Traveling in Randomly Embedded Random Graphs

Alan Frieze[1] and Wesley Pegden[2]

1 **Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA, USA**
2 **Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA, USA**

Abstract

We consider the problem of traveling among random points in Euclidean space, when only a random fraction of the pairs are joined by traversable connections. In particular, we show a threshold for a pair of points to be connected by a geodesic of length arbitrarily close to their Euclidean distance, and analyze the minimum length Traveling Salesperson Tour, extending the Beardwood-Halton-Hammersley theorem to this setting.

1998 ACM Subject Classification G.2.1 Combinatorics, G.3 Probability and Statistics

Keywords and phrases Traveling Salesman, Euclidean, Shortest Path

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2017.45

1 Introduction

The classical Beardwood-Halton-Hammersley theorem [1] (see also Steele [12] and Yukich [13]) concerns the minimum cost Traveling Salesperson Tour through n random points in Euclidean space. In particular, it guarantees the existence of an absolute (though still unknown) constant β_d such that if $X_1, X_2 \dots,$ is a random sequence of points, uniformly distributed in the d-dimensional cube $[0,1]^d$, the length $T(\mathcal{X}_{n,1})$ of a minimum tour through $X_1, \dots, X_n$ satisfies

$$T(\mathcal{X}_{n,1}) \sim \beta_d n^{\frac{d-1}{d}} \ a.s. \tag{1}$$

The present paper is concerned still with the problem of traveling among random points in Euclidean space. In our case, however, we suppose that only a (random) subset of the pairs of points are joined by traversable connections, independent of the geometry of the point set.

In particular, we study random embeddings of the Erdős-Rényi-Gilbert random graph $G_{n,p}$ into the d-dimensional cube $[0,1]^d$. We let $\mathcal{X}_n$ denote a uniformly random set of points $X_1, X_2, \dots, X_n \in [0,1]^d$, and we denote by $\mathcal{X}_{n,p}$ the random graph whose vertex set is $\mathcal{X}_n$ and whose pairs of vertices are joined by edges each with independent probability p. Edges are weighted by the Euclidean distance between their points, and we are interested in the total edge-weight required to travel about the graph.

This model has received much less attention than the standard model of a random geometric graph, defined as the intersection graph of unit balls with random centers $X_i, i \in [n]$, see Penrose [9]. We are only aware of the papers by Mehrabian [7] and Mehrabian and Wormald [8] who studied the *stretch factor* of $\mathcal{X}_{n,p}$. In particular, let $||x-y||$ denote the Euclidean distance between vertices x, y, and $\text{dist}(x,y)$ denote their distance in $\mathcal{X}_{n,p}$. They showed (considering the case $d=2$) that unless p is close to 1, the stretch factor

$$\sup_{x,y\in\mathcal{X}_{n,p}} \frac{\text{dist}(x,y)}{||x-y||}$$

tends to ∞ with n.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2017).
Editors: Klaus Jansen, José D. P. Rolim, David Williamson, and Santosh S. Vempala; Article No. 45; pp. 45:1–45:17
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

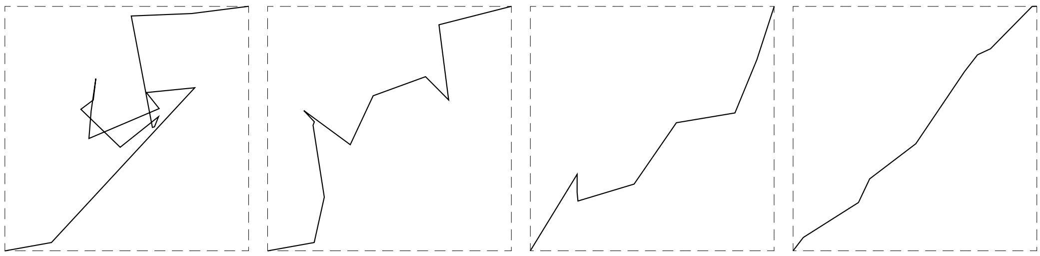

Figure 1 Paths in an instance of $\mathcal{X}_{n,p}$ for $d = 2$, $n = 2^{30}$, and $p = \frac{10}{n}, \frac{25}{n}, \frac{50}{n}$, and $\frac{200}{n}$, respectively. In each case, the path drawn is the shortest route between the vertices x and y which are closest to the SW and NE corners of the square. (See Q. 2, Section 5.)

As a counterpoint to this, our first result shows a very different phenomenon when we pay attention to additive rather than multiplicative errors. In particular, for $p \gg \frac{\log^d n}{n}$, the distance between a typical pair of vertices is arbitrarily close to their Euclidean distance, while for $p \ll \frac{\log^d n}{n(\log\log n)^{2d}}$, the distance between a typical pair of vertices in $\mathcal{X}_n$ is arbitrarily large (Figure 1). (We write $\log^k x$ for $(\log x)^k$.) In particular, this means that when $\frac{\log^d n}{n} \ll p < 1 - \varepsilon$, the supremum in the stretch factor theorem of Mehrabian and Wormald is due just to pairs of vertices which are very close together.

▶ **Theorem 1.** *Let $\omega = \omega(n) \to \infty$. We have for $d \geq 2$:*

(a) *For $p \leq \frac{1}{\omega^d(\log\log n)^{2d}} \frac{\log^d n}{n}$ and fixed $u = X_1$, $v = X_2$, say, we have*

$$\operatorname{dist}(u, v) \geq \frac{\omega}{8de^d} \qquad a.a.s.^1$$

(b) *For $p \geq \frac{\omega \log^d n}{n}$, we have a.a.s. that uniformly for all pairs of vertices $u, v \in \mathcal{X}_n$,*

$$\operatorname{dist}(u, v) = ||u - v|| + o(1).$$

Theorem 1 means that, even for p quite small, it is not that much more expensive to travel from one vertex of $\mathcal{X}_{n,p}$ to another than it is to travel directly between them in the plane. On the other hand, there is a dramatic dependence on p if the goal is to travel among *all* points. Let $T(\mathcal{X}_{n,p})$ denote the length of a minimum length tour in $\mathcal{X}_{n,p}$ hitting every vertex exactly once, i.e. a Traveling Salesperson tour.

▶ **Theorem 2.** *There exists a sufficiently large constant $K > 0$ such that for all $p = p(n)$ such that $p \geq \frac{K \log n}{n}$, $d \geq 2$, we have that*

$$T(\mathcal{X}_{n,p}) = \Theta\left(\frac{n^{\frac{d-1}{d}}}{p^{1/d}}\right) \qquad a.a.s. \tag{2}$$

(Recall that $f(n) = \Theta(g(n))$ means that $f(n)$ is bounded between positive constant multiples of $g(n)$ for sufficiently large n.) As the threshold for $G_{n,p}$ to be Hamiltonian is at $p = \frac{\log n + \log\log n + \omega(n)}{n}$ (see e.g. Bollobás [2]), this theorem covers nearly the entire range of p for which a TSP tour exists a.a.s.

Finally, we extend the asymptotically tight BHH theorem [1] to the case of $\mathcal{X}_{n,p}$ for any constant p. To formulate an "almost surely" statement, we let $\mathcal{X}_{\mathcal{N},p}$ denote a random graph on a random embedding of $\mathcal{N} = \{1, 2, \ldots, \}$ into $[0,1]^d$, where each pair $\{i, j\}$ is independently present as an edge with probability p, and consider $\mathcal{X}_{n,p}$ as the restriction of $\mathcal{X}_{\mathcal{N},p}$ to the first n vertices $\{1, \ldots, n\}$.

▶ **Theorem 3.** *If $d \geq 2$ and $p > 0$ is constant, then there exists $\beta_{d,p} > 0$ such that*

$$T(\mathcal{X}_{n,p}) \sim \beta_{d,p} n^{\frac{d-1}{d}} \qquad a.s.$$

Karp's algorithm [6] for a finding an approximate tour through $\mathcal{X}_n$ extends to the case $\mathcal{X}_{n,p}$, p constant as well:

▶ **Theorem 4.** *For fixed $d \geq 2$ and p constant, then there is an algorithm that a.s. finds a tour in $\mathcal{X}_{n,p}$ of value $(1 + o(1))\beta_{d,p} n^{(d-1)/d}$ in polynomial time, for all $n \in \mathcal{N}$.*

2 Traveling between pairs

2.1 Proof of Theorem 1(a)

Outline of proof

This is straightforward. We show by the first moment method that any path between u and v with "many" edges must contain a significant number of "long" edges and hence must be as long as claimed. We then show that a.a.s. there are no paths between u and v without many edges.

Proof proper

Let ν_d denote the volume of a d-dimensional unit ball; recall that ν_d is bounded ($\nu_d \leq \nu_5 < 6$ for all d).

