Spanners in randomly weighted graphs: Euclidean case

Alan Frieze*and Wesley Pegden†
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA 15213

Abstract

Given a connected graph $G = (V, E)$ and a length function $\ell : E \to \mathbb{R}$ we let $d_{v,w}$ denote the shortest distance between vertex v and vertex w. A t-spanner is a subset $E' \subseteq E$ such that if $d'_{v,w}$ denotes shortest distances in the subgraph $G' = (V, E')$ then $d'_{v,w} \leq t d_{v,w}$ for all $v, w \in V$. We study the size of spanners in the following scenario: we consider a random embedding $\mathcal{X}_p$ of $G_{n,p}$ into the unit square with Euclidean edge lengths. For $\epsilon > 0$ constant, we prove the existence w.h.p. of $(1+\epsilon)$-spanners for $\mathcal{X}_p$ that have $O_\epsilon(n)$ edges. These spanners can be constructed in $O_\epsilon(n^2 \log n)$ time. (We will use O_ϵ to indicate that the hidden constant depends on ε.) There are constraints on p preventing it going to zero too quickly.

1 Introduction

Given a connected graph $G = (V, E)$ and a length function $\ell : E \to \mathbb{R}$ we let $d_{v,w}$ denote the shortest distance between vertex v and vertex w. A t-spanner is a subset $E' \subseteq E$ such that if $d'_{v,w}$ denotes shortest distances in the subgraph $G' = (V, E')$ then $d'_{v,w} \leq t d_{v,w}$ for all $v, w \in V$. We say that the *stretch* of E' is at most t. In general, the closer t is to one, the larger we need E' to be relative to E. Spanners have theoretical and practical applications in various network design problems. For a recent survey on this topic see Ahmed et al [1]. Work in this area has in the main been restricted to the analysis of the worst-case properties of spanners. In this note, we assume that edge lengths are random variables and do a probabilistic analysis.

We consider the case where $\ell_{i,j} = |X_i - X_j|$, where $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ are n randomly chosen points from $[0,1]^2$. The case where the n points are arbitrarily chosen is the subject of the book [10] by Narasimham and Smid. Section 15.1.2 of this book considers the random model where all $\binom{n}{2}$ edges between points are available. We denote this mode by $\mathcal{X}_1$. In this paper we consider a model where only a specified subgraph of the possible edges are available. In particular, we assume that edges exist between the points in $\mathcal{X}$, independently with probability p. We denote this model by $\mathcal{X}_p$. It constitutes a random embedding of the random graph $G_{n,p}$ into $[0,1]^2$. In the open problem session of CCCG 2009 [11], O'Rourke asked the following question: for what values of p is it true that w.h.p. $\mathcal{X}_p$ is a t-spanner for $\mathcal{X}_1$, where $t = O(1)$. Mehrabian and Wormald [7] showed that there is no choice of p with this property. Frieze and Pegden [3] proved a related negative result and also considered the increase in the shortest path length when going from $\mathcal{X}_1$ to $\mathcal{X}_p$,

Now $d_{i,j} = |X_i - X_j|$ when $\{i, j\} \in \mathcal{X}_p$ implies that with probability one, a 1-spanner contains all $\approx \binom{n}{2}p$ edges. We prove the following: We write $O_{\varepsilon,\theta}(\cdot)$ if the hidden constant in the big O notation depends on ε, θ. At the moment, in some places, these constants can grow rather fast, for example the dependence on ε is only bounded by $\varepsilon^{-O(1/\varepsilon)}$.

*Research supported in part by NSF grant DMS1952285
†Research supported in part by NSF grant DMS1363136

Theorem 1. *Suppose that the edges of $\mathcal{X}_p$ are given their Euclidean length. Let $\varepsilon, \theta > 0$ be arbitrary fixed constants. We describe the construction of a $(1+\varepsilon)$-spanner E_ε for $\mathcal{X}_p$.*

(a) If $np^{1+\theta} \to \infty$ then $\mathbb{E}(|E_\varepsilon|) = O_{\varepsilon,\theta}(p^{-\theta} n)$.

(b) If $\frac{1}{p \log 1/p} = o(\log^{1/2} n)$ then $|E_\varepsilon| \leq \mathbb{E}(|E_\varepsilon|) + O(n)$ w.h.p.

The definition of E_ε is given below in (7). On the other hand,

Theorem 2. *Suppose that the edges of $\mathcal{X}_p$ are given their Euclidean length. Let $\varepsilon > 0$ be an arbitrary fixed constant. If $np^2 \to \infty$ then w.h.p. any $(1+\varepsilon)$-spanner for $\mathcal{X}_p$ requires $\Omega(\varepsilon^{-1/2} n)$ edges.*

Remark 1. *We stress that we describe a $(1+\varepsilon)$-spanner for $\mathcal{X}_p$ and not for $\mathcal{X}_1$. The results of [7] and [3] rule out $O(1)$-spanners for $\mathcal{X}_1$ that only use edges of $\mathcal{X}_p$. This is because there will w.h.p. be pairs of points that are close together in Euclidean distance, but relatively far apart in $\mathcal{X}_p$.*

Remark 2. *We have assumed in Theorem 1 that $np^{1+\theta} \to \infty$. If we were to allow $np^{1+\theta} = o(1)$ then we would find that $np^{-\theta} \gg n^2 p$ and so the claimed size of our spanner is more than the likely number of edges in $\mathcal{X}_p$.*

Remark 3. *The constant θ is an artifact of our proof and we conjecture that it can be removed so that w.h.p. there is a $(1+\varepsilon)$-spanner of size $O_\varepsilon(n)$.*

We note that when points are placed arbitrarily and *all pairs of points* are connected by an edge then the so-called Θ-graph (defined below) produces a $(1+\varepsilon)$-spanner with $O(n/\varepsilon)$ edges. See Theorem 4.1.5 of [10].

The argument we present for Theorem 1 can be easily adapted to deal with random geometric graphs $G_{\mathcal{X},r}$ for sufficiently large radius r. Here we generate $\mathcal{X}$ as in Theorem 1 and now we join two vertices/points X, Y by an edge if $|X - Y| \leq r$. See Penrose [12] for an early book on this model.

Theorem 3. *If $r^2 \gg \frac{\log n}{n}$ then w.h.p. there is a $(1+\varepsilon)$-spanner using $O(n\varepsilon^{-2})$ edges.*

We note finally that Frieze and Pegden [4] have also considered the case where edge lengths are independently exponential mean one. The results there are much tighter.

