Traveling salesman problem across dense cities

Ghurumuruhan Ganesan *

New York University, Abu Dhabi

Abstract

Consider n nodes $\{X_i\}_{1 \leq i \leq n}$ distributed independently across N cities contained with the unit square S according to a distribution f. Each city is modelled as an $r_n \times r_n$ square contained within S and let $TSPC_n$ denote the length of the minimum length cycle containing all the n nodes, corresponding to the traveling salesman problem (TSP). We obtain variance estimates for $TSPC_n$ and prove that if the cities are well-connected and densely populated in a certain sense, then $TSPC_n$ appropriately centred and scaled converges to zero in probability. We also obtain large deviation type estimates for $TSPC_n$. Using the proof techniques, we alternately obtain corresponding results for the length TSP_n of the minimum length cycle in the unconstrained case, when the nodes are independently distributed throughout the unit square S.

Key words: Traveling salesman problem, dense cities.

AMS 2000 Subject Classification: Primary: 60J10, 60K35; Secondary: 60C05, 62E10, 90B15, 91D30.

1 Introduction

The Traveling Salesman Problem (TSP) is the study of finding the minimum weight cycle containing all the nodes of a graph where each edge is assigned

^{*}E-Mail: gganesan82@gmail.com

a certain weight. In this paper, we consider the case of random Euclidean TSP, henceforth referred to simply as TSP, where the nodes are distributed randomly across the unit square S with origin as centre. The weight of an edge between two nodes is the Euclidean distance between them and the goal is to find the cycle of shortest length containing all the nodes. For more material on the TSP, we refer to the books by Gutin and Punnen (2006), Cook (2011) and references therein.

The analytical study of the random TSP problem originated in Beardwood et al (1959). The main result there is that if n nodes are randomly and uniformly distributed across the unit square S, then with high probability (i.e., with probability converging to one as $n \to \infty$), the length TSP_n of the minimum length spanning cycle grows roughly as $\beta \sqrt{n}$ for some constant $\beta > 0$. Equivalently, TSP_n appropriately scaled and centred converges to zero a.s. and in mean as $n \to \infty$. Subadditive ergodic type theorems are used for obtaining the convergence results and for a comprehensive survey, we refer to Steele (1981, 1993).

Since then there has been a lot of work focused on obtaining better bounds for the constant $\beta > 0$. Beardwood et al originally established that $0.625 \le \beta \le 0.922$. Recently, Steinerberger (2015) has obtained slightly improved bounds by estimating the probability of certain configurations that are avoided by the optimal cycle.

Because of its practical importance, there has also been a lot of work devoted to obtaining optimal and near optimal algorithms for obtaining the minimum length cycle. Arora (1998), Vazirani (2001), Karpinski et al (2015) develop and analyse polynomial time approximation schemes (PTAS) that determine near minimal spanning cycles for large vertex sets. Snyder and Daskin (2006) have used genetic algorithms to provide heuristic solutions for the generalized TSP problem, where the nodes are split into clusters and the objective is to find a minimum cost tour passing through exactly one node from each cluster. Recently, Pintea et al (2017) have proposed solutions to the generalized TSP problem using Ant algorithms.

The analytical literature above mainly consider nodes distributed in regular shapes like unit squares or circles. In this paper, we consider a slightly different scenario where cities (modelled as small squares) are spread across the unit square each containing a subset of the nodes. The cities are not necessarily regularly spaced and therefore the usual subadditive techniques to determine the convergence of TSP are not directly applicable here. Instead, we use approximation methods to find sharp upper and lower bounds

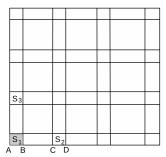


Figure 1: Tiling S into $r_n \times r_n$ squares with an inter-square distance of s_n .

for the optimal minimum spanning cycle and indirectly deduce convergence properties as the size of the vertex set $n \to \infty$.

Model Description

Structure of the cities

For integer $n \geq 1$, let r_n and s_n be real numbers such that $\frac{1-r_n}{r_n+s_n}$ is an integer. Tile the unit square S regularly into $r_n \times r_n$ size squares in such a way that the distance between any two squares is at least s_n as shown in Figure 1. In Figure 1, the grey square is of size $r_n \times r_n$, the segment AB has length r_n and the segment BC has length s_n . The $r_n \times r_n$ squares are called *cities* and the term s_n denotes the *intercity distance*.

Label the $r_n \times r_n$ squares (cities) as $\{S_l\}$ and identifying the centres of the squares $\{S_l\}$ with vertices in \mathbb{Z}^2 , we obtain a corresponding subset of vertices $\{z_l\} \subset \mathbb{Z}^2$. For example, in Figure 1, identify the centre of the square labelled S_1 with (0,0), the centre of S_2 with (1,0), the centre of S_3 with (0,1) and so on. Two vertices $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ are adjacent and connected by an edge if $|x_1 - x_2| + |y_1 - y_2| = 1$.

Fix N = N(n) cities $\{S_{j_1}, \ldots, S_{j_N}\}$ and let $\{z_{j_1}, \ldots, z_{j_N}\}$ be the vertices in \mathbb{Z}^2 corresponding to the centres of $\{S_{j_i}\}$. We say that the cities $\{S_{j_1}, \ldots, S_{j_N}\}$ are well-connected if the corresponding set of vertices $\{z_{j_i}\}$ form a connected subgraph of \mathbb{Z}^2 . Henceforth, we assume that $\{S_{j_1}, \ldots, S_{j_N}\}$ are well-connected and without loss of generality denote S_{j_i} by S_i for $1 \le i \le N$.

Nodes in the cities

Let f be any density on the unit square S satisfying the following conditions: There are constants $0 < \epsilon_1 \le \epsilon_2 < \infty$ such that

$$\epsilon_1 \le \inf_{x \in S} f(x) \le \sup_{x \in S} f(x) \le \epsilon_2$$
(1.1)

and

$$\int_{x \in S} f(x)dx = 1. \tag{1.2}$$

Define the density $g_N(.)$ on the N cities $\bigcup_{1 \le i \le N} S_i$ as

$$g_N(x) = \frac{f(x)}{\int_{\bigcup_{1 \le i \le N} S_i} f_j(x) dx}$$

$$\tag{1.3}$$

for all $x \in \bigcup_{1 \le j \le N} S_j$.

Let X_1, X_2, \ldots, X_n be n nodes independently and identically distributed (i.i.d.) in the N cities $\{S_j\}_{1 \leq j \leq N}$, each according to the density g_N . Define the vector (X_1, \ldots, X_n) on the probability space $(\Omega_X, \mathcal{F}_X, \mathbb{P})$. Let $K_n = K(X_1, \ldots, X_n)$ be the complete graph whose edges are obtained by connecting each pair of nodes X_i and X_j by the straight line segment (X_i, X_j) with X_i and X_j as endvertices. The line segment (X_i, X_j) is the edge between the nodes X_i and X_j and $d(X_i, X_j)$ denotes the (Euclidean) length of the edge (X_i, X_j) .

A cycle $C = (Y_1, Y_2, \dots, Y_t, Y_1)$ is a subgraph of K_n with vertex set $\{Y_j\}_{1 \leq j \leq t} \subset \{X_i\}$ and edge set $\{(Y_j, Y_{j+1})\}_{1 \leq j \leq t-1} \cup (Y_t, Y_1)$. The length of C is defined as the sum of the lengths of the edges in C; i.e.,

$$L(\mathcal{C}) = \sum_{i=1}^{t-1} d(Y_i, Y_{i+1}) + d(Y_t, Y_1) = \frac{1}{2} \sum_{i=1}^{t} l(Y_i, \mathcal{C}),$$
 (1.4)

where $l(Y_1, \mathcal{C}) = d(Y_1, Y_2) + d(Y_1, Y_t), l(Y_t, \mathcal{C}) = d(Y_t, Y_1) + d(Y_t, Y_{t-1})$ and for $2 \le i \le t$,

$$l(Y_i, C) = d(Y_i, Y_{i-1}) + d(Y_i, Y_{i+1})$$

is the sum of the length of the (two) edges in C containing Y_i as an endvertex. The cycle C is said to be a *spanning cycle* if C contains all the nodes $\{X_k\}_{1 \leq k \leq n}$. Let C_n be a spanning cycle satisfying

$$TSPC_n = L(\mathcal{C}_n) := \min_{\mathcal{C}} L(\mathcal{C}),$$
 (1.5)

where the minimum is taken over all spanning cycles C. If there is more than one choice for C_n , choose one according to a deterministic rule. The cycle C_n is defined to the *minimum spanning cycle* with corresponding length $TSPC_n$.

Letting

$$b_n := r_n \sqrt{nN},\tag{1.6}$$

we have the following result.