Let an edge be *long* if its length is at least $\ell_1 = \frac{\omega(\log\log n)^2}{4e^d \log n}$. Let $\varepsilon = \frac{1}{\log\log n}$ and let $\mathcal{A}_k$ be the event that there exists a path with k edges, $k \geq k_0 = \frac{\log n}{2d \log\log n}$ from u to v that uses at most εk long edges. Then

$$\mathbf{Pr}(\exists k : \mathcal{A}_k) \leq \sum_{k \geq k_0} (k-1)! \binom{n}{k-1} p^k \binom{k}{\varepsilon k} \left(\nu_d \left(\frac{\omega(\log\log n)^2}{4e^d \log n} \right)^d \right)^{(1-\varepsilon)k} \tag{3}$$

$$\leq \sum_{k \geq k_0} n^{k-1} p^k \left(\frac{e}{\varepsilon}\right)^{\varepsilon k} \left(\nu_d \left(\frac{\omega(\log\log n)^2}{4e^d \log n} \right)^d \right)^{(1-\varepsilon)k} \tag{4}$$

$$\leq \frac{1}{n} \sum_{k \geq k_0} \left(\frac{\nu_d \log^{d\varepsilon} n}{(4e^d)^{d(1-\varepsilon)}} \cdot \left(\frac{e}{\varepsilon}\right)^{\varepsilon} \right)^k \tag{5}$$

$$\leq \frac{1}{n} \sum_{k \geq k_0} \left(\frac{6e^{d+o(1)}}{4e^{d^2}} \right)^k = o(1),$$

after using $d \geq 2$ and $\log^{\varepsilon} n = e$.

Explanation of (3): Choose the $k-1$ interior vertices of the possible path and order them in one of $(k-1)!\binom{n}{k-1}$ ways as $(u_1, u_2, \ldots, u_{k-1})$. Then p^k is the probability that the edges exist in $G_{n,p}$. Now choose the short edges $e_i = (u_{i-1}, u_i), i \in I$ in one of $\binom{k}{(1-\varepsilon)k} = \binom{k}{\varepsilon k}$ ways and bound the probability that these edges are short by $\left(\nu_d \left(\frac{\omega(\log\log n)^2}{4e^d \log n} \right)^d \right)^{(1-\varepsilon)k}$ viz. the probability that u_i is mapped to the ball of radius ℓ_1, center u_{i-1} for $i \in I$.

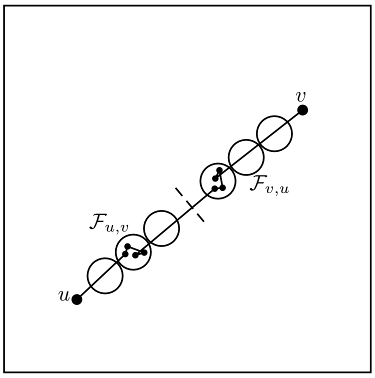

Figure 2 Finding a short path.

Now a.a.s. the shortest path in $G_{n,p}$ from u to v requires at least k_0 edges: Indeed the expected number of paths of length at most k_0 from u to v can be bounded by

$$\sum_{k=1}^{k_0}(k-1)!\binom{n}{k-1}p^k \leq \frac{1}{n}\sum_{k=1}^{k_0}\left(\frac{\log^d n}{\omega^d(\log\log n)^{2d}}\right)^k = o(1).$$

So a.a.s.

$$\text{dist}(u,v) \geq \varepsilon k_0 \ell_1 = \frac{\varepsilon \log n}{2d\log\log n}\cdot\frac{\omega(\log\log n)^2}{4e^d\log n} = \frac{\omega}{8de^d}.$$

◀

2.2 Proof of Theorem 1(b)

Outline of proof

We first consider two points u, v such that $||u-v|| \geq \gamma = \frac{1}{\log\log n}$. We then consider a set of 2β small disjoint balls with centers on the line joining u, v. We argue that a.a.s. (i) all of these balls contain (relatively) giant components, (ii) there is an edge joining the large components inside each ball, (iii) the diameter of each of these giant components is small and (iv) there is an edge between u and one of the g giant components X closest to u and an edge between v and one of the g giant components Y closest to v. This gives a path consisting of an edge from u to the giant component X plus a walk inside X plus an edge to the giant component Y plus an edge to v. Because the balls are small the length of this path is close to $||u-v||$. We reduce the case where $||u-v|| \leq \gamma$ to the first case.

Proof proper

We begin by considering the case of vertices u, v at distance $||u-v|| \geq \gamma$. Letting $\delta = \frac{1}{\log n}$, then, for sufficiently large n, we can find a set $\mathcal{B}$ of at least $\frac{2C}{\delta}$, $C = \frac{\gamma}{8}$, disjoint balls of radius δ centered on the line from u to v, such that $\frac{C}{\delta}$ of the balls are closer to u than v, and $\frac{C}{\delta}$ balls are closer to v than u (Figure 2). Denote these two families of $\frac{C}{\delta}$ balls by $\mathcal{F}_{u,v}$ and $\mathcal{F}_{v,u}$. (The sets $\mathcal{B}$, $\mathcal{F}_{u,v}$ and $\mathcal{F}_{v,u}$ are fixed for the rest of the argument.)

Given a ball $B \in \mathcal{F}_{\{u,v\}} = \mathcal{F}_{u,v} \cup \mathcal{F}_{v,u}$, the induced subgraph G_B on vertices of $\mathcal{X}$ lying in B is a copy of $G_{N,p}$, where $N = N(B)$ is the (random) number of vertices lying in B. Let

$$\mathcal{S}_B \text{ be the event that } N(B) \in \left[\frac{N_0}{2^{d+1}}, 2N_0\right] \text{ where } N_0 = \nu_d \delta^d n.$$

(Dividing by 2^{d+1} accounts for points close to the boundary of $[0,1]^d$.)

Now $N(B)$ is distributed as the binomial $Bin(n,q)$ where $q \in \nu_d\delta^d[2^{-d},1]$. The following Chernoff bounds will thus be useful:

$$\mathbf{Pr}(Bin(M,p) \le (1-\varepsilon)Mp) \le e^{-\varepsilon^2 Mp/2} \text{ for } 0 \le \varepsilon \le 1. \tag{6}$$

$$\mathbf{Pr}(Bin(M,p) \ge (1+\varepsilon)Mp) \le e^{-\varepsilon^2 Mp/3} \text{ for } 0 \le \varepsilon \le 1. \tag{7}$$

The bounds (6) and (7) imply that for $B \in \mathcal{F}_{\{u,v\}}$,

$$\mathbf{Pr}\left(\neg\mathcal{S}_B\right) \le e^{-\Omega(n\delta^d)} = e^{-n^{1-o(1)}}.$$

This gives us that a.a.s. $\mathcal{S}_B$ occurs for all pairs $u,v \in \mathcal{X}$ with $||u-v|| \ge \gamma$. We now argue that for all $B \in \mathcal{B}$:

(A) All subgraphs G_B for $B \in \mathcal{F}_{\{u,v\}}$ have a giant component X_B, containing at least $N_0/2^{d+2}$ vertices.
Indeed, the expected average degree in G_B is $Np = \Omega(\omega) \to \infty$ (and with probability $1-e^{-n^{1-o(1)}}$ we have $N = n^{1-o(1)}$) and at this value the giant component is almost all of B a.a.s. In particular, since $\mathcal{S}_B$ occurs, we have that

$$\mathbf{Pr}(|X_B| \le N_0/2^{d+2} \mid \mathcal{S}_B) \le e^{-\Omega(N_0)} \le e^{-\Omega(\delta^d n)} = o(1). \tag{8}$$

See [2] for the first inequality in (8). This can be inflated by $n^2 \cdot (2C\log n)$ to account for pairs u,v and the choice of $B \in \mathcal{F}_{\{u,v\}}$.

(B) There is an edge between X_B and $X_{B'}$ for all $B,B' \in \mathcal{F}_{\{u,v\}}$.
Indeed, the probability that there is no edge between $X_B, X_{B'}$, given (A), is at most

$$(1-p)^{N_0^2/2^{2d+2}} \le e^{-\Omega(\delta^{2d}n^2p)} \le e^{-n^{1-o(1)}}.$$

This can be inflated by $n^2 \cdot (C\log n)^2$ to account for all pairs u,v and all pairs B,B'.