2 Lower bound: the proof of Theorem 2

It is quite easy to prove the lower bound in Theorem 2., so we begin with this. Given an edge $\{A, B\} \in E(\mathcal{X}_p)$ we let $ellipse(A, B)$ be the ellipse with foci A, B defined by $|X - A| + |X - B| \leq (1+\varepsilon)r$. The edge $\{A, B\}$ is *lonely* if its length is r and there is no $X \in \mathcal{X} \cap ellipse(A, B)$ such that $\{A, X\}, \{B, X\}$ are edges of $\mathcal{X}_p$. Any $(1+\varepsilon)$-spanner must contain all of the lonely edges. Now $ellipse(A, B)$ has axes of size $a = (1+\varepsilon)r, b = (2\varepsilon + \varepsilon^2)^{1/2} r$ and so its volume is ψr^2 where $\psi = \pi(1+\varepsilon)(2\varepsilon+\varepsilon^2)^{1/2}/4$. By concentrating on points that are at least 0.1 from the boundary ∂D of $D = [0,1]^2$, we see that the expected number of lonely edges is at least

$$(0.64 - o(1))\binom{n}{2} p \int_{r=0}^{0.8\sqrt{2}} \left(1 - \psi r^2 p\right)^n \cdot 2\pi r dr \geq \frac{n^2 \pi}{2\psi} \int_{s=0}^{\psi p} (1-s)^n ds \geq \frac{n\pi}{3\psi}, \tag{1}$$

where we have used $(1-p)^n = o(1)$.

Concentration around the mean follows will follow from the Chebyshev inequality. In preparation for this, observe that if $r \geq \rho_\varepsilon = (20 \log n/(np\psi))^{1/2}$ then $\left(1 - \psi p r^2\right)^n = o(n^{-10})$ and so going back to the first integral in (1) we see that we can concentrate on lonely edges with $r \leq \rho_\varepsilon$. Next consider the event $\mathcal{R}$ that for each $A \in \mathcal{X}$ there are at most $100\psi^{-1}\log n$ $\mathcal{X}_p$ neighbors B such that $|A-B| \leq \rho_\varepsilon$. For a given A, the number of such close neighbors is distributed as a binomial with mean at most $20\pi\psi^{-1}\log n$. So the Chernoff bounds imply that $\mathcal{R}$ occurs with probability $1-o(n^{-10})$. So we let Z denote the number of lonely edges AB such that $|A-B| \leq \rho_\varepsilon$ and observe that $\mathbb{E}(Z) = \Omega(n/\varepsilon^{1/2}p)$.

Observe also that given an edge AB there are at most $O(\varepsilon^{-1}\log^2 n)$ edges CD for which $ellipse(A,B) \cap ellipse(C,D) \neq \emptyset$, assuming the occurrence of $\mathcal{R}$. Write $AB \sim CD$ to denote a non-empty intersection of ellipses. Thus, if $\mathcal{L}_{A,B}$ is the event that AB is lonely, then

$$\mathbb{E}(Z^2 \mid \mathcal{R}) \leq \sum_{AB}\sum_{CD\sim AB} \mathbb{P}(\mathcal{L}_{A,B} \mid \mathcal{R}) + \sum_{AB}\sum_{CD\not\sim AB} \mathbb{P}(\mathcal{L}_{A,B}, \mathcal{L}_{C,D} \mid \mathcal{R})$$
$$\leq O(\mathbb{E}(Z)\varepsilon^{-1}\log^2 n) + (1+o(1))\mathbb{E}(Z)^2 = (1+o(1))\mathbb{E}(Z)^2.$$

The Chebyshev inequality implies that Z is concentrated around its mean. This completes the proof of the lower bound in Theorem 1.

3 Upper bound: the proof of Theorem 1

Suppose that $0 < \varepsilon \ll 1$. It is perhaps instructive to consider the case where $p = 1$ i.e. where K_n is being embedded. In this case there are known, simple algorithms for finding a $(1+\varepsilon)$-spanner. For each $A \in \mathcal{X}$ we define τ cones $K_p(i,A), 0 \leq i < \tau$ with apex A and whose boundary rays make angles $i\varepsilon$ and $(i+1)\varepsilon$ with the horizontal. We then let $Y(i,A)$ denote the closest point in Euclidean distance to A in $K_p(i,A)$ that is adjacent to A in $\mathcal{X}_p$. We put $Y(i,A) = \perp$ if there is no such Y and let $d_{A,\perp} = \infty$. Also, define $i = i_{A,B}$ by $B \in K_p(i,A)$. When $p = 1$, the Yao graph [13] consists of the edges $(A, Y(i,A)), 0 \leq i < \tau, A \in \mathcal{X}$.

Remark 4. *It is known that the path $P(A,B) = (Z_0 = A, Z_1, \ldots, Z_m = B)$, where $Z_{i+1} = Y(i_{Z_i,B}, Z_i)$ has length at most $(\cos\varepsilon - \sin\varepsilon)^{-1}|A-B|$ and so the Yao graph has stretch factor $1+\varepsilon+O(\varepsilon^2)$.*

When $p < 1$, $P(A,B)$ may not exist in $\mathcal{X}_p$ and we show below how to overcome this problem.

We should also mention the very similar Θ-graph [9]. Here we replace $Y(i,A)$ by the point in $K(i,A)$ whose projection onto the bisector of $K(i,A)$ is closest to A. The Θ-graph also has a stretch factor of at most $(\cos\varepsilon - \sin\varepsilon)^{-1}$.

Let

$$r_\varepsilon = \left(\frac{M_{\theta,\varepsilon}}{np^{1+\theta}}\right)^{1/2} \text{ and } R_\varepsilon = \left(\frac{K_\theta \log n}{np^{1+\theta}}\right)^{1/2}. \tag{2}$$

where $M_{\theta,\varepsilon}$ is sufficiently large to justify some inequalities claimed below.

Let

$$E_1 = \{\{A,B\} \in \mathcal{X}_p : |A-B| \leq r_\varepsilon\}.$$

We have

$$\mathbb{E}(|E_1|) \leq \binom{n}{2}\pi r_\varepsilon^2 p \leq \frac{M_{\theta,\varepsilon} n}{2p^\theta} \tag{3}$$

and then we can assert that

$$|E_1| \leq \frac{M_{\theta,\varepsilon} n}{p^\theta} \text{ } w.h.p. \tag{4}$$

using the Chebyshev inequality. Here we can use the fact that the events of the form $\{|A-B| \leq r_\varepsilon\}$ are pair-wise independent.

Let

$$E_2 = \{(A, Y(i,A)) : A \in \mathcal{X}, i \in \{0,1,\ldots,\tau-1\}\} \text{ so that } |E_2| = O(n/\varepsilon). \tag{5}$$

The next two lemmas will discuss the case where A, B are sufficiently distant.

Lemma 4. *If $|A-B| \geq R_\varepsilon$ then with probability $1-o(n^{-10})$, $|A-Y| \leq \varepsilon|A-B|$, where $Y = Y(i_{A,B}, A)$.*

Proof. We have

$$\mathbb{P}(|A-Y| > \varepsilon|A-B|) \leq (1-\varepsilon\pi(\varepsilon R_\varepsilon)^2 p/2)^{n-1} \leq n^{-\varepsilon^3\pi M_{\theta,\varepsilon}/3p^\theta}.$$

The 2 in the middle expression allows half the cone to be outside $[0,1]^2$. □

Lemma 5. *If $r \geq R_\varepsilon$ then with probability $1-o(n^{-10})$, $d_{A,B} \leq (1+4\varepsilon)|A-B|$.*

