacian and the Laplace–Beltrami operator on the underlying manifold (e.g. [22]). This approach yields bounded distortion for random dot-product graphs in certain regimes.

We assume that parameters of the model are known, including the underlying space and its metric structure (in particular, its curvature and the number of dimensions). Thus we avoid

V. Dani, J. Díaz, T.P. Hayes et al.

questions of model selection or hypothesis testing, for which some lovely techniques have been proposed (e.g. [23-25]). We also assume that the parameter *r* is known, since this is easy to estimate from the typical degree.

1.5. Organization of the paper

In Section 2 we present some concentration results and define the concept of a *deep* vertex. Intuitively, a vertex is deep if it is more than r from the boundary of the square, so that its ball of potential neighbors is entirely in the interior. However, since we are only given the adjacency matrix of the graph, we base our definition on the number of vertices two steps away from v, and show that these graph neighborhood and geometric properties are closely related.

Section 3 shows that we can closely approximate Euclidean distances ||u-v|| given the adjacency matrix whenever v is deep. We do this in two steps: we give a precise short-range estimate of ||u-v|| when u, v have graph distance 2, and also a long-range estimate that uses the existence of a greedy path. By "hybridizing" these two distance estimates, switching from long to short range at a carefully chosen intermediate point, we obtain a significantly better estimate of ||u-v|| than was given in [17]. We believe these distance estimation techniques may be of interest in themselves.

In Section 4, we use this new estimate of Euclidean distances to reconstruct the vertex positions up to a symmetry of the square, by starting with a few deep "landmarks" and then triangulating to the other vertices. This gives smaller distortion than the algorithm in [18], achieving the scaling shown in Fig. 1.

In Section 5, we extend our method to random geometric graphs in the m-dimensional hypercube for fixed m and the 3-dimensional sphere.

In Section 6, we discuss extensions to "soft" random geometric graphs, in which a fraction of the edges have been randomly deleted from the graph before its adjacency matrix is shown to us.

Finally, in Section 7 we give some conclusions and pose a set of open research problems.

2. Preliminaries

2.1. Concentration of measure

We will use the following version of Chernoff's bound, which applies to both Poisson and binomial random variables, and in which the statement of the result has been "inverted" to describe a confidence interval for the outcome of the experiment, for a given confidence parameter δ .

Lemma 2.1. Let X be either a Poisson or binomial random variable with mean μ . Then, for every $\delta > 0$,

$$\mathbb{P}\left[|X-\mu| > \max\left\{3\log\frac{1}{\delta}, \sqrt{3\mu\log\frac{1}{\delta}}\right\}\right] \leq \delta.$$

Proof. We use standard tail bound arguments. The moment generating function of a binomial random variable is bounded above by that of a Poisson variable with the same mean,

$$\mathbb{E}_{X\sim \operatorname{Bin}(n,p=\mu/n)}\left[e^{\lambda X}\right] = \left(1 + \frac{\mu}{n}(e^{\lambda} - 1)\right)^n \leq e^{\mu(e^{\lambda} - 1)} = \mathbb{E}_{X\sim \operatorname{Poi}(\mu)}\left[e^{\lambda X}\right].$$

We then apply Markov's inequality. For the upper tail bound,

$$\mathbb{P}[X > \mu + t] = \mathbb{P}\left[e^{\lambda X} > e^{\lambda(\mu+t)}\right] \le \mathbb{E}[e^{\lambda X}]/e^{\lambda(\mu+t)} \le e^{\mu(e^{\lambda}-1)-\lambda(\mu+t)}.$$

The right-hand side is minimized by setting $\lambda = \log(1 + t/\mu)$, which gives

$$\mathbb{P}[X > \mu + t] \le e^{\mu f(t/\mu)}$$
 where $f(x) = x - (1+x)\log(1+x)$.

By inspection we have

 $f(x) \leq \max(-x/3, -x^2/3),$

and setting $x = (3/\mu)\log(1/\delta)$ completes the proof of the upper tail bound; the two cases correspond to $x \ge 1$ and $0 \le x \le 1$. The proof of the lower tail bound is similar. \Box

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

We will apply this lemma to derive a confidence interval for the number of vertices within a given region of the square $[0, \sqrt{n}]^2 = [0, \sqrt{n}] \times [0, \sqrt{n}]$.

Corollary 2.2. Let $R \subseteq [0, \sqrt{n}]^2$ be a region of area A. Then, for any $C \ge 1$, with probability at least $1 - 1/n^C$, the number X of vertices in R satisfies

$$|X-A| \leq \max\left\{4C\log n, 2\sqrt{CA\log n}\right\}.$$

Furthermore, for sufficiently large n, this bound still holds even if we have conditioned on the positions of a constant number of vertices, and R is allowed to depend on these.

2.2. Conditioning on positions

In several places we find it convenient to condition on the event that a vertex appears at a specific position. Readers accustomed to discrete probability may find the idea of conditioning on a zero-probability event jarring. However, in a continuous measure this makes perfect sense, since expectations can be written as integrals over a given vertex's position. Formally, the conditional probability density with which an event *E* occurs given the value of a continuous variable *X* is the derivative of the probability conditioned on *X* lying in a suitable interval:

$$\mathbb{P}[E \mid X = x] = \frac{\partial}{\partial x} \mathbb{P}[E \mid X \in [0, x)].$$

In our setting, we can define conditional versions of the model where there is a vertex at a specific position v. In the uniform case, this model consists of scattering n-1 additional vertices uniformly in the square (or whatever manifold we are working in). The probability of any event in the uniform model is then the uniform average over v of its probability in these conditional models. In the Poisson case, the conditional model consists of v and additional vertices generated by a Poisson point process with intensity 1 as before.

2.3. Deep vertices

In what follows we assume that we are working with random geometric graphs in the square $[0, \sqrt{n}]^2$. Because some of our arguments will break down for vertices near the boundary of this square, it will be useful to have an easy way to tell these vertices apart from the rest. To this end, we introduce the following notion of "deep" vertex.

Definition 2.3. Let *r* be fixed. We say that a vertex $v \in V$ is *deep* if at least $11r^2$ vertices have graphical distance 2 or less from *v*.

Note that being deep is a property of the graph, rather than its embedding in the plane. This will be important for our algorithms, which are only given access to the adjacency matrix. On the other hand, the following observation shows that being deep with high probability implies being fairly far from the boundary,

Lemma 2.4. For every $n \ge 1$, r > 0, with probability at least $1 - n \exp(-r^2/50)$, every deep vertex of *G* is located in the square $[r, \sqrt{n} - r]^2$, that is, the entire ball of radius *r* centered at the vertex is contained in $[0, \sqrt{n}]^2$.

Corollary 2.5. For every C > 0, there exists C' > 0 such that, whenever $r > C'\sqrt{\log n}$, with probability at least $1 - n^{-C}$, all deep vertices of *G* are located in the square $[r, \sqrt{n} - r]^2$.

Proof. Suppose *v* is a vertex located within distance *r* of the boundary of $[0, \sqrt{n}]^2$. Let B = B(v, 2r) be the ball of radius 2*r* centered at *v*. Then $B \setminus [0, \sqrt{n}]^2$ contains a circular segment of height at least *r*, and therefore central angle at least $2\pi/3$, and area at least $\left(\frac{4}{3}\pi - \sqrt{3}\right)r^2$. It follows that $B \cap [0, \sqrt{n}]^2$ has area at most $\left(\frac{8}{3}\pi + \sqrt{3}\right)r^2 \approx 10.1r^2$.

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx



Fig. 2. The "flower" of the proof of Lemma 2.6. The hexagon is centered at v, its side length is 0.99r, and its corners are x_1, \ldots, x_6 . With high probability, each of the balls of radius 0.01r (not shown to scale) centered on the x_j contains at least one point w_j ; these are neighbors of v since 0.99r + 0.01r = r. Moreover, each w_j is a neighbor of every point in the ball of radius 0.99r centered at x_j . Therefore, every point in the gray region has graph distance at most 2 from v.

Since every vertex within distance 2 of v is located within $B \cap [0, \sqrt{n}]^2$, in expectation the number of such vertices is at most $10.1r^2$. Since each vertex other than v is independently in B with probability $\frac{\text{vol}(B \cap [0, \sqrt{n}]^2)}{\text{vol}([0, \sqrt{n}]^2)}$, we can apply Chernoff's bound to conclude that, with probability at least $1 - \exp(-r^2/50)$, fewer than $11r^2$ vertices are in $B \cap [0, \sqrt{n}]^2$, and therefore v is not deep. A union bound over the n possibilities for vertex v concludes the proof. \Box

On the other hand, we will also need a large supply of deep vertices. Fortunately, this follows from another observation, which we now state and prove.

Lemma 2.6. There exist constants $C_1, C_2 > 0$ such that, if $r \ge C_1 \sqrt{\log n}$, with probability at least $1 - C_2/n^2$ every vertex located in $[2r, \sqrt{n} - 2r]^2$ is deep.

Proof. Let *v* be located in $[2r, \sqrt{n} - 2r]^2$. Let x_1, \ldots, x_6 be the corners of a regular hexagon whose center is at *v*, and whose side length is 0.99*r*.

We claim that, with high probability, there exist vertices $w_1, \ldots, w_6 \in V$ such that $||w_j - x_j|| \le 0.01r$ for each *j*. To see this, note that the ball of radius 0.01*r* centered at x_j has area $A = 0.0001\pi r^2$. Since $r \ge C_1 \sqrt{\log n}$, the probability that a given one of these balls is empty is at most $e^{-A} \le 1/6n^3$ if C_1 is sufficiently large. We then apply a union bound over these six events.

By the triangle inequality, $||w_j - v|| \le 0.99r + 0.01r = r$, so these w_j are neighbors of v. To see that the w_j in turn have many neighbors, which have graph distance 2 from v, we apply Corollary 2.2 to the "flower" formed by the union of the six balls of radius 0.99r centered at each of the x_j 's, shown in gray in Fig. 2. Any point in one of these balls is a neighbor of the corresponding w_j , again by the triangle inequality. Thus this region contains only vertices at distance 2 or less from v.

An easy exercise in geometry shows that the area of this flower equals $(0.99)^2(2\pi + 3\sqrt{3})r^2 \approx 11.25r^2$. Corollary 2.2 then implies that, with high probability, $\left|\bigcup_{j=1}^6 N(w_j)\right| \ge 11r^2$ and hence v is deep. Finally, we take the union bound over all vertices in $[2r, \sqrt{n} - 2r]^2$, of which there are at most n. \Box

3. Distance estimation: Breaking the $\Omega(r)$ barrier

3.1. Estimating short-range distances: Lunes and lenses

In this section given two vertices v, u with $d_G(v, u) \le 2$, we show how to estimate their Euclidean distance $x = ||v - u|| \le 2r$. In Fig. 3, we have two balls centered at v and u and with radius r,

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx



Fig. 3. We can estimate the Euclidean distance ||u - v|| = x of two vertices with $d_G(u, v) \le 2$ using the area of the lune $L = B(u, r) \setminus B(v, r) \ne \emptyset$. Denoting this area F(x), we can estimate x by applying the inverse F^{-1} to the number of points in $N(v) \setminus N(u)$. For the case x > r we consider the lens $B(u, r) \cap B(v, r)$, which has area $\pi r^2 - F(x)$.

B(v, r) and B(u, r). In the case x < r we will consider the *lune* defined by each half of the symmetric difference between B(u, r) and B(v, r) (that is the case reflected in the figure). In the case $r \le x < 2r$, we will consider the *lens* defined by the intersection of the balls.