(C) For each $B \in \mathcal{F}_{\{u,v\}}$, the graph diameter $\text{diam}(X_B)$ (the maximum number of edges in any shortest path in X_B) satisfies

$$\mathbf{Pr}\left(\text{diam}(X_B) > \frac{100\log N_0}{\log(N_0p)}\right) \le n^{-3}. \tag{9}$$

This can be inflated by $n^2 \cdot (2C\log n)$ to account for pairs u,v and the choice of $B \in \mathcal{F}_{\{u,v\}}$. Fernholz and Ramachandran [4] and Riordan and Wormald [11] gave tight estimates for the diameter of the giant component, but we need this cruder estimate with a lower probability of being exceeded. We prove this later in Lemma 5. It will be convenient for the proof of Lemma 5 to assume that $N_0p = O(\log N_0)$. There is no loss in generality because Theorem 1(b) holds a fortiori for larger p. This follows from a standard coupling argument, involving adding random edges to increase the edge probability.

Part (C) implies that with high probability, for any u,v at distance $\ge \gamma$ and all $B \in \mathcal{F}_{\{u,v\}}$ and vertices $x,y \in X_B$,

$$\text{dist}(x,y) \le 200\delta \times \frac{\log N_0}{\log(N_0p)} \le \frac{200}{\log n} \times \frac{\log n - d\log\log n + \log\nu_d}{\log\omega + \log\nu_d} = o(1). \tag{10}$$

As the giant components X_B $(B \in \mathcal{F}_{u,v})$ contain in total at least $\frac{C}{\delta}\frac{N_0}{2^{d+2}} = \frac{C}{2^{d+2}}\nu_d n\delta^{d-1}$ vertices, the probability that u has no neighbor in these giant components is at most

$$(1-p)^{C\nu_d n\delta^{d-1}/2^{d+2}} \le e^{-C\nu_d np\delta^{d-1}/2^{d+2}} = n^{-\omega C\nu_d/2^{d+2}}.$$

In particular, the probability is small after multiplication by n^2, and thus a.a.s., for all pairs $x, y \in X_{n,p}$, x has a neighbor in X_B for some $B \in \mathcal{F}_{u,v}$ and y has a neighbor in $X_{B'}$ for some $B' \in \mathcal{F}_{v,u}$. Now by part (B) and equation (10), we can find a path

$$u, w_0, w_1, \dots, w_s, z_t, z_{t-1}, \dots, z_1, z_0, v$$

from u to v where the w_i's are all in some X_B for $B \in \mathcal{F}_{u,v}$ and the total Euclidean length of the path $w_0, \dots, w_s$ tends to zero with n, and the z_i's are all in some $X_{B'}$ for some $B' \in \mathcal{F}_{v,u}$, and the total Euclidean length of the path $z_0, \dots, w_t$ tends to zero with n. Meanwhile, the Euclidean segments corresponding to the three edges u, w_0, w_s, z_t, and z_0, v lie within δ of disjoint segments of the line segment from u to v, and thus have total length $\leq ||u - v|| + 6\delta$, giving

$$\text{dist}(u, v) \leq ||u - v|| + 6\delta + o(1) = ||u - v|| + o(1). \tag{11}$$

We must also handle vertices $u, v \in \mathcal{X}_{n,p}$ with $||u - v|| < \gamma$. Given such a pair, we let B_u, B_v denote any choice of balls of radius γ such $\text{dist}(B_u, B_v) \geq \gamma$, $\text{dist}(B_u, u), \text{dist}(B_v, v) \leq \gamma(\sqrt{d} + 2)$. (These bounds are chosen to make such a choice trivially possible, even when u, v are close to a corner.) Observe that we have: where C_u, C_v denote the giant components of B_u, B_v,

$$\mathbf{Pr}(\forall u, v \in \mathcal{X}_{n,p}, \exists w \in C_u, z \in C_v \text{ such that } u \sim w, v \sim z) \to 1 \tag{12}$$

with n since a.a.s we have that B_u and B_v contain at least $\nu_d n \gamma^d / 2^{d+2}$ points for all $u, v \in \mathcal{X}_{n,p}$ and we have that $1 - 2n^2(1-p)^{n \cdot \nu_d \gamma^d / 2^{d+2}} \to 1$. In particular, we can a.a.s for all pairs $u, v \in \mathcal{X}_{n,p}$ find $w \sim u$ within distance $\gamma(\sqrt{d} + 4)$ of u, $z \sim v$ within Euclidean distance $\gamma(\sqrt{d} + 4)$ of v, such that

$$\gamma \leq ||w - z|| \leq (2\sqrt{d} + 8)\gamma.$$

Now, we can use the previous case (11) to see that

$$\text{dist}(u, v) \leq (2\sqrt{d} + 9)\gamma + 6\delta + o(1) = o(1). \tag{13}$$

In particular, $\text{dist}(u, v) - ||u - v|| = o(1)$. ◀

We complete the proof of Theorem 1 by proving

▶ **Lemma 5.** *Suppose that $Np = \omega \to \infty, \omega = O(\log N)$ and let C_1 denote the unique giant component of size $N - o(N)$ in $G_{N,p}$, that q.s.*[2] *exists. Then for L large,*

$$\mathbf{Pr}\left(\text{diam}(C_1) \geq \frac{L \log N}{\log Np}\right) \leq O(N^{-L/10}).$$

Proof. See appendix. ◀

3 Traveling among all vertices

Our first aim is to prove Theorem 3; this will be accomplished in Section 3.2, below. In fact, we will prove the following general statement, which will also be useful in the proof of Theorem 2:

[2] A sequence of events $\mathcal{E}_n$ occurs *quite surely* q.s. if $\mathbf{Pr}(\neg\mathcal{E}_n) = O(n^{-K})$ for all positive constants K.

▶ **Theorem 6.** *Let $\mathcal{Y}_1^d \subset [0,1]^d$ denote a set of points chosen from any fixed distribution, such that the cardinality $Y = |\mathcal{Y}_1^d|$ satisfies $\mathbf{E}(Y) = \mu > 0$ and $\mathbf{Pr}(Y \geq k) \leq C\rho^k$ for all k, for some $C > 0, \rho < 1$. For $t > 0$ let $\mathcal{Y}_t^d$ denote a random set of points in $[0,t]^d$ obtained from the union of t^d independent copies $\mathcal{Y}_1^d + x$ ($x \in \{0, \cdots, t-1\}^d$).*

If $p > 0$ is constant, $d \geq 2$, and $\mathcal{Y}_{t,p}^d$ denotes the random graph on $\mathcal{Y}_t^d$ with independent edge probabilities p, then $\exists \beta > 0$ (depending on p and the process generating $\mathcal{Y}_1^d$) such that

(i) $T(\mathcal{Y}_{t,p}^d) \approx \beta t^d$ *a.a.s., and*

(ii) $T(\mathcal{Y}_{t,p}^d) \leq \beta t^d + o(t^d)$ *q.s.*[3]

Note that as a probabilistic statement, Part (i) above asserts that there exists a choice for $o(1)$ (a function of t, say, tending to 0) such that $(1-o(1))\beta t^d \leq T(\mathcal{Y}_{t,p}^d) \leq (1+o(1))\beta t^d$ holds a.a.s. Similarly for Part (ii), the statement asserts the existence of a suitable fixed choice of $o(t^d)$ (a function of t, whose ratio to t^d tends to 0).

The restriction $\mathbf{Pr}\left(|\mathcal{Y}_1^d| \geq k\right) \leq C\rho^k$ simply ensures that we have exponential tail bounds on the number of points in a large number of independent copies of $\mathcal{Y}_1^d$:

▶ **Observation 7.** *For the total number T_n of points in n independent copies of $\mathcal{Y}_1^d$, we have for some absolute constant $A_{C,\rho} > 0$,*

$$\mathbf{Pr}(|T_n - \mu n| > \delta\mu n) < e^{-A_{C,\rho}\delta^2\mu^2 n}. \tag{14}$$

Note that the conditions on the distribution of $\mathcal{Y}_t^d$ are satisfied for a Poisson cloud of intensity 1, and it is via this case that we will derive Theorem 3. Other examples for which these conditions hold include the case where $\mathcal{Y}_t^d$ is simply a suitable grid of points, or is a random subset of a suitable grid of points in $[0,t]^d$, and we will make use of this latter case of Theorem 6 in our proof of Theorem 2.

Outline of proof of Theorem 6

Our proof uses subadditivity, but some of the standard properties of the classical case (e.g., monotonicity) fail in our setting, requiring us to use induction on d to achieve the result. For technical reasons (see also Question 4 of Section 5) Theorems 6 and 3 are given just for $d \geq 2$, and before beginning with the induction, we must carry out a separate argument to bound the length of the tour in 1 dimension.

When $d = 1$ all we can prove is an $O(n)$ bound on the length of the minimum tour. We do this by examining a natural greedy algorithm for finding a tour. This is the content of Lemma 8. After this we prove a sort of Lipschitz condition for the tour length, see Lemma 10. This will substitute for monotonicity. After this we can push ahead using subadditivity.