Proof. Let X_1, X_2 be points on the line segment AB at distance $|A-B|/3, 2|A-B|/3$ from A respectively. Let $B_i, i = 1,2$ be the ball of radius εr centred at X_i. Let A_1 be the set of $\mathcal{X}_p$ neighbors of A in X_1 and let A_2 be the set of $\mathcal{X}_p$ neighbors of B in X_2. $\mathcal{E}_i, i = 1,2$ be the event that $|A_i| \geq \pi r^2 np/10$. Then the Chernoff bounds imply that

$$\mathbb{P}(\mathcal{E}_1 \wedge \mathcal{E}_2) \geq 1 - 2e^{-\pi r^2 np/1000} = 1 - O(n^{-\pi M_{\theta,\varepsilon}/1000p^\theta}).$$

Let $\mathcal{E}_3$ be the event that there is an $\mathcal{X}_p$ edge between A_1 and A_2. Then

$$\mathbb{P}(\mathcal{E}_3 \mid \mathcal{E}_1 \wedge \mathcal{E}_2) \geq 1 - (1-p)^{r^4 n^2 p^2/100} = 1 - O(n^{-K^2_{\theta,\varepsilon}/100p^\theta}).$$

Finally note that if $\mathcal{E}_i, i = 1,2,3$ all occur then $d_{A,B} \leq (1+4\varepsilon)|A-B|$. (4 is trivial and avoids any computation.) □

For $A, B \in \mathcal{A}$ we let $P_{A,B}$ denote the shortest path between A, B in $\mathcal{X}_p$ and we let $d_{A,B}$ denote the length of $P_{A,B}$.

Let

$$\mathcal{B}_\varepsilon = \{(A,B) : \ d_{A,B} \geq (1+\varepsilon)|B-A| \text{ and } r = |A-B| \geq r_\varepsilon\} \tag{6}$$

and

$$E_3 = \bigcup_{(A,B)\in\mathcal{B}_\varepsilon} E(P_{A,B}).$$

Let

$$\mathcal{C}_\varepsilon = \{(A,B) : \ d_{A,B} \leq (1+\varepsilon)|B-A| \text{ and } r = |A-B| \in [r_\varepsilon, R_\varepsilon] \text{ and } |A-Y| \geq \varepsilon|A-B|\},$$

where $Y = Y(i_{A,B}, A)$. Let

$$E_4 = \bigcup_{(A,B)\in\mathcal{C}_\varepsilon} E(P_{A,B}).$$

We show in Lemmas 8 and 11 that the expected sizes of the sets E_3, E_4 are $O_\varepsilon(n)$. Let

$$E_\varepsilon = \bigcup_{i=1}^{4} E_i. \tag{7}$$

Time: The construction of E_ε can obviously be done in polynomial time. The most time consuming parts being solving the all pairs shortest path problems defined by E_3, E_4. We show below that these sets consist of $O_\varepsilon(n)$ edges in expectation. So the expected time to solve these $O(n)$ single source problems via Dijkstra's algorithm is $O_\varepsilon(n^2 \log n)$, see Fredman and Tarjan [2].

For $X, Y \in \mathcal{X}$ we let $\widehat{d}_{X,Y}$ denote the length of the path from X to Y constructed by the following procedure: Given $A, B \in \mathcal{X}$ where $\{A, B\} \notin E$ we construct a path $A = Z_0, Z_1, \ldots, Z_k = B$ as follows: in the following, $Y_j = Y(i, Z_j)$ for $B \in K(i, Z_j), j \geq 0$.

CONSTRUCT:

D1 If $\{Z_j, B\} \in E_1$ then use $P_{Z_j,B}$ to complete the path, otherwise,

D2 If $|Z_j - Y_j| > \varepsilon|Z_j - B|$ then use $P_{Z_j,B}$ to complete the path, otherwise,

D3 If $d_{Y_j,B} \geq (1 + 5\varepsilon)|Y_j - B|$ then use $P_{Z_j,B}$ to complete the path, otherwise

D4 $Z_{j+1} \leftarrow Y_j$.

Remark 5. *We observe that Lemma 4 implies that with probability $1 - o(n^{-10})$ we do not use $P_{Z_j,B}$ for $|Z_j - B| \geq R_\varepsilon$. Denote the corresponding event by $\mathcal{U}$.*

The next lemma is used to estimate the quality of the path built by CONSTRUCT. (We can obviously replace 8ε by ε in order to get a $(1 + \varepsilon)$-spanner.)

Lemma 6. CONSTRUCT *produces a path of length at most $(1 + 7\varepsilon)d_{A,B}$.*

Proof. Let $A = Z_0, Z_1, \ldots, Z_k = B$ be the sequence defined by CONSTRUCT. If $k = 1$ then CONSTRUCT uses that path $P_{A,B}$ which has stretch one. Otherwise, let $d_j = |Z_j - B|$ for $0 \leq j \leq k$ and observe that it is a monotone decreasing sequence. Define $\bar{Z}_{j+1}$ to the point on the segment $Z_j Z_k$ such that $|\bar{Z}_{j+1} - Z_k| = |Z_{j+1} - Z_k|$. The assumption that $|Z_j - Z_{j+1}| \leq \varepsilon|Z_j - Z_k|$ implies that $\angle Z_{j+1} Z_k \bar{Z}_{j+1} < \pi/2$, and thus that the ratio

$$\frac{|Z_{j+1} - Z_j|}{d_j - d_{j+1}} \tag{8}$$

can be bounded by considering the case where $\angle Z_{j+1} Z_k \bar{Z}_{j+1} = \pi/2$, as it is drawn in Figure 1.

We have in that case that $\sin \varepsilon = \frac{d_{j+1}}{|Z_j - Z_{j+1}|}$ and $\cos \varepsilon = \frac{d_j}{|Z_j - Z_{j+1}|}$, giving $d_j - d_{j+1} = (\cos \varepsilon - \sin \varepsilon)|Z_j - Z_{j+1}|$. So, if CONSTRUCT only uses D4 then the length $L_{A,B}$ of the path constructed satisfies

$$L_{A,B} = \sum_{j=0}^{k-1} |Z_{j+1} - Z_j| \leq (\cos \varepsilon - \sin \varepsilon) \sum_{j=1}^{k} (d_j - d_{j+1}) = (\cos \varepsilon - \sin \varepsilon)|A - B| \leq (\cos \varepsilon - \sin \varepsilon) d_{A,B}.$$

Suppose that CONSTRUCT uses a path in D1,D2 or D3. If $k = 1$ then CONSTRUCT uses a shortest path from A to B in $\mathcal{X}_p$. Assume then that $k \geq 2$. It follows from the above argument that

$$\sum_{j=0}^{k-2} |Z_{j+1} - Z_j| \leq (\cos \varepsilon - \sin \varepsilon)||A - Z_{k-1}|.$$

Now,

$$d_{Z_{k-1},B} \leq |Z_{k-2} - Z_{k-1}| + d_{Z_{k-2},B} \leq \varepsilon|Z_{k-2} - B| + (1 + 5\varepsilon)|Z_{k-2} - B|$$

So,

$$L_{A,B} \leq (\cos \varepsilon - \sin \varepsilon)||A - Z_{k-1}| + (1 + 6\varepsilon)|Z_{k-2} - B|$$

$$\leq (1+6\varepsilon)(|A - Z_{k-2}| + |Z_{k-2} - B|)$$
$$\leq (1+6\varepsilon)(\cos\varepsilon - \sin\varepsilon)|A - B|.$$

