Spanners in randomly weighted graphs: independent edge lengths

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 \( d'_{v,w} \leq td_{v,w} \) for all \( v, w \in V \). We show that for a large class of graphs with suitable degree and expansion properties with independent exponential mean one edge lengths, there is w.h.p. a 1-spanner that uses \( \approx \frac{1}{2}n \log n \) edges and that this is best possible. In particular, our result applies to the random graphs \( G_{n,p} \) for \( np \gg \log n \).

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 \( d'_{v,w} \leq td_{v,w} \) 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.

Suppose that \( G = ([n], E) \) is almost regular in that
\[
(1 - \theta)dn \leq \delta(G) \leq \Delta(G) \leq (1 + \theta)dn
\]
where \( 1 \geq d \gg \frac{\log \log n}{\log^{1/2} n} \) and \( \theta = \frac{1}{\log^{1/2} n} \). Here \( \delta, \Delta \) refer to minimum and maximum degree respectively.

We will also assume either that \( d > 1/2 \) or
\[
|E(S, T)| \geq \psi |S| |T| \text{ for all } |S|, |T| \geq \theta n.
\]
Here \( \psi = \frac{\omega \log \log n}{\log^{1/2} n} \leq d \) where \( \omega = \omega(n) \to \infty \) as \( n \to \infty \) and \( E(S, T) \) denotes the set of edges of \( G \) with one end in \( S \subseteq [n] \) and the other end in \( T \subseteq [n], S \cap T = \emptyset \).

Let \( \mathcal{G}(d) \) denote the set of graphs satisfying the stated conditions, (1) and (2). We observe that \( K_n \in \mathcal{G}(1) \) and that w.h.p. \( G_{n,p} \in \mathcal{G}(p) \), as long as \( np \gg \log n \). The weighted perturbed model of Frieze [5] where randomly weighted edges are added to a randomly weighted \( dn \)-regular graph also lies in \( \mathcal{G}(d) \).

∗Research supported in part by NSF grant DMS1952285
†Research supported in part by NSF grant DMS1363136
Suppose that the edges \{i, j\} of \(G\) are given independent lengths \(\ell_{i,j}, 1 \leq i < j \leq n\) that are distributed as the exponential mean one random variable, denoted by \(E(1)\). In general we let \(E(\lambda)\) denote the exponential random variable with mean \(1/\lambda\).

When \(G = K_n\), Janson [9] proved the following: W.h.p. and in expectation
\[
d_{1,2} \approx \frac{\log n}{n}; \quad \max_{j>1} d_{1,j} \approx \frac{2\log n}{n}; \quad \max_{i,j} d_{i,j} \approx \frac{3\log n}{n}.
\]
Here (i) \(A_n \approx B_n\) if \(A_n = (1+o(1))B_n\) and (ii) \(A_n \gg B_n\) if \(A_n/B_n \to \infty\), as \(n \to \infty\).

It follows that w.h.p. the length of the longest edge in any shortest path is at most \(L = (3+o(1))\log n\). It follows further that w.h.p. if we let \(E'\) denote the set of edges of length at most \(L\) then this is a 1-spanner of size \(O(n\log n)\).

Theorem 1. Let \(G \in G(d)\) or let \(G\) be a \(dn\)-regular graph with \(d > 1/2\) where the lengths of edges are independent exponential mean one. The following holds w.h.p.

(a) The minimum size of a 1-spanner is asymptotically equal to \(\frac{1}{2}n \log n\).

(b) If \(2 \leq \lambda = O(1)\) then a \(\lambda\)-spanner requires at least \(\frac{n\log n}{601d\lambda}\) edges.

A companion paper deals with \((1+\varepsilon)\)-spanners in embeddings of \(G_{n,p}\) in \([0,1]^2\) as studied by Frieze and Pegden [7]. Here we choose \(n\) random points \(X = \{X_1, X_2, \ldots, X_n\}\) in \([0,1]^2\) and connect a pair \(X_i, X_j\) with probability \(p\) by an edge of length \(|X_i - X_j|\).

