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 of $G_{n,p}$ into the unit square with Euclidean edge lengths.

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 $v, w \in V$. 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 $[0, 1]^2$. In addition we assume that edges exist between the points in $\mathcal{X}$, independently with probability p. This model, denoted by the graph $\mathcal{X}_p$ of a random embedding of $G_{n,p}$ into $[0, 1]^2$ has been considered in Frieze and Pegden [7], Mehrabian and Wormald [11]. The case of $p = 1$ (i.e. randomly chosen points) is discussed in Section 15.1.2 of Narasimham and Smid [13].

Now $d_{i,j} = |X_i - X_j|$ when $\{i, j\} \in \mathcal{X}_p$ implies that with probability one, a 1-spanner contains $\approx \binom{n}{2} p$ edges. We prove the following:

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

(a) If $np/ \log n \to \infty$ then $\mathbb{E}(|E_\varepsilon|) = O(\varepsilon^{-3}(\log 1/\varepsilon)p^{-1}n)$.

(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.

(c) If $np \to \infty$ then w.h.p. any $(1 + \varepsilon)$-spanner requires $\Omega(n\varepsilon^{-1}p^{-1})$ edges.

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

Assuming $\Omega(n \log n)$ edges we get a much simpler proof. Also, if we only want an $O(n)$ bound on the expected number of edges in our spanner then we can relax the constraint on p to $np/\log n \to \infty$. I.e. the extra restrictions are needed to prove the high probability result.

Theorem 2. *Suppose that $np^2/\log n \to \infty$. Then w.h.p. there is a $(1+\varepsilon)$-spanner with $O(\varepsilon^{-2}p^{-1}n\log n)$ edges.*

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 [14] 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 [8] have also considered the case where edge lengths are independently exponential mean one. The results there are much tighter.

2 Proof of Theorem 1

Suppose that $0 < \varepsilon \ll 1$. Let

$$r_\varepsilon = \frac{1000 \log^{1/2} 1/\varepsilon}{n^{1/2}\varepsilon^{3/2}p} \text{ and } R_\varepsilon = \left(\frac{10^6 \log n}{\pi n p \varepsilon^3}\right)^{1/2}. \tag{1}$$

The constraint on p means that

$$nR_\varepsilon^2 = n^{o(1)}.$$

Let

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

We have

$$\mathbb{E}(|E_1|) \leq \binom{n}{2}\pi r_\varepsilon^2 p \leq 10^6 \pi n \log(1/\varepsilon)/(2\varepsilon^3 p) \tag{2}$$

and then we can assert that

$$|E_1| \leq \frac{(10^6\pi \log 1/\varepsilon)n}{\varepsilon^3 p} \quad w.h.p. \tag{3}$$

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

For each $A \in \mathcal{X}$ we define τ cones $K(i, A), 0 \leq i < \tau$ 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 $K(i, A)$ 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 $i = i_{A,Y}$ by $Y \in K(i, A)$. Let

$$E_2 = \{(A, Y_{i,A}) : A \in \mathcal{X}, i \in \{0, 1, \ldots, \tau - 1\}\} \text{ so that } |E_2| = O(n/\varepsilon). \tag{4}$$

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 $P_{A,B}$. 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 $1 - o(n^{-100})$, $|A - Y| \leq \varepsilon|A - B|$, where $Y = Y(i_{A,B}, A)$.*

Proof. We have

$$\mathbb{P}(|A-Y|>\varepsilon|A-B|)\leq(1-\varepsilon\pi(\varepsilon R_\varepsilon)^2p/2)^{n-1}\leq e^{-10^6\log n/(\pi\varepsilon)}=o(n^{-100}).$$

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

Lemma 5. *If $r_\varepsilon \leq r=|A-B|\geq R_\varepsilon$ then with probability $1-o(n^{-2})$, $d_{A,B}\leq(1+4\varepsilon)|A-B|$.*

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 $|A_i|\geq r^2np/10$. Then the Chernoff bounds imply that

$$\mathbb{P}(\mathcal{E}_1\wedge\mathcal{E}_2)\geq 1-2e^{-\pi r^2np/1000}=1-o(n^{-100}).$$

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

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

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.) $\square$

Let

$$\mathcal{B}_\varepsilon=\{(A,B):\ d_{A,B}\geq(1+\varepsilon)|B-A| \text{ and } r=|A-B|\geq r_\varepsilon\} \tag{5}$$

and

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

Let

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

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 9 and 12 that the sets E_3,E_4 have linear size w.h.p. Let $E_\varepsilon=\bigcup_{i=0}^4E_i$.