Let $0 \le x \le 2r$. We define the function F(x) = F(x, r) to be one half of the area of the symmetric difference of two balls of radius r whose centers are at distance x, that is, the area of the lune $L = B((x, 0), r) \setminus B((0, 0), r)$, here for concreteness u and v are located at (0, 0) and (x, 0)) respectively. A little calculus shows that F is given by the following formula:

$$F(x) = \pi r^2 - 2r^2 \arccos \frac{x}{2r} + \frac{x}{2}\sqrt{4r^2 - x^2}.$$
(2)

There are several ways to derive this. One is to integrate the following differential equation:

$$F'(x) = \sqrt{4r^2 - x^2},$$
(3)

with boundary condition F(0) = 0. This differential equation follows because increasing x by an infinitesimal dx increases L with an arc-shaped area of width dx, stretching along the right side of the red circle between the two intersections between the red and blue circles in Fig. 3. This arc has height $2\sqrt{r^2 - (x/2)^2} = \sqrt{4r^2 - x^2}$ and width dx, so the area increases by $\sqrt{4 - x^2} dx$. Since F'(x) is decreasing on the interval [0, 2r), (3) also implies that, if $x \in [0, r]$, we have

$$F'(x) \ge F'(r) = \sqrt{3}r \,. \tag{4}$$

We will need the following concavity-like fact about *F*.

Lemma 3.1. *Let* $0 \le x_1 \le x_2 \le 2r$ *. Then*

$$F(x_2) \ge F(x_1) + (x_2 - x_1) \frac{F'(x_1) + F'(x_2)}{2}$$

Proof. Since $F'(x) = \sqrt{4r^2 - x^2}$, differentiating twice more, we find that

$$F'''(x) = \frac{-4r^2}{(4r^2 - x^2)^{3/2}} \le 0,$$

so F' is concave.

Sample *X* uniformly from the interval $[x_1, x_2]$, then sample *Y* uniformly from the two endpoints, $\{x_1, x_2\}$, such that its conditional expectation is *X*. Note that this implies that *Y* is uniform over

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

$$\{x_1, x_2\}$$
. Then we have:

 $\frac{F(x_2)-F(x_1)}{x_2-x_1}$

 $\begin{array}{ll} \frac{\langle x_1 \rangle}{1} &= \mathbb{E}(F'(X)) & \text{by the Fundamental Theorem of Calculus} \\ &= \mathbb{E}(F'(\mathbb{E}\ (Y \mid X))) & \text{by definition of Y} \\ &\geq \mathbb{E}(\mathbb{E}\ (F'(Y) \mid X)) & \text{by Jensen's Inequality} \\ &= \mathbb{E}(F'(Y)) & \text{by the Law of Total Probability} \\ &= \frac{F'(x_1) + F'(x_2)}{2} & \text{since Y is uniform over } \{x_1, x_2\}. \end{array}$

Rearranging, this completes the proof. \Box

Now, for a random geometric graph, *G*, and two adjacent vertices v, w, the neighbors of v that are non-neighbors of w are exactly the vertices in the lune $B(v, r) \setminus B(w, r)$, which has area F(||v - w||). Assuming v is deep, this lune is entirely within the domain $[0, \sqrt{n}]^2$, so, intuitively, $|N(v) \setminus N(w)|$ should be close to F(||v - w||). With this in mind, we define

$$d(v, w) = F^{-1}(\min\{|N(v) \setminus N(w)|, F(r)\}),$$
(5)

with the hope that $\tilde{d}(v, w)$ will be a good approximation to the true Euclidean distance, ||v - w||, most of the time. Note that $\tilde{d}(v, w)$ is a function of the adjacency matrix and r.

The following Lemma makes the above intuition precise.

Lemma 3.2. Let *G* be a random geometric graph with parameters *r*, *n*. With probability at least $1-1/n^2$ we have, for all $(v, w) \in E(G)$ such that v is deep,

$$\left| \|v - w\| - \tilde{d}(v, w) \right| \le 100 \max\left\{ \frac{\log n}{r}, \sqrt{\frac{\|v - w\| \log n}{r}} \right\}.$$
(6)

Proof. Fix a pair of vertices v, w, and condition on their positions. Let d = ||v - w||. Assume that $v \in [r, \sqrt{n} - r]^2$, so that the ball of radius r centered at v is entirely within the domain, $[0, \sqrt{n}]^2$. By Lemma 2.4, with high probability, this assumption will not cause us to miss any deep vertices. Also assume that $d \le r$, so that v and w are neighbors, since otherwise there is nothing to prove.

Observe that $N(v) \setminus N(w)$ consists of w, together with exactly those vertices located within the lune $L = B(v, r) \setminus B(w, r)$. Since L has area F(d) and is contained within the domain, $[0, \sqrt{n}]^2$, it follows by Corollary 2.2 that with probability at least $1 - \frac{1}{\sqrt{4}}$, the number of vertices, X, in L satisfies

$$|X - F(d)| \le \max\left\{16\log n, 4\sqrt{F(d)\log n}\right\},\$$

and hence

$$||N(v) \setminus N(w)| - F(d)| \le 1 + \max\left\{16 \log n, 4\sqrt{F(d) \log n}\right\}.$$

Now, since we saw in (4) that *F* satisfies $F'(x) \ge r\sqrt{3}$ for $0 \le x \le r$, it follows by the Inverse Function Theorem that $0 \le (F^{-1})'(y) \le \frac{1}{\sqrt{3}r}$ for $0 \le y \le F(r)$, and hence that

$$|\tilde{d}(v,w)-d| \leq \frac{1+16\left(\log n + \sqrt{F(d)\log n}\right)}{\sqrt{3}r}$$

Since $F(d) \leq 2dr$, this implies

$$|\tilde{d}(v,w)-d| \leq \frac{1+\max\left\{16\log n, 4\sqrt{2dr\log n}\right\}}{\sqrt{3}r},$$

which is (6), only with slightly better constants. Taking a union bound over all possible choices of v, w completes the proof. \Box

Remark 3.3. Notice Eq. (6) is only useful when *r* is at least a constant factor greater than the connectivity threshold. In particular, for $r \le 10\sqrt{\log n}$, the conclusion of Lemma 3.2 becomes trivial, because the right-hand side of (6) exceeds *r*.

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

So far we have only defined $\tilde{d}(v, w)$ for adjacent pairs of vertices where one is deep. We now extend this definition to pairs of vertices at graphical distance 2. Once again, we will require that at least one of them is deep.

Let $v, w \in V$ be such that v is deep, and $d_G(v, w) = 2$. Let d = ||v - w||. Note that $r < d \le 2r$. This time, rather than using the number of vertices in $B(v, r) \setminus B(w, r)$ to estimate the distance, we will instead use the intersection of the balls, which has area $\pi r^2 - F(d)$.

 $N(v)\cap N(w)$ is the set of vertices of $V \setminus \{v, w\}$ that are located within the lens-shaped intersection $B(v, r)\cap B(w, r)$. Since v is deep, $B(v, r)\cap B(w, r) \subset [0, \sqrt{n}]^2$. Therefore, conditioned on the positions of v and w,

$$\mathbb{E}\left[|N(v) \cap N(w)|\right] = \frac{(n-2)\left(\pi r^2 - F(d)\right)}{n}.$$

With this in mind, we will write down the distance estimate as

$$\tilde{d}(v,w) = F^{-1}\left(\max\left\{F(r), \pi r^2 - |N(v) \cap N(w)|\right\}\right)$$
(7)

and we will now prove that in this case also, $\tilde{d}(v, w) \approx ||v - w||$ with high probability.

Lemma 3.4. Let G = (V, E) be a random geometric graph with parameters r, n. Then, with probability at least $1 - 1/n^2$ we have, for all $v, w \in V$ such that v is deep and $d_G(v, w) = 2$,

$$\left| \|v - w\| - \tilde{d}(v, w) \right| \le 100 \max\left\{ \frac{(\log n)^{2/3}}{r^{1/3}}, \left(\frac{2r - \|v - w\|}{r} \right)^{1/4} \sqrt{\log n} \right\}.$$
(8)

Remark 3.5. Similarly to Lemma 3.2, we note that (8) is only useful when *r* is at least a constant factor greater than the connectivity threshold. In particular, for $r \le 10\sqrt{10 \log n}$, the conclusion of Lemma 3.4 becomes trivial, because the right-hand side of (8) exceeds *r*.

Proof. In light of the Remark, assume $r \ge 10\sqrt{10 \log n}$, since otherwise there is nothing to prove. Fix a pair of vertices v, w, and condition on their positions. Let d = ||v - w||. Assume that $B(v, r) \subset [0, \sqrt{n}]^2$, since otherwise by Lemma 2.4 v is not deep with probability at least $1 - n \exp(-r^2/50) \ge 1 - n^{-19}$, so we would have nothing to prove. Also assume that $r < ||v - w|| \le 2r$, since otherwise $d_G(v, w)$ cannot be 2, and we would again have nothing to prove.

Observe that, since d(v, w) > r, $N(v) \cap N(w)$ consists of exactly those vertices located within the lens $L = B(v, r) \cap B(w, r)$. Since *L* has area $\pi r^2 - F(d)$ and is contained within the domain, $[0, \sqrt{n}]^2$, it follows by Corollary 2.2 that with probability at least $1 - \frac{1}{n^4}$, the number of vertices, *X*, in *L* satisfies

$$|X - (\pi r^2 - F(d))| \le \max\left\{ 16 \log n, 4\sqrt{(\pi r^2 - F(d)) \log n} \right\},\$$

and hence

$$||N(v) \cap N(w)| - (\pi r^2 - F(d))| \le \max\left\{16\log n, 4\sqrt{(\pi r^2 - F(d))\log n}\right\}.$$

By Lemma 3.1 applied with $\{x_1, x_2\} = \{d, \tilde{d}\}$, we have

$$|F(d) - F(\tilde{d})| \ge |d - \tilde{d}| \frac{F'(d) + F'(d)}{2}$$

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

We change variables, using $\varepsilon = 2r - d$, $\tilde{\varepsilon} = 2r - \tilde{d}$, and apply the definition of F' to find

$$\begin{split} |d - \tilde{d}| &\leq \frac{2|F(d) - F(d)|}{F'(d) + F'(\tilde{d})} \\ &= \frac{2|F(d) - \max\{F(r), \pi r^2 - |N(v) \cap N(w)|\}|}{F'(d) + F'(\tilde{d})} \\ &= \frac{2|F(d) - \max\{F(r), \pi r^2 - |N(v) \cap N(w)|\}|}{\sqrt{4r\varepsilon - \varepsilon^2} + \sqrt{4r\widetilde{\varepsilon} - \widetilde{\varepsilon}^2}} \\ &\leq \frac{2\max\left\{16\log n, 4\sqrt{(\pi r^2 - F(d))\log n}\right\}}{\sqrt{4r\varepsilon - \varepsilon^2} + \sqrt{4r\widetilde{\varepsilon} - \widetilde{\varepsilon}^2}} \\ &\leq \frac{2\max\left\{16\log n, C\sqrt{\varepsilon^{3/2}r^{1/2}\log n}\right\}}{\sqrt{4r\varepsilon - \varepsilon^2} + \sqrt{4r\widetilde{\varepsilon} - \widetilde{\varepsilon}^2}} \\ &= O\left(\frac{\varepsilon}{r}\right)^{1/4}\sqrt{\log n}, \end{split}$$

assuming $\varepsilon^3 r > C(\log n)^2$. For the other case, we may as well treat ε as zero, in which case we get

$$|d - \tilde{d}| = \widetilde{\varepsilon} \le rac{16 \log n}{\sqrt{4r\widetilde{\varepsilon} - \widetilde{\varepsilon}^2}} \ pprox rac{16 \log n}{\sqrt{3r\widetilde{\varepsilon}}},$$

which implies

$$\widetilde{\varepsilon} \leq \left(\frac{16\log n}{\sqrt{3r}}\right)^{2/3},$$

as desired. \Box

The results of Lemmas 3.2 and 3.4 are summarized in the following theorem.