Theorem 1. Suppose r_n, s_n and N = N(n) satisfy

$$r_n^2 \ge \frac{M \log n}{n}, \frac{n}{N^2} \longrightarrow 0 \text{ and } \frac{Ns_n}{b_n} \longrightarrow 0$$
 (1.7)

as $n \to \infty$, for some constant M > 0. If $M = M(\epsilon_1, \epsilon_2) > 0$ is large, then

$$\frac{1}{b_n} (TSPC_n - \mathbb{E}TSPC_n) \longrightarrow 0 \text{ in probability}$$
 (1.8)

as $n \to \infty$. In addition, there are positive constants $\{\theta_i\}_{1 \le i \le 6}$ such that

$$\theta_1 b_n \le \mathbb{E}TSPC_n \le \theta_2 b_n, \tag{1.9}$$

$$\mathbb{P}\left(TSPC_n \ge \theta_3 b_n\right) \ge 1 - e^{-\theta_4 N} \tag{1.10}$$

and

$$\mathbb{P}\left(TSPC_n \le \theta_5 b_n\right) \ge 1 - \exp\left(-\theta_6 \frac{n}{N}\right) \tag{1.11}$$

for all n large.

In words, if the cities are wide and dense enough, then the centred and scaled minimum length of the traveling salesman cycle converges to zero in probability.

Unconstrained TSP

There are n nodes $\{X_i\}_{1\leq i\leq n}$ independently distributed in the unit square S each according to the distribution f satisfying (1.1). As in (1.5), let TSP_n be the length of the minimum spanning cycle containing all the nodes $\{X_i\}_{1\leq i\leq n}$.

Beardwood et al (1959) use subadditive techniques to study the convergence of the ratio $\frac{TSP_n}{\sqrt{n}} \longrightarrow \beta$ for some constant $\beta > 0$, a.s. as $n \to \infty$. Another approach involves the study of concentration of TSP_n around its mean via concentration inequalities (see Steele (1993)). Here we use the techniques used in the proof of Theorem 1 to obtain the following result.

Theorem 2. The variance

$$\mathbb{E}\left(TSP_n - \mathbb{E}TSP_n\right)^2 \le Cn^{2/3} \tag{1.12}$$

for some constant C > 0 and for all $n \ge 1$ and so in particular,

$$\frac{1}{\sqrt{n}}(TSP_n - \mathbb{E}TSP_n) \longrightarrow 0 \text{ in probability}$$

as $n \to \infty$. Also there are positive constants $\{\theta_i\}_{1 \le i \le 3}$ such that

$$\theta_1 \sqrt{n} \le \mathbb{E}TSP_n \le 5\sqrt{n},\tag{1.13}$$

$$\mathbb{P}\left(TSP_n \le 5\sqrt{n}\right) = 1\tag{1.14}$$

and

$$\mathbb{P}\left(TSP_n \ge \theta_2 \sqrt{n}\right) \ge 1 - \exp\left(-\frac{\theta_3 n}{\log n}\right) \tag{1.15}$$

for all n large.

The paper is organized as follows. In Section 2, we obtain preliminary estimates needed for the proofs of main Theorems. In Section 3, we prove Theorem 1 and in Section 4, we prove Theorem 2.

2 Preliminary estimates

We first describe the strips method used throughout to find an upper bound for the length of minimum length cycles.

Strips method

Suppose there are $a \geq 3$ nodes $\{x_i\}_{1 \leq i \leq a}$ placed in a square R of side length b. For $3 \leq j \leq a$ let $K(x_1, \ldots, x_j)$ be the complete graph with vertex set $\{x_i\}_{1 \leq i \leq j}$ and let \mathcal{C}_j be a spanning cycle of $K(x_1, \ldots, x_j)$ such that

$$L(\mathcal{C}_j) = \min_{\mathcal{C}} L(\mathcal{C}) =: TSP(x_1, \dots, x_j; R), \tag{2.1}$$

where the minimum is taken over all spanning cycles of $K(x_1, \ldots, x_j)$ and $L(\mathcal{C})$ is the length of \mathcal{C} (see (1.4)).

For any $3 \le j \le a$,

$$TSP(x_1, \dots, x_j; R) \le TSP(x_1, \dots, x_a; R) \le 5b\sqrt{a}. \tag{2.2}$$

Proof of (2.2): The first estimate in (2.2) is obtained by monotonicity as follows. Let $C = (y_1, \ldots, y_{j+1}, y_1)$ be any cycle in $K(x_1, \ldots, x_a)$ with vertex set $\{y_i\}_{1 \leq i \leq j+1} = \{x_i\}_{1 \leq i \leq j+1}$ and without loss of generality suppose that $y_{j+1} = x_{j+1}$. Recall that (y_j, y_{j+1}) is the edge with y_j and y_{j+1} as endvertices. Removing the edges (y_j, y_{j+1}) and (y_{j+1}, y_1) , and adding the edge (y_1, y_j) we get a new cycle C' with vertex set $\{x_i\}_{1 \leq i \leq j}$ (see Figure 2(a)).

By triangle inequality, the lengths

$$d(y_j, y_1) \le d(y_j, y_{j+1}) + d(y_{j+1}, y_1). \tag{2.3}$$

and therefore the length L(C') of C' (see (1.4) for definition) is

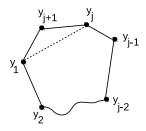
$$L(\mathcal{C}') = \sum_{i=1}^{j-1} d(y_i, y_{i+1}) + d(y_j, y_1) \le \sum_{i=1}^{j} d(y_i, y_{i+1}) + d(y_{j+1}, y_1) = L(\mathcal{C}).$$

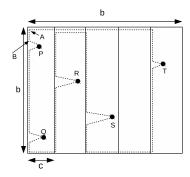
But by definition $TSP(x_1, ..., x_j; b) \leq L(\mathcal{C}')$ and so $TSP(x_1, ..., x_j; b) \leq L(\mathcal{C})$. Taking minimum over all cycles \mathcal{C} with vertex set $\{x_i\}_{1 \leq i \leq j+1}$, we get

$$TSP(x_1,\ldots,x_j;R) \leq TSP(x_1,\ldots,x_{j+1};R).$$

For the second estimate in (2.2), divide R into vertical rectangles (strips) each of size $c \times b$ so that the number of strips is $\frac{b}{c}$ as shown in Figure 2(b). Here a = 5 and without loss of generality suppose that $P = x_1, Q = x_2, R = x_3, S = x_4$ and $T = x_5$. The dotted line corresponds to a cycle containing all the nodes P, Q, R, S and T. Starting from close to the top left corner at point A, we go vertically down and encounter the nodes P, Q, R, S and T in that order. Each time we are close to a node, we "reach" for the node by a slightly inclined line. For example, the node P is joined to the vertical dotted line AB by the inclined line BP.

After the final node T is encountered, we join it to the starting point A by inclined, vertical and horizontal lines as shown in Figure 2. The cycle \mathcal{D} constructed above consists of vertical, horizontal and inclined lines. The number of strips is $\frac{b}{c}$ and the sum of the lengths of the vertical lines in a particular strip is at most the height of the strip b. Therefore the total length of vertical lines in \mathcal{D} is at most $\frac{b}{c}b$.





(a) Removing the vertex y_{j+1} from the (b) Estimating minimum length using cycle \mathcal{C} .

Figure 2: (a) Monotonicity in the TSP length. (b) Estimating the length of the TSP using strips method.

The total length of the horizontal lines in \mathcal{D} before encountering the final node T is at most b. Since T is joined to A by a curve consisting of a horizontal line, the total length of horizontal lines in \mathcal{D} is at most 2b.

Finally, each inclined line in \mathcal{D} has length at most $\frac{c}{\sqrt{2}}$, since the corresponding slope is at most 45 degrees. There are a nodes and there are exactly two inclined lines containing any particular node. Therefore the total length of the inclined lines in \mathcal{D} is at most $ac\sqrt{2}$.

Summarizing, the total length of edges in \mathcal{D} is at most $\frac{b^2}{c} + ac\sqrt{2} + 2b$. By construction, the cycle \mathcal{D} encounters the nodes x_1, \ldots, x_a in that order and so applying triangle inequality as before, the cycle $\mathcal{C} = (x_1, x_2, \ldots, x_a, x_1)$ with edges being the straight lines $(x_1, x_2), (x_2, x_3), \ldots, (x_a, x_1)$, has total length no more than the sum of length of edges in \mathcal{D} . Thus

$$TSP(x_1, ..., x_a; R) \le L(C) \le \frac{b^2}{c} + ac\sqrt{2} + 2b.$$
 (2.4)

Setting $c = \frac{b}{\sqrt{a}}$ in (2.4), we get

$$TSP(x_1, \dots, x_a; R) \le b\sqrt{a} + \sqrt{2}b\sqrt{a} + 2b \le 5b\sqrt{a},$$

since
$$a \ge 1$$
.