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 $A=Z_0,Z_1,\dots,Z_k=B$ as follows: in the following, $Y_j=Y(i,Z_j)$ for $B\in K(i,Z_j), j\geq 0$.

Construct:

D1 If $\{Z_j,B\}\in E_1$, use the edge $\{Z_j,B\}$, otherwise

D2 If $|Z_j-Y_j|>|Z_j-B|$ 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 1. *We observe that Lemmas 4 and 5 imply that with probability $1-o(n^{-100})$ we do not use $P_{Z_j,B}$ for $|Z_j-B|\geq 2R_\varepsilon$. 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 5ε by ε in order to get a $(1+\varepsilon)$-spanner.)

Lemma 6. *Let $A = Z_0, Z_1, \ldots, Z_k, Z_{k+1} = B$ be a sequence of points where*

(i) $Z_{j+1} \in K(i_{Z_j,B}, Z_j)$ *for all* $0 \le j < k$*,*

(ii) $|Z_j - Z_{j+1}| \le |Z_j - B|$ *for* $0 \le j < k$*,*

(iii) $d_{Z_k,B} \le (1 + 5\varepsilon)|Z_k - B|$.

Then

$$d_{Z_k,B} + \sum_{j=0}^{k-1} |Z_{j+1} - Z_j| \le (1 + 5\varepsilon)|A - B|. \tag{6}$$

Proof. Let $d_j = |Z_j - B|$ for $0 \le j \le k$. To prove the lemma, it suffices to show for all j that

$$|Z_{j+1} - Z_j| \le (1 + 5\varepsilon)(d_j - d_{j+1}). \tag{7}$$

This implies that

$$d_{Z_k,B} + \sum_{j=0}^{k-1} |Z_{j+1} - Z_j| \le (1 + 5\varepsilon)d_k + (1 + 5\varepsilon)\sum_{j=0}^{k-1}(d_j - d_{j+1}) = (1 + 5\varepsilon)d_0.$$

To this end, define $\bar{Z}_{j+1}$ to the point on the segment Z_jZ_k such that $|\bar{Z}_{j+1} - Z_k| = |Z_{j+1} - Z_k|$. By the assumption that $|Z_j - Z_{j+1}| \le |Z_j - Z_k|$, we have that $\angle Z_{j+1}Z_k\bar{Z}_{j+1} < \pi/2$, and thus that the ratio

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

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 $d_j - d_{j+1} = (\cos\varepsilon - \sin\varepsilon)|Z_j - Z_{j+1}|$ which implies (7). $\square$

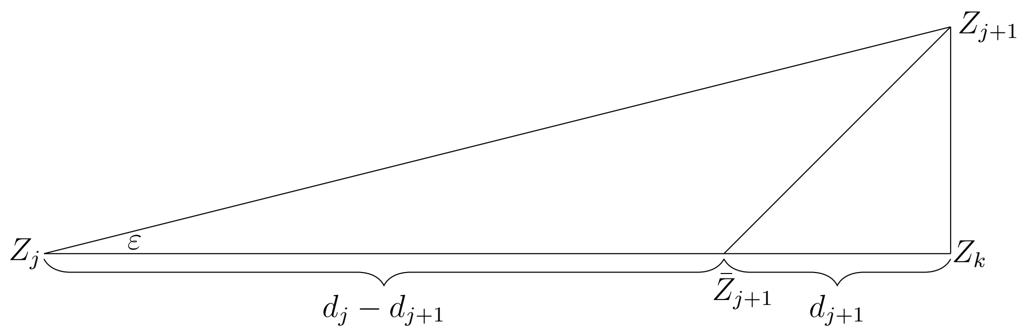

Figure 1:

Assuming (6) we see immediately that the path built by CONSTRUCT has a length within a $1 + 5\varepsilon$ factor of the minimum. We argue next that

Lemma 7. *The edges of the paths $P_{Z_j,B}$ used in* CONSTRUCT *are contained in $E_3 \cup E_4$.*

Proof. First 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 $|Y_j - B| \leq |Z_j - B|/2$. But then $|Z_j - Y_j| \geq |Z_j - B| - |Y_j - B| \geq |Z_j - B|/2$, a contradiction. □

Lemma 8. CONSTRUCT *produces a path of length at most* $(1+6\varepsilon)d_{A,B}$ *and only edges of length at most* $\varepsilon R_\varepsilon$ *contribute to* E_3, E_4.

Proof. We can assume that $|A - B| \geq R_\varepsilon$. Let $A = Z_0, Z_1, \ldots, Z_k, Z_{k+1} = B$ be the path constructed and let Z_ℓ be the first point within $\varepsilon R_\varepsilon$ of B. Note that this means $Z_\ell, Z_{\ell+1}, \ldots, Z_k$ are all within $\varepsilon R_\varepsilon$ of B. This yields the second claim of the lemma.

Let $\widehat{d}_\ell = \sum_{i=0}^{\ell-1} |Z_{i+1} - Z_i|$ be the length of the subpath from A to Z_ℓ. It follows from Lemma 6 that $\widehat{d}_\ell \leq (1+5\varepsilon)|A - Z_\ell|$. Because $|Z_{\ell-1} - B| \geq R_\varepsilon$ and $|Z_{\ell-1} - Z_\ell| \leq \varepsilon|Z_{\ell-1} - B|$ we see that $|Z_\ell - B| \geq (1-\varepsilon)R_\varepsilon$. It follows from Lemma 5 that $d_{Z_\ell,B} \leq (1+4\varepsilon)|Z_\ell - B|$. The path constructed by SECURE therefore has length at most

$$(1+5\varepsilon)|A - Z_\ell| + (1+5\varepsilon)(1+4\varepsilon)\varepsilon R_\varepsilon \leq (1+6\varepsilon)|A - B|.$$

□

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

2.1 $\mathbb{E}(|E_3|)$

Lemma 9. $\mathbb{E}(|E_3|) = O(n/\varepsilon^2)$.

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 (5)). 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 at most $(1+(K+1)\varepsilon)r$.

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

$$|E_3| \leq \sum_{A,B\in\mathcal{X}} \sum_{k,K=1}^{\infty} \sum_{P\in\mathcal{L}_{K,k,A,B}} |P| \cdot 1_{d_{A,B}\geq(1+K\varepsilon)r}. \tag{8}$$

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 $L \geq 1$,

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

Claim 1. *There exists an absolute constant* Λ *such that for* $K \geq 1$,

$$\mathbb{E}\left(L_{K,k,A,B} \cdot 1_{d_{A,B}\geq(1+\varepsilon K)r} \big| |A - B| = r\right) \leq \left(\frac{\Lambda F(K,\varepsilon)(1+K\varepsilon)r^2 np(1-p)^{(k-1)/2}}{k^2}\right)^k e^{-cF(K/4,\varepsilon)(1+K\varepsilon/4)r^2np}. \tag{9}$$

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 $b = r\sqrt{(1+L\varepsilon)^2 - 1} = rF(L,\varepsilon)$. 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 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. 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

$$\frac{(x_1-x_2)^2}{\xi^2} + \frac{(y_1-y_2)^2}{\eta^2} \leq \frac{2(x_1^2+x_2^2)}{\xi^2} + \frac{2(y_1^2+y_2^2)}{\eta^2} = 2\sum_{i=1}^{2}\left(\frac{x_i^2}{\xi^2} + \frac{y_i^2}{\eta^2}\right) \leq 4. \tag{10}$$

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

The following is well-known:

Lemma 10. *Z_1 is distributed as $U^{1/2}(a^2\cos^2(2\pi V) + b^2\sin^2(2\pi V))^{1/2}$ where U, V are independent uniform $[0,1]$ random variables.*

Proof. □

Now $(2\xi + \xi^2)^{1/2} \leq (1+\xi)$ and so $F(K+1,\varepsilon) \leq 1 + (K+1)\varepsilon$. So, Z_1 is dominated by $U^{1/2}(1+(K+1)\varepsilon)r$. It follows from Lemma 10 that $\ell(P)$ is dominated by $(1+(K+1)\varepsilon)r$ times the sum of k independent copies of $U^{1/2}$. Lemma 9 of Frieze and Tkocz [9] shows that

$$\mathbb{P}(Z_1 + Z_2 + \cdots + Z_k \leq (1+(K+1)\varepsilon)r) \leq \frac{2^k}{(2k)!} \leq \frac{e^{2k}}{k^{2k}2^k}. \tag{11}$$

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

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

(In equation (9), the ratio $\frac{F(K+1,\varepsilon)(1+(K+1)\varepsilon)}{F(K,\varepsilon)(1+K\varepsilon)} \leq 4$ has been absorbed into Λ.)

To get (9), we need to also make make use of the term $1_{d_{A,B} \geq (1+\varepsilon K)r}$ in (9).

Case 1: $K\varepsilon \leq 1$**:** We define two rhombi, R_A, R_B. Let M denote the middle of the segment AB. Then R_A has one diagonal AM and another diagonal $W_A W_A'$ of length $(K\varepsilon)^{1/2}r/10$ that is orthogonal to AM and bisects it. The rhombus R_A is defined similarly. Finally let $\widehat{R}_A = R_A \cap [0,1]^2$ and $\widehat{R}_B = R_B \cap [0,1]^2$. Note that at $\widehat{R}_A$ has area at least 1/2 of the area of $\widehat{R}_A$ and similarly for $\widehat{R}_B$. Thus if $K \geq 1$ then

$$\text{area}(\widehat{R}_A) \geq \frac{(K\varepsilon)^{1/2}r^2}{20} \geq \frac{F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2}{100}. \tag{12}$$

Here we have used $K\varepsilon \leq 1$ to justify the second inequality.

For a pair of points A, B let $d^*_{A,B}(X)$ denote the minimum length of a path $Q = (A, S, T, B)$ in $\mathcal{X}_p$ where $S \in \widehat{R}_A \setminus X$ and $T \in \widehat{R}_B \setminus X$. We wish to show that

$$\ell(Q) \leq (1+K\varepsilon)r \text{ for all choices of } S, T. \tag{13}$$

Now fix S and consider the function $f(T) = |ST| + |TB|$. This is a convex function and so it is maximised at an extreme point of $\widehat{R}_B \setminus X$. Then for a fixed T we find that maximising over S, we have that S must be an extreme point of $\widehat{R}_A \setminus X$. To verify (13), it is enough to check the length

$$\ell(AW_AW'_BB) \leq \left(2\left(\left(\frac{1}{4}\right)^2 + \left(\frac{(K\varepsilon)^{1/2}}{20}\right)^2\right)^{1/2} + \left(\left(\frac{1}{2}\right)^2 + \left(\frac{(K\varepsilon)^{1/2}}{10}\right)^2\right)^{1/2}\right) r$$
$$\leq \left(2\left(\frac{1}{4}\left(1+\frac{K\varepsilon}{400}\right) + \left(\frac{1}{2}\left(1+\frac{K\varepsilon}{100}\right)\right)\right)\right) r < \left(1+\frac{K\varepsilon}{40}\right) r.$$

This dominates the other 15 possibilities for a pair S, T.