2 Proof of Theorem 1

The proof of Theorem 1 uses a few parameters. We will list some of them here for easy reference:
\[
\theta = \frac{1}{\log^{1/2} n}; \quad k_0 = \log n; \quad k_1 = \theta n; \quad \alpha = 1 - 2\theta.
\]
\[
\ell_0 = \frac{(1+\theta) \log n}{dn}; \quad \ell_1 = \frac{5 \log n}{dn}; \quad \ell_2 = \ell_0 - \frac{(\log \log n)^2}{dn}; \quad \ell_3 = \frac{\log n}{200\lambda dn}.
\]

We also use the Chernoff bounds for the binomial \(B(n, p)\): for \(0 \leq \varepsilon \leq 1\),
\[
\mathbb{P}(B(n, p) \leq (1 - \varepsilon)np) \leq e^{-\varepsilon^2 np/2}.
\]
\[
\mathbb{P}(B(n, p) \geq (1 + \varepsilon)np) \leq e^{-\varepsilon^2 np/3}.
\]
\[
\mathbb{P}(B(n, p) \geq \alpha np) \leq \left(\frac{e}{\alpha}\right)^{\alpha np}.
\]

It will only be in Section 2.2 that we will need to use condition [2].

2.1 Lower bound for part (a)

We identify sets \(X_v\) (defined below) of size \(\approx \log n\) such that w.h.p. a 1-spanner must contain \(X_v\) for \(n - o(n)\) vertices \(v\). The sets \(X_v\) are the edges from \(v\) to its nearest neighbors. If an edge \(\{v, x\}\) is missing from a set \(S \subseteq E(K_n)\) then a path from \(v\) to \(x\) must go to a neighbor \(y\) of \(v\) and then traverse \(K_n - v\) to reach \(x\). Such a path is likely to have length at least the distance promised by [3], scaled by \(d^{-1}\).

We first prove the following:
Lemma 2. Fix $v, w_1, w_2, \ldots, w_\ell$ for $\ell = O(\log n)$ and let $\alpha = 1 - 2\theta$. Then,

$$\mathbb{P}\left( \exists 1 \leq i \leq \ell : d_{v, w_i} \leq \frac{\alpha \log n}{dn} \right) = o(1).$$

Proof. There are at most $((1 + \theta)dn)^{k-1}$ paths using $k$ edges that go from vertex $v$ to vertex $w_i, 1 \leq i \leq \ell$. The random variable $E(1)$ dominates the uniform $[0, 1]$ random variable $U_1$. We write this as $E(1) \succ U_1$. As such we can couple each edge weight with a lower bound given by a copy of $U_1$. The length of one of these $k$-edge paths is then at least the sum of $k$ independent copies of $U_1$. The fraction $xk/k!$ is an upper bound on the probability that this sum is at most $x$ (tight if $x \leq 1$). Therefore,

$$\mathbb{P}\left( \exists 1 \leq i \leq \ell : d_{v, w_i} \leq x = \frac{\alpha \log n}{dn} \right) \leq \ell \sum_{k=1}^{n-1} ((1 + \theta)dn)^{k-1} \frac{x^k}{k!} \leq \frac{\ell}{dn} \sum_{k=1}^{10\log n} \left( \frac{e^{1+\theta} \alpha \log n}{k} \right)^k + O(n^{-10}) \leq \frac{10\ell \log n}{dn^{1-\alpha e^\theta}} + o(1) = o(1).$$

For a vertex $v \in [n]$, let

$$A_v = \left\{ w \neq v : \ell_{v, w} \leq \frac{\log n}{dn} \right\}.$$

Lemma 3. W.h.p. $|A_v| \leq 4 \log n$ for all $v \in [n]$.