Length of TSP within cities

Recall from discussion prior to (1.7) that $n \geq 1$ nodes $\{X_k\}_{1 \leq k \leq n}$ are distributed across the $r_n \times r_n$ squares $\{S_j\}_{1 \leq j \leq N}$ according to a Binomial process with intensity g_N as defined in (1.3). In this subsection, we obtain estimates for the length T_l of the minimum length cycle containing all the nodes of the square S_l .

If p_l denotes the probability that a node of $\{X_j\}$ occurs inside S_l , then

$$\frac{\eta_1}{N} \le p_l := \frac{\int_{S_l} f(x) dx}{\int_{\bigcup_i S_i} f(x) dx} \le \frac{\eta_2}{N},\tag{2.5}$$

where $\eta_1 = \frac{\epsilon_1}{\epsilon_2} \le \frac{\epsilon_2}{\epsilon_1} = \eta_2$ (see (1.1)). Therefore if

$$N_l = \sum_{i=1}^n \mathbf{1}(X_i \in S_l)$$
 (2.6)

denotes the number of nodes of $\{X_j\}$ in the square S_l , then N_l is Binomially distributed with parameters n and p_l ; i.e., for any $1 \le k \le n$,

$$\mathbb{P}(N_l = k) = B(k; n, p_l) := \binom{n}{k} p_l^k (1 - p_l)^{n-k}, \tag{2.7}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the Binomial coefficient. Moreover,

$$\frac{\eta_1 n}{N} \le \mathbb{E} N_l = n p_l \le \frac{\eta_2 n}{N} \tag{2.8}$$

by (2.5).

Let $\{Y_j\}_{1\leq j\leq N_l}$ be the nodes of $\{X_j\}$ present in the square S_l . Formally, if $N_l=0$, set $\{Y_j\}_{1\leq j\leq N_l}:=\emptyset$. If $N_l\geq 1$, define N_l indices j_1,\ldots,j_{N_l} as follows. Let

$$j_1 = j_1(X_1, \dots, X_n) := \min\{1 \le k \le n : X_k \in S_l\}$$

be the least indexed node of $\{X_k\}$ present in S_l . Let

$$j_2 = \min\{j_1 + 1 \le k \le n : X_k \in S_l\}$$

be the next least indexed node of $\{X_k\}$ present in S_l and so on. Set $Y_i = X_{j_i}$ for $1 \le j \le N_l$.

Set $T_l = 0$ if $N_l \le 2$ and if $N_l \ge 3$ set

$$T_l := TSP(Y_1, \dots, Y_{N_l}; S_l)$$
 (2.9)

where TSP(.;.) is as defined in (2.1). The following is the main lemma proved in this subsection.

Lemma 3. If M > 0 is arbitrary and (1.7) holds, the following is true: There are positive constants $\{\delta_i\}_{1 \le i \le 3}$ such that for all $n \ge 2$ and for any $1 \le l \le N$,

$$\delta_1 r_n \sqrt{\frac{n}{N}} \le \mathbb{E} T_l \le \delta_2 r_n \sqrt{\frac{n}{N}} \quad and \quad \mathbb{E} T_l^2 \le \delta_3 \left(r_n \sqrt{\frac{n}{N}} \right)^2.$$
 (2.10)

Moreover, if

$$U_l = U_l(n) := \left\{ \frac{\eta_1 n}{2N} \le N_l \le \frac{2\eta_2 n}{N} \right\},$$
 (2.11)

where η_1 and η_2 are as in (2.5), then there are positive constants $\{\delta_i\}_{i=4,5}$ such that for all $n \geq 2$ and for any $1 \leq l \leq N$,

$$\mathbb{P}(U_l) \ge 1 - \exp\left(-\delta_4 \frac{n}{N}\right) \tag{2.12}$$

and

$$T_l \mathbf{1}(U_l) \le \delta_5 r_n \sqrt{\frac{n}{N}}.$$
 (2.13)

To prove the above Lemma, we perform some preliminary computations. We first derive bounds for the total number of squares N. From (1.7) we have that $r_n^2 \geq \frac{M \log n}{n}$ and since all the $r_n \times r_n$ squares $\{S_l\}_{1 \leq l \leq N}$ are contained within the unit square S, we also have $Nr_n^2 \leq 1$ and therefore $N \leq \frac{n}{M \log n}$. Similarly from (1.7) we also have that $\frac{n}{N^2} \longrightarrow 0$ as $n \to \infty$ and so $N \geq \sqrt{n}$ for all n large. Combining we get

$$\sqrt{n} \le N \le \frac{n}{M \log n} \text{ and } \frac{n}{N} \ge M \log n$$
(2.14)

for all n large.

For $k \geq 2$, let $D_l(k)$ be the expected minimum distance between the node Y_k and every other node in S_l , given that there are $N_l = k$ nodes in S_l ; i.e.,

$$D(k) = D_l(k) := \mathbb{E}\left(d(Y_k, \{Y_u\}_{1 \le u \le k-1}) | N_l = k\right), \tag{2.15}$$

where $d(A, B) = \min_{x \in A, y \in B} d(x, y)$ is the minimum distance between finite sets A and B. We have the following properties.

(b1) For any $k \geq 2$ and $1 \leq l \leq N$, the term

$$D_l(k) \ge \int_0^{\frac{r_n}{\sqrt{\delta}}} \left(1 - \pi \eta_2 \left(\frac{r}{r_n} \right)^2 \right)^{k-1} dr \tag{2.16}$$

where $\eta_2 = \frac{\epsilon_2}{\epsilon_1}$ is as in (2.5).

(b2) There are positive constants $\gamma_i, 1 \leq i \leq 3$ such that for any $k \geq 2$ and $1 \leq l \leq N$, the minimum distance

$$\gamma_1 \frac{r_n}{\sqrt{k}} \le D_l(k) \le \gamma_2 \frac{r_n}{\sqrt{k}} \text{ and } \mathbb{E}\left(d^2(Y_k, \{Y_u\}_{1 \le u \le k-1}) | N_l = k\right) \le \gamma_3 \frac{r_n^2}{k}.$$
(2.17)

Proof of (b1) - (b2): Given $N_l = k$, the nodes in S_l are independently distributed in S_l with distribution f; i.e.,

$$D_l(k) = \mathbb{E}d(Z_k, \{Z_i\}_{1 \le j \le k-1})$$
(2.18)

where $\{Z_i\}_{1 \leq i \leq k}$ are i.i.d. with distribution

$$\mathbb{P}(Z_1 \in A) = \frac{\int_{A \cap S_l} f(x) dx}{\int_{S_l} f(x) dx}.$$
 (2.19)

Use Fubini's theorem and (2.19) to write

$$D_l(k) = \frac{1}{\int_{S_l} f(x) dx} \int_{S_l} \mathbb{E}d(x, \{Z_j\}_{1 \le j \le k-1}) f(x) dx, \qquad (2.20)$$

where

$$\mathbb{E}d(x, \{Z_j\}_{1 \le j \le k-1}) = \int_0^\infty \mathbb{P}(d(x, \{Z_j\}_{1 \le j \le k-1}) \ge r) dr. \tag{2.21}$$

For any $x \in S_l$, the minimum distance from x to $\{Z_1, \ldots, Z_{k-1}\}$ is at least r if and only if $B(x,r) \cap S_l$ contains no point of $\{Z_j\}_{1 \leq j \leq k-1}$. Here B(x,r) is the ball of radius r centred at x. Wherever the point $x \in S_l$, the area

of $B(x,r) \cap S_l$ is at most πr^2 and so together with (1.1), we then get

$$\mathbb{P}(d(x, \{Z_j\}_{1 \le j \le k-1}) \ge r) = \left(1 - \frac{\int_{B(x,r) \cap S_l} f(x) dx}{\int_{S_l} f(x) dx}\right)^{k-1} \\
\ge \left(1 - \pi \eta_2 \frac{r^2}{r_n^2}\right)^{k-1}, \tag{2.22}$$

where $\eta_2 = \frac{\epsilon_2}{\epsilon_1}$ is as in (2.5).

To prove the lower bound for $D_l(k)$ in (2.17) of (b2), fix $k \geq 2$ and use (2.16) to get that

$$D_l(k) \ge \int_0^{\frac{r_n}{\sqrt{\delta k}}} \left(1 - \delta \left(\frac{r}{r_n} \right)^2 \right)^{k-1} dr \ge \int_0^{\frac{r_n}{\sqrt{\delta k}}} \left(1 - \frac{1}{k} \right)^{k-1} dr \ge \frac{e^{-1} r_n}{\sqrt{\delta k}}$$

for all n large. The final estimate is obtained by using $\left(1 - \frac{1}{r}\right)^{r-1} \ge e^{-1}$ for all $r \ge 2$.