□

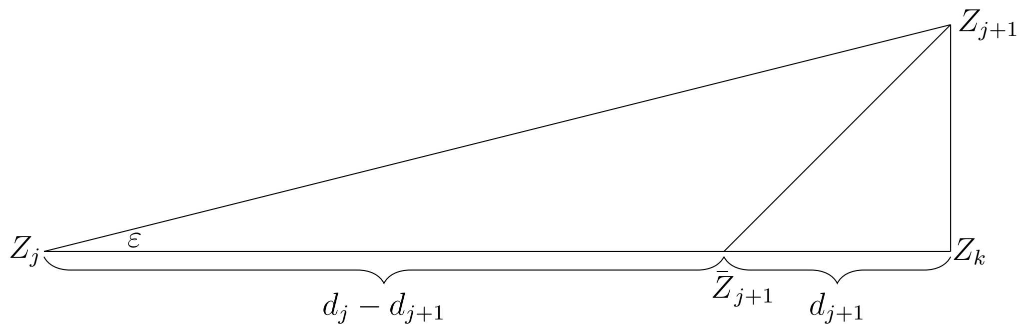

Figure 1: Extreme case for (8)

We argue next that

Lemma 7. *The edges of the paths $P_{Z_j,B}$ used in* CONSTRUCT *are contained in $E_1 \cup E_3 \cup E_4$. Furthermore, only edges of length at most R_ε contribute to E_3, E_4.*

Proof. First consider the path $P = P_{Z_j,B}$ used in D1. Because $\{Z_j, B\} \in E_1$, we have that $d_{Z_j,B} \leq r_\varepsilon$ and so all the edges of $P_{Z_j,B}$ are also in E_1.

Next consider the path $P = P_{Z_j,B}$ used in D2. If $d_{Z_j,B} \geq (1+\varepsilon)|Z_j - B|$ then $E(P) \subseteq E_3$. Otherwise, $E(P) \subseteq E_4$.

Now consider the path $P = P_{Z_j,B}$ used in D3. If $d_{Z_j,B} \geq (1+\varepsilon)|Z_j - B|$ then $E(P) \subseteq E_3$. So assume that $d_{Z_j,B} \leq (1+\varepsilon)|Z_j - B|$. If $|Z_j - Y_j| \geq \varepsilon|Z_j - B|$ then $E(P) \subseteq E_4$. So assume that $|Z_j - Y_j| \leq \varepsilon|Z_j - B|$. At this point we have

$$(1+5\varepsilon)\,|Y_j - B| \leq d_{Y_j,B} \leq |Z_j - Y_j| + d_{Z_j,B} \leq (1+2\varepsilon)|Z_j - B| \leq (1+2\varepsilon)\,(|Z_j - Y_j| + |Y_j - B|).$$

This implies that $|Z_j - Y_j| \geq 3\varepsilon|Y_j - B|/(1+2\varepsilon)$. If $|Y_j - B| \geq |Z_j - B|/2$ then we have $E(P) \subseteq E_4$. So assume that $|Y_j - B| < |Z_j - B|/2$. But then $|Z_j - Y_j| \geq |Z_j - B| - |Y_j - B| \geq |Z_j - B|/2$, a contradiction. □

The next two lemmas bound the expected number of edges in the sets E_3, E_4.

3.1 $\mathbb{E}(|E_3|)$

Lemma 8. $\mathbb{E}(|E_3|) = O_{\theta,\varepsilon}\left(\frac{n}{p^\theta}\right)$.

Proof. Fix a pair of points $A, B \in \mathcal{X}$ and let $r = |A - B|$ where $r_\varepsilon \leq r \leq R_\varepsilon$ ($r_\varepsilon, R_\varepsilon$ defined in (6)). Note next that shortest paths are always induced paths. We let $\mathcal{L}_{K,k,A,B}$ denote the set of induced paths from A to B with $k+1 \geq 2$ edges in $\mathcal{X}_p$, of total length in $[(1+K\varepsilon)r, (1+(K+1)\varepsilon)r]$.

We let $L_{K,k,A,B} = |\mathcal{L}_{K,k,A,B}|$. Then we have

$$|E_3| \leq \sum_{A,B\in\mathcal{X}} \sum_{k,K=1}^{\infty} k|\left\{P \in \mathcal{L}_{K,k,A,B}\right\}|. \tag{9}$$

This is because if $d_{A,B} \geq (1+\varepsilon)|A-B|$ then the shortest path from A to B has its length in $J_{K,r} = [(1+K\varepsilon)r, (1+(K+1)\varepsilon)r]$, for some $K \geq 1$. Next define, for $L \geq 1$,

$$F(L,\varepsilon) := (2L\varepsilon + L^2\varepsilon^2)^{1/2}.$$

Claim 1. *There are constants Λ, c such that for $K \geq 1$,*

$$\mathbb{E}\left(L_{K,k,A,B} \| A-B| = r\right) \leq \left(\frac{\Lambda F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2 np(1-p)^{(k-1)/2}}{k^2(K\varepsilon(1+K\varepsilon))^{1/4}}\right)^k e^{-cF(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2 np}. \quad (10)$$

Proof of Claim 1: Let $E_{A,B}(L)$ denote the ellipse with centre the midpoint of AB, foci at A, B so that one axis is along the line through AB and the other is orthogonal to it. The axis lengths a, b being given by $a = (1+L\varepsilon)r$ and $b = r((1+L\varepsilon)^2-1)^{1/2} = rF(L,\varepsilon)$. Thus $E_{A,B}(L)$ is the set of points whose sum of distances to A, B is at most $(1+L\varepsilon)r$.

Given k points $P_1, \ldots, P_k$, the path $P = (A = P_0, P_1, \ldots, P_k, P_{k+1} = B)$ is of length at most $(1+(K+1)\varepsilon)r$ only if all these points lie in $E_{A,B}(K+1)$. Thus for all i the point P_{i+1} lies in an ellipse with axes $2a, 2b$ centred at P_i. Here we are using the fact that if a point x lies in an ellipse E then E is contained in a copy of $2E$ centered at x. Indeed, suppose that $(x_i, y_i), i = 1,2$ are two points in the ellipse $E = \left\{\frac{x^2}{\xi^2} + \frac{y^2}{\eta^2} \leq 1\right\}$. Then

$$\frac{(x_1-x_2)^2}{\xi^2} + \frac{(y_1-y_2)^2}{\eta^2} \leq \frac{2(x_1^2+x_2^2)}{\xi^2} + \frac{2(y_1^2+y_2^2)}{\eta^2} = 2\sum_{i=1}^{2}\left(\frac{x_i^2}{\xi^2} + \frac{y_i^2}{\eta^2}\right) \leq 4. \quad (11)$$

It follows that (x_1, y_1) is contained in a copy of $2E$ centered at (x_2, y_2).