We have that for any X, we have

$$1_{d_{A,B} \geq (1+\varepsilon K)r} \leq 1_{d^*_{A,B}(X) \geq (1+\varepsilon K)r}.$$

The number of points ν_A in $\widehat{R}_A \setminus \{P_i\}$ that are adjacent to A in $\mathcal{X}_p$ is distributed as a binomial with mean at least $c_A F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2(n-k)p$ for some absolute constant $c_A > 0$. This follows from (12). Because $k = o(n)$, w.h.p., as will be discussed below, see (23), we have that $\mathbb{P}(\nu_A \leq c_A F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2np/2) \leq e^{-c_A F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2np/20}$. Similarly, $\mathbb{P}(\nu_B \leq c_B F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2np/2) \leq e^{-c_B F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2np/20}$ where ν_B is the number of points in $\widehat{R}_B$ that are adjacent to B in $\mathcal{X}_p$. An edge between $\widehat{R}_A$ and $\widehat{R}_B$ will show that $d^*_{A,B} < (1+K\varepsilon)r$, see (13). We see that there will be at least $\nu_A\nu_B/2$ possible edges. It follows that if $K \geq 1$ then

$$\rho_{k,K,\varepsilon} = \mathbb{P}(d^*_{A,B} \geq (1+K\varepsilon)r \mid |A-B| = r, P_1, \ldots, P_k) \leq$$
$$e^{-c_A F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2np/20} + e^{-c_B F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2np/20} + (1-p)^{c_Ac_B F((K+1),\varepsilon)^2(1+(K+1)\varepsilon)^2 r^4n^2p^3/4}$$
$$\leq e^{-c_A F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2np/20} + e^{-c_B F(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2np/20} + (1-p)^{c_Ac_B F(K+1,\varepsilon)^2(1+(K+1)\varepsilon)^2 r^2r_\varepsilon^2n^2p^3/4}$$
$$e^{-cF(K+1,\varepsilon)(1+(K+1)\varepsilon)r^2np}, \quad (14)$$

for some absolute constant $c > 0$, since $r_\varepsilon^2np^2 \geq 10^4\varepsilon^{-2}$.

Case 2: $K\varepsilon \geq 1$**:** We replace R_A, R_B by two halves S_A, S_B of the rectangle with center M 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. Putting $\widehat{S}_A = S_A \cap [0,1]^2$ and $\widehat{S}_B = S_B \cap [0,1]^2$ we see that all we need do now is to prove the equivalent of (12) and (13). Then,

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

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

We further have that for all $S \in \widehat{S}_A, T \in \widehat{S}_B$ that, using the triangle inequality,

$$\ell(ASTB) \leq \left(1+\frac{(K+1)\varepsilon}{10}\right) r + 4\left(\frac{K\varepsilon}{10}+\frac{(K+1)\varepsilon}{10}\right) r < (1+(K+1)\varepsilon)r.$$

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

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

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,

$$
\begin{aligned}
&\mathbb{E}(|E_3|)\\
&\le \binom{n}{2} \sum_{\rho=\rho_{\min}}^{\infty} \sum_{K=1}^{\infty} \sum_{k=1}^{n-2} k \left(\frac{\Lambda F(K,\varepsilon)(1+K\varepsilon) r^2 np (1-p)^{(k-1)/2}}{k^2} \right)^k e^{-cF(K+1,\varepsilon)(1+(K+1)\varepsilon) r^2 np} \mathbb{P}(|A-B| \in J_\rho)\\
&\le \binom{n}{2} \pi \sum_{\rho=\rho_{\min}}^{\infty} \sum_{K=1}^{\infty} \sum_{k=1}^{n-2} k \left(\frac{\Lambda F(K,\varepsilon)(1+K\varepsilon) r^2 np (1-p)^{(k-1)/2}}{k^2} \right)^k e^{-cF(K+1,\varepsilon)(1+(K+1)\varepsilon) r^2 np} \left(\frac{2\rho+1}{n} \right)\\
&\le 2\pi n \sum_{k=1}^{n-2} k \sum_{K=1}^{\infty} \left(\frac{\Lambda F(K,\varepsilon)(1+K\varepsilon)(1-p)^{(k-1)/2}}{k^2} \right)^k \sum_{\rho=\rho_{\min}}^{\infty} e^{-cF(K+1,\varepsilon)(1+(K+1)\varepsilon)\rho^2} \rho^{2k+1}\\
&\le 2\pi n \sum_{k=1}^{n-2} k \sum_{K=1}^{\infty} \left(\frac{\Lambda F(K,\varepsilon)(1+K\varepsilon)(1-p)^{(k-1)/2}}{k^2} \right)^k \int_{s=0}^{\infty} e^{-cF(K+1,\varepsilon)(1+(K+1)\varepsilon)s} s^k ds\\
&= 2\pi n \sum_{k=1}^{n-2} k \sum_{K=1}^{\infty} \left(\frac{\Lambda F(K,\varepsilon)(1+K\varepsilon)(1-p)^{(k-1)/2}}{k^2} \right)^k \left(\frac{1}{cF(K+1,\varepsilon)(1+(K+1)\varepsilon)} \right)^{k+1} k!\\
&\le 2\pi n \sum_{k=1}^{n-2} k \left(\frac{64\Lambda(1-p)^{(k-1)/2}}{k} \right)^k \sum_{K=1}^{\infty} \left(\frac{1}{cF(K+1,\varepsilon)(1+(K+1)\varepsilon)} \right) \qquad (15)\\
&= O\left(\frac{n}{\varepsilon^2} \right).
\end{aligned}
$$