3.1 Bounding the expected tour length in 1 dimension

▶ **Lemma 8.** *Consider the random graph $G = G_{n,p}$ on the vertex set $[n]$ with constant p, where each edge $\{i,j\} \in E(G)$ is given length $|i-j| \in \mathbb{N}$. Let Z denote the minimum length of a Hamilton cycle in G starting at vertex 1, assuming one exists. If no such cycle exists let $Z = n^2$. Then there exists a constant A_p such that*

$$\mathbf{E}(Z) \leq A_p n \text{ and } Z \leq A_p n, \text{ q.s.}$$

[3] In this context $O(n^{-\omega(1)})$ is replaced by $O(t^{-\omega(1)})$.

We omit the proof due to space limitations.

Let us observe now that we get an upper bound $\mathbf{E}(T(\mathcal{Y}_{t,p}^1)) \leq A_p t$ on the length of a tour in 1 dimension. We have

$$\mathbf{E}(T(\mathcal{Y}_{t,p}^1)) = \sum_{n=0}^{\infty} \mathbf{E}\left(T(\mathcal{Y}_{t,p}^1) \middle| |\mathcal{Y}_{t,p}^1| = n\right) \mathbf{Pr}(|\mathcal{Y}_{t,p}^1| = n).$$

When conditioning on $|\mathcal{Y}_{t,p}^1| = n$, we let $P_1 < P_2 < \cdots < P_n \subset [0,t]$ be the points in $\mathcal{Y}_{t,p}^1$. We choose $k \in \{0, n-1\}$ uniformly randomly and let $\xi_i = ||P_{k+i+1} - P_{k+i}||$, where the indices of the P_j are evaluated modulo n. We now have $\mathbf{E}(\xi_i) \leq \frac{2t}{n}$ for all i, and

$$\mathbf{E}\left(T(\mathcal{Y}_{t,p}^1) \middle| |\mathcal{Y}_{t,p}^1| = n\right) \leq A_p n \cdot \frac{2t}{n},$$

and thus

$$\mathbf{E}\left(T(\mathcal{Y}_{t,p}^1)\right) \leq 2A_p t. \tag{15}$$

3.2 The asymptotic tour length

Our proof of Theorem 6 uses recursion, by dividing the $[t]^d$ cube into smaller parts. However, since our divisions of the cube must not cross boundaries of the elemental regions $\mathcal{Y}_1^d$, we cannot restrict ourselves to subdivisions into perfect cubes (in general, the integer t may not have the divisors we like).

To this end, if $L = T_1 \times T_2 \times \cdots \times T_d$ where each T_i is either $[0,t]$ or $[0,t-1]$, we say L is a d-dimensional *near-cube* with sidelengths in $\{t-1, t\}$. For $0 \leq d' \leq d$, we define the canonical example $L_d^{d'} := [0,t]^{d'} \times [0,t-1]^{d-d'}$ for notational convenience, and let

$$\Phi_p^{d,d'}(t) = \mathbf{E}\left(T(\mathcal{Y}_{t,p}^d \cap L_d^{d'})\right).$$

so that

$$\Phi_p^d(t) := \Phi_p^{d,d}(t) = \Phi_p^{d,0}(t+1).$$

In the unlikely event that $\mathcal{Y}_{t,p}^d \cap L_d^{d'}$ is not Hamiltonian, we take $T(\mathcal{Y}_{t,p}^d \cap L_d^{d'}) = t^{d+1}\sqrt{d}$, for technical reasons.

Our first goal is an asymptotic formula for Φ:

▶ **Lemma 9.** *There exists $\beta > 0$ such that*

$$\Phi_p^{d,d'}(t) \sim \beta t^d.$$

The proof of this is deferred until after the proof of Corollary 12 below.

The proof is by induction on $d \geq 2$. We prove the base case $d = 2$ along with the general case. We begin with a technical lemma.

▶ **Lemma 10.** *For every fixed p, d, there is a constant $F_{p,d} > 0$ such that*

$$\Phi_p^{d,d'}(t) \leq \Phi_p^{d,d'-1}(t) + F_{p,d} t^{d-1} \tag{16}$$

for all t sufficiently large. In particular, this implies that there is a constant $A_{p,d} > 0$ such that

$$\Phi_p^d(t+h) \leq \Phi_p^d(t) + A_{p,d} h t^{d-1} \tag{17}$$

for sufficiently large t and $1 \leq h \leq t$.

Proof. See appendix. ◀

Our argument is an adaptation of that in Beardwood, Halton and Hammersley [1] or Steele [12], with modifications to address difficulties introduced by the random set of available edges. First we introduce the concept of a decomposition into near-cubes. (Allowing near-cube decompositions is necessary for the end of the proof, beginning with Lemma 13). Simplifications relying on *Boundary Functionals* as in Yukich [13] do not appear to be available due to missing edges.

We say that a partition of $L_d^{d'}$ into m^d near-cubes S_α with sidelengths in $\{u, u+1\}$ indexed by $\alpha \in [m]^d$ is a *decomposition* if for each $1 \le b \le d$, there is an integer M_b such that, letting

$$f_b(a) = \begin{cases} au \text{ if } a < M_b \\ (a - M_b)(u+1) + M_b u \text{ if } a \ge M_b. \end{cases} .$$

we have that

$$S_\alpha = [f_1(\alpha_1 - 1), f_1(\alpha_1)] \times [f_2(\alpha_2 - 1), f_2(\alpha_2)] \times \cdots \times [f_d(\alpha_d - 1), f_d(\alpha_d)].$$

Observe that so long as $u \ll t$, $L_d^{d'}$ always has a decomposition into near-cubes with sidelengths in $\{u, u+1\}$. Indeed, if $t = ru - s$ for $0 \le s < u$ then we can take $M_b = s$ for $b \le d'$ and $M_b = s - 1$ for $b > d'$, unless $s = 0$, in which case $M_b = u - 1$.

First we note that tours in not-too-small near-cubes of a decomposition can be pasted together into a large tour at a reasonable cost:

▶ **Lemma 11.** *Fix $\delta > 0$, and suppose $t = mu$ for $u = t^\gamma$ for $\delta < \gamma \le 1$ ($m, u \in \mathbb{Z}$), and suppose S_α ($\alpha \in [m]^d$) is a decomposition of $L_d^{d'}$. We let $\mathcal{Y}_{t,p}^{d,\alpha} := \mathcal{Y}_{t,p}^d \cap S_\alpha$. We have*

$$T(\mathcal{Y}_{t,p}^d \cap L_d^{d'}) \le \sum_{\alpha \in [m]^d} T(\mathcal{Y}_{t,p}^{d,\alpha}) + 4m^d u\sqrt{d} \qquad \textit{with probability at least} \quad 1 - e^{-\Omega(u^d p^2)}.$$

Proof. See appendix. ◀

Linearity of expectation (and the upper bound $t^{d+1}\sqrt{d}$ on $T(\mathcal{Y}_{t,p}^d)$ when there is no tour) now gives a short-range recursive bound on $\Phi_p^d(t)$ when t factors reasonably well:

▶ **Corollary 12.** *For all large u and $1 \le m \le u^{10}$ ($m, u \in \mathcal{N}$),*

$$\Phi_p^d(mu) \le m^d(\Phi_p^d(u) + B_{p,d}u)$$

for some constant B_d. ◀

Proof of Lemma 9. Note that here we are using a decomposition of $[mu]^d$ into m^d subcubes with sidelength u; near-cubes are not required.

To get an asymptotic expression for $\Phi_p^d(t)$ we now let

$$\beta = \liminf_t \frac{\Phi_p^d(t)}{t^d}.$$

Choose u_0 large and such that

$$\frac{\Phi_p^d(u_0)}{u_0^d} \le \beta + \varepsilon$$

and then define the sequence $u_k, k \geq -1$ by $u_{-1} = u_0$ and $u_{k+1} = u_k^{10}$ for $k \geq 0$. Assume inductively that for some $i \geq 0$ that for $A_{p,d}$ as in Lemma 10 and $B_{p,d}$ as in Corollary 12,

$$\frac{\Phi_p^d(u_i)}{u_i^d} \leq \beta + \varepsilon + \sum_{j=-1}^{i-2} \left(\frac{A_{p,d}}{u_j} + \frac{B_{p,d}}{u_j^{d-1}} \right). \tag{18}$$