Theorem 3.6. Let G = (V, E) be a random geometric graph with parameters r, n, where $r \ge 100\sqrt{\log n}$. With probability at least $1 - 2/n^2$ we have, for all vertices $v \ne w$ such that $d_G(v, w) \le 2$ and v is deep,

$$\left| \|v - w\| - \tilde{d}(v, w) \right| \le 100\eta (\|v - w\|) \sqrt{\log n} ,$$
(9)

where $\eta : [0, 2r] \rightarrow [0, 1]$ is defined by

$$\eta(x) = \begin{cases} \frac{\sqrt{\log n}}{r} & \text{for } 0 \le x \le \frac{\log n}{r}, \\ \sqrt{\frac{x}{r}} & \text{for } \frac{\log n}{r} \le x \le r, \\ \left(\frac{2r-x}{r}\right)^{1/4} & \text{for } r \le x \le 2r - \frac{(\log n)^{2/3}}{r^{1/3}}, \\ \frac{(\log n)^{1/6}}{r^{1/3}} & \text{for } 2r - \frac{(\log n)^{2/3}}{r^{1/3}} \le x \le 2r. \end{cases}$$

3.2. Estimating long-range distances

Next we show a fairly tight relationship between Euclidean and graph distance for all pairs of vertices, including distant ones. This is a slightly sharper version of [17, Thm 1.1]. The main difference in the proof is that, where before, a short path between two given vertices is found by finding vertices close to a straight line between the endpoints, our proof instead analyzes a greedy algorithm generating a path that may deviate further from the straight line. We start with the following geometrical lemma.

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

Lemma 3.7. Let B_1 , B_2 be two overlapping Euclidean balls in \mathbb{R}^2 of radius r_1 and r_2 respectively, and let d be the distance between their centers. Consider the lens $L = B_1 \cap B_2$. Let δ denote the width of L, that is,

$$\delta = \min\{r_1 + r_2 - d, 2r_1, 2r_2\}$$

Then the area A of L satisfies

$$A = \Theta \left(\delta^{3/2} \min\{r_1, r_2\}^{1/2} \right)$$

Proof. There are many ways to see this, and we prove an *m*-dimensional generalization in Lemma 5.1 below. Here we give a proof using the function F(x) from the previous section.

Let $r = \min\{r_1, r_2\}$. Without loss of generality, we can assume $\delta < r/10$. Otherwise, *L* includes a constant fraction of the ball of radius *r*, and $A = \Theta(\delta^{3/2} r^{1/2}) = \Theta(r^2)$ as stated.

Now note that *A* is between 1/2 and 1 times the area of a lens of width δ formed by two balls B_1 , B_2 of radius *r* whose centers are $x = 2r - \delta$ apart. This lens is B_1 minus the lune $B_1 \setminus B_2$, so its area is

$$A_{\emptyset}(\delta) = \pi r^2 - F(2r - \delta).$$

where *F* is defined as in (2). From (3) we have the following differential equation for A_0 ,

$$A'_{\delta}(\delta) = -F'(2r-\delta) = \sqrt{4r^2 - x^2} = \sqrt{4\delta r - \delta^2} = \Theta(\delta^{1/2} r^{1/2}).$$

Integrating this from the boundary condition $A_{\emptyset}(0) = 0$ to a given δ gives $A = \Theta(\delta^{3/2} r^{1/2})$ and completes the proof. \Box

Remark 3.8. Note in the definition of δ the cases $2r_1$ and $2r_2$ refer to the situation when one ball is totally inside the other ball. In our random geometric graphs, all the neighborhood balls have the same radius. But as it will be the case in the proof of Theorem 3.9, we also argue about lunes between balls with different radii, therefore the need to define δ in full generality.

The main result of this section is the following theorem,

Theorem 3.9. There exist absolute constants C_1 , C_2 , C_3 such that, for all $n \ge 1$, all $r \ge C_1 \sqrt{\log n}$, with probability at least $1 - C_2/n$, all pairs of vertices u, v satisfy

$$\left\lceil \frac{\|u - v\|}{r} \right\rceil \le d_G(u, v) \le \left\lceil \frac{\|u - v\| + \kappa}{r} \right\rceil, \tag{10}$$

where

$$\kappa = C_3 \left(\frac{\|u - v\|}{r^{4/3}} + \frac{\log n}{r^{1/3}} \right).$$
(11)

Proof. The first inequality is easy: since every edge has Euclidean length at most r, we have $||u - v|| \le d_G(u, v)r$ by the Triangle Inequality. Since $d_G(u, v)$ is an integer, rounding it up does not change its value, so

$$\left\lceil \frac{\|u-v\|}{r} \right\rceil \leq d_G(u,v).$$

We now turn our attention to the second inequality of (10), the upper bound on $d_G(u, v)$. In an attempt to find a short path from u to v, we consider the following greedy algorithm, (see Fig. 4). Let $x_0 = u$. For $i \ge 1$, we define x_{i+1} to be the neighbor of x_i that has minimal Euclidean distance to v (note that x_{i+1} is unique with probability 1). The algorithm terminates if no neighbor of x_i is closer to v than x_i is. Hopefully this is because $x_i = v$, in which case we say there exists a greedy path from u to v. However, it may instead happen that the greedy algorithm gets stuck in a local minimum, and never reaches v.

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx



Fig. 4. The greedy routing analysis of Theorem 3.9. At each step we go from x_i to the neighbor x_{i+1} closest to v. In the analysis, we consider the intersections of x_i 's neighborhood with balls centered at v, with the radii of the latter chosen so that these intersections have area ln 2, 2 ln 2, 3 ln 2, and so on. Each of these intersections contains a point with constant probability, so that most steps make significant progress towards v.

We will prove that, with probability $1 - O(n^{-3})$, a greedy path exists from *u* to *v*, and its length is at most the desired upper bound. Taking a union bound over all pairs *u*, *v* completes the proof. Notice this greedy algorithm has nothing to do with our reconstruction algorithm: it is purely for our analysis, i.e., to prove that short paths probably exist.

We will discuss our solution in terms of a sequence of independent fair coin flips. These coin flips are the outcomes of a sequence of experiments of the form "Is $||x_{i+1} - v|| \le a$?", where x_1, \ldots, x_ℓ is our greedy walk from u to v. The values a are chosen adaptively to make each of these coin flips fair, conditioned on previously revealed steps of the greedy walk, and on previous coin flips.

Specifically, assume that x_i has just been revealed. Our first coin flip asks whether x_{i+1} lies in the lens L_1 defined as the intersection of the ball of radius r centered at x_i with the ball of radius a_1 centered at v, with a_1 chosen so that L_1 has area ln 2. Working in the Poisson model, the probability that L_1 contains at least one vertex is exactly 1/2. If the coin comes up "no", i.e., L_1 is empty, we ask the same question, while increasing the radius of the ball around v to a_2 , and then a_3 , and so on, with a_t chosen so that the resulting lens L_t has area $t \ln 2$. Each time we increment t, we gain an additional region $L_t \setminus L_{t-1}$ of area ln 2, which corresponds to another coin flip. We continue until the tth coin comes up "yes", i.e., L_t contains at least one point. At that point we know that $x_{i+1} \in L_t \setminus L_{t-1}$ and therefore $||x_{i+1} - v|| \le a_t$ as shown in Fig. 5. We then reveal x_{i+1} (note that it might be one of several points in that region).

We then repeat the whole process with x_{i+1} in place of x_i . This continues until x_i is within the disk of radius r around v, after which $x_{i+1} = v$ and we are done.

Note that the above argument breaks down if the sequence of lenses examined in stage i + 1 overlaps the lenses examined in stage i. In this case, since the planar regions being examined are not disjoint, the second sequence of coin flips is no longer conditionally independent of the first. Fortunately, this can only happen if there is a very long sequence of consecutive "no" answers in stage i, which is unlikely in a sequence of fair coin flips. Specifically, we are only worried about the event that $||x_i - v|| \ge ||x_{i-1} - v|| - r/2$, which, since the area of the lens of width r/2 with centers at x_{i-1} and v and radii r and $||x_i - v|| - r/2$, respectively, is $\Theta(r^2)$ by Lemma 3.7. Choosing the constants carefully, we can ensure that the probability of this lens being empty is $O(1/n^4)$, and hence with high probability it does not happen at any stage of the greedy algorithm.

The next step in the analysis is to relate the number of "no" answers received in the aforementioned sequence of fair coin flips to the progress made by the greedy algorithm. Note that each sequence of k - 1 "no" answers followed by one "yes", corresponds to x_{i+1} being in the lens of area $k \ln 2$, which has, by Lemma 3.7, width $\Theta(k^{2/3}r^{-1/3})$, which is $O(kr^{-1/3})$. Thus

$$||x_{i+1} - v|| \le a_k = ||x_i - v|| - r + O(kr^{-1/3})$$

It follows that, if the total number of coins flipped on the journey from u to v is m, we necessarily have $||u - v|| \ge r(d_G(u, v) - 1) - O(mr^{-1/3})$.

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx



Fig. 5. The analysis of greedy routing in Theorem 3.9. We consider the intersections of the neighborhood of the current point x_i with balls centered at v, with the radius of the latter chosen so that these intersections have area ln 2, 2 ln 2, 3 ln 2, and so on. Each additional region of area ln 2 is nonempty with probability 1/2, creating a fair coin flip. When this coin gives "yes", i.e., that region is nonempty, we reveal x_{i+1} in that region.

Setting $m = \frac{4||u-v||}{r} + 48 \log n$, Chernoff's bound tells us that the probability of getting fewer than m/4 heads in a sequence of m coin flips is at most $\exp(-m/16) = O(n^{-3})$. The second kind of bad event is getting a sequence of more than $4 \log n$ consecutive tails. Since the chance of having a particular set of k coin flips be all tails is 2^{-k} , and there are at most $m = O(\sqrt{n}/r)$ possible positions for a run to start, a union bound shows that the combined probability of having such a run is $O(n^{-4}\sqrt{n}/r) = O(n^{-3})$.