□

2.2 $\mathbb{E}(|E_4|)$

Lemma 11. *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*

$$
\left(n\pi r^2 p (1-p)^{(k-1)/2} \frac{\varepsilon(1+\varepsilon)^3 e^2}{16k^2} \right)^k (1-\pi\varepsilon^3 r^2 p)^{n-2}. \qquad (16)
$$

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 $(1+\varepsilon)r$.

By (10), this is the same as the probability that the the path has total length at most $(1+\varepsilon)r$. Applying (11), we have that

$$
\rho_k \le (2\pi\varepsilon(1+\varepsilon) r^2 p)^k (1-p)^{k(k-1)/2} \left(\frac{e^2(1+\varepsilon)^2}{16k^2} \right)^k.
$$

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

$$n^k(2\pi\varepsilon(1+\varepsilon)r^2p)^k(1-p)^{k(k-1)/2}\left(\frac{e^2(1+\varepsilon)^2}{16k^2}\right)^k(1-\pi\varepsilon^3r^2p)^{n-k-2} =$$

$$\left(\frac{n\pi r^2p(1-p)^{(k-1)/2}}{1-\pi\varepsilon^3r^2p}\frac{\varepsilon(1+\varepsilon)^3e^2}{16k^2}\right)^k(1-\pi\varepsilon^3r^2p)^{n-2} \leq$$

$$\left(n\pi r^2p(1-p)^{(k-1)/2}\frac{\varepsilon(1+\varepsilon)^3e^2}{16k^2}\right)^k(1-\pi\varepsilon^3r^2p)^{n-2}.$$

□

Lemma 12. $\mathbb{E}(|E_4|) = O\left(\frac{n}{\varepsilon^3p}\right)$.

Proof. We have

$$\mathbb{E}(|E_4|) \leq 2\pi\int_{r=r_\varepsilon}^{\infty}\binom{n}{2}\sum_{k=1}^{\infty}k\left(n\pi r^2p(1-p)^{(k-1)/2}\frac{\varepsilon(1+\varepsilon)^3e^2}{16k^2}\right)^k(1-\pi\varepsilon^3r^2p)^{n-2}rdr \tag{17}$$

$$\leq 2\pi\binom{n}{2}\sum_{k=1}^{\infty}k\int_{r=r_\varepsilon}^{\infty}\left(\frac{\pi\varepsilon r^2np(1-p)^{(k-1)/2}}{k^2}\right)^k e^{-\pi\varepsilon^3r^2np}rdr$$

$$\leq \frac{n}{\varepsilon^3p}\sum_{k=1}^{\infty}k\int_{s=A}^{\infty}\left(\frac{\varepsilon(1-p)^{(k-1)/2}s}{\varepsilon^3k^2}\right)^k e^{-s}ds, \tag{18}$$

where $A = \pi\varepsilon^2r_\varepsilon^2np = 10^6\varepsilon^{-1}\pi\log(1/\varepsilon)/p$. Now,

$$I_k = \int_{s=A}^{\infty}s^ke^{-s} = k!\sum_{\ell=0}^{k}\frac{e^{-A}A^\ell}{\ell!} \leq 2e^{-A}A^k, \qquad \text{if } k \leq A/2. \tag{19}$$