So, the probability of the event that $(A = P_0, P_1, \ldots, P_k)$ is in $E_{A,B}(K+1)$ is at most $\prod_{i=1}^k \mathbb{P}(\mathcal{P}_i)$ where $\mathcal{P}_i$ is the event that P_{i+1} is in the ellipse congruent to $2E_{A,B}(K+1)$, centred at P_i. So,

$$\mathbb{P}((A = P_0, P_1, \ldots, P_k, B) \text{ is in } E_{A,B}(K+1)) \leq (\pi r^2 F(K+1,\varepsilon)(1+(K+1)\varepsilon))^k p. \quad (12)$$

The final p factor is $\mathbb{P}(\{P_k, B\} \in E)$. Given $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_k$ the length of P is at most the sum $Z_1 + \cdots + Z_k$ of independent random variables where Z_i is the distance to the origin of a random point in an ellipse with axes $2a, 2b$ centred at the origin.

Lemma 9. *(a) Z_1 is distributed as $2(U(a^2\cos^2(2\pi V) + b^2\sin^2(2\pi V)))^{1/2}$ where U, V are independent uniform $[0,1]$ random variables.*

(b) Z_1 stochastically dominates $\zeta^{-1/2}U^{1/2}(K\varepsilon(1+K\varepsilon))^{1/4}r$ for some $\zeta > 0$.

Proof. (a) This follows from the fact that a point in E is of the form $(a\cos 2\pi\theta, b\sin 2\pi\theta)u$ where $0 \leq u, \theta \leq 1$.

(b) We have

$$\begin{aligned}\mathbb{P}(Z_1 \leq x) &= \mathbb{P}\left(U \leq \frac{x^2}{4(a^2\cos^2(2\pi V) + b^2\sin^2(2\pi V))}\right) \\ &= \mathbb{E}\left(\min\left\{1, \tfrac{1}{4}x^2(a^2\cos^2(2\pi V) + b^2\sin^2(2\pi V))^{-1}\right\}\right) \\ &\leq \min\left\{1, \mathbb{E}\left(\frac{x^2}{a^2\cos^2(2\pi V) + b^2\sin^2(2\pi V)}\right)\right\}.\end{aligned}$$

Now

$$\mathbb{E}\left(\frac{1}{a^2\cos^2(2\pi V) + b^2\sin^2(2\pi V)}\right) = \frac{2}{\pi}\int_{z=0}^{\pi/2} \frac{dz}{a^2\cos^2(z) + b^2\sin^2(z)} = \frac{2}{\pi}\int_{z=0}^{\pi/2} \frac{dz}{a^2\sin^2(z) + b^2\cos^2(z)}$$

$$
\begin{aligned}
&= \frac{2}{\pi}\int_{z=0}^{\pi/2}\frac{dz}{(a^2-b^2)\sin^2(z)+b^2} \\
&\leq \frac{4}{\pi}\int_{z=0}^{1/2}\frac{dz}{(a^2-b^2)z^2+b^2}+O\left(\frac{1}{a^2}\right) \\
&= \frac{4}{\pi r^2}\int_{z=0}^{1/2}\frac{dz}{z^2+2K\varepsilon+K^2\varepsilon^2}+O\left(\frac{1}{(1+(K+1)\varepsilon)^2r^2}\right). \\
&= \frac{4}{\pi r^2}\frac{\arctan\left(\frac{1}{2(2K\varepsilon+K^2\varepsilon^2)^{1/2}}\right)}{(2K\varepsilon+K^2\varepsilon^2)^{1/2}}++O\left(\frac{1}{(1+(K+1)\varepsilon)^2r^2}\right).
\end{aligned}
$$

So

$$
\mathbb{P}(Z_1\leq x)\leq\frac{\zeta x^2}{(K\varepsilon(1+K\varepsilon))^{1/2}r^2}
$$

for some $\zeta>0$.

This implies that Z_1 dominates $\zeta^{-1/2}U^{1/2}(K\varepsilon(1+K\varepsilon))^{1/4}r$. □

Lemma 9 of Frieze and Tkocz [5] implies that if $U_1,U_2,\ldots,U_k$ are independent copies of $U^{1/2}$ then

$$
\mathbb{P}(U_1^{1/2}+U_2^{1/2}+\cdots+U_k^{1/2}\leq u)\leq\frac{(2u)^{2k}}{(2k)!}.
$$

Putting $u=\frac{\alpha^{1/2}}{(K\varepsilon(1+K\varepsilon))^{1/4}r}$, we see that

$$
\mathbb{P}(Z_1+Z_2+\cdots+Z_k\leq(1+(K+1)\varepsilon)r)\leq\left(\frac{\alpha(1+(K+1)\varepsilon)}{(K\varepsilon(1+K\varepsilon))^{1/4}}\right)^k\frac{2^k}{(2k)!}\leq\left(\frac{\alpha(1+(K+1)\varepsilon)}{(K\varepsilon(1+K\varepsilon))^{1/4}}\right)^k\frac{e^{2k}}{k^{2k}2^k}. \tag{13}
$$

Thus, given k random points $P_1,\ldots,P_k$, the probability that $A,P_1,\ldots,P_k$ is an induced path of length $\leq(1+(K+1)\varepsilon)r$ is at most

$$
\left(\frac{\Lambda F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2np(1-p)^{(k-1)/2}}{k^2(K\varepsilon(1+K\varepsilon))^{1/4}}\right)^k.
$$

To get the exponential term in (10), we need to also make make use of the fact that $d_{A,B}\geq(1+\varepsilon K)r$.

Case 1: $K\varepsilon\leq 1$**:** Let $\gamma=\lceil 1+\theta^{-1}\rceil$. We define γ rhombi, $R_i,i=1,2,\ldots,\gamma$. We partition AB into γ segments $L_1,L_2,\ldots,L_\gamma$ of length r/γ. The rhombus R_i has one diagonal L_i and another diagonal of length $h=((K+1)\varepsilon)^{1/2}r/10\gamma$ that is orthogonal to AB and bisects it. Finally let $\widehat{R}_i=R_i\cap[0,1]^2$. Note that $\widehat{R}_i$ has area at least 1/2 of the area of R_i. Thus if $K\geq 1$ then since $K\varepsilon\leq 1$,

$$
\alpha\geq\alpha_i=\text{area}(\widehat{R}_i)\geq\frac{((K+1)\varepsilon)^{1/2}r^2}{20\gamma}\geq\frac{\alpha}{100} \tag{14}
$$

where

$$
\alpha=\frac{F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2}{\gamma}.
$$

For a pair of points A,B and set $X\subseteq\mathcal{X}$, let $d^*_{A,B}(X)$ denote the minimum length of a path $Q=(A,S_1,S_2,\ldots,S_\gamma,B)$ in $\mathcal{X}_p$ where $S_i\in\widehat{R}_i\setminus X$. Here X will stand for $P_1,P_2,\ldots,P_k$ in the analysis below. Furthermore we can restrict our attention to $|X|=k=o(n)$, as shown in (26) below. We first wish to show that

$$
\ell(Q)<(1+K\varepsilon)r \text{ for all choices of } S_1,S_2,\ldots,S_\gamma. \tag{15}
$$

Now fix i and consider the function $f(S) = \ell(A, S_1, S_2, \ldots, S_{i-1}, S, S_{i+1}, \ldots, , S_\gamma, B)$. This is a convex function of S and so it is maximised at an extreme point of $\widehat{R}_i \setminus X$. Thus to verify (15), it is enough to check paths that only use the vertices of the rhombi. We claim that

$$\ell(Q) \leq \gamma \left(4h^2 + \frac{1}{\gamma^2}\right)^{1/2} r \leq r\gamma \left(2h + \frac{1}{\gamma}\right) \leq (1 + (K+1)\varepsilon)r \tag{16}$$

where we have used $K\varepsilon \leq 1$ for the last inequality. Equation (16) follows from the fact that $\left(4h^2 + \frac{1}{\gamma^2}\right)^{1/2} r$ maximises the distance between points in adjacent rhombi.