Putting all of the above together, except for an $O(n^{-3})$ probability of failure, we have

$$\|u - v\| \ge r(d_G(u, v) - 1) - O(mr^{-1/3})$$

= $r(d_G(u, v) - 1) - O\left(\frac{\|u - v\|}{r^{4/3}} + r^{-1/3}\log n\right)$

or equivalently, for some constant C_3

$$\begin{aligned} d_G(u, v) &\leq 1 + \frac{\|u - v\|(1 + O(r^{-4/3})) + O(r^{-1/3}\log n)}{r} \\ &\leq 1 + \frac{\|u - v\|}{r} + C_3 \left(\frac{\|u - v\|}{r^{7/3}} + \frac{\log n}{r^{4/3}}\right) \\ &= 1 + \frac{\|u - v\| + \kappa}{r} \,. \end{aligned}$$

Since $d_G(u, v)$ is an integer, and $\frac{||u-v||+\kappa}{r}$ being an integer is a measure zero event, it follows that

$$d_G(u, v) \leq \left\lceil \frac{\|u - v\| + \kappa}{r}
ight
ceil.$$

Let us discuss how we will use Theorems 3.6 and 3.9 to break the $\Omega(r)$ barrier in distance estimation, and thus in reconstruction. Suppose $r = n^{\alpha}$ where $0 < \alpha < 1/2$ is a constant. Then since $||u - v|| = O(n^{1/2})$, we have from (11)

$$\frac{\kappa}{r} = O\left(\max\left(n^{\frac{1}{2} - \frac{7}{3}\alpha}, n^{-\frac{4}{3}\alpha}\log n\right)\right),\tag{12}$$

and since $\frac{1}{2} - \frac{7}{3}\alpha > -\frac{4}{3}\alpha$ we have

$$\kappa = O(n^{\beta}), \quad \text{where} \quad \beta = \frac{1}{2} - \frac{4}{3} \alpha.$$
(13)

V. Dani, J. Díaz, T.P. Hayes et al.

If $\alpha > 3/14$, then $\beta < \alpha$ and $\kappa = o(r)$. In this case the upper and lower bounds on $d_G(u, v)$ differ by at most 1, and moreover are equal for most pairs of vertices, making $d_G(u, v)$ a nearly-deterministic function of ||u - v||. Using $\lceil x \rceil \le x + 1$ and multiplying through by r gives the bounds

$$d_G(u, v)r - (r + \kappa) \le ||u - v|| \le d_G(u, v)r$$
,

so that $d_G(u, v)r$ is an estimate of ||u - v|| with error $r + \kappa = (1 + o(1))r$. Previous work [17,18] used this bound to reconstruct the graph with a distortion of $(1 + \varepsilon)r$ for arbitrarily small constant ε . This gives the performance shown by the dotted line in Fig. 1.

But in fact $d_G(u, v)r$ is a much more accurate estimate of ||u - v|| for certain pairs of vertices. Namely, if ||u - v|| is just below a multiple of r, then rounding up the left and right sides of (10) does not change either very much. We state this with the following corollary.

Corollary 3.10. With $\kappa = \kappa(||u-v||)$ defined as in (11), suppose that for some δ and some non-negative integer t we have

$$tr - (\kappa + \delta) < \|u - v\| < tr - \kappa.$$

Then

$$d_G(u, v)r - (\kappa + \delta) \le ||u - v|| \le d_G(u, v)r$$

Proof. We have

$$\left\lceil \frac{\|u-v\|+\kappa}{r} \right\rceil \leq \frac{\|u-v\|+\kappa+\delta}{r}$$

and (10) then gives the stated result. \Box

Thus, if ||u - v|| is in one of these intervals, Theorem 3.9 lets us estimate ||u - v|| from the adjacency matrix with error $\delta + \kappa$ instead of $r + \kappa$. Below we will combine this with the more precise estimate of short-range distances from Lemma 3.2 to achieve this error for all pairs u, v of vertices where v is deep, not just those for which ||u - v|| is almost a multiple of r. As a result, the error in our distance estimates and the distortion of our reconstruction is $O(r^{\beta})$ where β decreases from 1/2 to 0 as α increases as shown by the solid line in Fig. 1. Specifically, we obtain a nontrivial result for any $\alpha > 0$ and a more accurate reconstruction than [18] in the range $\alpha > 3/14$ where their theorem applies. At $\alpha = 3/8$, where $\beta = 0$ another source of error takes over, leaving us with $O(\sqrt{\log n})$ distortion.

3.3. Hybrid estimates of long-range distances

In order to combine the long-range estimates of Theorem 3.9 with the more precise short-range estimates of Lemma 3.2, it will be helpful to set up some general machinery.

Definition 3.11. Suppose $V \subset \mathbb{R}^2$. Let $d : V^2 \to [0, \infty]$ and $\varepsilon : \mathbb{R} \to [0, \infty]$ be two functions satisfying, for all $u, v \in V$,

 $d(u, v) - \varepsilon(u, v) \le ||u - v|| \le d(u, v).$

Then we say *d* is an upper bound on Euclidean distance with error function ε .

These error functions will often be bounded by functions of the Euclidean distance, in which case we will write $\varepsilon(||u - v||)$ rather than $\varepsilon(u, v)$. In our application, *V* consists of all vertices in the geometric graph, but we will achieve particularly small $\varepsilon(u, v)$ when *v* is deep.

A basic tool for combining distance estimates is the following:

Lemma 3.12. If d_1 and d_2 are upper bounds on Euclidean distance with error functions $\varepsilon_1, \varepsilon_2$ respectively, then $\min\{d_1, d_2\}$ is an upper bound on Euclidean distance with error $\min\{\varepsilon_1, \varepsilon_2\}$.

V. Dani, J. Díaz, T.P. Hayes et al.



Fig. 6. For any intermediate point w, the hybrid distance estimate $d_1(u, w) + d_2(w, v)$ is an upper bound on ||u - v|| with error bounded by Lemma 3.13.



Fig. 7. Figure for the proof in Lemma 3.13.

Proof. For any u, v we have $||u - v|| \le \min\{d_1, d_2\}$. On the other side, assume without loss of generality that $\varepsilon_1 \le \varepsilon_2$. Then $||u - v|| \ge d_1 - \varepsilon_1 \ge \min\{d_1, d_2\} - \varepsilon_1$. \Box

The next lemma shows another way to combine two upper bounds on ||u - v||. We choose a vertex w between u and v and use the triangle inequality, using d_1 to bound ||u - w|| and d_2 to bound ||w - v|| (see Fig. 6). Finally, we minimize over all intermediate vertices w. This hybrid is especially useful when, as with our long-range and short-range estimates, d_1 and d_2 have different ranges of ||u - v|| in which they achieve small error.

Lemma 3.13. Suppose d_1 and d_2 are upper bounds on Euclidean distance with error functions ε_1 and ε_2 respectively. Define the hybrid distance estimate \hat{d} by

$$d = \min(d_1(u, w) + d_2(w, v)),$$

where w ranges over all vertices. Then \hat{d} is an upper bound on Euclidean distance with error

$$\widehat{\varepsilon}(u,v) \leq \min_{w} \left[\varepsilon_1(u,w) + \varepsilon_2(w,v) + \|u-w\| + \|w-v\| - \|u-v\| \right].$$

Proof. Fix a pair of vertices *u*, *v*. For any vertex *w*, by the triangle inequality we have

$$||u - v|| \le ||u - w|| + ||v - w|| \le d_1(u, w) + d_2(w, v),$$

so \hat{d} is an upper bound on Euclidean distance. On the other hand, as shown in Fig. 7 we have

$$\begin{aligned} \|u - v\| &= \|u - w\| + \|v - w\| - \left(\|u - w\| + \|w - v\| - \|u - v\|\right) \\ &\geq d_1(u, w) + d_2(v, w) - \left(\varepsilon_1(u, w) + \varepsilon_2(w, v) + \|u - w\| + \|w - v\| - \|u - v\|\right). \end{aligned}$$

Using Lemma 3.12 to minimize the error over w completes the proof. \Box

V. Dani, J. Díaz, T.P. Hayes et al.



Fig. 8. The lens L(x) of Lemma 3.14. If δ is large enough, this lens is nonempty with high probability, in which case we can use any point w in it as an intermediate point for Lemma 3.13.

Next, we will use the fact that if a lens is large enough to contain at least one point w with high probability, this yields an upper bound on the minimum in Lemma 3.13.

Lemma 3.14. Suppose G is a random geometric graph, and that with high probability, d_1 and d_2 are upper bounds on Euclidean distance with errors $\varepsilon_1(u, v) = \varepsilon_1(||u - v||)$ and $\varepsilon_2(u, v) = \varepsilon_2(||u - v||)$. Define the hybrid distance d as in Lemma 3.13. Then there is a constant C such that, with high probability, \hat{d} is also an upper bound on Euclidean distance, with error $\hat{\varepsilon}(u, v) = \hat{\varepsilon}(||u - v||)$ where

$$\widehat{\varepsilon}(\|u-v\|) \le \min_{0 < x < \|u-v\|} \max_{0 \le \delta_1, \delta_2 \le \delta(x)} \Big[\varepsilon_1(x+\delta_1) + \varepsilon_2(\|u-v\| - x - \delta_2) + \delta(x)\Big],\tag{14}$$

where

$$\delta(x) = C(\log n)^{2/3} \left(\min\{x, \|u - v\| - x\}\right)^{-1/3}.$$
(15)

Proof. Fix a pair u, v with 0 < x < ||u - v||, and consider the lens L(x) of width $\delta(x)$ consisting of the intersection of the balls centered at u and v of radius $r_1 = x + \delta(x)$ and $r_2 = ||u - v|| - x$ respectively as shown in Fig. 8. Lemma 3.7 and $x \le r_1$ implies that the area of L(x) is proportional to $\delta(x)^{3/2} r_1^{1/2} \ge C^{3/2}(\log n) x^{-1/2} r_1^{1/2} \ge C^{3/2} \log n$, which is w.h.p. nonempty if C is sufficiently large. While *x* takes an infinite number of values, for all *x* we have

$$\delta(x) \ge \delta_{\min} = \Omega\left((\log n)^{3/2} \|u - v\|^{-1/3}\right).$$

Now consider $2||u-v||/\delta_{\min}$ lenses of width $\delta_{\min}/2$, where *x* is an integer multiple of $\delta_{\min}/2$. For any x, L(x) contains one of these smaller lenses, so if they are all nonempty, so is L(x) for all x. Each of these small lenses also has area $\Omega(\log n)$, and there are $O(||u - v||/\delta_{\min}) = O(||u - v||^{4/3}) = O(n^{2/3})$ of them. Thus if we set the constant C large enough for the area of the smaller lenses to be $3 \log n$, say, the probability that any of them are empty, for any pair u, v, is o(1).

Hence L contains at least one vertex w with high probability. Applying Lemma 3.13 to w, writing $||u-w|| = x + \delta_1$ and $||w-v|| = ||u-v|| - x - \delta_2$ as shown in Fig. 8, and pessimistically maximizing $\varepsilon_1(u, w)$ and $\varepsilon_2(w, v)$ over the lens yields the desired upper bound on the error of d.

We use the previous lemma to break the $\Omega(r)$ barrier for the error in estimating Euclidean distances in $G \in \mathcal{G}(n, r)$.

Now assume v is deep, and define d_1 and d_2 as follows:

 $d_1(u, v) = rd_G(u, v),$

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

$$d_2(u, v) = \begin{cases} \tilde{d}(u, v) + C_2 \sqrt{\log n} & \text{if } d_G(u, v) \le 2, \\ +\infty & \text{otherwise}, \end{cases}$$

where *d* is defined as in Eq. (5) and C_2 is the constant in Eq. (6). Thus d_1 is the upper bound of Corollary 3.10, and d_2 is the precise short-range estimate \tilde{d} of Lemma 3.2 with a small increment to make it an upper bound on Euclidean distance with high probability.

Remark 3.15. Given this choice of d_1 and d_2 , the hybrid estimate $\widehat{d}(u, v)$ is the shortest path distance from u to v in a weighted graph G_v with vertices V and edges $E_v = E \cup \{(w, v) : d_G(w, v) \le 2\}$ where each edge (w, v) with $d_G(w, v) \le 2$ has weight $\widetilde{d}(w, v) + C_2 \sqrt{\log n}$ and all other edges have weight r. Below we invoke Dijkstra's algorithm to compute these shortest paths. Alternatively, for any fixed v, we can compute $\widehat{d}(u, v)$ for all u using a modified breadth-first search algorithm in $O(n + m + \Delta \log \Delta)$ steps, where Δ is the degree of vertex v. The only modification to the standard queue-based implementation of BFS is to sort the level-one nodes by increasing order of \widetilde{d} , before placing them in the queue. Since all other edges have the same weight, r, which exceeds the difference between any two vertices in the queue, it follows by induction that all vertices will be discovered in increasing order of \widehat{d} .