(Use $I_k = kA^{k-1}e^{-A} + kI_{k-1}$ to obtain the equation.)
Using (19) in (18) we get, for small ε and $k_0 = 10\log_b 1/\varepsilon$ where $b = 1/(1-p)$,

$$\sum_{k=1}^{k_0}k\int_{s=A}^{\infty}\left(\frac{(1-p)^{(k-1)/2}s}{\varepsilon^2k^2}\right)^k e^{-s}ds \leq Ae^{-A}\sum_{k=1}^{k_0}\left(\frac{A}{\varepsilon^2k^2}\right)^k \leq$$

$$Ak_0\exp\left\{-\frac{10^6(\log 1/\varepsilon)\pi}{\varepsilon p}+\left(\frac{10^6(\log 1/\varepsilon)\pi}{\varepsilon^2p}\right)^{1/2}\right\} \leq \exp\left\{-\frac{10^5(\log 1/\varepsilon)\pi}{\varepsilon p}\right\}, \tag{20}$$

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

$$\sum_{k=k_0+1}^{\infty}k\int_{s=A}^{\infty}\left(\frac{(1-p)^{(k-1)/2}s}{\varepsilon^2k^2}\right)^k e^{-s}ds \leq \int_{s=A}^{\infty}e^{-s}\sum_{k=k_0+1}^{\infty}\left(\frac{2\varepsilon^3s}{k^2}\right)^k ds \leq \int_{s=A}^{\infty}e^{-(1-\varepsilon)s}ds \leq e^{-A/2}. \tag{21}$$

Substituting (20), (21) into (18) we see that $\mathbb{E}(|E_4|) = O\left(\frac{n}{\varepsilon^3p}\right)$. □

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 (2), (4), Lemma 9 and Lemma 12.

2.3 Concentration of measure

Theorem 1 claims a high probability result. We apply McDiarmid's inequality [12] 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 $1 - o(n^{-101})$ 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

$$\mathbb{P}(\mathcal{B}) = 1 - o(n^{-100}). \tag{22}$$

Let

$$\kappa_1 = \frac{100 \log^{1/2} n}{p}. \tag{23}$$

The significance of κ_1 is that the factors $(1-p)^{k(k-1)/2}$ in equations (15) and (17) imply that

with probability $1 - o(n^{-100})$, no path contributing to E_3 or E_4 has more than κ_1 edges. (24)

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 9 and 12 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 1) occur and these two events occur with probability $1 - o(n^{-100})$. 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 $\times_{i=1}^{m^2} \Omega_i \times \times_{C_j \sim C_k} \Omega_{j,k}$ 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 verifies the existence of the claimed $(1+\varepsilon)$-spanner and now we argue that w.h.p. we need $\Omega(n(\varepsilon p)^{-1})$ edges. We say that an edge $\{A,B\}$ is *lonely* if its length is r and there are no $\mathcal{X}_p$-adjacent points in the ellipse C with foci A,B defined by $|X-A|+|X-B| \leq (1+\varepsilon)r$. Any $(1+\varepsilon)$-spanner must contain all lonely edges. Now the volume of C is $\pi\varepsilon(1+\varepsilon)r^2/2$. 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

$$(0.64 - o(1))\binom{n}{2} p \int_{r=0}^{0.8\sqrt{2}} \left(1 - \frac{\pi\varepsilon(1+\varepsilon)pr^2}{2}\right)^n 2\pi r dr \geq \pi\binom{n}{2} \cdot \frac{1}{\pi\varepsilon(1+\varepsilon)p}\int_{s=0}^{\varepsilon p}(1-s)^n ds \approx \frac{n}{2\varepsilon(1+\varepsilon)p}.$$

This completes the proof of Theorem 1.