For the upper bound for $D_l(k)$ in (2.17), again use (2.22) and the fact that $B(x,r) \cap S_l$ has area at least $\frac{\pi r^2}{4}$ no matter where the position of x, to get

$$\mathbb{P}(d(x, \{Z_j\}_{1 \le j \le k-1}) \ge r) \le \left(1 - \frac{\pi}{4\epsilon_1} \left(\frac{r}{r_n}\right)^2\right)^{k-1} \le \exp\left(-\frac{\pi(k-1)}{4\epsilon_1 r_n^2} r^2\right)$$

and so

$$D_l(k) \le \int_0^\infty \exp\left(-\frac{\pi(k-1)}{4\epsilon_1 r_n^2} r^2\right) dr \le \frac{Cr_n}{\sqrt{k-1}} \le \frac{2Cr_n}{\sqrt{k}}$$

for all $k \geq 2$ and for some positive constant C, not depending on k or l.

Finally for the second moment estimate in (2.17), we argue analogous to (2.15) and get that the term $\mathbb{E}(d^2(Y_k, \{Y_u\}_{1 \leq u \leq k-1}) | N_l = k)$ equals

$$\mathbb{E}d^{2}(Z_{k}, \{Z_{j}\}_{1 \leq j \leq k-1}) = \frac{1}{\int_{S_{l}} f(x) dx} \int_{S_{l}} \mathbb{E}d^{2}(x, \{Z_{j}\}_{1 \leq j \leq k-1}) f(x) dx \quad (2.23)$$

where $\{Z_i\}_{1\leq i\leq k}$ are i.i.d. with distribution as in (2.19). Arguing as in the

previous paragraph we get

$$\mathbb{E}(d^{2}(x, \{Z_{j}\}_{1 \leq j \leq k-1})) = \int r \mathbb{P}(d(x, \{Z_{j}\}_{1 \leq j \leq k-1}) \geq r) dr$$

$$\leq \int_{0}^{\infty} r \exp\left(-\frac{\pi(k-1)}{4\epsilon_{1}r_{n}^{2}}r^{2}\right) dr$$

$$\leq \frac{Cr_{n}^{2}}{k}$$

$$(2.24)$$

for some constant C > 0 not depending on k or x. Substituting (2.24) into (2.23) gives the desired bound for the second moment in (2.17).

Proof of Lemma 3: The proof of (2.12) follows from standard Binomial estimates and the estimate for $\mathbb{E}N_l$ in (2.8). The proof of (2.13) follows from the strips estimate (2.2) with $a = \frac{2\eta_2 n}{N}$ and $b = r_n$.

To prove the first estimate of (2.10) assume $N_l \geq 3$ and recall that $\{Y_u\}_{1 \leq u \leq N_l}$ are the nodes of the Binomial process in the square S_l (see paragraph prior to (2.15)). Let C_l denote the minimum length cycle of length T_l containing the nodes $\{Y_u\}_{1 \leq u \leq N_l}$. If $l(Y_u, C_l)$, $1 \leq u \leq N_l$ is the sum of length of the two edges containing Y_u as an endvertex then

$$l(Y_u, \mathcal{C}_l) \ge 2d(Y_u, \{Y_v\}_{v \ne u}),$$

the minimum distance of Y_u from all the other nodes in S_l as defined in (2.15). From (1.4),

$$T_l = L(\mathcal{C}_l) = \frac{1}{2} \left(\sum_{u=1}^{N_l} l(Y_u, \mathcal{C}_l) \right) \ge \left(\sum_{u=1}^{N_l} d(Y_u, \{Y_v\}_{v \ne u}) \right)$$

and so

$$\mathbb{E}T_{l} = \sum_{k \geq 3} \mathbb{E}T_{l}\mathbf{1}(N_{l} = k) \geq \mathbb{E}\sum_{k \geq 3} \sum_{u=1}^{k} d(Y_{u}, \{Y_{v}\}_{v \neq u})\mathbf{1}(N_{l} = k).$$
 (2.25)

Recalling the definition of $D_l(k)$ in (2.15) we further get

$$\mathbb{E}T_l = \sum_{k \ge 3} \mathbb{P}(N_l = k) k D_l(k) \ge \sum_{\frac{\eta_1 n}{2N} \le k \le \frac{2\eta_2 n}{N}} \mathbb{P}(N_l = k) k D_l(k), \qquad (2.26)$$

provided n is large enough so that

$$\frac{\eta_1 n}{2N} \ge \frac{\eta_1}{2} M \log n \ge 3,$$

the middle estimate being true because of (2.14).

Using the estimate $D_l(k) \geq \frac{\gamma_1 r_n}{\sqrt{k}}$ (see (2.17)) in (2.26) we then get

$$\mathbb{E}T_{l} \geq \gamma_{1}r_{n} \sum_{\frac{\eta_{1}n}{2N} \leq k \leq \frac{2\eta_{2}n}{N}} \mathbb{P}(N_{l} = k)\sqrt{k}$$

$$\geq \gamma_{1}r_{n}\sqrt{\frac{\eta_{1}n}{2N}} \sum_{\frac{\eta_{1}n}{2N} \leq k \leq \frac{2\eta_{2}n}{N}} \mathbb{P}(N_{l} = k)$$

$$\geq \gamma_{1}r_{n}\sqrt{\frac{\eta_{1}n}{2N}} \left(1 - \exp\left(-C\frac{n}{N}\right)\right), \qquad (2.27)$$

for some constant C > 0, by (2.12). Since $\frac{n}{N} \longrightarrow \infty$ as $n \to \infty$, (see (2.14)), we get the lower bound for $\mathbb{E}T_l$ from (2.27).

For the upper bound of $\mathbb{E}T_l$ in (2.10), we argue as follows. Recall that $T_l = L(\mathcal{C}_l)$ is the length of the minimum length cycle \mathcal{C}_l containing all the N_l nodes of $\{X_k\}$ in S_l . If the number of nodes $N_l \leq \frac{2\eta_2 n}{N}$, then from (2.13), we have that $T_l \leq Cr_n\sqrt{\frac{n}{N}}$ for some constant C > 0. If $N_l \geq \frac{2\eta_2 n}{N}$, then use the fact that T_l is bounded above by $N_l r_n \sqrt{2}$, since each edge in \mathcal{C}_l has both endvertices in the $r_n \times r_n$ square S_l and therefore has length at most $r_n\sqrt{2}$. Thus

$$\mathbb{E}T_{l} \leq Cr_{n}\sqrt{\frac{n}{N}} + r_{n}\sqrt{2}\mathbb{E}\left(N_{l}\mathbf{1}\left(N_{l} > \frac{2\eta_{2}n}{N}\right)\right) \leq Cr_{n}\sqrt{\frac{n}{N}} + r_{n}\sqrt{2}\mathbb{E}(N_{l}\mathbf{1}(U_{l}^{c})),$$
(2.28)

where U_l is as defined in (2.11).

Recall from discussion following (2.6) that N_l is Binomially distributed with parameters n and p_l and so by standard Binomial estimates

$$\mathbb{E}N_l^2 \le C(np_l)^2 \le \frac{Cn^2}{N^2} \tag{2.29}$$

for some constant C > 0, where the final estimate in (2.29) follows from the estimate for p_l in (2.5). Using Cauchy-Schwarz inequality we therefore get

$$\mathbb{E}N_l \mathbf{1}(U_l^c) \le \left(\mathbb{E}N_l^2\right)^{\frac{1}{2}} \left(\mathbb{P}(U_l^c)\right)^{\frac{1}{2}} \le C_1 \frac{n}{N} \exp\left(-C_2 \frac{n}{N}\right) \le \sqrt{\frac{n}{N}}, \tag{2.30}$$

for all n large and for some positive constants C_1, C_2 . The middle inequality in (2.30) follows from (2.12) and the final inequality in (2.30) is true since $\frac{n}{N} \longrightarrow \infty$ as $n \to \infty$ (see (2.14)). Substituting (2.30) into (2.28) gives the upper bound for $\mathbb{E}T_l$ in (2.10). The proof of the bound for $\mathbb{E}T_l^2$ is analogous as above.

Define the covariance between T_{l_1} and T_{l_2} for distinct l_1 and l_2 as

$$cov(T_{l_1}, T_{l_2}) = \mathbb{E}T_{l_1}T_{l_2} - \mathbb{E}T_{l_1}\mathbb{E}T_{l_2}.$$
(2.31)

We need the following result for future use. Recall the definition of ϵ_1 and ϵ_2 in (1.1).