We bound the errors ε_1 and ε_2 as follows. As discussed after Theorem 3.9, for most values of ||u - v|| we have $\varepsilon_1(||u - v||) = \Theta(r)$. However, we will choose the lens in Lemma 3.14 so that ||u - w|| is almost a multiple of r, in which case Corollary 3.10 shows that $\varepsilon_1(||u - w||)$ is much smaller.

For ε_2 , Lemma 3.2 implies that, for some absolute constant C_4 , with high probability

$$\varepsilon_{2}(\|u-v\|) \leq \begin{cases} C_{4}\sqrt{\log n} & \text{if } \|u-v\| \leq 2r - C_{4}r^{-1/3}\log n, \\ +\infty & \text{otherwise}. \end{cases}$$
(16)

Here we used equation (11) and the upper bound of Theorem 3.9 to show that with high probability $d_G(u, v) \le 2$ whenever

$$||u - v|| \le 2r - \kappa(2r) = 2r - C_3 r^{-1/3}(2 + \log n)$$

and we set $C_4 > \max\{C_3, 2C_2\}$.

Having gathered these facts, we will apply Lemma 3.14 to d_1 and d_2 with a judicious choice of lens L(x). First note that, since $d_2(w, v) = +\infty$ if $d_G(w, v) > 2$, we can write the hybrid distance estimate as,

$$\hat{d}(u,v) = \min_{w: d_G(w,v) \le 2} d_1(u,w) + d_2(w,v).$$
(17)

Theorem 3.16. Let $r = n^{\alpha}$ for a constant $0 < \alpha < 1/2$. For all pairs u, v where v is deep, define $\widehat{d}(u, v)$ as in Eq. (17). Then with high probability, \widehat{d} is an upper bound on the Euclidean distance ||u - v|| with error

$$\widehat{\varepsilon}(u,v) \le C' \begin{cases} n^{\frac{1}{2} - \frac{4}{3}\alpha} & \alpha < 3/8, \\ \sqrt{\log n} & 3/8 \le \alpha < 1/2, \end{cases}$$
(18)

for some absolute constant C'. That is,

 $\widehat{d}(u, v) - \widehat{\varepsilon}(u, v) \le ||u - v|| \le \widehat{d}(u, v).$

Proof. Below we will bound the error of \hat{d} as $\hat{\varepsilon} = \kappa + O(\sqrt{\log n})$. Recall that if $0 < \alpha < 1/2$, Eq. (11) gives

$$\kappa = (1 + o(1))C_3 n^{\beta}$$
 where $\beta = \frac{1}{2} - \frac{4}{3} \alpha$.

If $\alpha < 3/8$ then $\beta > 0$ and $\hat{\varepsilon} = (1 + o(1))\kappa$. If $\alpha \ge 3/8$ then $\beta \le 0$, $\kappa = O(1)$, and $\hat{\varepsilon} = O(\sqrt{\log n})$.

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

We can upper bound equation (14) with any choice of *x* between 0 and ||u - v||, i.e. any lens between *u* and *v* that w.h.p. contains an intermediate point *w*. We choose this lens as follows. First let *t* be the largest integer such that ||u - v|| > tr + r/2. We assume without loss of generality that $t \ge 1$, since if t = 0 we have ||u - v|| < 3r/2 and Eq. (16) gives $\hat{\varepsilon} \le \varepsilon \le \varepsilon_2 = O(\sqrt{\log n})$.

Let *C* be the constant required by Lemma 3.14. If we define

$$\delta = C(\log n)^{2/3} (r/2)^{-1/3} \tag{19}$$

and set $x = tr - (\kappa + \delta)$, we have min $\{x, ||u - v|| - x\} \ge r/2$, and δ satisfies the condition (15). That is, the lens L(x) of width δ shown in Fig. 8 contains at least one point w with high probability.

Now we can bound the errors in $d_1(u, w)$ and $d_2(w, v)$. First, we have chosen L(x) so that ||u - w|| is almost an integer, i.e. $tr - (\kappa + \delta) < ||u - w|| < tr - \kappa$. Then Corollary 3.10 tells us that $d_1(u, w)$ has error bounded by $\varepsilon_1 \le \kappa + \delta$. Thus, for all $0 \le \delta_1 \le \delta$, the first term of Eq. (14) is bounded by $\varepsilon_1(x + \delta_1) \le \kappa + \delta$.

Second, we have $||v - w|| \le ||u - v|| - x \le 3r/2 + \kappa + \delta$, so w.h.p. $d_G(w, v) \le 2$ and Eq. (16) gives $\varepsilon_2(w, v) \le C_4\sqrt{\log n}$. Thus, for all $0 \le \delta_2 \le \delta$, the second term of Eq. (14) is bounded by $\varepsilon_2(||u - v|| - x - \delta_2) \le C_4\sqrt{\log n}$.

Combining all of this, Lemma 3.14 tells us that with high probability

$$\widehat{\varepsilon}(\|u-v\|) \leq \kappa + \delta + C_4 \sqrt{\log n} + \delta = \kappa + C_4 \sqrt{\log n} + o(1).$$

Setting $C' > \max\{C_3, C_4\}$ completes the proof. \Box

4. The reconstruction algorithm

In this section we use our distance estimates to reconstruct the positions of the points up to a symmetry of the square. Our global strategy is similar to [18]: we first fix a small number of "landmark" vertices v whose positions can be estimated accurately up to a symmetry of the plane. Then for each vertex u we use the estimated distances $\hat{d}(u, v)$ to reconstruct u's position by triangulation. In [18], the landmarks are vertices close to the corners of the square. Here they will instead be a set of three deep vertices that are far from collinear, forming a triangle which is acute and sufficiently large.

Lemma 4.1. Let x, y, z, u be four points in the plane. Suppose x, y, z form an acute triangle with minimum side length at least ℓ . Then, if we know the positions of x, y, z with error at most η , and we have upper bounds $\widehat{d}(u, v)$ on the Euclidean distances ||u - v|| for all $v \in \{x, y, z\}$ with error $\widehat{\varepsilon}$, and all of these distances are at most D, we can determine the position of u relative to x, y, z with error at most

$$C_5 \frac{D(\hat{\varepsilon} + \eta)}{\ell},\tag{20}$$

for an absolute constant C₅.

Proof. First, let us assume fixed positions for x, y, z within η of their estimated positions (which we can always do so that they form an acute triangle). By the triangle inequality, this changes the distances ||u - v|| for $v \in \{x, y, z\}$ by at most $\pm \eta$. Thus u is in the intersection U of three annuli,

$$U = \bigcap_{v \in \{x, y, z\}} B(v, \widehat{d}(u, v) + \eta) \setminus B(v, \widehat{d}(u, v) - \eta - \widehat{\varepsilon}).$$
(21)

Any point u' in U gives an approximation of u's position with error at most the Euclidean diameter of U, namely $\max_{u,u' \in U} ||u - u'||$. We will show this diameter is bounded by Eq. (20).

We use some basic vector algebra. Let $\varepsilon' = \widehat{\varepsilon} + 2\eta \le 2(\widehat{\varepsilon} + \eta)$. For any $u, u' \in U$ we have, for all $v \in \{x, y, z\}$,

$$-\varepsilon' \le \|u - v\| - \|u' - v\| \le \varepsilon'.$$

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

Since the triangle *x*, *y*, *z* is acute, at least one of its sides makes an angle φ with the vector u - u' where $0 \le \varphi \le \pi/4$. Taking this side to be (*x*, *y*) we have, without loss of generality,

$$(y-x) \cdot (u-u') = ||y-x|| ||u-u'|| \cos \varphi \ge \sqrt{\frac{1}{2}} ||y-x|| ||u-u'||.$$

Next, we rewrite this dot product as follows,

$$\begin{aligned} 2(y-x)\cdot(u-u') &= \|x-u\|^2 - \|x-u'\|^2 - \|y-u\|^2 + \|y-u'\|^2 \\ &= (\|x-u\| - \|x-u'\|)(\|x-u\| + \|x-u'\|) \\ &- (\|y-u\| - \|y-u'\|)(\|y-u\| + \|y-u'\|) \\ &\leq \varepsilon'(\|x-u\| + \|x-u'\| + \|y-u\| + \|y-u'\|), \end{aligned}$$

where the first line is a classical polarization identity. Putting these together, we have

$$\|u-u'\| \leq \frac{\sqrt{2}}{\|y-x\|} \varepsilon' \big(\|x-u\| + \|x-u'\| + \|y-u\| + \|y-u'\| \big) \leq \frac{4\sqrt{2}R\varepsilon'}{\ell},$$

completing the proof with $C_5 = 8\sqrt{2}$. \Box

Remark 4.2. In our application, *D* is at most the diameter $\sqrt{2n}$ of the square, $\ell = \Omega(\sqrt{n})$, and $\eta = O(\hat{\epsilon})$. Thus we can reconstruct *u*'s position relative to *x*, *y*, *z* with error $O(\hat{\epsilon})$.

Theorem 4.3. Let $r = n^{\alpha}$ for a constant $0 < \alpha < 1/2$. There is an algorithm with running time $O(n^2)$ that with high probability reconstructs the vertex positions of a random geometric graph, modulo symmetries of the square, with distortion an absolute constant times $\hat{\varepsilon}$ as defined in (18), i.e.

$$d^* = C'' \begin{cases} n^{\frac{1}{2} - \frac{4}{3}\alpha} & \text{if } \alpha < 3/8, \\ \sqrt{\log n} & \text{if } 3/8 \le \alpha < 1/2, \end{cases}$$

for some absolute constant C".

Proof. We use the fact, proved in [18], that w.h.p. the true positions of the lowest-degree vertices are within $\sqrt{\log n}$ of the corners of the square, and we can find those lowest degree vertices in $O(n^2)$. Call these vertices a, b, c, d.

We construct a good triple, i.e., deep vertices x, y, z that form an acute triangle with a minimum side length at least $\ell = 0.1\sqrt{n}$. Recall that by the bounds of Theorem 3.9, with high probability any triple of deep vertices whose graph distances are in the range $[0.1\sqrt{n}/r, 0.14\sqrt{n}/r]$ qualifies.

There are many ways to find such a triple. One is to find a vertex *x* deep inside the square, e.g., with a graph distance at least $0.65\sqrt{n}/r$ from *a*, *b*, *c*, *d*. We can then take *y* to be any vertex such that $d_G(x, y)$ is in the interval $[0.1\sqrt{n}/r, 0.14\sqrt{n}/r]$, and *z* to be any vertex such that $d_G(x, z)$ and $d_G(y, z)$ are both in this interval. At each stage of this process such a vertex exists with very high probability, and by Theorem 3.9 and Lemma 2.6 all of them are deep. Finding *x* takes O(n) time by breadth-first search from *a*, *b*, *c*, *d*, and finding *y* and *z* similarly takes O(n) time each. (A randomized algorithm can simply sampling triples uniformly at random: since a constant fraction of triples qualify, this succeeds with high probability within $O(\log n)$ tries.)

Theorem 3.16 gives us estimated side lengths $\hat{d}(x, y)$, $\hat{d}(y, z)$, $\hat{d}(x, z)$ that have error $\hat{\varepsilon}$. This lets us estimate the positions of x, y, z modulo an isometry of \mathbb{R}^2 , i.e. a translation, rotation, or reflection of the plane. Equivalently, it lets us construct a triangle x, y, z which is congruent to their true positions up to distortion $\eta = O(\hat{\varepsilon})$. Then we use Lemma 4.1 to reconstruct the position of each vertex u relative to this triangle. Given $\hat{d}(u, v)$ for all $v \in \{x, y, z\}$, in (20) we have $D \le \sqrt{2n}$ and $\ell \ge 0.1\sqrt{n}$, Lemma 4.1 gives us u's position relative to x, y, z with error $10\sqrt{2}C_5(\varepsilon + \eta) = O(\hat{\varepsilon})$.