This is true for $i = 0$, and then for $i \geq 0$ and $0 \leq u \leq u_i$ and $d \leq m \in [u_{i-1}, u_{i+1}]$ we have

$$\begin{aligned}
\frac{\Phi_p^d(mu_i + u)}{(mu_i + u)^d} &\leq \frac{\Phi_p^d(mu_i) + A_{p,d}u(mu_i)^{d-1}}{(mu_i)^d}, && \text{from Lemma 10,} \\
&\leq \frac{m^d(\Phi_p^d(u_i) + B_{p,d}u_i) + A_{p,d}u(mu_i)^{d-1}}{(mu_i)^d}, && \text{from Corollary 12,} \qquad (19) \\
&\leq \beta + \varepsilon + \sum_{j=-1}^{i-2} \left(\frac{A_{p,d}}{u_j} + \frac{B_{p,d}}{u_j^{d-1}} \right) + \frac{B_{p,d}}{u_i^{d-1}} + \frac{A_{p,d}}{m}, && \text{by induction,} \\
&\leq \beta + \varepsilon + \sum_{j=-1}^{i-1} \left(\frac{A_{p,d}}{u_j} + \frac{B_{p,d}}{u_j^{d-1}} \right). && (20)
\end{aligned}$$

Putting $m = u_{i+1}/u_i$ and $u = 0$ into (20) completes the induction. We deduce from (18) and (20) that for $i \geq 0$ we have

$$\frac{\Phi_p^d(t)}{t^d} \leq \beta + \varepsilon + \sum_{j=-1}^{\infty} \left(\frac{A_{p,d}}{u_j} + \frac{B_{p,d}}{u_j^{d-1}} \right) \leq \beta + 2\varepsilon \qquad \text{for } t \in J_i = [u_{i-1}u_i, u_i(u_{i+1}+1)] \tag{21}$$

Now $\bigcup_{i=0}^{\infty} J_i = [u_0^2, \infty]$ and since ε is arbitrary, we deduce that

$$\beta = \lim_{t \to \infty} \frac{\Phi_p^d(t)}{t^d}, \tag{22}$$

We can conclude that

$$\Phi_p^d(t) \sim \beta t^d,$$

which, together with Lemma 10, completes the proof of Lemma 9, once we show that $\beta > 0$ in (22). To this end, we let ρ denote $\mathbf{Pr}(|\mathcal{Y}_1^d| \geq 1)$, so that $\mathbf{E}(|\mathcal{Y}_t^d|) \geq \rho t^d$. We say $x \in \{0, \ldots, t-1\}^d$ is *occupied* if there is a point in the copy $\mathcal{Y}_1^d + x$. Observing that a unit cube $[0,1]^d + x$ ($x \in \{0, \ldots, t-1\}^d$) is at distance at least 1 from all but $3^d - 1$ other cubes $[0,1]^d + y$, we certainly have that the minimum tour length through $\mathcal{Y}_t^d$ is at least $\frac{\mathcal{O}}{3^d-1}$, where where $\mathcal{O}$ is the number of occupied x. Linearity of expectation now gives that $\beta > \rho/(3^d - 1)$, completing the proof of Lemma 9. ◀

Before continuing, we prove the following much cruder version of Part (ii) of Theorem 6:

▶ **Lemma 13.** *For any fixed $\varepsilon > 0$, $T(\mathcal{Y}_{t,p}^d) \leq t^{d+\varepsilon}$ q.s.*

Proof. We let $m = \lfloor t^{1-\varepsilon/2} \rfloor$, $u = \lfloor t/m \rfloor$, and let $\{\mathcal{Y}_{\tau,p}^{d,\alpha}\}$ be a decomposition of $\mathcal{Y}_{t,p}^d$ into m^d near-cubes with sidelengths in $\{u, u+1\}$. We have that q.s. each $\mathcal{Y}_{\tau,p}^{d,\alpha}$ has (i) $\approx u^d$ points, and (ii) a Hamilton cycle H_α. We can therefore q.s. bound all $T(\mathcal{Y}_{\tau,p}^{d,\alpha})$ by $du \cdot u^d$, and Lemma 11 gives that q.s. $T(\mathcal{Y}_{t,p}^d) \leq 4dut^d + 4m^d u\sqrt{d}$. ◀

Proof of Theorem 6. We consider a decomposition $\{S_\alpha\}$ $(\alpha \in [m]^d)$ of $\mathcal{Y}_t^d$ into m^d near-cubes of side-lengths in $\{u, u+1\}$, for $\gamma = 1 - \frac{\varepsilon}{2}$, $m = \lfloor t^\gamma \rfloor$, and $u = \lfloor t/m \rfloor$.

Lemma 9 gives that

$$\mathbf{E}\, T(\mathcal{Y}_{t,p}^{d,\alpha}) \sim \beta u^d \sim \beta t^{(1-\gamma)d}.$$

Let

$$\mathcal{S}_\gamma(\mathcal{Y}_{t,p}^d) = \sum_{\alpha \in [m]^d} \min\left\{T(\mathcal{Y}_{t,p}^{d,\alpha}), 2dt^{(1-\gamma)(d+\varepsilon)}\right\}.$$

Note that $\mathcal{S}_\gamma(\mathcal{Y}_{t,p}^d)$ is the sum of $t^{\gamma d}$ identically distributed bounded random variables.

Now, since q.s. $T(\mathcal{Y}_{t,p}^{d,\alpha}) \leq 2dt^{(1-\gamma)(d+\varepsilon)}$ for all α by Lemma 13, we have that q.s. $\mathcal{S}_\gamma(\mathcal{Y}_{t,p}^d) = \sum_\alpha T(\mathcal{Y}_{t,p}^{d,\alpha})$. Applying Hoeffding's theorem we see that for any $\xi > 0$, we have

$$\mathbf{Pr}(|\mathcal{S}_\gamma(\mathcal{Y}_{t,p}^d) - m^d\,\mathbf{E}(T(\mathcal{Y}_{u,p}^d))| \geq \xi) \leq 2\exp\left(-\frac{2\xi^2}{4m^d d^2 t^{2(1-\gamma)(d+\varepsilon)}}\right).$$

Putting $\xi = t^{d\varepsilon}$ for small ε, we see that

$$\mathcal{S}_\gamma(\mathcal{Y}_{t,p}^d) = \beta t^d + o(t^d) \qquad q.s. \tag{23}$$

Note next that Lemma 11 implies that

$$T(\mathcal{Y}_{t,p}^d) \leq \mathcal{S}_\gamma(\mathcal{Y}_{t,p}^d) + \delta_2 \text{ where } \delta_2 = o(t^d) \qquad q.s. \tag{24}$$

It follows from (23) and (24) and the fact that $\mathbf{Pr}(|\mathcal{Y}_t^d| = t^d) = \Omega(t^{-d/2})$ that

$$T(\mathcal{Y}_{t,p}^d) \leq \beta t^d + o(t^d) \qquad q.s. \tag{25}$$

which proves part (ii) of Theorem 6.

Of course, we have from Lemma 9 that

$$\mathbf{E}(T(\mathcal{Y}_{t,p}^d)) = \beta t^d + \delta_1 \text{ where } \delta_1 = o(t^d), \tag{26}$$

and we show next that that this together with (24) implies part (i) of Theorem 6, that:

$$T = T(\mathcal{Y}_{t,p}^d) = \beta t^d + o(t^d) \qquad a.a.s. \tag{27}$$

We choose $0 \leq \delta_3 = o(t^d))$ such that $0 \leq \delta_2, |\delta_1| = o(\delta_3)$. Let $I = [\beta t^d - \delta_3, \beta t^d + \delta_2]$. Then we have

$$\begin{aligned}\beta t^d + \delta_1 = \mathbf{E}(T(\mathcal{Y}_{t,p}^d) \mid T(\mathcal{Y}_{t,p}^d) \geq (\beta t^d + \delta_2)\,\mathbf{Pr}(T(\mathcal{Y}_{t,p}^d) \geq \beta t^d + \delta_2)\\ + \mathbf{E}(T(\mathcal{Y}_{t,p}^d) \mid T(\mathcal{Y}_{t,p}^d) \in I)\,\mathbf{Pr}(T(\mathcal{Y}_{t,p}^d) \in I) +\\ \mathbf{E}(T(\mathcal{Y}_{t,p}^d) \mid T(\mathcal{Y}_{t,p}^{d,\alpha}) \leq \beta t^d - \delta_3)\,\mathbf{Pr}(T(\mathcal{Y}_{t,p}^d) \leq \beta t^d - \delta_3).\end{aligned}$$

Now $\varepsilon_1 = \mathbf{E}(T(\mathcal{Y}_{t,p}^d) \mid T(\mathcal{Y}_{t,p}^d) \geq \beta t^d + \delta_2)\,\mathbf{Pr}(T(\mathcal{Y}_{t,p}^d) \geq \beta t^d + \delta_2) = O(t^{-\omega(1)})$ since $|\mathcal{Y}_{t,p}^d| \leq 2d^{1/2}t^d$ and $\mathbf{Pr}(T(\mathcal{Y}_{t,p}^d) \geq \beta t^d + \delta_2) = O(t^{-\omega(1)})$, from (25).