Lemma 4. There is a positive constant $M_0 = M_0(\epsilon_1, \epsilon_2)$ large so that the following holds if (1.7) is satisfied with $M > M_0$: There are positive constants C_1, C_2 such that for all $n \geq 2$ and for any $1 \leq l_1 \neq l_2 \leq N$,

$$|cov(T_{l_1}, T_{l_2})| \le C_1 \left(\mathbb{E} T_{l_1} \mathbb{E} T_{l_2} \right) \frac{n}{N^2} \le C_2 \frac{r_n^2 n^2}{N^3}.$$
 (2.32)

To prove Lemma 4, we use Poissonization described in the next subsection.

Poissonization

Recall from discussion prior to (1.7) that $n \geq 1$ nodes $\{X_k\}_{1 \leq k \leq n}$ are distributed across the $r_n \times r_n$ squares $\{S_j\}_{1 \leq j \leq N}$ according to a Binomial process with intensity $g_N(.)$ as defined in (1.3). Throughout, we use Poissonization as a tool to obtain estimates for probabilities of events for the corresponding Binomial process. We make precise the notions in this subsection.

Let \mathcal{P} be a Poisson process on the squares $\bigcup_{j=1}^{N} S_j$ with intensity function $ng_N(.)$ defined on the probability space $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$. If $N_l^{(P)}$ be the number of nodes of \mathcal{P} present in the square S_l , $1 \leq l \leq N$, then

$$\mathbb{P}_0(N_l^{(P)} = k) = Poi(k; np_l) := e^{-np_l} \frac{(np_l)^k}{k!}, \tag{2.33}$$

where p_l is as defined in (2.5). Moreover,

$$\frac{\eta_1 n}{N} \le \mathbb{E}_0 N_l^{(P)} = n p_l \le \frac{\eta_2 n}{N} \tag{2.34}$$

by (2.5).

Let $\{Y_j\}_{1 \leq j \leq N_l^{(P)}}$ be the nodes of \mathcal{P} present in the square S_l . Analogous to (2.9), set $T_l^{(P)} = 0$ if $N_l^{(P)} \leq 2$ and if $N_l^{(P)} \geq 3$ set

$$T_l^{(P)} := TSP(Y_1, \dots, Y_{N_l^{(P)}}; S_l)$$
 (2.35)

where TSP(.;.) is as defined in (2.1). The following result is analogous to Lemma 3.

Lemma 5. If M > 0 is arbitrary and (1.7) holds, then the following is true: There are positive constants $\{\delta_i\}_{1 \leq i \leq 3}$ such that for all $n \geq 2$ and for any $1 \leq l \leq N$,

$$\delta_1 r_n \sqrt{\frac{n}{N}} \le \mathbb{E}_0 T_l^{(P)} \le \delta_2 r_n \sqrt{\frac{n}{N}}, \quad \mathbb{E}_0 \left(T_l^{(P)} \right)^2 \le \delta_3 \left(r_n \sqrt{\frac{n}{N}} \right)^2$$
 (2.36)

and

$$\mathbb{P}_0\left(T_l^{(P)} \ge \delta_4 r_n \sqrt{\frac{n}{N}}\right) \ge \delta_5. \tag{2.37}$$

Proof of Lemma 5: The proof of (2.36) is analogous as in the Binomial case and proceeds as follows. Define

$$U_l^{(P)} = U_l^{(P)}(n) := \left\{ \frac{\eta_1 n}{2N} \le N_l^{(P)} \le \frac{2\eta_2 n}{N} \right\}, \tag{2.38}$$

where η_1 and η_2 are as in (2.5). Analogous to (2.12), the following bound is obtained by standard Poisson distribution estimates: There is a positive constant γ such that for all $n \geq 2$ and for any $1 \leq l \leq N$,

$$\mathbb{P}_0\left(U_l^{(P)}\right) \ge 1 - \exp\left(-\gamma \frac{n}{N}\right). \tag{2.39}$$

As in the Binomial case, given $N_l^{(P)}=k$, the nodes of \mathcal{P} are i.i.d. distributed according to distribution (2.19). Therefore for $k\geq 2$ we let

$$D_l^{(P)}(k) = \mathbb{E}_0 \left(d(Y_k, \{Y_j\}_{1 \le j \le k-1}) | N_l^{(P)} = k \right)$$

and as in (2.15) obtain that

$$D_l^{(P)}(k) = \mathbb{E}d(Z_k, \{Z_j\}_{1 \le j \le k-1}) = D_l(k), \tag{2.40}$$

where $D_l(k)$ is as defined in (2.15), the random variables $\{Z_j\}_{1 \leq j \leq k}$ are i.i.d. with distribution (2.19) and the final equality in (2.40) is true because of (2.18). Consequently $D_l^{(P)}(k)$ also satisfies properties (b1) - (b2) and the rest of the proof of (2.36) is analogous to the Binomial case.

Finally, the estimate in (2.37) is obtained by using (2.36) and the Paley-Zygmund inequality

$$\mathbb{P}_0\left(T_l^{(P)} \ge \lambda \mathbb{E}_0 T_l^{(P)}\right) \ge (1 - \lambda)^2 \frac{(\mathbb{E}_0 T_l^{(P)})^2}{\mathbb{E}_0\left(T_l^{(P)}\right)^2}$$

for
$$0 < \lambda < 1$$
.

We now use Poissonization and obtain intermediate estimates needed to prove Lemma 4. Recall from (2.9) and (2.35) that T_l and $T_l^{(P)}$ are the lengths of the minimum length cycles containing all the nodes in the $r_n \times r_n$ square S_l , $1 \le l \le N$ in the Binomial and the Poisson process, respectively. Recall the definition of ϵ_1 and ϵ_2 in (1.1).

Lemma 6. There is a positive constant $M_0 = M_0(\epsilon_1, \epsilon_2)$ large so that the following holds if (1.7) is satisfied with $M > M_0$: There are positive constants C_0, C_1 and C_2 such that for all $n \ge C_0$ and for any $1 \le l \le N$,

$$|\mathbb{E}T_l - \mathbb{E}_0 T_l^{(P)}| \le C_1 (\mathbb{E}T_l) \left(\frac{n}{N^2}\right) \le C_2 \left(\frac{r_n n^{3/2}}{N^{5/2}}\right).$$
 (2.41)

Moreover, for any $1 \le l_1 \ne l_2 \le N$

$$|\mathbb{E}(T_{l_1}T_{l_2}) - \mathbb{E}_0(T_{l_1}^{(P)}T_{l_2}^{(P)})| \le C_1 \left(\mathbb{E}T_{l_1}\mathbb{E}T_{l_2}\right) \left(\frac{n}{N^2}\right) \le C_2 \left(\frac{r_n^2 n^2}{N^3}\right). \quad (2.42)$$

To prove Lemma 6, we need estimates on the difference between Binomial and Poisson distributions. For $k, l \geq 1$ recall the Binomial distribution $B(k; n, p_l)$ and the Poisson distribution $Poi(k; np_l)$ as defined in (2.7) and (2.33), respectively. For $k_1, k_2, l_1, l_2 \geq 1$, let

$$B(k_1, k_2; n, p_{l_1}, p_{l_2}) := \binom{n}{k_1, k_2} p_{l_1}^{k_1} p_{l_2}^{k_2} (1 - p_{l_1} - p_{l_2})^{n - k_1 - k_2}, \tag{2.43}$$

where $\binom{n}{k_1,k_2} = \frac{n!}{k_1!k_2!(n-k_1-k_2)!}$. We have the following properties. (c1) There is a constant C>0 such that for all $n\geq 3,\ 1\leq l\leq N$ and $\frac{\eta_1n}{2N}\leq 1$

 $k \leq \frac{2\eta_2 n}{N}$,

$$|B(k; n, p_l) - Poi(k; np_l)| \le Poi(k; np_l) \left(1 + \frac{Cn}{N^2}\right). \tag{2.44}$$

(c2) There is a constant C>0 such that for all $n\geq 3$, and for any $1\leq l_1, l_2\leq N$ and $\frac{\eta_1 n}{2N}\leq k_1, k_2\leq \frac{2\eta_2 n}{N},$

$$|B(k_1, k_2; n, p_{l_1}, p_{l_2}) - Poi(k_1; np_{l_1}) Poi(k_2; np_{l_2})|$$

$$\leq Poi(k_1; np_{l_1}) Poi(k_2; np_{l_2}) \left(1 + \frac{Cn}{N^2}\right). \tag{2.45}$$

Proof of (c1) - (c2): To prove (2.44) in (c1), we write $p_l = p$ for simplicity. Use $\binom{n}{k} \leq \frac{n^k}{k!}$ and $1 - x \leq e^{-x}$ for 0 < x < 1 to get

$$\binom{n}{k} p^k (1-p)^{n-k} \le \frac{(np)^k}{k!} e^{-p(n-k)} = Poi(k; np) e^{kp}.$$