This gives us a reconstruction up to an isometry as shown in Fig. 9. Finally, we rotate and translate this reconstruction to the square $[0, \sqrt{n}]^2$. It is easy to compute an isometry that sends $\{a, b, c, d\}$ to the corners $[0, \sqrt{n}]^2$ with error $d^* = O(\sqrt{\log n})$: for instance, translate *a* to (0, 0) and

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx



Fig. 9. Our reconstruction is built around a triangle x, y, z of deep vertices. It may be translated, rotated, or reflected in \mathbb{R}^2 by an isometry, but it can then be shifted to the square $[0, \sqrt{n}]^2$. Then it will be a good reconstruction up to a rotation or reflection of the square.

then rotate one of the closer corners to $(0, \sqrt{n})$. This gives a reconstruction which, up to a rotation or reflection of the square, has distortion $d^* = O(\hat{\varepsilon} + \sqrt{\log n}) = O(\hat{\varepsilon})$.

Step 1 can be done by breadth-first search, first from a, b, c, d and then from x and y, and thus takes O(n) time. The bulk of the running time comes from computing the graph distances. Notice we have a few real-valued and geometric calculations to do. In order to compute $\hat{d}(u, v)$ for all u and all $v \in \{x, y, z\}$, we need $\tilde{d}(w, v)$ for all $v \in \{x, y, z\}$ and all w with $d_G(w, v) \le 2$. There are $O(n^2)$ such w. Inverting the function F in Eq. (5) is a matter of arithmetic; we need to do this to $O(\log r) \le O(\log n)$ bits of precision, which can be done in polylog(n) time. Thus computing these \tilde{d} can be done in time $O(r^2 \operatorname{polylog}(n))$, which is o(n) since $\alpha < 1/2$.

Once we have the distance estimates $\widehat{d}(u, v)$, finding estimated positions u in each region U defined in Eq. (21) is a geometric calculation which can be done to $O(\log n)$ bits of precision in polylog(n) time as in [18]. The same is true of computing the angle by which we need to rotate the reconstruction to $[0, \sqrt{n}]^2$ after translating one corner to the origin.

Since the typical degree of a vertex is $\pi r^2 = \Theta(n^{2\alpha})$, the number of edges in *G* is w.h.p. $m = \Theta(n^{2\alpha+1})$, which gives $m = \omega(n)$ and $m = o(n^2)$ since $0 < \alpha < 1/2$. Dijkstra's algorithm in a graph with *n* vertices and *m* edges runs in time $O(m + n \log n)$, so the running time is w.h.p. O(m). Thus the total running time is dominated by that of Dijkstra's algorithm, which for simplicity we bound as $O(n^2)$. \Box

Remark 4.4. After we have reconstructed the positions of all vertices, i.e. their coordinates, we can extend our distance estimates to all the pairs u, v of vertices that are not deep, simply by estimating ||u - v|| as the distance between the reconstructed positions of u and v.

5. Extensions to other domains

Our results can be generalized from the square to a number of alternative domains for random geometric graphs, including to higher-dimensional Euclidean space and to some curved manifolds. Here we show that our results extend to the *m*-dimensional hypercube and the 3-dimensional sphere, solving an open problem posed in [18].



Fig. 10. The figure illustrates how we can bound the area of a lens using inscribed and circumscribed cylinders.

5.1. Reconstruction in higher-dimensional Euclidean space

The simplest generalization is where the underlying domain is $[0, n^{1/m}]^m \subset \mathbb{R}^m$, i.e., an *m*-dimensional hypercube with volume *n*. Where *m* is fixed, i.e., it does not vary with *n*. Let the radius of *G* be *r*. Then, *n* points are selected at random from the hypercube and pairs of them are adjacent if they are within Euclidean distance *r*. As before, the goal is to reconstruct the points' positions based on the adjacency matrix of the graph.

The following lemma generalizes Lemma 3.7, and relates the width of a lens defined as the intersection of two balls of radius r to its m-dimensional volume.

Lemma 5.1. Let B_1 , B_2 be two overlapping Euclidean balls of radius $r_1 \le r_2$, respectively, in \mathbb{R}^m . Consider the lens $L = B_1 \cap B_2$. Let ε denote the width of L, that is,

 $\varepsilon = \min\{r_1 + r_2 - d, 2r_1\},\$

where d is the distance between the two centers. Then the volume V of L satisfies

$$V = \Theta\left(\sqrt{\varepsilon^{m+1}r_1^{m-1}}\right),\,$$

where the hidden constant depends only on m.

Proof. Assume $\varepsilon \le r_1$, since otherwise the upper bound follows trivially from the volume of the ball. Our lower bound for the case $\varepsilon > r_1$ will follow from the case $\varepsilon \le r_1$.

Let *C* be the cylinder circumscribed around *L*, with axis of symmetry the line joining the centers of *B*₁ and *B*₂. Then *C* has height ε , and radius $\sqrt{r_1^2 - (r_1 - \varepsilon_1)^2} = \sqrt{r_2^2 - (r_2 - \varepsilon_2)^2} = O(\sqrt{r_1\varepsilon_1}) = O(\sqrt{r_2\varepsilon_2})$, where $\varepsilon_1, \varepsilon_2$ are the heights of the two spherical segments comprising *L*. Since $\varepsilon_1 \le \varepsilon$, we find that the volume of this cylinder is $O(\varepsilon(\sqrt{r_1\varepsilon})^{m-1})$, as desired.

Similarly, if we scale down the height and radius of *C* by a factor of 2, and translate it appropriately, we can place this re-scaled cylinder completely within the lens *L*. This follows from the fact that half the circumscribed rectangle can fit inside a two-dimensional lens (see Fig. 10) and that the lens and cylinders in question are volumes of revolution. Since this changes the volume of the cylinder by a factor of 2^{-m} this establishes the desired lower bound with a constant that depends only on *m*. \Box

Given this relation between the width and volume of the lens, analogously to the work in Section 3, we can compute both short and long range estimates of the distance, and then combine them into a hybrid estimate. The error in the hybrid estimate is given by the following theorem.

Theorem 5.2. Let $r = n^{\alpha}$ for a constant $0 < \alpha < 1/m$. For all pairs u, v where v is deep, define d(u, v) be the hybrid estimate of the distance. Then with high probability, \hat{d} is an upper bound on the Euclidean

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

distance ||u - v|| with error

$$\widehat{\varepsilon}(u,v) \le C_m \begin{cases} n^{\frac{1}{m} - \frac{2m}{m+1}\alpha} & \alpha < \frac{m+1}{2m^2}, \\ \sqrt{\log n} & \frac{m+1}{2m^2} \le \alpha < \frac{1}{m}, \end{cases}$$
(22)

for some dimension-dependent constant C_m.

We omit the details of the proof except to comment that it closely follows the steps in Section 3. In order to use the hybrid estimates for reconstruction, we need to find an appropriate number of deep landmarks. By standard linear algebra, it suffices to have m + 1 landmarks that form a non-degenerate simplex. As in the proof of Theorem 4.3, we can find an approximately equilateral m-simplex in quadratic time, directly from the adjacency matrix of the graph. Finally, once we have reconstructed the relative positions of all the points, it is easy to compute an isometry that shifts the reconstructed hypercube to the origin.

In order to use the hybrid estimates for reconstruction, we need to find an appropriate number of deep landmarks. By standard linear algebra, it suffices to have m + 1 landmarks that form a non-degenerate simplex. As in the proof of Theorem 4.3, we can find an approximately equilateral m-simplex in quadratic time, directly from the adjacency matrix of the graph. Finally, once we have reconstructed the relative positions of all the points, it is easy to compute an isometry that shifts the reconstructed hypercube to the origin.

Putting it all together we have the following theorem for reconstruction of RGGs in $[0, n^{1/m}]^m$.

Theorem 5.3. Let $r = n^{\alpha}$ for a constant $0 < \alpha < 1/m$. There is an algorithm with running time $O(n^2)$ that with high probability reconstructs the vertex positions of a random geometric graph, modulo symmetries of the hypercube, with distortion

$$d^* \leq C_m \begin{cases} n^{\frac{1}{m} - \frac{2m}{m+1}\alpha} & \text{for } \alpha < \frac{m+1}{2m^2}, \\ \sqrt{\log n} & \text{for } \frac{m+1}{2m^2} \leq \alpha < \frac{1}{m} \end{cases}$$

for some dimension-dependent constant C_m .

5.2. Reconstruction on the sphere

Finally, we argue that our algorithm also works on some curved manifolds and submanifolds where the geometric graph is defined in terms of geodesic distance. In particular we claim this for the *m*-dimensional spherical (hyper)surface S_m of a ball in \mathbb{R}^{m+1} . Here we sketch the proof for the two-dimensional surface of a sphere in \mathbb{R}^3 (see Fig. 11). Note that the distortion is now defined by minimizing over the sphere's continuous symmetry group, i.e., over all rotations and reflections of the sphere.

In previous work, the authors of [25] gave a procedure to distinguish random geometric graphs on S^m from Erdős-Rényi random graphs. In addition, [22] gave a spectral method for reconstructing random graphs generated by a sparsified graphon model on the sphere, but this does not include the geodesic disk model we study here.

We define random geometric graphs on the sphere as follows. We scale the sphere so that its surface area is n, setting its radius to $R = \sqrt{n/(4\pi)}$. We scatter n points uniformly at random on it, or generate them with a Poisson point process with intensity 1. In either model the expected number of points in a patch of surface is equal to its area. We define the graph as $(u, v) \in E$ if and only if $||u - v||_g < r$ where $||u - v||_g$ is the geodesic distance, i.e., the length of the shorter arc of a great circle that connects u and v. If we associate each point u with a unit vector $\vec{u} \in \mathbb{R}^3$ that points towards it from the center of the sphere, $||u - v||_g$ is *R* times the angle between \vec{u} and \vec{v} .

Theorem 5.4. Let $r = n^{\alpha}$ for a constant $0 < \alpha < 1/2$. There is an algorithm with running time $O(n^2)$ that with high probability reconstructs the vertex positions of a random geometric graph, modulo a rotation or reflection of the sphere, with distortion an absolute constant times $n^{\frac{1}{2}-\frac{4}{3}\alpha}$ if $\alpha < 3/8$ and $\sqrt{\log n}$ if $\alpha \geq 3/8$.

V. Dani, J. Díaz, T.P. Hayes et al.

Proof Sketch. First note that objects of size o(R) act nearly as they do in the flat plane. In particular, the neighborhood B(u, r) of each point u is a spherical cap with angular radius r/R. The area of this cap is

$$2\pi R^2 \left(1 - \cos(r/R)\right) = \pi r^2 \left(1 - O(r^2/R^2)\right) = \pi r^2 \left(1 - O(r^2/n)\right) \,.$$

If $r = n^{\alpha}$ where $\alpha < 1/2$, the area of this neighborhood is 1 - o(1) times the area πr^2 of the corresponding ball in the flat Euclidean plane. Similarly, Lemma 3.7 for the area of a lens-shaped intersection between two caps holds as long as $r_1, r_2 = o(R)$.