Proof. We have, from the Chernoff bounds and $E(1) \succ U_1$ that

$$\mathbb{P}(|A_v| \geq 4 \log n) \leq \mathbb{P}\left( \text{Bin } \left( (1 + \theta)dn, \frac{\log n}{dn} \right) \geq 4 \log n \right) \leq \left( \frac{e(1 + \theta)}{4} \right)^{4 \log n} = o(n^{-1}).$$

The lemma follows from the union bound, after multiplying the RHS of (5) by $n$. \qed

For $v \in [n]$, let $\delta_v$ be the distance from $v$ to its nearest neighbor. Let

$$B = \left\{ v : \delta_v \geq \frac{\log^{1/2} n}{dn} \right\}.$$

Lemma 4. $|B| \leq ne^{-\log^{1/3} n}$ w.h.p.

Proof. We have

$$\mathbb{E}(|B|) \leq n \left( \exp \left\{ -\frac{\log^{1/2} n}{dn} \right\} \right)^{(1-\theta)dn} = ne^{-(1-\theta)\log^{1/2} n}.$$

The lemma follows from the Markov inequality. \qed

Let

$$X_v = \left\{ e = \{v, x\} : \ell(e) \leq \delta_v + \frac{\alpha \log n}{dn} \right\}.$$
Lemma 5. Let $S \subseteq E(K_n)$ define a 1-spanner. Then w.h.p. $S \supseteq X_v$ for all but $o(n)$ vertices $v$.

Proof. Let $G_S = ([n], S)$ and suppose that $v \notin B$. Then
\[
\delta_v + \frac{\alpha \log n}{dn} < \frac{\log^{1/2} n}{dn} + \frac{\alpha \log n}{dn} < \frac{\log n}{dn}
\]
and so $X_v \subseteq \{v\} \times A_v$ and in particular $|X_v| \leq 4 \log n$ w.h.p. by Lemma 3.

If $G_S$ does not contain an edge $e = \{v, x\} \in X_v$, then the $G_S$-distance from $v$ to $x$ is then w.h.p. at least
\[
\delta_v + \frac{\alpha \log n}{dn} > d_{v,x}.
\]

To obtain (7) we have used Lemma 2 applied to $K_n - v$ with $x$ replacing $v$ and $w_1, w_2, \ldots, w_{\ell}$ being the remaining neighbors of $v$ in $K_n$.

So, if
\[ C = \{v \notin B : \exists 1\text{-spanner } S \supseteq X_v\} \]
then $\mathbb{E}(|C|) = o(n)$.

Any 1-spanner must contain $X_v$, $v \in [n] \setminus (B \cup C)$ and the lemma follows from the Markov inequality. \hfill \square

Now $|X_v|$ dominates $Bin \left((1 - \theta)dn, 1 - \exp \left\{-\frac{\alpha \log n}{dn}\right\}\right)$ and so by the Chernoff bounds
\[
\mathbb{P}\left(|X_v| \leq (1 - \varepsilon)\alpha \log n + O\left(\frac{\log^2 n}{n}\right)\right) \leq e^{-\varepsilon^2 \alpha \log n/(2 + o(1))} = o(1) \text{ for } \varepsilon = \log^{-1/3} n.
\]

Applying Lemma 5 we see that w.h.p. a 1-spanner contains at least $\frac{1-o(1)}{2} n \log n$ edges. The factor 2 comes from the fact that $\{v, w\}$ can be in $X_v \cap X_w$. (In this case the edge $\{v, w\}$ contributes twice to the sum of the $|A|_{v'}$'s.) Note that we do not need (2) to prove the lower bound.

2.2 Upper bound for part (a)

Let $\ell_0 = \frac{\sqrt{n} \log n}{dn}$ and $\ell_1 = \frac{5 \log n}{n}$ and $E_0 = \{e : \ell(e) \leq \ell_0\}$. Now $|E(G)| \in (1 \pm \theta)dn^2/2$ and so the Chernoff bounds imply that w.h.p. $|E_0| \approx \frac{1}{2} n \log n$ and our task is to show that adding $o(n \log n)$ edges to $E_0$ gives us a 1-spanner w.h.p. We will do this by showing that w.h.p. there are only $o(n \log n)$ edges $e$ with $\ell(e) > \ell_0$ that are the shortest path between their endpoints. Adding these $o(n \log n)$ edges to $E_0$ creates a 1-spanner, since every edge on a shortest path in a graph is itself a shortest path between its endpoints.