Let Z denote the number of paths Q such that all edges exist in $\mathcal{X}_p$. We use Janson's inequality [6] to bound the probability that $Z = 0$. We have, with $\nu = n - |X| = n - o(n)$,

$$\mathbb{E}(Z) = \nu(\nu-1)\cdots(\nu-\gamma+1)p^{\gamma+1} \prod_{i=1}^{\gamma} \alpha_i \geq \left(\frac{\alpha np}{100}\right)^\gamma \frac{p}{2}.$$

Then for a pair of paths Q, Q' let $\rho(Q, Q'), \sigma(Q, Q')$, denote the number of vertices and edges the Q, Q' have in common. (Exclude A, B from this count.) We write $Q \sim Q'$ to mean that $\rho(Q, Q') > 0$. Then,

$$\bar{\Delta} = \sum_{Q \sim Q'} \mathbb{P}(Q, Q') \leq 2^{2\gamma} \sum_{\substack{1 \leq \sigma \leq \gamma+1 \\ \sigma \leq \rho \leq 2\sigma}} (\alpha n)^{2\gamma - \rho} p^{2\gamma+2-\sigma} \leq 2^{2\gamma+1} (\alpha n)^{2\gamma-1} p^{2\gamma+1}. \tag{17}$$

Explanation for (17) Because $r \geq r_\varepsilon$, we have $\alpha np \gg 1$. Thus the sum in (17) is dominated by the term $\rho = \sigma = 1$ where Q, Q' only share an edge incident to A or B. The factor $2^{2\gamma}$ accounts for the places on Q, Q' that share a common vertex.

It follows that if $K \geq 1$ then

$$\rho_{k,K,\varepsilon} = \mathbb{P}(d^*_{A,B} \geq (1 + K\varepsilon)r \mid |A - B| = r, P_1, \ldots, P_k) \leq \exp\left\{-\frac{\mathbb{E}(Z)^2}{2\bar{\Delta}}\right\} \leq$$
$$\exp\left\{-\frac{F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2 np}{2^{2\gamma+4}10^{4\gamma}\gamma}\right\} \leq \exp\left\{-\frac{M_{\theta,\varepsilon} F(K+1,\varepsilon)(1+(K+1)\varepsilon)}{2^{2\gamma+4}10^{4\gamma}\gamma p^\theta}\right\}$$

Case 2: $K\varepsilon \geq 1$**:** Let R be the rectangle with center the midpoint of AB and one side of length $(1 + (K+1)\varepsilon/10)r$ parallel to AB and the other of side $K\varepsilon/10$ orthogonal to AB. We partition R into rectangles $W_1, W_2, \ldots, W_\gamma$ where each W_i has side lengths $(1+(K+1)\varepsilon/10)r/\gamma$ and $K\varepsilon/10$. Putting $\widehat{W}_i = W_i \cap [0,1]^2, i = 1, 2, \ldots, \gamma$ we see that all we need do now is to prove the equivalent of (14) and (15). Then,

$$\text{area}(\widehat{W}_i) \geq \left(1 + \frac{(K+1)\varepsilon}{10}\right) \frac{K\varepsilon}{20\gamma} r^2 \geq \frac{F(K+1,\varepsilon)(1+(K+1)\varepsilon)}{1000\gamma} r^2.$$

We have used $K\varepsilon \geq 1$ to justify the second inequality.

We further have that for all $S_i \in \widehat{S}_i, i = 1, 2, \ldots, \gamma$ that, using the triangle inequality,

$$\ell(A, S_1, \ldots, S_\gamma, B) \leq \gamma \left(1 + \frac{(K+1)\varepsilon}{10}\right) \frac{r}{\gamma} + \gamma \left(\frac{K\varepsilon}{10} + \frac{4(K+1)\varepsilon}{10}\right) \frac{r}{\gamma} < (1 + (K+1)\varepsilon)r.$$

Thus, the probability $\rho_{k,K,\varepsilon}$ defined above satisfies

$$\rho_{k,K,\varepsilon} \leq \left(\frac{\Lambda F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2 np(1-p)^{(k-1)/2}}{k^2}\right)^k e^{-cF(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2 np},$$

and the claim follows by linearity of expectation.

End of proof of Claim 1

It will be convenient to replace r by $\frac{\rho}{(np)^{1/2}}$ and write $J_\rho = [\frac{\rho}{n^{1/2}}, \frac{\rho+1}{n^{1/2}}]$ and let $\rho_{\min} = r_\varepsilon (np)^{1/2}$. Then,

$$\begin{aligned}
&\mathbb{E}(|E_3|)\\
&\leq \binom{n}{2} \sum_{\rho=\rho_{\min}}^{\infty} \sum_{K=1}^{\infty} \sum_{k=1}^{n-2} k \left(\frac{\Lambda F(K+1,\varepsilon)(1+(K+1)\varepsilon) r^2 np (1-p)^{(k-1)/2}}{k^2 (K\varepsilon(1+K\varepsilon))^{1/4}} \right)^k \\
&\qquad\qquad \times e^{-cF(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2 np} \mathbb{P}(|A-B| \in J_\rho)\\
&\leq \binom{n}{2} \pi \sum_{\rho=\rho_{\min}}^{\infty} \sum_{K=1}^{\infty} \sum_{k=1}^{n-2} k \left(\frac{\Lambda F(K+1,\varepsilon) 1+(K+1)\varepsilon) r^2 np (1-p)^{(k-1)/2}}{k^2 (K\varepsilon(1+K\varepsilon))^{1/4}} \right)^k \\
&\qquad\qquad \times e^{-cF(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2 np} \left(\frac{2\rho+1}{n} \right)\\
&\leq 2\pi n \sum_{k=1}^{n-2} k \sum_{K=1}^{\infty} \left(\frac{\Lambda F(K+1,\varepsilon) 1+(K+1)\varepsilon)(1-p)^{(k-1)/2}}{k^2 (K\varepsilon(1+K\varepsilon))^{1/4}} \right)^k \sum_{\rho=\rho_{\min}}^{\infty} e^{-cF(K+1,\varepsilon)(1+(K+1)\varepsilon)\rho^2} \rho^{2k+1}\\
&\leq 2\pi n \sum_{k=1}^{n-2} k \sum_{K=1}^{\infty} \left(\frac{\Lambda F(K+1,\varepsilon) 1+(K+1)\varepsilon)(1-p)^{(k-1)/2}}{k^2 (K\varepsilon(1+K\varepsilon))^{1/4}} \right)^k \int_{s=0}^{\infty} e^{-cF(K+1,\varepsilon)(1+(K+1)\varepsilon)s} s^k ds\\
&= 2\pi n \sum_{k=1}^{n-2} k \sum_{K=1}^{\infty} \left(\frac{\Lambda F(K+1,\varepsilon) 1+(K+1)\varepsilon)(1-p)^{(k-1)/2}}{k^2 (K\varepsilon(1+K\varepsilon))^{1/4}} \right)^k \left(\frac{1}{cF(K+1,\varepsilon)(1+(K+1)\varepsilon)} \right)^{k+1} k!\\
&\leq 2\pi n \sum_{k=1}^{n-2} k \left(\frac{\Lambda (1-p)^{(k-1)/2}}{k\varepsilon^{1/4}} \right)^k \sum_{K=1}^{\infty} \left(\frac{1}{cF(K+1,\varepsilon)(1+(K+1)\varepsilon)} \right) \frac{1}{(K(1+K\varepsilon))^{k/4}} \qquad (18)\\
&= O_\varepsilon(n).
\end{aligned}$$