3 Proof of Theorem 2

Let $K(i,X), Y(i,X)$ be as in Section 2 and let $M = 4/\varepsilon^2$ and

$$E_5 = \left\{X, Y \in E : |X-Y| \leq \left(\frac{M\log n}{np}\right)^{1/2}\right\}.$$

The Chebyshev inequality implies that w.h.p. $|E_5| = O((\varepsilon^2 p)^{-1} n \log n)$. Let E_1, E_2 be as in Section 2.

Suppose first that X is at least $\left(\frac{M\log n}{np}\right)^{1/2}$ from the boundary.

$$\mathbb{P}\left(\nexists Y(i,X) \text{ or } |X - Y(i,X)| \geq \left(\frac{M\log n}{np}\right)^{1/2}\right) \leq \left(1 - \frac{\pi\varepsilon M\log n}{2np}\right)^{n-1} = o(n^{-2}). \tag{25}$$

If X is closer to the boundary, then (25) may not be true for some i. This could be the case where X,Y are both close to ∂D and $\{X,Y\} \notin E_1$. In these cases we merge $K(i,X)$ with $K(i+1,X)$ or $K(i-1,X)$, which ever is appropriate. The net effect is to replace ε by 2ε for this particular cone. This has negligible effect.

Assume that $|A-B| \geq \left(\frac{M\log n}{np}\right)^{1/2}$. It follows that w.h.p. we can go $A \to Y(i,B) \to Y(i,Y(i,B)) \to \cdots$, at least until we are at Z within $\left(\frac{M\log n}{np}\right)^{1/2}$ of B. Then with probability at least $1 - \left(1 - \frac{\pi M\log n}{4n}\right)^{n-2}$ we can find a path ZXB from Z to B using only edges from E_1, of total length at most $2\left(\frac{M\log n}{np}\right)^{1/2}$. The length of the path from A to B is then at most

$$\left(1+\frac{\varepsilon}{2}\right)|A-B| + 2\left(\frac{M\log n}{np}\right)^{1/2} = |A-B|\left(1+\frac{\varepsilon}{2}+\frac{2}{M}\right) \leq (1+\varepsilon)d_{A,B}.$$

The validity of the first term $\left(1+\frac{\varepsilon}{2}\right)|A-B|$ follows from Lemma 6.

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

We have considered a Euclidean version, asking for a $(1+\varepsilon)$-spanner and random geometric graphs. We could perhaps extend the results of Theorems 1, 2,3 to $[0,1]^d, d \geq 3$. This does not seem impossible. 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.

There are a number of related questions one can tackle:

1. We could replace edge lengths by E_2^s where $s < 1$. This would allow us to generalise edge lengths to distributions with a density f for which $f(x) \approx x^{1/s}$ as $x \to 0$. This is a more difficult case than $s = 1$ and it was considered by Bahmidi and van der Hofstadt [3]. They prove that w.h.p. $d_{1,2}$ grows like $\frac{n^s}{\Gamma(1+1/s)^s}$ where Γ denotes Euler's Gamma function. The analysis is more complex than that of [10] and it is not clear that our proof ideas can be generalised to handle this situation.

2. Can we reduce the reliance on ε in the upper bound to ε^{-1}?

References

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

[2] N. Alon and J. Spencer, The Probabilistic Method, Third Edition, Wiley and Sons, 2008

[3] S. Bahmidi and R. van der Hofstadt, Weak disorder asymptotics in the stochastic mean-field model of distance, *Annals of Applied Probability* 22 (2012) 29-69.

[4] A note on two problems in connexion with graphs, *Numerische Mathematik* 1 (1959) 269-271.

[5] A.M. Frieze, The effect of adding randomly weighted edges.

[6] A.M. Frieze and M. Karoński, Introduction to Random Graphs, Cambridge University Press, 2015.

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

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

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

[10] S. Janson, One, two and three times $\log n/n$ for paths in a complete graph with random weights, *Combinatorics, Probability and Computing* 8 (1999) 347-361.

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

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

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

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