So, if $\lambda = \mathbf{Pr}(T(\mathcal{Y}_{t,p}^d) \in I)$ then we have

$$\beta t^d + \delta_1 \leq \varepsilon_1 + (\beta t^d + \delta_2)\lambda + (\beta t^d - \delta_3)(1 - \lambda)$$

or

$$\lambda \geq \frac{\delta_1 - \varepsilon_1 + \delta_3}{\delta_2 + \delta_3} = 1 - o(1),$$

and this proves (27) completing the proof of Theorem 6. ◀

Proof of Theorem 3. We now let $\mathcal{W}_{t,p}^d$ be the graph on the set of points in $[0,t]^d$ which is the result of a Poisson process of intensity 1. Our first task is to bound the variance $\mathcal{V}(t)$ of $T(\mathcal{W}_{t,p}^d)$. Here we follow Steele's argument [12] with only small modifications. We approximate $T(\mathcal{W}_{2t,p}^d)$ as the sum over 2^d half-size cubes of $T(\mathcal{W}_{t,p}^d)$ and use this to show that $\sum_{k=1}^{\infty} \frac{\mathcal{V}(2^k t)}{(2^k t)^{2d}} \leq \infty$. This deals with n of the form $2^k t$ for some value of t and we then have to fill in the gaps.

Let $\mathcal{E}_t$ denote the event that

$$T(\mathcal{W}_{2t,p}^d) \leq \sum_{\alpha \in [2]^d} T(\mathcal{W}_{t,p}^{d,\alpha}) + 2^{d+2} t\sqrt{d}. \tag{28}$$

Observe that Lemma 11 implies that

$$\mathbf{Pr}(\neg \mathcal{E}_t) \leq e^{-\Omega(t^d p)}. \tag{29}$$

We define the random variable $\lambda(t) = T(\mathcal{W}_{t,p}^d) + 10t\sqrt{d}$, and let λ_i denote independent copies of $\lambda(t)$. Conditioning on $\mathcal{E}_t$, we have from (28) that

$$\lambda(2t) \leq \sum_{i=1}^{2^d} \lambda_i(t) - 4t\sqrt{d} \leq \sum_{i=1}^{2^d} \lambda_i(t). \tag{30}$$

In particular, (29) implies that letting $\Upsilon(t) = \mathbf{E}(\lambda(t)) = \Omega(t^d)$ (see (26)) and $\Psi(t) = \mathbf{E}(\lambda(t)^2)$, we have for sufficiently large t that

$$\begin{aligned}
\Psi(2t) &\leq \mathbf{E}\left(\left(\sum_{\alpha \in [2]^d} T(\mathcal{W}_{t,p}^{d,\alpha}) + 2^{d+2} t\sqrt{d} + 21t\sqrt{d}\right)^2\right) \\
&= \sum_{i=1}^{2^d} \mathbf{E}((\lambda_i(t) - 10t\sqrt{d})^2) + \sum_{i \neq j}^{2^d} \mathbf{E}(\lambda_i(t) - 10t\sqrt{d})\, \mathbf{E}(\lambda_j(t) - 10t\sqrt{d}) + \\
&\qquad + (2^{d+2} + 21)t\sqrt{d} \sum_{i=1}^{2^d} \mathbf{E}(\lambda_i(t) - 10t\sqrt{d}) + ((2^{d+2} + 21)t\sqrt{d})^2 \\
&= 2^d\, \mathbf{E}((\lambda(t) - 10t\sqrt{d})^2) + 2^d(2^d - 1)\, \mathbf{E}(\lambda(t) - 10t\sqrt{d})^2 + \\
&\qquad + 2^d(2^{d+2} + 21)t\sqrt{d}\, \mathbf{E}(\lambda(t) - 10t\sqrt{d}) + ((2^{d+2} + 21)t\sqrt{d})^2 \\
&= 2^d \Psi(t) + 2^d(2^d - 1)\Upsilon(t)^2 - \Omega(t\, \mathbf{E}(\lambda(t)) + O(t^2)) \\
&\leq 2^d \Psi(t) + 2^d(2^d - 1)\Upsilon(t)^2.
\end{aligned}$$

For

$$\mathcal{V}(t) := \mathbf{Var}(T(\mathcal{W}_{t,p}^d)) = \Psi(t) - \Upsilon(t)^2,$$

we have

$$\frac{\mathcal{V}(2t)}{(2t)^{2d}} - \frac{1}{2^d} \frac{\mathcal{V}(t)}{t^{2d}} \leq \frac{\Upsilon(t)^2}{t^{2d}} - \frac{\Upsilon(2t)^2}{(2t)^{2d}}.$$

Now with $t \geq 1$ arbitrary, summing over $2^k t$ for $k = 0, \ldots, M-1$ gives

$$\sum_{k=1}^{M} \frac{\mathcal{V}(2^k t)}{(2^k t)^{2d}} - \frac{1}{2^d} \sum_{k=0}^{M-1} \frac{\mathcal{V}(2^k t)}{(2^k t)^{2d}} \leq \frac{\Upsilon(t)^2}{t^{2d}} - \frac{\Upsilon(2^M t)^2}{(2^M t)^{2d}} \leq \frac{\Upsilon(t)^2}{t^{2d}}$$

and so, solving for the first sum, we find

$$\sum_{k=1}^{M} \frac{\mathcal{V}(2^k t)}{(2^k t)^{2d}} \leq \left(1 - \frac{1}{2^d}\right)^{-1} \left(\frac{\mathcal{V}(t)}{t^{2d}} + \frac{\Upsilon(t)^2}{t^{2d}}\right) < \infty. \tag{31}$$

Still following Steele, we let $N(t)$ be the Poisson counting process on $[0, \infty)$. We fix a random embedding $\mathcal{U}$ of $\mathcal{N}$ in $[0,1]^d$ as $u_1, u_2, \dots$ and a random graph $\mathcal{U}_p$ where each edge is included with independent probability p. We let $\mathcal{U}_{n,p}$ denote the restriction of this graph to the first n natural numbers. In particular, note that $\mathcal{U}_{N(t^d),p}$ is equivalent to $\mathcal{W}_{t,p}$, scaled from $[0,t]^d$ to $[0,1]^d$. Thus, applying Chebychev's inequality to (31) gives, in conjunction with Lemma 9, that

$$\sum_{k=0}^{\infty} \mathbf{Pr}\left(\left|\frac{t2^k T(\mathcal{U}_{N((t2^k)^d),p})}{(t2^k)^d} - \beta_{p,d}\right| > \varepsilon\right) < \infty \tag{32}$$

and so for $t > 0$ that

$$\lim_{k\to\infty} \frac{T(\mathcal{U}_{N((t2^k)^d),p})}{(t2^k)^{d-1}} = \beta_{p,d} \qquad a.s. \tag{33}$$

Now choosing some large integer ℓ, we have that (33) holds simultaneously for all the (finitely many) integers $t \in S_P = [2^\ell, 2^{\ell+1})$; and for $2^\ell \leq r \in \mathbb{R}$, we have that

$$r \in [2^k t, 2^k(t+1)) \text{ for } t \in S_\ell \text{ and some } k. \tag{34}$$

(We simply choose k such that $2^\ell \leq 2^{-k} r < 2^{\ell+1}$.)

◀

Unlike the classical case $p = 1$, in our setting, we do not have monotonicity of $T(\mathcal{U}_{n,p})$. Nevertheless, we show a kind of continuity of the tour length through $T(\mathcal{U}_{n,p})$:

▶ **Lemma 14.** *For all $\varepsilon > 0$, $\exists \delta > 0$ such that for all $0 \leq k < \delta n$, we have*

$$T(\mathcal{U}_{n+k,p}) < T(\mathcal{U}_{n,p}) + \varepsilon n^{\frac{d-1}{d}}, \qquad q.s. \tag{35}$$

Proof. See appendix. ◀

Applying Lemma 14 with $\delta = (1+\frac{1}{t})^d - 1 = O(\frac{d}{t})$ so that we have $(2^k t)^d \leq r^d \leq (2^k t)^d(1+\delta)$ by (34), and using the fact that

$$(1-2\delta)N(r^d) < N((1-\delta)r^d) < N((1+\delta)r^d) < (1+2\delta)N(r^d) \; q.s. \text{ (with respect to } r),$$

gives that for some $\varepsilon_\ell > 0$ which can be made arbitrarily small by increasing ℓ, we have q.s.

$$T(\mathcal{U}_{N(((t+1)2^k)^d),p}) - \varepsilon_\ell r^{d-1} < T(\mathcal{U}_{N(r^d),p}) < T(\mathcal{U}_{N((t2^k)^d),p}) + \varepsilon_\ell r^{d-1},$$

and so dividing by r^{d-1} and using (33) and taking limits we find that a.s.

$$\beta_{p,d} - 2\varepsilon_\ell \leq \liminf_{r\to\infty} \frac{T(\mathcal{U}_{N(r^d)})}{r^{d-1}} \leq \limsup_{r\to\infty} \frac{T(\mathcal{U}_{N(r^d)})}{r^{d-1}} \leq \beta_{p,d} + 2\varepsilon_\ell.$$

Since ℓ may be arbitrarily large, we find that

$$\lim_{r\to\infty} \frac{T(\mathcal{U}_{N(r^d)})}{r^{d-1}} = \beta_{p,d}.$$

Now the elementary renewal theorem guarantees that

$$N^{-1}(n) \sim n, \qquad a.s.$$

So we have a.s.

$$\lim_{r\to\infty} \frac{T(\mathcal{U}_{n,p})}{n^{\frac{d-1}{d}}} = \lim_{r\to\infty} \frac{T(\mathcal{U}_{N(N^{-1}(n)),p})}{(N^{-1}(n))^{\frac{d-1}{d}}} \frac{(N^{-1}(n))^{\frac{d-1}{d}}}{n^{\frac{d-1}{d}}} = \beta_{p,d} \cdot 1 = \beta_{p,d}.$$

4 The case $p(n) \to 0$

This is omitted due to space restrictions.