Using (2.5) and the fact that $k \leq \frac{2\eta_2}{N}$ we get

$$e^{kp} \le \exp\left(\frac{k\eta_2 n}{N}\right) \le \exp\left(2\eta_2 \frac{n}{N^2}\right)$$

and since

$$e^x = 1 + x + \sum \frac{x^k}{k!} \le 1 + x + \sum_{k>2} x^k \le 1 + 2x$$
 (2.46)

for all x small, we get $e^{kp} \le 1 + \frac{4\eta_2 n}{N^2}$, proving the upper bound in (2.44). To obtain a lower bound, we use the estimate

$$1 - x \ge e^{-x - x^2} \tag{2.47}$$

for all $0 < x < \frac{1}{2}$. To prove (2.47), write $\log(1-x) = -x - R(x)$ where

$$R(x) = \sum_{k>2} \frac{x^k}{k} \le \frac{1}{2} \sum_{k>2} x^k = \frac{x^2}{2(1-x)} \le x^2$$

since $x < \frac{1}{2}$. Use $\binom{n}{k} \ge \frac{(n-k)^k}{k!}$ and (2.47) to get

$$B(k; n, p) \ge \frac{1}{k!} (n - k)^k p^k e^{-p(n-k) - p^2(n-k)} = Poi(k; np) \left(1 - \frac{k}{n}\right)^k e^{kp - (n-k)p^2}$$
(2.48)

As before, using the fact that $\frac{\eta_1 n}{2N} \leq k \leq \frac{2\eta_2 n}{N}$ we get

$$\left(1 - \frac{k}{n}\right)^k \ge 1 - \frac{k^2}{n} \ge 1 - \frac{4\eta_2^2 n}{N^2} \tag{2.49}$$

and using (2.5) we get

$$kp - (n-k)p^2 \ge kp - np^2 \ge \frac{\eta_1 n}{2N} \frac{\eta_1}{N} - n\left(\frac{\eta_2}{N}\right)^2 = -\eta \frac{n}{N^2}$$
 (2.50)

where $\eta = \eta_2^2 - \frac{\eta_1^2}{4} > 0$, since $\epsilon_1 \leq \epsilon_2$ and so $\eta_1 = \frac{\epsilon_1}{\epsilon_2} \leq \frac{\epsilon_2}{\epsilon_1} = \eta_2$. Using (2.49) and (2.50) into (2.48) gives

$$B(k; n, p) \geq Poi(k; np) \left(1 - \frac{\eta_1^2}{4} \frac{n}{N^2}\right) \exp\left(-\eta \frac{n}{N^2}\right)$$
$$\geq Poi(k; np) \left(1 - \frac{\eta_1^2}{4} \frac{n}{N^2}\right) \left(1 - \eta \frac{n}{N^2}\right),$$

since $e^{-x} \ge 1 - x$ for 0 < x < 1. This proves (2.44).

To prove (2.45), write $p_{l_1} = p_1, p_{l_2} = p_2$ and $B_{12} = B(k_1, k_2; n, p_1, p_2)$ for simplicity. Use

$$\binom{n}{k_1, k_2} = \frac{1}{k_1! k_2!} n(n-1) \dots (n-k_1 - k_2 + 1) \le \frac{n^{k_1 + k_2}}{k_1! k_2!}$$
(2.51)

to get

$$B_{12} \le \frac{(np_1)^{k_1}}{k_1!} \frac{(np_2)^{k_2}}{k_2!} e^{-(p_1+p_2)n} e^{(p_1+p_2)(k_1+k_2)}. \tag{2.52}$$

Using (2.5), we get $p_1 + p_2 \leq \frac{2\eta_2}{N}$ and since $k_1, k_2 \leq \frac{2\eta_2 n}{N}$ we get using (2.46) that

$$e^{(p_1+p_2)(k_1+k_2)} \le \exp\left(\frac{4\eta_2^2 n}{N^2}\right) \le 1 + \frac{8\eta_2^2 n}{N^2}$$
 (2.53)

for all n large, since $\frac{n}{N^2} \longrightarrow 0$ as $n \to \infty$ (see (1.7)). Substituting (2.53) into (2.52), we get the upper bound for B_{12} in (2.45).

For the lower bound for B_{12} again use (2.51) to get

$$\binom{n}{k_1, k_2} \ge \frac{1}{k_1! k_2!} (n - k_1 - k_2)^{k_1 + k_2} = \frac{n^{k_1 + k_2}}{k_1! k_2!} \left(1 - \frac{k_1 + k_2}{n} \right)^{k_1 + k_2}.$$

Using $(1-x)^r \ge 1 - rx$ for r, x > 0 we further get

$$\binom{n}{k_1, k_2} \ge \frac{n^{k_1 + k_2}}{k_1! k_2!} \left(1 - \frac{(k_1 + k_2)^2}{n} \right) \ge \frac{n^{k_1 + k_2}}{k_1! k_2!} \left(1 - \frac{4\eta_2^2 n}{N^2} \right) \tag{2.54}$$

since $k_1, k_2 \leq \frac{2\eta_2 n}{N}$. Substituting (2.54) into (2.43) we get

$$B_{12} \ge \frac{(np_1)^{k_1}}{k_1!} \frac{(np_2)^{k_2}}{k_2!} (1 - p_1 - p_2)^{n - k_1 - k_2} \left(1 - \frac{4\eta_2^2 n}{N^2} \right). \tag{2.55}$$

To evaluate $(1 - p_1 - p_2)^{n-k_1-k_2}$, we use the estimate (2.47) which is applicable since from (2.5), we have

$$p_1 + p_2 \le \frac{2\eta_2}{N} \le \frac{2\eta_2}{\sqrt{n}} \longrightarrow 0$$

as $n \to \infty$ (see (2.14)). Using (2.47), we get

$$(1 - p_1 - p_2)^{n - k_1 - k_2} \ge e^{-(p_1 + p_2)(n - k_1 - k_2) - (p_1 + p_2)^2(n - k_1 - k_2)} = e^{-np_1} e^{-np_2} e^{I_1 - I_2},$$
(2.56)

where

$$I_1 = (p_1 + p_2)(k_1 + k_2) \ge 0 (2.57)$$

and

$$I_2 = (p_1 + p_2)^2 (n - k_1 - k_2) \le n(p_1 + p_2)^2 \le \frac{\eta_2^2 n}{N^2}$$
 (2.58)

for some constant $C_1 > 0$. The final estimate in (2.58) follows from the fact that $p_1 + p_2 \leq \frac{2\eta_2 n}{N}$ (see (2.5)). Using $e^{-x} \geq 1 - x$ we get

$$e^{I_1 - I_2} \ge e^{-I_2} \ge 1 - \frac{\eta_2^2 n}{N^2}$$
 (2.59)

and substituting (2.59) into (2.56), we

$$(1 - p_1 - p_2)^{n - k_1 - k_2} \ge e^{-np_1} e^{-np_2} \left(1 - \frac{\eta_2^2 n}{N^2} \right). \tag{2.60}$$

Using (2.60) in (2.55), we get the lower bound for B_{12} in (2.45).

Using properties (c1) - (c2) we prove Lemma 6. Proof of (2.41) in Lemma 6: Recall from (2.6) that N_l is the number of nodes of the Binomial process $\{X_k\}$ in the square S_l and let U_l be the event as defined in (2.11). Write

$$\mathbb{E}T_l = I_1 + I_2 \tag{2.61}$$

where

$$I_1 = \mathbb{E}T_l \mathbf{1}(U_l) = \sum_{\frac{\eta_1 n}{2N} \le k \le \frac{2\eta_2 n}{N}} \mathbb{E}T_l \mathbf{1}(N_l = k),$$

 $I_2 = \mathbb{E}T_l \mathbf{1}(U_l^c)$ and η_1, η_2 are as in (2.5). Similarly

$$\mathbb{E}_0 T_I^{(P)} = I_1^{(P)} + I_2^{(P)} \tag{2.62}$$

where

$$I_1^{(P)} = \mathbb{E}_0(T_l^{(P)}\mathbf{1}(U_l^{(P)})),$$

$$I_2^{(P)} = \mathbb{E}_0(T_l^{(P)}\mathbf{1}(U_l^{(P)})^c), \ U_l^{(P)} = \left\{\frac{\eta_1 n}{2N} \le N_l^{(P)} \le \frac{2\eta_2 n}{N}\right\}$$
 is as defined in (2.38)

and $N_l^{(P)}$ is the number of nodes of the Poisson process \mathcal{P} inside the square S_l (see discussion prior to (2.33)).