The area of the lune $B(u, r) \setminus B(v, r)$ is a continuous function F(||u - v||) analogous to Eq. (2), and if we define the short-range distance estimate \tilde{d} as in Eq. (5) then Lemma 3.2 holds unchanged. Theorem 3.9 holds as well: the only significant change to the regions shown in Fig. 5 occurs if v is nearly antipodal to $x_0 = u$, but this only helps us since the greedy routing algorithm can get closer to v by moving in any direction on its first step. Thus we can compute hybrid distance estimates $\hat{d}(u, v)$ as before, with error $\hat{\varepsilon}$ as defined in (18).

If anything, the triangulation of Lemma 4.1 is easier on the sphere than on the plane. We can easily find three points x, y, z whose pairwise geodesic distances are within $O(\sqrt{\log n})$ of $(\pi/2)R$, so that the vectors $\vec{x}, \vec{y}, \vec{z}$ are nearly perpendicular. We can start with any point x (conveniently, all vertices are deep) and finding y and z takes O(n) time each: they exist with high probability since, as in flat space, a spherical cap with radius $\Omega(\sqrt{\log n})$ is nonempty with high probability.

We start our reconstruction by assuming that $\vec{x}, \vec{y}, \vec{z}$ are the three orthonormal basis vectors in \mathbb{R}^3 . Then we can estimate \vec{u} as

$$\vec{u} \approx \sum_{v \in \{x,y,z\}} \vec{v}(\vec{u} \cdot \vec{v}) = \sum_{v \in \{x,y,z\}} \vec{v} \cos \frac{\|u - v\|}{R} \approx \sum_{v \in \{x,y,z\}} \vec{v} \cos \frac{d(u,v)}{R} ,$$

normalizing if we like so that $\|\vec{u}\| = 1$. This gives \vec{u} with error $O(\hat{\varepsilon}/R)$, and mapping to the sphere of radius *R* reconstructs *u*'s position relative to *x*, *y*, *z* with error $O(\hat{\varepsilon})$. The right spherical triangle formed by our assumed positions for *x*, *y*, *z* is congruent to their true positions up to $O(\sqrt{\log n})$. So, up to some rotation or reflection of the sphere, this yields a reconstruction with distortion $O(\hat{\varepsilon} + \sqrt{\log n}) = O(\hat{\varepsilon})$. \Box

Remark 5.5. We have taken advantage of the fact that the 2-sphere has a convenient embedding in \mathbb{R}^3 . A more general approach, which we claim applies to any compact Riemannian submanifold with bounded curvature, would be to work entirely within the manifold itself, building a sufficient mesh of landmarks and then triangulating using geodesic distance. In particular, in the popular model of hyperbolic embeddings (e.g. [26–28]) where the submanifold is a ball of radius ℓ in a negatively curved space with radius of curvature *R*, we believe similar algorithms will work as long as $\ell/R = O(1)$. We leave this for future work.

6. Missing edges

In this section, we will briefly address a variant of the random geometric graph model, in which, before we are given the adjacency matrix, each edge is, independently, kept with probability *p*, and otherwise dropped. We will show that, in this setting, we can still accurately reconstruct the positions of the vertices, as long as *p* is not too small.

There are several reasons for considering this sort of generalization of random geometric graphs. It is a special case of a more general model in which, after placing the vertices in the plane, each edge (v, w) is included with probability f(||v - w||) where f is an arbitrary function, often required to be nonincreasing. For example

- The usual RGG model is a special case where $f(x) = \mathbb{H}(x \le r)$.
- The model currently under discussion corresponds to $f(x) = p \not\Vdash (x \leq r)$. This so-called "soft RGG" model was studied by Penrose in [29].

V. Dani, J. Díaz, T.P. Hayes et al.



Fig. 11. Using three landmarks x, y, z on the sphere to triangulate to other points in Theorem 5.4.

• Kleinberg's [30] small-world graph corresponds to $f(x) = \max\{1, x^{-C}\}$, where *C* is a constant, although in his model only the edges are random; the vertices form an $n \times n$ square array. He also uses ℓ_1 distance rather than ℓ_2 . An earlier small-world graph model due to Watts and Strogatz [31] corresponds to $f(x) = \mathbb{W}(x \le r) + p \mathbb{W}(x > r)$, although in that case, the underlying graph is an *n*-cycle.

In the soft RGG model, since the average degree is now $p\pi r^2$ except for vertices close to the boundary of the square, and we clearly cannot reconstruct the location of an isolated vertex, we can only hope to reconstruct locations when $pr^2 = \Omega(\log n)$. Let us make the somewhat stronger assumption that

$$p^2 r^2 = \Omega(\log n). \tag{23}$$

This will ensure that whenever $||u - v|| \leq r$ they almost surely have a common neighbor, so $d_G(u, v) \leq 2$. Moreover, this condition will be sufficient to allow our short-range distance reconstruction procedure to succeed, with slight modifications. Under these assumptions, our main arguments still work, with only minor modifications. The main technical issue is that, whereas before, for a region $A \subset B(v, r) \cap B(w, r)$, we could confidently predict that

$$|N(v) \cap N(w) \cap A| \approx |N(v) \cap A| \approx |V \cap A| \approx area(A)$$

after a 1 - p fraction of the edges have been dropped, we expect instead that

$$|N(v) \cap N(w) \cap A| \approx p|N(v) \cap A| \approx p^2|V \cap A| \approx p^2 \operatorname{area}(A).$$
(24)

By our assumption, (23), the above expectation is large enough to ensure that as long as the area of A is not too much smaller than πr^2 , $|N(v) \cap N(w) \cap A|$ is concentrated near its expectation. For the larger quantity $|N(v) \cap A|$, the expected size is much larger, and, broadly speaking, we can handle much smaller regions, A.

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

Proposition 6.1. Given the adjacency matrix of a graph $G = \mathcal{G}(n, r, p)$, where p and r are not given, there is an efficient algorithm which estimates p and r accurately with error probability $O(n^{-C'})$, for a constant C' > 0.

Proof Sketch. We can use the fact that for a deep vertex v, the expected degree is $p\pi r^2$. And for a pair of deep vertices whose distance is $x \le r$, their expected co-degree, $|N(v) \cap N(w)|$ is $p^2F(x)$, where F is the area of $B(v, r) \cap B(w, r)$, for which a formula was given in the proof of Lemma 3.2. This can be used to deduce that the average co-degree of two adjacent vertices is, with high probability, $Cp^2\pi r^2$, where C is a known constant. Combining this with the above prediction for the average degree, we can recover accurate estimates for p and r. \Box

Short range distance estimates

Let $G = \mathcal{G}(n, r, p)$, where $r p \ge 100\sqrt{\log n}$. Suppose v, w are vertices for which $d_G(v, w) \le 2$. We want to accurately estimate ||v - w|| as a function of $|N(v) \cap N(w)|$. In light of (24), we can set

$$\tilde{d}(v,w) = F^{-1}\left(\max\left\{F(r),\pi r^2 - \frac{|N(v) \cap N(w)|}{p^2}\right\}\right)$$

where *F*, defined just before the statement of Lemma 3.1, is the function determining the area of the lune $B(v, r) \setminus B(w, r)$, so that $\pi r^2 - F$ is the area of the corresponding lens $B(v, r) \cap B(w, r)$.

Theorem 6.2. Let $G = \mathcal{G}(n, r, p)$ where $r p \ge 100\sqrt{\log n}$. With probability at least $1 - 2/n^2$ we have, for all vertices $v \ne w$ such that $d_G(v, w) \le 2$ and v is deep,

$$\left| \|v - w\| - \tilde{d}(v, w) \right| \le C \eta(\|v - w\|) \log n,$$
⁽²⁵⁾

where C > 0 is a constant and $\eta : [0, 2r] \rightarrow [0, 1]$ is defined by

$$\eta(x) = \begin{cases} \frac{1}{p} \left(\frac{2r-x}{r}\right)^{1/4} & \text{for } 0 < x \le 2r - \frac{(\log n)^{2/3}}{r^{1/3}}, \\ \frac{(\log n)^{1/6}}{p r^{1/3}} & \text{for } 2r - \frac{(\log n)^{2/3}}{r^{1/3}} \le x \le 2r. \end{cases}$$

In particular, $\eta(x) \leq \frac{2^{1/4}}{p} < \frac{1.2}{p}$ for all $0 \leq x \leq 2r$.

Note that, because the asymmetric difference sets, $N(v) \setminus N(w)$ are no longer contained within the lune $B(v, r) \setminus B(w, r)$, we are now basing our distance estimations only on the intersections $N(v) \cap N(w)$, which are still contained within the lens $B(v, r) \cap B(w, r)$. Consequently, our error estimate $\eta(x)$ no longer approaches zero as $x \to 0$.

Long range distance estimates:

Here is our main theorem about long-range distance estimates with missing edges.

Theorem 6.3. Let $G = \mathcal{G}(n, r, p)$, where $r p \ge 100\sqrt{\log n}$. There exist absolute constants C_1, C_2, C_3 such that, for all $n \ge 1$, all $r \ge C_1\sqrt{\log n}$, with probability at least $1 - C_2/n^2$, all pairs of vertices u, v satisfy

$$\left\| \frac{\|u - v\|}{r} \right\| \le d_G(u, v) \le \max\left\{ 2, \left\lceil \frac{\|u - v\| + \kappa p^{-2/3} + \kappa_2 p^{-4/3}}{r} \right\rceil \right\}$$
(26)

where κ is the same as in Theorem 3.9, namely,

$$\kappa = C_3 \left(\frac{\|u - v\|}{r^{4/3}} + \frac{\log(n)}{r^{1/3}} \right)$$
(27)

and κ_2 is given by

$$\kappa_2 = \frac{C_4 (\log n)^{2/3}}{r^{1/3}}.$$

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

There are three differences between the conclusion of Theorem 6.3 and the original Theorem 3.9. Firstly, the original error term, κ , is now scaled by a factor of $p^{-2/3}$, because the greedy algorithm now makes less progress per step on average, due to the culled edges. There is also a second error term, κ_2 due to the possibility that the final edge for the greedy algorithm may have been dropped. To avoid this, we stop running the greedy algorithm once the distance to the target vertex gets a little smaller than 2r, and directly verify that almost surely only two more hops are required. Finally, in case we start out at Euclidean distance less than r, we still may need two hops, hence the "max" in the formula.

Hybrid distance estimates:

Finally we use the framework of Section 3.3 to create hybrid estimates of long-range distances. That is, we set

$$\widehat{d}(u,v) = \min_{w:d_G(w,v) \le 2} d_1(u,w) + d_2(w,v)$$
(28)

where d_1 is the long range estimate and d_2 is the short-range estimate defined above. Then, our main result is to compute the region of r and p for which reconstruction is possible. Specifically,

Theorem 6.4. Let $r = n^{\alpha}$ and $p = n^{-\gamma}$ for constants $0 \le \gamma \le \alpha \le 1/2$. For all pairs u, v where v is deep, define $\widehat{d}(u, v)$ as in Eq. (17). Then with high probability, \widehat{d} is an upper bound on the Euclidean distance ||u - v|| with error

$$\widehat{\varepsilon}(u,v) \le C n^{\lambda/6} \log(n), \text{ where } \lambda = \max\left\{6\gamma, \ 3 - 8\alpha + 4\gamma\right\}$$
(29)

and C is a constant.