5 Further questions

Theorem 1 shows that there is a definite qualitative change in the diameter of $\mathcal{X}_{n,p}$ at around $p = \frac{\log^d n}{n}$, but our methods leave a $(\log \log n)^{2d}$ size gap for the thresholds.

▶ **Question 1.** *What is the precise threshold for there to be distances in $\mathcal{X}_{n,p}$ which tend to ∞? What is the precise threshold for distance in $\mathcal{X}_{n,p}$ to be arbitrarily close to Euclidean distance? What is the behavior of the intermediate regime?*

One could also analyze the geometry of the geodesics in $\mathcal{X}_{n,p}$ (Figure 1). For example:

▶ **Question 2.** *Let ℓ be the length of a random edge on the geodesic between fixed points at at constant distance in $\mathcal{X}_{n,p}$. What is the distribution of ℓ?*

Improving Theorem 2 to give an asymptotic formula for $T(\mathcal{X}_{n,p})$ is another obvious target. It may seem unreasonable to claim such a formula for all (say, decreasing) functions p; in particular, in this case, the constant in the asymptotic formula would necessarily be universal. The following, however, seems reasonable:

▶ **Conjecture 15.** *If $p = \frac{1}{n^\alpha}$ for some constant $0 < \alpha < 1$ then there exists a constant $\beta_{\alpha,d}$ such that a.a.s. $T(\mathcal{X}_{n,p}) \sim \beta_{\alpha,d} \frac{n^{\frac{d-1}{d}}}{p^{1/d}}$.*

We note that $T(\mathcal{X}_{n,1})$ is known to be remarkably well-concentrated around its mean; see, for example, the sharp deviation result of Rhee and Talagrand [10].

▶ **Question 3.** *How concentrated is the random variable $T(\mathcal{X}_{n,p})$?*

The case of where $p = o(1)$ may be particularly interesting.

Even for the case $p = 1$ covered by the BHH theorem, the constant $\beta_{1,d}$ $(d \geq 2)$ from Theorem 6 is not known. Unlike the case of $p = 1$, the 1-dimensional case is not trivial for our model. In particular, we have proved Theorems 3 and 2 only for $d \geq 2$. We have ignored the case $d = 1$ not because we consider the technical problems insurmountable, but because we hope that it may be possible to prove a stronger result for $d = 1$, at least for the case of constant p.

▶ **Question 4.** *Determine an explicit constant $\beta_{p,1}$ as a function of (constant) p such that for $d = 1$,*

$$\lim_{n\to\infty} T(\mathcal{X}_{n,p}) = \beta_{p,1}.$$

Our basic motivation has been to understand the constraint imposed on travel among random points by the restriction set of traversable edges which is chosen randomly independently of the geometry of the underlying point-set. While the Erdős-Rényi-Gilbert model is the prototypical example of a random graph, other models such as the Barabási-Albert preferential attachment graph have received wide attention in recent years, due to properties (in particular, the distribution of degrees) they share with real-world networks.

▶ **Question 5.** *If the preferential attachment graph is embedded randomly in the unit square (hypercube), what is the expected diameter? What is the expected size of a minimum-length spanning tree?*

References

1 J. Beardwood, J. H. Halton, and J. M. Hammersley. The shortest path through many points. *Mathematical Proceedings of the Cambridge Philosophical Society*, 55:299–327, 1959.
2 B. Bollobás. *Random Graphs, Second Edition*. Cambridge University Press, 2001.
3 B. Bollobás, T. Fenner, and A. M. Frieze. An algorithm for finding hamilton paths and cycles in random graphs. *Combinatorica*, 7:327–341, 1987.
4 D. Fernholz and V. Ramachandran. The diameter of sparse random graphs. *Random Structures and Algorithms*, 31:482–516, 2007.
5 Y. Gurevich and S. Shelah. Expected computation time for hamiltonian path problem. *SIAM Journal on Computing*, 16:486–502, 1987.
6 R. M. Karp. Probabilistic analysis of partitioning algorithms for the traveling-salesman problem in the plane. *Mathematics of Operations Research*, 2:209–244, 1977.
7 A. Mehrabian. A randomly embedded random graph is not a spanner. In *Proceedings of the 23rd Canadian Conference on Computational Geometry (CCCG 2011)*, pages 373–374, 2011.
8 A. Mehrabian and N. Wormald. On the stretch factor of randomly embedded random graphs. *Discrete & Computational geometry*, 49:647–658, 2013.
9 P. M. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.
10 W. Rhee and M. Talagrand. A sharp deviation inequality for the stochastic traveling salesman problem. *The Annals of Probability*, 17:1–8, 1989.
11 O. Riordan and N. Wormald. The diameter of sparse random graphs. *Combinatorics, Probability and Computing*, 19:835–926, 2010.
12 J. M. Steele. Subadditive euclidean functionals and nonlinear growth in geometric probability. *The Annals of Probability*, 9:365–376, 1981.
13 J. Yukich. *Probability Theory of Classical Euclidean Optimization Problems*. Springer, 1991.

A Proof of Lemma 5

Let $\mathcal{B}(k)$ be the event that there exists a set S of k vertices in $G_{N,p}$ that induces a connected subgraph and in which more than half of the vertices have less than $\omega/2$ neighbors outside S. Then for $k = o(N)$ we have

$$\mathbf{Pr}(\mathcal{B}(k)) \leq \binom{N}{k} p^{k-1} k^{k-2} 2^k \, \mathbf{Pr}(Bin(N-k,p) \leq \omega/2)^{k/2} \tag{36}$$

$$\leq \frac{e^k \omega^k}{pk^2} 2^k \left(e^{-((N-k)p-\omega/2)^2/(2(N-k)p)} \right)^{k/2}, \text{ from (6) with } \varepsilon = 1 - \frac{\omega}{2(N-k)p}, \tag{37}$$

$$\leq \frac{e^k \omega^k}{pk^2} 2^k \left(e^{-(.99\omega-\omega/2)^2/2\omega} \right)^{k/2} \tag{38}$$

$$\leq p^{-1} (2e\omega e^{-\omega/20})^k \leq N e^{-k\omega/21}. \tag{39}$$

$$\tag{40}$$

Explanation of (36): $\binom{N}{k}$ bounds the number of choices for S. We then choose a spanning tree T for S in k^{k-2} ways. We multiply by p^{k-1}, the probability that T exists. We then choose half the vertices X of S in at most 2^k ways and then multiply by the probability that each $x \in X$ has at most $\omega/2$ neighbors in $[N] \setminus S$.

If $\kappa = \kappa(L) = \frac{L \log N}{\log Np}$ then (39) implies that $\mathbf{Pr}(\mathcal{B}(\kappa)) \leq N^{1-L}$.