From (2.61) and (2.62), we therefore get

$$|\mathbb{E}T_l - \mathbb{E}_0 T_l^{(P)}| \le |I_1 - I_1^{(P)}| + I_2 + I_2^{(P)}.$$
 (2.63)

The remainder terms I_2 and $I_2^{(P)}$ satisfy

$$\max(I_2, I_2^{(P)}) \le C(\mathbb{E}T_l) \frac{n}{N^2}$$
 (2.64)

for some constant C > 0. We prove (2.64) for I_2 and an analogous proof holds for $I_2^{(P)}$. Indeed, every edge in the minimum length cycle \mathcal{C}_l containing all the nodes in the $r_n \times r_n$ square S_l has both endvertices within S_l and so has length at most $r_n \sqrt{2}$. Since there are N_l nodes in the square S_l , we must have $T_l \leq N_l r_n \sqrt{2}$ and so

$$I_2 = \mathbb{E}T_l \mathbf{1}(U_l^c) \le r_n \sqrt{2} \mathbb{E}N_l \mathbf{1}(U_l^c). \tag{2.65}$$

Using the third expression in (2.30) to estimate $\mathbb{E}N_l\mathbf{1}(U_l^c)$ we get

$$I_2 \le C_1 r_n \sqrt{2} \frac{n}{N} \exp\left(-C_2 \frac{n}{N}\right) = C_1 \sqrt{2} \left(r_n \sqrt{\frac{n}{N}}\right) \left(\sqrt{\frac{n}{N}} \exp\left(-C_2 \frac{n}{N}\right)\right) \tag{2.66}$$

for some constants $C_1, C_2 > 0$. From the lower bound in (2.10) we have $\mathbb{E}T_l \ge C_3 r_n \sqrt{\frac{n}{N}}$ and so

$$I_{2} \leq C_{4} (\mathbb{E}T_{l}) \left(\sqrt{\frac{n}{N}} \exp\left(-C_{2} \frac{n}{N}\right) \right)$$

$$= C_{4} (\mathbb{E}T_{l}) \left(\frac{n}{N^{2}}\right) \left(\left(\frac{N^{3}}{n}\right) \exp\left(-\frac{C_{2}n}{2N}\right)\right)^{\frac{1}{2}}$$
(2.67)

Using the upper bound $N \leq \frac{n}{M \log n}$ from (2.14), we have

$$\left(\frac{N^3}{n}\right) \exp\left(-\frac{C_2 n}{2N}\right) \le \frac{n^2}{M^3 (\log n)^3} \exp\left(-\frac{C_2 M}{2} \log n\right) \le 1 \tag{2.68}$$

for all n large, provided M > 0 large. Fixing such an M and using (2.68) in (2.67), we get (2.64).

To estimate the difference $I_1 - I_1^{(P)}$ in (2.63), recall that given $N_l = k$, the nodes in S_l are independently distributed in S_l with distribution $\frac{f(.)}{\int_{S_l} f(x) dx}$ (see (2.19)) and so

$$I_{1} = \sum_{\frac{\eta_{1}n}{2N} \le k \le \frac{2\eta_{2}n}{N}} \mathbb{P}(N_{l} = k) \mathbb{E}(T_{l} | N_{l} = k) = \sum_{\frac{\eta_{1}n}{2N} \le k \le \frac{2\eta_{2}n}{N}} B(k; n, p_{l}) \Delta(k, q_{l})$$
(2.69)

where $B(k; n, p_l)$ is the Binomial probability distribution as defined in (2.7), $q_l = \int_{S_l} f(x) dx$,

$$\Delta(k, q_l) = \mathbb{E}(T_l | N_l = k) = \int_{S_l} TSP(z_1, \dots, z_k; S_l) \frac{f(z_1)}{q_l} \dots \frac{f(z_k)}{q_l} dz_1 \dots dz_k$$
(2.70)

and $TSP(z_1, \ldots, z_k; S_l)$ is the minimum length of a cycle containing all the nodes $z_1, \ldots, z_k \in S_l$ (see (2.1)).

Similarly, as argued in (2.40), given $N_l^{(P)} = k$, the nodes of the Poisson process \mathcal{P} are also distributed in S_l according to distribution $\frac{f(.)}{\int_{S_l} f(x) dx}$. Therefore

$$\mathbb{E}(T_l^{(P)}|N_l^{(P)}=k)=\Delta(k,q_l)$$

as defined in (2.70) and so

$$I_1^{(P)} = \sum_{\frac{\eta_1 n}{2N} \le k \le \frac{2\eta_2 n}{N}} \Delta(k, q_l) Poi(k; np_l),$$
 (2.71)

where $Poi(k; np_l)$ is the Poisson distribution as defined in (2.33). From (2.69) and (2.71), we therefore get

$$|I_1 - I_1^{(P)}| \le \sum_{\frac{\eta_1 n}{2N} \le k \le \frac{2\eta_2 n}{N}} \Delta(k, q_l) |B(k; n, p_l) - Poi(k; np_l)|.$$
 (2.72)

Using estimate (2.44) of property (c1) to approximate the Binomial distribution with the Poisson distribution, we get

$$|I_{1} - I_{1}^{(P)}| \leq C_{1} \left(\sum_{\frac{\eta_{1}n}{2N} \leq k \leq \frac{2\eta_{2}n}{N}} Poi(k; np_{l}) \Delta(k, q_{l}) \right) \frac{n}{N^{2}}$$

$$\leq C_{1} \left(\sum_{k \geq 0} Poi(k; np_{l}) \Delta(k, q_{l}) \right) \frac{n}{N^{2}}$$

$$= C_{1} \left(\mathbb{E}_{0}(T_{l}^{(P)}) \right) \frac{n}{N^{2}}$$

$$(2.73)$$

for some constant $C_1 > 0$. Finally, from (2.10) and (2.36), we obtain that both $\mathbb{E}_0(T_l^{(P)})$ and $\mathbb{E}T_l$ are bounded above and below by constant multiples of $r_n\sqrt{\frac{n}{N}}$ and so $\mathbb{E}_0(T_l^{(P)}) \leq C_2\mathbb{E}T_l$ for some constant $C_2 > 0$ and from (2.73), we therefore get

$$|I_1 - I_1^{(P)}| \le C_3 \left(\mathbb{E}T_l\right) \frac{n}{N^2}$$
 (2.74)

for some constant $C_3 > 0$. Substituting (2.74) and (2.64) into (2.63) gives

$$|\mathbb{E}T_l - \mathbb{E}_0 T_l^{(P)}| \le C_4 (\mathbb{E}T_l) \frac{n}{N^2} \le C_5 \left(\frac{r_n n^{3/2}}{N^{5/2}}\right),$$

for some positive constants C_4, C_5 , again using the upper bound for $\mathbb{E}T_l$ from (2.10). This proves (2.41).

Proof of (2.42) of Lemma 6: Recall the definition of U_l in (2.11) and write

$$\mathbb{E}T_{l_1}T_{l_2} = J_1 + J_2, \tag{2.75}$$

where $J_1 = \mathbb{E}T_{l_1}T_{l_2}\mathbf{1}(U_{l_1}\cap U_{l_2})$ and $J_2 = \mathbb{E}T_{l_1}T_{l_2}\mathbf{1}(U_{l_1}^c \cup U_{l_2}^c)$. Similarly, for the Poisson case let $U_l^{(P)}$ be the event defined in (2.38) and write

$$\mathbb{E}_0 T_{l_1}^{(P)} T_{l_2}^{(P)} = J_1^{(P)} + J_2^{(P)}, \tag{2.76}$$

where $J_1^{(P)} = \mathbb{E}_0 T_{l_1}^{(P)} T_{l_2}^{(P)} \mathbf{1}(U_{l_1}^{(P)} \cap U_{l_2}^{(P)})$ and $J_2^{(P)} = \mathbb{E}_0 T_{l_1}^{(P)} T_{l_2}^{(P)} \mathbf{1}(U_{l_1}^{(P)} \cup U_{l_2}^{(P)})^c$. From (2.75) and (2.76), we get

$$|\mathbb{E}T_{l_1}R_{l_2} - \mathbb{E}_0 T_{l_1}^{(P)} T_{l_2}^{(P)}| \le |J_1 - J_1^{(P)}| + J_2 + J_2^{(P)}. \tag{2.77}$$