Janson [9] analysed the performance of Dijkstra’s [4] algorithm on the complete graph $K_n$ with exponential edge-weights; we will adapt his argument to our setting on a graph $G$ satisfying conditions (1) and (2).

In particular, we analyze Dijkstra’s algorithm for shortest paths from vertex 1 where edges have exponential weights. Recall that after $i$ steps of the algorithm we have a tree $T_i$ and a set of values $d_v, v \in [n]$ such that for $u \in T_i$, $d_u$ is the length of the shortest path from 1 to $u$. For $v \notin T_i$, $d_v$ is the length of the shortest path from 1 to $v$ that follows a path from 1 to $u \in T_i$ and then uses the edge $\{u, v\}$. Let $\delta_i = \max \{v \in T_i : d_v\}$.

The constraints on the length $l(u, v)$ of the edge $\{u, v\}$ for $u \in T_i, v \notin T_i$ are that $d_u + l(u, v) \geq \delta_i$ or equivalently that $l(u, v) \geq \delta_i - d_u$. Fixing $T_i$ and the lengths of edges within $T_i$ or its complement, every set of lengths $\{l(u, v)\}_{u \in T_i, v \notin T_i}$ satisfying these constraints would give the same history of the algorithm to this point.

Due to the memoryless property of the exponential distribution we then have that $l(u, v) = \delta_i - d_u + E_{u,v}$ where $E_{u,v}$ is a mean-1 exponential, independent of all other $E(u', v')$.

Thus the Dijkstra algorithm is equivalent in distribution to the following discrete-time process:
• Set \( v_1 = 1, T_1 = \{1\} \).

• Having defined \( T_i \), associate a mean-1 exponential \( E_{u,v} \) to each edge \( \{u,v\} \in E(T_i, \bar{T}_i) \) that is independent of the process to this point. Define \( e_{i+1} \) to be the edge \( \{u,v\} \in E(T_i, \bar{T}_i) \) minimizing \( \delta_i + E_{u,v} \), and define \( v_{i+1} \) to be the vertex for which \( e_{i+1} = \{v_j, v_{i+1}\} \) for some \( v_j \in T_i \). Finally define \( d_{v_{i+1}} \) by \( \delta_i + E_{u,v} \).

Finally, note that, as the minimum of \( r \) rate-1 exponentials is an exponential of rate \( r \), this is equivalent in distribution to the following process:

• Set \( v_1 = 1, T_1 = \{1\} \).

• Having defined \( v_i, T_i \), define a vertex \( v_{i+1} \) by choosing an edge \( e_{i+1} = \{v_j, v_{i+1}\} \) (\( j \leq i \)) uniformly at random from \( E(T_i, \bar{T}_i) \), set \( T_{i+1} = T_i \cup \{v_{i+1}\} \), and define \( d_{1,v_{i+1}} = d_{1,v_i} + \gamma_i \) where \( \gamma_i \) is an (independent) exponential random variable of rate \( \gamma_i = E(T_i, \bar{T}_i) \).

It follows that
\[
\mathbb{E}(d_{1,m}) = S_m := \sum_{i=1}^{m-1} \mathbb{E}\left( \frac{1}{\gamma_i} \right) \quad \text{and} \quad \mathbb{V}(d_{1,m}) = \sum_{i=1}^{m-1} \mathbb{E}\left( \frac{1}{\gamma_i^2} \right).
\]

Observe that we have
\[
(1-\theta)i(dn-i) \leq \gamma_i \leq (1+\theta)idn \quad \text{w.h.p.}
\]
and so for \( 1 \leq i \leq \theta n \) we have
\[
\gamma_i = idn(1+\zeta_i) \quad \text{where} \quad |\zeta_i| = O(\theta) \quad \text{w.h.p.}
\]
Also, we have
\[
\gamma_i = (n-i)dn(1+\zeta_i) \quad \text{where} \quad |\zeta_i| = O(\theta) \quad \text{w.h.p.}
\]
for \( n-\theta n \leq i \leq n \).
It follows that
\[
S_m = (1+O(\theta)) \sum_{i=1}^{\theta n} \frac{1}{dn} = \frac{\log n}{dn} + O\left( \frac{\log^{1/2} n}{n} \right) \quad \text{w.h.p.} \tag{8}
\]