□

3.2 $\mathbb{E}(|E_4|)$

Lemma 10. *The expected number of $(k+1)$-edge induced paths of length at most $(1+\varepsilon)r$ from A to B in $\mathcal{X}_p$ can be bounded by*

$$\left(n\pi r^2 p (1-p)^{(k-1)/2} \frac{\varepsilon(1+\varepsilon)^3 e^2}{2k^2} \right)^k (1-\pi\varepsilon^3 r^2 p)^{n-k-2} p. \qquad (19)$$

Proof. Let ρ_k denote the probability that k fixed points $X_1, \ldots, X_k$ satisfy that:

- $A = X_0, X_1, \ldots, X_k$ is an induced path
- For all $i = 1, \ldots, k$, X_i lies in a copy of the ellipse $2 \cdot E_{A,B}$, translated to be centered at X_{i-1}, and
- The total length of the path has total length at most $(1+\varepsilon)r$.
- $\{X_k, B\} \in \mathcal{X}_p$.

From the discussion immediately prior to (11), we see that ρ_k bounds the probability that the the path has total length at most $(1+\varepsilon)r$. So we have that

$$\rho_k \leq (2\pi\varepsilon(1+\varepsilon)r^2 p)^k (1-p)^{k(k-1)/2} \left(\frac{e^2(1+\varepsilon)^2}{2k^2} \right)^k p.$$

Thus, by linearity of expectation, the number of induced paths $A = X_0, \dots, X_k$ such that

- the total length of the path is at most $(1+\varepsilon)r$, and
- no point off the path lies within distance εr of A in the cone $K(i, A)$

is at most

$$n^k(2\pi\varepsilon(1+\varepsilon)r^2p)^k(1-p)^{k(k-1)/2}\left(\frac{e^2(1+\varepsilon)^2}{2k^2}\right)^k(1-\pi\varepsilon^3r^2p)^{n-k-2}p =$$
$$\left(\frac{n\pi r^2p(1-p)^{(k-1)/2}}{1-\pi\varepsilon^3r^2p}\frac{\varepsilon(1+\varepsilon)^3e^2}{2k^2}\right)^k(1-\pi\varepsilon^3r^2p)^{n-k-2}p \le$$
$$\left(n\pi r^2p(1-p)^{(k-1)/2}\frac{\varepsilon(1+\varepsilon)^3e^2}{3k^2}\right)^k(1-\pi\varepsilon^3r^2p)^{n-k-2}p.$$

□

Lemma 11. $\mathbb{E}(|E_4|) = O_\varepsilon(n)$.

Proof. We have

$$\mathbb{E}(|E_4|) \le 2\pi\int_{r=r_\varepsilon}^{R_\varepsilon}\binom{n}{2}p\sum_{k=1}^{\infty}k\left(n\pi r^2p(1-p)^{(k-1)/2}\frac{\varepsilon(1+\varepsilon)^3e^2}{3k^2}\right)^k(1-\pi\varepsilon^3r^2p)^{n-k-2}rdr \tag{20}$$
$$\le 2\pi\binom{n}{2}p\sum_{k=1}^{\infty}k\int_{r=r_\varepsilon}^{R_\varepsilon}\left(\frac{e\pi\varepsilon r^2np(1-p)^{(k-1)/2}}{k^2}\right)^k e^{-\pi\varepsilon^3r^2np}rdr$$
$$\le \frac{n}{\varepsilon^3}\sum_{k=1}^{\infty}k\int_{s=A}^{\infty}\left(\frac{e\varepsilon(1-p)^{(k-1)/2}s}{\varepsilon^3k^2}\right)^k e^{-s}ds, \tag{21}$$

where $A = \pi\varepsilon^2r_\varepsilon^2np = M_{\theta,\varepsilon}p^{-\theta}$. Now,

$$I_k = \int_{s=A}^{\infty}s^ke^{-s} = k!\sum_{\ell=0}^{k}\frac{e^{-A}A^\ell}{\ell!} \le 2e^{-A}A^k, \qquad \text{if } k \le A/2. \tag{22}$$

(Use $I_k = kA^{k-1}e^{-A} + kI_{k-1}$ to obtain the equation.)
Using (22) in (21) we get, for small ε and $k_0 = 10\log_b 1/\varepsilon$ where $b = 1/(1-p)$,

$$\sum_{k=1}^{k_0}k\int_{s=A}^{\infty}\left(\frac{e(1-p)^{(k-1)/2}s}{\varepsilon^2k^2}\right)^k e^{-s}ds \le e^{-A}\sum_{k=1}^{k_0}\left(\frac{eA}{\varepsilon^2k^2}\right)^k \le$$
$$Ak_0\exp\left\{-M_{\theta,\varepsilon}p^{-\theta} + (M_{\theta,\varepsilon}p^{-\theta})^{1/2}\right\} \le \exp\left\{-\frac{M_{\theta,\varepsilon}}{2p^\theta}\right\}, \tag{23}$$

where we have used $(eC/x^2)^x \le e^{2C^{1/2}}$ for $C > 0$.
Finally,

$$\sum_{k=k_0+1}^{\infty}k\int_{s=A}^{\infty}\left(\frac{e(1-p)^{(k-1)/2}s}{\varepsilon^2k^2}\right)^k e^{-s}ds \le \int_{s=A}^{\infty}e^{-s}\sum_{k=k_0+1}^{\infty}\left(\frac{2e\varepsilon^3s}{k^2}\right)^k ds \le \int_{s=A}^{\infty}e^{-(1-\varepsilon)s}ds \le e^{-A/2}. \tag{24}$$

Substituting (23), (24) into (21) we see that $\mathbb{E}(|E_4|) = O\left(\frac{n}{\varepsilon^3}\right)$. □

We have argued that CONSTRUCT builds a $(1+\varepsilon)$-spanner w.h.p. The set of edges in this spanner is that of $\bigcup_{i=0}^4 E_i$. Part (a) of Theorem 1 now follows from (3), (5), Lemma 8 and Lemma 11.