Next let $\mathcal{D}(k) = \mathcal{D}_N(k)$ be the event that there exists a set S of size k for which the number of edges $e(S)$ contained in S satisfies $e(S) \geq 2k$. Then,

$$\mathbf{Pr}(\mathcal{D}(k)) \leq \binom{N}{k}\binom{\binom{k}{2}}{2k}p^{2k} \leq \left(\frac{Ne}{k}\cdot\left(\frac{ke\omega}{2N}\right)^2\right)^k = \left(\frac{ke^3\omega^2}{4N}\right)^k.$$

Since $\omega = O(\log n)$ we have that q.s.

$$\nexists k \in [\kappa(1), N^{3/4}] \text{ such that } \mathcal{D}(k) \text{ occurs.} \tag{41}$$

Now let $\mathcal{B}(k_1,k_2) = \bigcup_{k=k_1}^{k_2}\mathcal{B}(k)$ and $\mathcal{D}(k_1,k_2) = \bigcup_{k=k_1}^{k_2}\mathcal{D}(k)$, and suppose that

$$\mathcal{B}(k_1,k_2) \cup \mathcal{D}(k_1,k_2) \text{ does not occur,} \tag{42}$$

where $k_1 = \kappa(L/4)$ and $k_2 = N^{3/4}$. Fix a pair of vertices v, w and define sets $S_0, S_1, S_2, \ldots$ where S_i is the set of vertices at distance i from v. If there is no $i \leq k_1$ with $w \in S_i$ then we must have $S_{k_1} \neq \emptyset$ and $|S_{\leq k_1}| \geq k_1$ where $S_{\leq t} = \bigcup_{i=0}^{t} S_i$ for $t \geq 0$. This is because $v, w \in C_1$ and C_1 is connected and so $|S_{\leq i+1}| \geq |S_{\leq i}| + 1$. We also see that $k_1 \leq |S_{\leq t}| \leq N^{3/4}$ implies that $|S_{t+1}| \geq \omega|S_{\leq t}|/10$. Indeed, if $|S_{t+1}| < \omega|S_{\leq t}|/10$ then $S_{\leq t+1}$ has at most $(\omega+10)|S_{\leq t}|/10$ vertices and more than $\omega|S_{\leq t}|/4$ edges, contradiction.

Thus if L is large, then we find that there exists $t \leq k_1 + \kappa(3/4) \leq N^{3/4}$ such that $|S_{\leq t}| \geq N^{3/4}$ and so also that $|S_t| \geq (1-o(1))N^{3/4}$. Now apply the same argument from w to create sets $T_0, T_1, \ldots, T_s$, where either we reach v or find that $|T_s| \geq N^{3/4}$ where $s \leq k_1 + \kappa(3/4)$. At this point the edges between S_t and T_s are unconditioned and the probability there is no $S_t : T_s$ edge is at most $(1-p)^{N^{3/2}} = O(e^{-\Omega(N^{1/2})})$.

B Proof of Lemma 10

We let S denote the subgraph of $\mathcal{Y}_{t,p}^d \cap L_d^{d'}$ induced by the difference $L_d^{d'} \setminus L_d^{d'-1}$.

By ignoring the d'th coordinate of S, we obtain the $(d-1)$ dimensional set $\pi(S)$, for which induction on d (or equation (15) if $d = 2$) implies an expected tour $T(S)$ of length $\Phi_p^{d-1,d'-1}(t) \leq \beta_p^{d-1}t^{d-1}$, and so changing notation, we can write

$$\Phi_p^{d-1,d'-1}(t) \leq D_{p,d-1}t^{d-1}.$$

We have that

$$\mathbf{E}(T(S)) \leq \mathbf{E}(T(\pi(S)) + d^{1/2}\,\mathbf{E}(|\pi(S)|) \leq D_{p,d-1}t^{d-1} + d^{1/2}t^{d-1}.$$

The first inequality stems from the fact that the points in $L_d^{d'} \setminus L_d^{d'-1}$ have a d' coordinate in $[t-1, t]$.

Now if $\mathcal{Y}_{t,p}^d \cap L_d^{d'-1}$ and S are both Hamiltonian, then we have

$$T(\mathcal{Y}_{t,p}^d \cap L_d^{d'}) \leq T(\mathcal{Y}_{t,p}^d \cap L_d^{d'-1}) + T(S) + O_d(t) \tag{43}$$

which gives us the Lemma, by linearity of expectation. We have (43) because we can patch together the minimum cost Hamilton cycle H in $\mathcal{Y}_{t,p}^d \cap L_d^{d'-1}$ and the minimum cost path P in S as follows: Let u_1, v_1 be the endpoints of P. If there is an edge u, v of H such that $(u_1, u), (v_1, v)$ is an edge in $\mathcal{Y}_{t,p}^d$ then we can create a cycle H_1 through $\mathcal{Y}_{t,p}^d \cap L_d^{d'-1} \cup P$ at an extra cost of at most $2d^{1/2}t$. The probability there is no such edge is at most $(1-p^2)^{t/2}$, which is negligible given the maximum value of $T(\mathcal{Y}_{t,p}^d \cap L_d^{d'})$.

On the other hand, because p is a constant, the probability that either of $\mathcal{Y}_{t,p}^d \cap L_d^{d'-1}$ or S is not Hamiltonian is exponentially small in t, (see for example [5]), which is again negligible given the maximum value of $T(\mathcal{Y}_{t,p}^d \cap L_d^{d'})$. This completes the proof of (16).

To obtain (17) we use (16) to write

$$\Phi_p^{d,d}(t+h) \leq \Phi_p^{d,0}(t+h) + dF_{p,d}(t+h)^{d-1} = \Phi_p^d(t+h-1) + dF_{p,d}(t+h)^{d-1}$$
$$\leq \Phi_p^d(t) + dF_{p,d}\sum_{i=0}^{h}(t+i)^{d-1}.$$

C Proof of Lemma 11

Let $\mathcal{B}, \mathcal{C}$ denote the events

$$\mathcal{B} = \left\{\exists \alpha : \mathcal{Y}_{t,p}^{d,\alpha} \text{ is not Hamiltonian}\right\},$$
$$\mathcal{C} = \left\{\exists \alpha : \left| |\mathcal{Y}_{t,p}^{d,\alpha}| - u^d \right| \geq \delta u^d \right\},$$

and let $\mathcal{E} = \mathcal{B} \cup \mathcal{C}$.

Now $\mathbf{Pr}(\mathcal{B}) \leq m^d e^{-\Omega(u^d p)}$ and, by Observation 7, $\mathbf{Pr}(\mathcal{C}) \leq m^d e^{-\Omega(u^d)}$ and so $\mathbf{Pr}(\mathcal{E}) \leq e^{-\Omega(u^d p)}$. Assume therefore that $\neg\mathcal{E}$ occurs. Each subcube S_α will contain a minimum length tour H_α. We now order the subcubes $\{S_\alpha\}$ as $T_1, \ldots, T_{m^d}$, such that for $S_\alpha = T_i$ and $S_{\alpha'} = T_{i+1}$, we always have that the Hamming distance between α and α' is 1. Our goal is to inductively assemble a tour through the subcubes $T_1, T_2, \ldots, T_j$ from the smaller tours H_α with a small number of additions and deletions of edges.

Assume inductively that for some $1 \leq j < m^d$ we have added and deleted edges and found a single cycle C_j through the points in $T_1, \ldots, T_j$ in such a way that (i) the added edges have total length at most $4\sqrt{d}ju$ and (ii) we delete one edge from $\tau(T_1)$, $\tau(T_j)$ and two edges from each $\tau(T_i), 2 \leq i \leq j-1$. To add the points of T_{j+1} to create C_{j+1} we delete one edge (u,v) of $\tau(T_j) \cap C_j$ and one edge (x,y) of $\tau(T_{j+1})$ such that both edges $\{u,x\}, \{v,y\}$ are in the edge set of $\mathcal{Y}_{t,p}^d$. Such a pair of edges will satisfy (i) and (ii) and the probability we cannot find such a pair is at most $(1-p^2)^{(u^d/2-1)u^d/2}$. Thus with probability at least $1 - e^{\Omega(u^d p^2)}$ we build the cycle C_{m^d} with a total length of added edges $\leq 4\sqrt{d}m^d u$.

D Proof of Lemma 14

We consider cases according to the size of k.

Case 1: $k \leq n^{\frac{1}{3}}$. Note that we have $T(\mathcal{U}_{n+1,p}) < T(\mathcal{U}_{n,p}) + \sqrt{d}$ q.s., since we can q.s. find an edge in the minimum tour though $\mathcal{U}_{n,p}$ whose endpoints are both adjacent to $(n+1)$. $n^{\frac{1}{3}}$ applications of this inequality now give (35).

Case 2: $k > n^{\frac{1}{3}}$. In this case the restriction $\mathcal{R}$ of $\mathcal{U}_{n+k,p}$ to $\{n+1, \ldots, k\}$ is q.s. (with respect to n) Hamiltonian [3]. In particular, by Theorem 6, we can q.s. find a tour T though $\mathcal{R}$ of length $\leq 2\beta_p^d k^{\frac{d-1}{d}}$. Finally, there are q.s., edges $\{x,y\}$ and $\{w,z\}$ on the minimum tours through $\mathcal{U}_{n,p}$ and $\mathcal{R}$, respectively, such that $x \sim w$ and $y \sim z$ in $\mathcal{U}_{n+k,p}$, giving a tour of length

$$T(\mathcal{U}_{n+k,p}) \leq T(\mathcal{U}_{n,p}) + 2\beta_{p,d}k^{\frac{d-1}{d}} + 4\sqrt{d}.$$