**Lemma 6.** W.h.p. \( \max_{i,j} d_{i,j} \leq \ell_1 = \frac{5\log n}{dn} \).

**Proof.** Following [9], let \( k_1 = \theta n \) and \( Y_i = E_i^{\gamma_i}, 1 \leq i < n \) so that \( Z_1 = d_{1,k_1} = Y_1 + Y_2 + \cdots + Y_{k_1-1} \). For \( t < 1 - \frac{1+o(1)}{dn} \) we have implies that w.h.p. for \( m = k_1 - 1 \),
\[
\mathbb{E}(e^{tdnZ_1}) = \mathbb{E}\left( \prod_{i=1}^{m} e^{tdnY_i} \right) = \sum_x \mathbb{E}\left( \prod_{i=1}^{m} e^{tdnY_i} \mid \gamma_m = x \right) \mathbb{P}(\gamma_m = x) = \mathbb{E}\left( \prod_{i=1}^{m-1} e^{tdnY_i} \right) \sum_x \mathbb{E}(e^{tdY_m} \mid \gamma_m = x) \mathbb{P}(\gamma_m = x) = \mathbb{E}\left( \prod_{i=1}^{m-1} e^{tdnY_i} \right) \sum_x \frac{x}{x-tdn} \mathbb{P}(\gamma_m = x) = \mathbb{E}\left( \prod_{i=1}^{m-1} e^{tdnY_i} \right) \left( 1 - \frac{(1+o(1))t}{i} \right)^{-1}.
\]

Here the term in (9) stems from the fact that given \( \gamma_m, Y_m \) is independent of \( Y_1, Y_2, \ldots, Y_{m-1} \).
Then for any \( \beta > 0 \) we have
\[ P\left( Z_1 \geq \frac{\beta \log n}{d_n} \right) \leq E(e^{t dn Z_1 - t^2 \log n}) \leq e^{-t^2 \log n} \prod_{i=1}^{k_1-1} \left( 1 - \frac{(1 + o(1))t}{i} \right)^{-1} \]
\[ = e^{-t^2 \log n} \exp \left\{ \sum_{i=1}^{k_1-1} \left( \frac{(1 + o(1))t}{i} + O\left( \frac{1}{i^2} \right) \right) \right\} = \exp \left\{ (1 + o(1) - \beta) t \log n \right\}. \]

It follows, on taking \( \beta = 2 + o(1) \) that w.h.p.
\[ d_{j,k_1} \leq \frac{(2 + o(1)) \log n}{dn} \text{ for all } j \in [n]. \]

Letting \( \hat{T}_{k_1} \) be the set corresponding to \( T_{k_1} \) when we execute Dijkstra’s algorithm starting at vertex 2. First consider the case where \( d \leq 1/2 \) and (2) holds. Then, using (2), we have that either \( T_{k_1} \cap \hat{T}_{k_1} \neq \emptyset \) or,
\[ P\left( \exists e \in T_{k_1} : \hat{T}_{k_1} : X(e) \leq \frac{1}{n} \right) \leq \exp \left\{ -\psi^2 n^2 \right\} = o(n^{-2}) \]

This shows that we fail to find a path of length \( \leq \frac{(4 + o(1)) \log n}{dn} + \frac{1}{n} \) between a fixed pair of vertices with probability \( o(n^2) \). In particular, taking a union bound over all pairs of vertices, we obtain that w.h.p. \( \max_{i,j} d_{i,j} \leq \frac{(4 + o(1)) \log n}{dn} + \frac{1}{n} \).