3.3 Concentration of measure

Theorem 1 claims a high probability result. We apply McDiarmid's inequality [8] to prove that $|E_3|, |E_4|$ are within range w.h.p. We do not seem to be able to apply the inequality directly and so a little preparation is necessary. We first let $m = \lfloor 1/R_\varepsilon \rfloor$ and divide $[0,1]^2$ into a grid of m^2 subsquares $\mathcal{C} = (C_1, C_2, \ldots, C_{m^2})$ of size $1/m \geq R_\varepsilon$. The Chernoff bounds imply that with probability $1 - o(n^{-10})$ each $C \in \mathcal{C}$ contains at most $\rho_0 = 2nR_\varepsilon^2$ randomly chosen points of $\mathcal{X}$. Suppose that we generate the points one by one and color a point blue if it is one of the first ρ_0 points in its subsquare. Otherwise, color it red. Let $\mathcal{B}$ be the event that all points of $\mathcal{X}$ are blue and we note that

$$\mathbb{P}(\mathcal{B}) = 1 - o(n^{-10}). \tag{25}$$

Let

$$\kappa_1 = \frac{100 \log^{1/2} n}{p}. \tag{26}$$

The significance of κ_1 is that the factors $(1-p)^{k(k-1)/2}$ in equations (18) and (20) imply that

with probability $1 - o(n^{-2})$, no path contributing to E_3 or E_4 has more than κ_1 edges. (27)

We let Z_3 denote the number of edges $e = \{A, B\}$ that satisfy

(i) A, B are blue.

(ii) $r_\varepsilon \leq |A - B| \leq 2R_\varepsilon$ and $|Y(i_{A,B}, A) - A| \geq \varepsilon |A - B|$..

(iii) e is on an induced path in $\mathcal{X}_p$ that has length at least $(1+\varepsilon)|A - B|$ and at most κ_1 edges, each of length at most R_ε.

Similarly, let Z_4 denote the number of edges $e = \{A, B\}$ that satisfy

(i) A, B are blue.

(ii) $r_\varepsilon \leq |A - B| \leq 2R_\varepsilon$.

(iii) e is on an induced path in $\mathcal{X}_p$ that has length at most $(1+\varepsilon)|A - B|$ and at most κ_1 edges, each of length at most R_ε.

Let $Z_i', i = 3, 4$ be defined as for Z_i, without (i). Note that Lemma's 8 and 11 estimate $|E_i|$ through $|E_i| \leq Z_i'$ and showing $\mathbb{E}(Z_i') = O(n)$. Furthermore, $Z_i = Z_i', i = 3, 4$ if $\mathcal{U}, \mathcal{B}$ (see Remark 5) occur and these two events occur with probability $1 - o(n^{-10})$. Thus we have for $i = 3, 4$,

$$|E_i| \leq Z_i, \text{ w.h.p.}$$

and

$$E(Z_i) \leq \mathbb{E}(Z_i' \mid \mathcal{B} \cap \mathcal{U}) \mathbb{P}(\mathcal{B} \cap \mathcal{U}) + n^2 \mathbb{P}(\neg \mathcal{B} \vee \neg \mathcal{U}) \leq \mathbb{E}(Z_i') + n^2 \mathbb{P}(\neg \mathcal{B} \vee \neg \mathcal{U}) = O(n).$$

We will therefore bound the probability that either Z_3 or Z_4 exceeds its mean by n. We let $W = Z_3 + Z_4$. To apply McDiarmid's Inequality we have to establish a Liptschitz bound for W. Our probability space consists of $\bigtimes_{i=1}^{m^2} \Omega_i \times \bigtimes_{C_j \sim C_k} \Omega_{j,k}$ where Ω_i is a set of at most ρ_0 random points in subsquare C_i together with a list of all of the edges inside C_i. We say that $C_j \sim C_k$ if there boundaries share a common point. Thus for a fixed C_j there are usually 8 subsquares C_k such that $C_j \sim C_k$. The set $\Omega_{j,k}$ determines the edges between points in C_j and C_k. It can be represented by a $\rho_0 \times \rho_0$ $\{0,1\}$-matrix in which each entry appears independently with probability p. All in all there are $n^{1-o(1)}$ components of this probability space.

A point $X \in \mathcal{X}$ is in at most $\nu_0 = (9\rho_0)^{\kappa_1} = n^{o(1)}$ of the paths counted by W. So, changing an Ω_i or an $\Omega_{i,j}$ can only change W by at most $\nu_1 = 2\rho_0\nu_0\kappa_1 = n^{o(1)}$ and so the random variable W is ν_1-Liptschitz.. It then follows from McDiarmid's inequality that

$$\mathbb{P}(W \geq \mathbb{E}(W) + n) \leq \exp\left\{-\frac{n^2}{2n^{1-o(1)}\nu_1^2}\right\} = e^{-n^{1-o(1)}}.$$

This completes the proof of Theorem 1.

4 Proof of Theorem 3

For this we only have to observe that w.h.p. $K(X, i)$ exists for all X, i. This follows from the Chernoff bounds and the fact that the expected number of vertices in $K(X, i)$ grows faster than $\log n$. We can therefore use Lemma 6 to prove the existence of the required spanner.

5 Summary and open questions

There is a significant gap between the upper and lower bounds of Theorems 1 and 2, in their dependence on ε, p. Closing this gap is our greatest interest.

We have considered a Euclidean version, asking for a $(1 + \varepsilon)$-spanner and random geometric graphs. We could probably extend the results of Theorems 1, 2,3 to $[0, 1]^d, d \geq 3$. This does not seem difficiult. There is a slight problem in that the cones $K(i, X)$ intersect in sets of positive volume. The intersection volumes are relatively small and so the problems should be minor. We do not claim to have done this.

References

[1] R. Ahmed, G. Bodwin, F. Sahneh, K. Hamm, M. Javad, S. Kobourov and R. Spence, Graph Spanners: A Tutorial Review.

[2] M. Fredman and R. Tarjan, Fibonacci heaps and their uses in improved network optimization algorithms, *25th Annual Symposium on Foundations of Computer Science. IEEE* (1984) 338-346.

[3] A.M. Frieze and W. Pegden, Travelling in randomly embedded random graphs, *Random Structures and Algorithms* 55 (2019) 649-676

[4] A.M. Frieze and W. Pegden, Spanners in randomly weighted graphs: independent edge lengths.

[5] A.M. Frieze and T. Tkocz, Shortest paths with a cost constraint: a probabilistic analysis.

[6] S. Janson, Poisson approximation for large deviations, *Random Structures and Algorithms* 1 (1990) 221-230.

[7] A. Mehrabian and N. Wormald, On the Stretch Factor of Randomly Embedded Random Graphs, *Discrete & Computational Geometry* 49 (2013) 647-658.

[8] C. McDiarmid, On the method of bounded differences, *Surveys in combinatorics* 141 (1989) 148-188.

[9] G. Narasimhan and Smid, Approximating the stretch of Euclidean graphs, *SIAM Journal on Computing* 30 (2000) 978-989.

[10] G. Narasimhan and Smid, Geometric Spanner Networks, Cambridge University Press, 2007.

[11] E. D. Demaine and J. O'Rourke, Open Problems from CCCG 2009, *In Proceedings of the 22nd Canadian Conference on Computational Geometry (CCCG 2010)*, 83–86.

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

[13] A.C. Yao, On constructing minimum spanning trees in k-dimensional spaces and related problems, *SIAM Journal on Computing* 11 (1982) 721-736.