For this result, we use our previously established hybrid distance construction, based on Theorems 6.2 and 6.3. For the short-distance estimates, we have error uniformly upper-bounded by 1/p, which is $O(r\sqrt{\log n})$. For the long-range estimate, the error is smallest when the Euclidean distance is close to a multiple of r, in which case it is approximately $\frac{\kappa}{p^{2/3}} + \frac{\kappa_2}{p^{4/3}}$. So the hybrid distance is likely to select a waypoint close to a multiple of r units from the landmark, and have total error

$$\frac{1.2}{p} + \frac{\kappa}{p^{2/3}} + \frac{\kappa_2}{p^{4/3}}.$$

Plugging in $r = n^{\alpha}$ and $p = n^{-\gamma}$, and doing a case analysis to identify the largest of the above three error terms, we deduce the result.

Corollary 6.5. For $p = \Theta(1)$ and $r \ge n^{3/8}$, \hat{d} is an upper bound on the Euclidean distance ||u - v|| with error $O(\log n)$. For $r = n^{\alpha}$ and $p = n^{-\gamma}$, where $3 + 4\gamma \le 14\alpha$ and $\gamma \le \alpha < 1/2$, \hat{d} is an upper bound on ||u - v|| with error $O(r \log n)$.

7. Conclusions and further work

We have shown how a combination of geometric and analytic ideas can be used to reconstruct random geometric graphs with lower distortion than in previous work [18], achieving a distortion of o(r) whenever $r = n^{\alpha}$ for $\alpha > 3/14$. Here we pose several questions for further work.

First, let us call a reconstruction ϕ consistent if its distances are consistent with the graph: that is, if $(u, v) \in E$ if and only if $\|\phi(u) - \phi(v)\| \leq r$. Even if ϕ has small distortion d^* , it might not be consistent: some edges $(u, v) \in E$ might have $\|\phi(u) - \phi(v)\|$ between r and $r + 2d^*$, and similarly some non-neighboring pairs might have $\|\phi(u) - \phi(v)\|$ between $r - 2d^*$ and r. To the best of our knowledge, even finding a single consistent embedding for random geometric graphs is an open question. It might be possible to refine our embedding to make it consistent, by using "forces" to move neighbors slightly closer together, and push non-neighbors farther away.

Secondly, a natural question is whether we can prove a significant lower bound on the distortion. An information-theoretic approach to this question would be to show that even the Bayesian algorithm, which chooses from the uniform measure on all consistent embeddings, has a typical

V. Dani, J. Díaz, T.P. Hayes et al.

distortion. We have been unable to prove this. However, here we sketch an argument that there *exist* consistent embeddings with a certain distortion by applying a continuous function f to the square $[0, \sqrt{n}]^2$ that "warps" the true embedding. If f's derivatives are at most δ in absolute value, then for each v, points close to the edge of v's neighborhood may move $O(\delta r)$ closer or farther away. However, a typical v has some $\varepsilon = O(1/r)$ for which there are no points whose distance is between $r - \varepsilon$ and $r + \varepsilon$, since the area of the corresponding annulus is O(1). This suggests that if $\delta = O(\varepsilon/r) = O(1/r^2)$, the warped embedding is still consistent (except for a few vertices where we need to be more careful). On other hand, even if f does not change the distance between nearby vertices very much, it can still move some vertices $\delta \sqrt{n}$ from their true positions, giving a distortion $d^* = \Omega(\sqrt{n}/r^2)$. If $r = n^{\alpha}$ this gives $\Omega(n^{1/2-2\alpha})$.

Even if this lower bound can be made rigorous, and even if it applies to typical consistent embeddings rather than just a few, there is a large gap between it and our upper bounds. Thus it is tempting to think that our algorithm can be improved, reducing the distortion still further. One approach would be to try to extend the geometry of overlapping disks in Theorem 3.6 to larger graph distances. Another would be to combine them with the spectral ideas of e.g. [22].

Thirdly, we would like to extend the results of Section 6 to other soft RGG models where $(u, v) \in E$ with probability f(||u - v||) for some known function f. It is fairly easy to generalize our arguments to families of functions where f(x) = 0 for x > r so that the bound $d_G(u, v) \ge ||u - v||/r$ still holds, although things change if f(x) tends to zero as x approaches r from below. For instance, if $f(x) = (1 - x/r)^{\zeta}$ for some constant $\zeta > 0$, the expected number of edges from u to a lens of width δ and area A just inside B(u, r) is $\Theta((\delta/r)^{\zeta}A) = \Theta(\delta^{3/2+\zeta} r^{1/2-\zeta})$. In order for there to be $\Theta(1)$ such edges, so that the greedy routing of Theorem 3.9 can succeed, we need to use lenses of width $\delta = \Theta(r^{-(1-2\zeta)/(3+2\zeta)})$ rather than $\Theta(r^{-1/3})$, altering the exponent in one of our error terms. Similarly, while we can carry out short-range estimates by replacing F(x) in (2) with an appropriate integral, the error bounds in Theorem 3.6 would change when ||u - v|| is close to zero or close to 2r.

Perhaps the more interesting question is how to reconstruct vertex positions when f(x) > 0for all $x \in \mathbb{R}^+$: for instance, if $f(x) = e^{-x/r}$. In that case, there are a small number of "longrange" edges (u, v) in the graph where $||u - v|| \gg r$. As a consequence, for some pairs we have $d_G(u, v) \ll ||u - v||/r$. Our intuition is that in this case, a greedy routing will not always give a good upper bound on $d_G(u, v)$, since it may be worth going "out of your way" to use these long-range edges in your path. At the same time, spectral methods may continue to perform well in this context. We leave this as a challenge for the future.

Finally, we would like to see how far these techniques can be extended to curved manifolds and submanifolds with boundary. In Theorem 5.4 we took advantage of the fact that the 2-sphere has a convenient embedding in \mathbb{R}^3 . A more general approach, which we claim applies to any compact Riemannian submanifold with bounded curvature, would be to work entirely within the manifold itself, building a sufficiently dense mesh of landmarks and then triangulating within mesh cells. In particular, in the popular model of hyperbolic embeddings (e.g. [26–28]) where the submanifold is a ball of radius ℓ in a negatively curved space with radius of curvature *R*, we believe similar algorithms will work as long as $\ell/R = O(1)$. We leave this for future work.

References

- Peter D. Hoff, Adrian E. Raftery, Mark S. Handcock, Latent space approaches to social network analysis, J. Amer. Statist. Assoc. 97 (2021) 1090–1099.
- [2] Yi-Jiao Zhang, Kai-Cheng Yang, F. Radicchi, Systematic comparison of graph embedding methods in practical tasks, Phys. Rev. E 104 (044315) (2021).
- [3] Heinz Breu, David G. Kirkpatrick, Unit disk graph recognition in NP-hard, Compt. Geometry Theory Appl. 9 (1–2) (1998) 3–24.
- [4] James Aspnes, David K. Goldenberg, Yang R. Yang, On the computational complexity of sensor network localization, in: Proceedings of the 1st ALGOSENSORS, 2004, pp. 32–44.
- [5] Jehoshua Bruck, Jie Gao, Anxiao Jiang, On the computational complexity of sensor network localization, in: Proceedings of the 6th ACM MOBIHOC, ACM, 2005, pp. 181–192.
- [6] Elchanan Mossel, Joe Neeman, Allan Sly, Reconstruction and estimation in the planted partition model, Probab. Theory Related Fields 162 (3-4) (2015) 431-461.

V. Dani, J. Díaz, T.P. Hayes et al.

European Journal of Combinatorics xxx (xxxx) xxx

- [7] Cristopher Moore, The computer science and physics of community detection: Landscapes, phase transitions, and hardness, Bull. EATCS 121 (2017) 19–55.
- [8] Emmanuel Abbe, Community detection and stochastic block models:Recent developments, J. Mach. Learn. Res. 18 (177) (2018).
- [9] Brent N. Clark, Charles J. Colbourn, David S. Johnson, Unit disk graphs, Discrete Math. 86 (1990) 165-177.
- [10] Edward. E. Gilbert, Random plane networks, J. Soc. Ind. Appl. Math. 9 (5) (1961) 533–543.
- [11] Mathew Penrose, Random Geometric Graphs, Oxford University Press, 2003.
- [12] Marc Walters, Random geometric graphs, Surv. Combin. (2011) 365–402.
- [13] Milan Bradonjić, Robert Elsässer, Tobias Friedrich, Thomas Sauerwald, Alexandre Stauffer, Efficient broadcast on random geometric graphs, in: Proceedings of the 21st. ACM-SIAM SODA, 2010, pp. 1412–1421.
- [14] Robert Ellis, Jeremy L. Martin, Catherine Yan, Random geometric graph diameter in the unit ball, Algorithmica 47 (4) (2007) 421–438.
- [15] Muthu Muthukrishnan, Gopal Pandurangan, The bin-covering technique for thresholding random geometric graph properties, in: Proceedings of the 16st. ACM-SIAM SODA, 2005, pp. 989–998.
- [16] E. Arias-Castro, A. Channarond, B. Pelletier, N. N. Verzelen, On the estimation of latent distances using graph distances, Electron. J. Stat. 15 (1) (2018) 722–747.
- [17] Josep Diaz, Dieter Mitsche, Guillem Perarnau, Xavier Pérez, On the relation between graph distance and Euclidean distance in random geometric graphs, Adv. Appl. Probab. 48 (3) (2016) 812–831.
- [18] Josep Diaz, Colin Mcdiarmid, Dieter Mitsche, Learning random points from geometric graphs or orderings, Random Struct. Algorithms 57 (2) (2019) 339–370.
- [19] Purnamrita Sarkar, Deepayan Chakrabarti, Andrew W. Moore, Theoretical justification of popular link prediction heuristics, in: Proceedings of the 22th AJCAI Conference, 2011, pp. 2722–2727.
- [20] Ittai Abraham, Shiri Chechik, David Kempe, Aleksandrs Slivkins, Low-distortion inference of latent similarities from a multiplex social network, SIAM J. Comput. (3) (2015) 617–668.
- [21] Cosma Rohilla Shalizi, Dena Asta, Consistency of maximum likelihood for continuous-space network models, 2019, arXiv:ArXive:1711.02123v2.
- [22] Ernesto Araya Valdivia, De Castro Yohann, Latent distance estimation for random geometric graphs, in: Advances in Neural Information Processing Systems, 2019.
- [23] S. Parthasarathy, D. Sivakoff, M. Tian, Y. Wang, A quest to unravel the metric structure behind perturbed networks, in: Proceedings of the 33rd Syymposium on Computational Geometry, 20175, pp. 53.1–53.16.
- [24] Shane Lubold, Arun G. Chandrasekhar, Tyler H. McCormick, Identifying the latent space geometry of network models through analysis of curvature, 2021, arXiv:ArXive:1711.02123.
- [25] Sebastien Bubeck, Jian Ding, Ronen Eldan, Miklós Rácz, Testing for high-dimensional geometry in random graphs, Random Struct. Algorithms 49 (3) (2016) 503–532.
- [26] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, Marián Boguñá, Hyperbolic geometry of complex networks, Phys. Rev. E 82 (036106) (2010).
- [27] Maksim Kitsak, Ivan Voitalov, Dmitri Krioukov, Link prediction with hyperbolic geometry, Phys. Rev. Res. 2 (043113) (2020).
- [28] Thomas Bläsius, Tobias Friedrich, Anton Krohmer, Sören Laue, Efficient embedding of scale-free graphs in the hyperbolic plane, IEEE/ACM Trans. Netw. 26 (2) (2018) 920–933.
- [29] Mathew Penrose, Connectivity of soft random geometric graphs, Ann. Appl. Probab. 26 (2016) 986–1028.
- [30] John Kleinberg, The small-world phenomenon: An algorithmic perspective, in: Proceedings of the 32nd ACM STOC, ACM, 2000, pp. 163–170.
- [31] Duncan J. Watts, Steven H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393.6684 (2011) 365–402.