If \( G \) has \( \delta(G) \geq (1 - \tau)dn \) with \( d = 1/2 + \varepsilon, \varepsilon > 0 \) constant, then any pair of vertices has at least \( (2\varepsilon - 2\theta)n \) common neighbors. We pair up the vertices of \( T_{k_1} \) and bound the probability that we cannot find a path of length 2 whose endpoints consist of one of our pairs, and which uses only edges of length at most \( \frac{\log n}{n \log \log n} \), as
\[ \left( e^{-\left( \frac{\log n}{n \log \log n} \right)^2} \right)^{-\theta n(2\varepsilon - 2\theta)n} = o(n^{-2}). \]

Again we are done by a union bound over possible pairs. \( \square \)

We now consider the probability that a fixed edge \( e \) satisfies that \( \ell(e) > \ell_0 \) and that \( e \) is a shortest path from 1 to \( n \).

**Lemma 7.** Let \( E(e) \) denote the event that \( \ell(e) > \ell_0 \) and \( e \) is a shortest path from 1 to \( n \).
\[ P\left( E \mid \max_j d_{1,j} \leq \ell_1 \right) = o\left( \frac{\log n}{n} \right). \]

**Proof.** Without loss of generality we write \( e = \{1, n\} \). If \( E = E(e) \) occurs then we have the occurrence of the event \( F \) where
\[ F = \{d_{1,m} + \ell(f_m) \geq \ell(e), m = 2, 3, \ldots, n - 1\} \]
and \( f_m \) denotes the edge joining vertex \( n \) to the vertex whose shortest distance from vertex 1 (in \( G - \{n\} \)) is the \( m \)th smallest. (If the edge does not exist then \( \ell(f_m) = \infty \) in the calculation below.) Indeed this follows from Dijkstra’s algorithm; the event \( F \) indicates that at every step of the algorithm, no path shorter than the edge \( \{1, n\} \) is found.

Let \( n_0 = n(1 - d/2) \). We need \( \ell(f_m) + d_m \geq \xi = \ell(e) \) for all \( m \) in order that \( F \) occurs. If \( d_{1,n_0} = x \) then this is implied by \( \bigcap_{m=1}^{n_0} \{ \ell(f_m) \geq \xi - x \} \). Using the independence of the \( \ell(f_m) \) and \( d_{1,i}, i = 2, \ldots, n_0 \), we bound
\[ P(F \mid \max_{1,j} d_{1,j} \leq \ell_1) \leq \frac{1}{P(\max_j d_{1,j} \leq \ell_1)} \int_{\xi = \ell_0}^{\ell_1} e^{-\xi} \int_{x=0}^{\infty} \mathbb{P} \left( \bigcap_{m=1}^{n_0} \{ \ell(f_m) \geq \xi - x \} \right) d\mathbb{P}\{d_{1,n_0} = x\} d\xi \]
and using the fact that there are at least \( dn/2 - 1 \) indices \( m \) for which \( \ell(f_m) < \infty \) we bound
\[ \mathbb{P}(\mathcal{F} \mid \max_{i,j} d_{1,j} \leq \ell_1) \leq (1 + o(1)) \int_{\xi=\ell_0}^{\ell_1} \int_0^\infty \min \{ 1, e^{-dn(\xi-x)/3} \} \, d\mathbb{P} \{ d_{1,n_0} = x \} \, d\xi. \]  

(12)

Now, if \( \ell_2 = \ell_0 - \frac{(\log \log n)^2}{dn} \) then

\[ \int_{\xi=\ell_0}^{\ell_1} \int_0^{\ell_2} \min \{ 1, e^{-dn(\xi-x)/3} \} \, d\mathbb{P} \{ d_{1,n_0} = x \} \, d\xi \leq \ell_1 \exp \left\{ -\frac{(\log \log n)^2}{3} \right\} = o \left( \frac{\log n}{n} \right). \]  

(13)

It remains to bound the same expression where the second integral goes from \( x = \ell_2 \) to \( \infty \).

First consider the case where \( d \leq 1/2 \) and (2) holds. We have from (8) that

\[ \mathbb{E}(d_{1,n_0}) = S_{n_0} \leq (1 + O(\theta)) \sum_{i=1}^{\theta n} \frac{1}{dn} + \sum_{i=\theta n+1}^{n_0} \frac{1}{\psi_i(n-i)} \]  

\[ \leq \frac{(1 + O(\theta)) \log n}{dn} + \frac{1}{\psi(n)} \sum_{i=\theta n+1}^{n_0} \left( \frac{1}{i} + \frac{1}{n-i} \right) = \frac{(1 + O(\theta)) \log n}{dn} + O \left( \frac{\log \log n}{n} \right) \]

\[ = \frac{\log n}{dn} + O \left( \frac{\log^{1/2} n}{n} \right) < \ell_2 = \frac{\sqrt{\theta}}{2dn} \]

and

\[ \text{Var}(d_{1,n_0}) \leq (1 + O(\theta)) \sum_{i=1}^{\theta n} \frac{1}{dn^2 i^2} + \sum_{i=\theta n+1}^{n_0} \frac{1}{\psi_i^2(n-i)^2} \leq \frac{\pi^2}{3d^2n^2}. \]  

(15)

Chebyshev’s inequality then gives that

\[ \mathbb{P}(d_{1,n_0} \geq S_{n_0} + x) \leq \frac{\pi^2}{3d^2x^2n^2}. \]

As a consequence of this we see that

\[ \int_{\xi=\ell_0}^{\ell_1} \int_0^{\ell_2} \min \{ 1, e^{-dn(\xi-x)/3} \} \, d\mathbb{P} \{ d_{1,n_0} = x \} \, d\xi \leq \frac{\ell_1\pi^2}{3d^2(\ell_2 - S_{n_0})^2n^2} \leq \frac{2\ell_1\pi^2}{\ell_2 \theta \log^2 n} = O \left( \frac{1}{n \log^{1/2} n} \right). \]  

(16)

The lemma follows for \( d \leq 1/2 \), from (13) and (16) and the Markov inequality.

When \( d > 1/2 \) we can replace the second sum in (14) by

\[ \sum_{i=\theta n+1}^{n_0} \frac{1}{\varepsilon n \min \{ i, n-i \}} = O \left( \frac{1}{n \log n} \right), \quad \text{where } \varepsilon = d - \frac{1}{2}. \]

By the same token, the second sum in (15) will be \( o(n^{-2}) \). The remainder of the proof will go as for the case \( d \leq 1/2 \).

Together with Lemma 6, Lemma 7 implies that w.h.p. the number of edges \( e \) for which \( \mathcal{E}(e) \) occurs is \( o(n \log n) \). Adding these to \( E_0 \) gives us a 1-spanner of size \( \approx \frac{1}{2} n \log n \).

2.3 Lower bound for part (b)

**Lemma 8.** Fix a set \( A \) such that \( |A| \leq a_0 = O(\log n) \). Let \( \mathcal{P} \) be the event that there exists a path \( P \) of length at most \( \ell_4 = \frac{\log n}{200dn} \) joining two distinct vertices of \( A \). Then \( \mathbb{P}(\mathcal{P}) = O(n^{o(1)-199/200}) \).
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 Hofstad [3]. They prove that w.h.p. $d_{1,2}$ grows like $\frac{n^{s}}{\Gamma(1+1/s)^{s}}$, where $\Gamma$ denotes Euler’s Gamma function. The analysis is more complex than that of [9] and it is not clear that our proof ideas can be generalised to handle this situation.

2. The results of Theorem 1 apply to $G_{n,p}$. It would be of some interest to consider other models of random or quasi-random graphs.

### 3 Summary and open questions

We have determined the asymptotic size of the smallest 1-spanner when the edges of a dense (asymptotically) regular graph $G$ are given independent lengths distributed as $E_2$, modulo the truth of [2] or the degree being $dn, d > 1/2$.

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 Hofstad [3]. They prove that w.h.p. $d_{1,2}$ grows like $\frac{n^{s}}{\Gamma(1+1/s)^{s}}$, where $\Gamma$ denotes Euler’s Gamma function. The analysis is more complex than that of [9] and it is not clear that our proof ideas can be generalised to handle this situation.

2. The results of Theorem 1 apply to $G_{n,p}$. It would be of some interest to consider other models of random or quasi-random graphs.
References