The remainder terms J_2 and $J_2^{(P)}$ satisfy

$$\max(J_2, J_2^{(P)}) \le C_1(\mathbb{E}T_{l_1}\mathbb{E}T_{l_2}) \frac{n}{N^2} \le C_2\left(\frac{r_n^2 n^2}{N^3}\right)$$
 (2.78)

for some constants $C_1, C_2 > 0$. We prove (2.78) for J_2 and an analogous proof holds for $J_2^{(P)}$. As argued in the proof of (2.64), every one of the N_{l_1} edges in the minimum length cycle C_{l_1} of length T_{l_1} has both endvertices within S_{l_1} and so has length at most $r_n\sqrt{2}$. Therefore

$$J_2 = \mathbb{E}T_{l_1}T_{l_2}\mathbf{1}(U_{l_1}^c \cup U_{l_2}^c) \le \left(r_n\sqrt{2}\right)^2 \mathbb{E}N_{l_1}N_{l_2}\mathbf{1}(U_{l_1}^c \cup U_{l_2}^c). \tag{2.79}$$

Using Cauchy-Schwarz inequality,

$$\mathbb{E}N_{l_1}N_{l_2}\mathbf{1}(U_{l_1}^c \cup U_{l_2}^c) \le \left(\mathbb{E}N_{l_1}^2 N_{l_2}^2\right)^{\frac{1}{2}} \mathbb{P}\left(U_{l_1}^c \cup U_{l_2}^c\right)^{\frac{1}{2}}$$
(2.80)

and using the estimate (2.12), we have

$$\mathbb{P}\left(U_{l_1}^c \cup U_{l_2}^c\right) \le \mathbb{P}\left(U_{l_1}^c\right) + \mathbb{P}\left(U_{l_2}^c\right) \le 2\exp\left(-4C\frac{n}{N}\right) \tag{2.81}$$

for some constant C > 0 and for all n large.

To evaluate $\mathbb{E}N_{l_1}^2N_{l_2}^2$, use $ab \leq \frac{a^2+b^2}{2}$ to get

$$\mathbb{E}N_{l_1}^2 N_{l_2}^2 \le \frac{1}{2} \left(\mathbb{E}N_{l_1}^4 + \mathbb{E}N_{l_2}^4 \right) \tag{2.82}$$

and use the fact that the term N_l is Binomially distributed with parameters n and p_l , where $p_l \leq \frac{\eta_2}{N}$ (see (2.5)) and η_2 does not depend on l or n. Therefore

$$\mathbb{E}N_l^4 \le C_1(np_l)^4 \le C_2\left(\frac{n}{N}\right)^4$$

for some constants C_1, C_2 not depending on l or n and so from (2.82) we get

$$\mathbb{E}N_{l_1}^2 N_{l_2}^2 \le C_3 \left(\frac{n}{N}\right)^4. \tag{2.83}$$

Using (2.83) and (2.81) in (2.80) we get

$$\mathbb{E}N_{l_1}N_{l_2}\mathbf{1}(U_{l_1}^c \cup U_{l_2}^c) \le C_4 \left(\frac{n}{N}\right)^2 \exp\left(-2C\frac{n}{N}\right). \tag{2.84}$$

Substituting (2.84) into (2.79) gives (2.42).

$$J_2 \le C_5 r_n^2 \left(\frac{n}{N}\right)^2 \exp\left(-2C\frac{n}{N}\right) = C_5 \left(\frac{r_n^2 n^2}{N^3}\right) N \exp\left(-2C\frac{n}{N}\right). \tag{2.85}$$

Since $N \leq \frac{n}{M \log n}$ (see (2.14)) we have that

$$N \exp\left(-2C\frac{n}{N}\right) \le \frac{n}{M \log n} \exp\left(-2CM \log n\right) \le 1$$

for all n large provided M > 0 is large. Fixing such an M, we get (2.78).

To evaluate the difference $J_1 - J_1^{(P)}$, recall from discussion prior to (2.69) that given $N_l = k$, the nodes of the Binomial process are distributed in the square S_l with distribution (2.19). Similarly, given $N_l^{(P)} = k$, the nodes of the Poisson process are also distributed according to (2.19). Therefore analogous to (2.72) we get

$$|J_1 - J_1^{(P)}| = \sum_{\frac{\eta_1 n}{2N} \le k_1, k_2 \le \frac{2\eta_2 n}{N}} |B_{l_1, l_2} - Poi(k_1; np_{l_1}) Poi(k_2; np_{l_2}) |\Delta(k_1, q_{l_1}) \Delta(k_2, q_{l_2})|$$

where q_{l_1}, q_{l_2} and $\Delta(., .)$ are as defined in (2.70) and $B_{l_1, l_2} = B(k_1, k_2; n, p_{l_1}, p_{l_2})$ is as defined in (2.43).

Since k_1 and k_2 are both of the order of $\frac{n}{N}$, we get from (2.45) that

$$|B_{l_1,l_2} - Poi(k_1; np_{l_1})Poi(k_2; np_{l_2})| \le \delta Poi(k_1, np_{l_1})Poi(k_2; np_{l_2})\frac{n}{N^2}$$

for some constant $\delta > 0$ not depending on n, k_1, k_2, l_1 or l_2 . Using this in (2.86) and arguing as in (2.73) we then get

$$|J_1 - J_1^{(P)}| \le C \mathbb{E}_0(T_{l_1}^{(P)}) \mathbb{E}_0(T_{l_2}^{(P)}) \left(\frac{n}{N^2}\right)$$

for some constant C > 0. Using the upper bound $\mathbb{E}_0(T_{l_1}^{(P)}) \leq C_1 r_n \sqrt{\frac{n}{N}}$ for some constant C_1 not depending on l_1 (see (2.36)), we then get

$$|J_1 - J_1^{(P)}| \le C_2 \frac{r_n^2 n^2}{N^3}. (2.87)$$

Substituting (2.87) and (2.78) into (2.76) gives the final estimate in (2.42). The middle estimate in (2.42) follows from the bounds for $\mathbb{E}T_l$ in (2.10).

Proof of Lemma 4: Since the Poisson process \mathcal{P} is independent on disjoint subsets, we have

$$cov_0(T_{l_1}^{(P)}, T_{l_2}^{(P)}) = \mathbb{E}_0(T_{l_1}^{(P)} T_{l_2}^{(P)}) - \mathbb{E}_0 T_{l_1}^{(P)} \mathbb{E}_0 T_{l_2}^{(P)} = 0.$$

Therefore write

$$|cov(T_{l_1}, T_{l_2})| = |cov(T_{l_1}, T_{l_2}) - cov_0(T_{l_1}^{(P)}, T_{l_2}^{(P)})| \le Z_1 + Z_2 + Z_3,$$

where

$$Z_{1} = |\mathbb{E}T_{l_{1}}T_{l_{2}} - \mathbb{E}_{0}T_{l_{1}}^{(P)}T_{l_{2}}^{(P)}| \leq C\left(\frac{r_{n}^{2}n^{2}}{N^{3}}\right),$$

$$Z_{2} = |\mathbb{E}_{0}^{(P)}T_{l_{1}}\mathbb{E}_{0}^{(P)}T_{l_{2}} - \mathbb{E}T_{l_{1}}\mathbb{E}T_{l_{2}}| \leq Z_{3} + Z_{4},$$

$$Z_{3} = |\mathbb{E}_{0}T_{l_{1}}^{(P)} - \mathbb{E}T_{l_{1}}|\mathbb{E}_{0}T_{l_{2}}^{(P)} \leq C\left(\frac{r_{n}n^{3/2}}{N^{5/2}}\right)\left(r_{n}\sqrt{\frac{n}{N}}\right) = C\left(\frac{r_{n}^{2}n^{2}}{N^{3}}\right)$$

and similarly,

$$Z_4 = \mathbb{E}T_{l_1}|\mathbb{E}_0T_{l_2}^{(P)} - \mathbb{E}T_{l_2}| \le C\frac{r_n^2n^2}{N^3},$$

for some constant C > 0. The estimate for Z_1 follows from (2.42) and the estimates for Z_3 and Z_4 follow from (2.41) and the estimates for $\mathbb{E}T_l$ and $\mathbb{E}_0T_l^{(P)}$ in (2.10) and (2.36), respectively.

3 Proof of Theorem 1

For $1 \leq l \leq N$, recall from (2.13) that T_l is the length of the minimum length cycle C_l containing all the nodes of $\{X_k\}$ contained in the square S_l . Also we have from Section 1 that s_n denotes the minimum distance between two squares in $\{S_l\}_{1\leq l\leq N}$. If the squares in $\{S_l\}$ are sufficiently far apart it is intuitive to expect that the overall minimum length cycle C_{tot} containing all the nodes of $\{X_k\}$ is simply obtained by merging together the cycles C_l . In other words, it is reasonable to expect that C_{tot} "covers" all nodes of a particular square before "proceeding" to the next square. However, we give

a small argument below to see that this is not necessarily true if the total number of nodes n is large enough.

Suppose the intercity distance $s_n = 10r_n$ and $r_n = \sqrt{M \frac{\log n}{n}}$ for some large constant M > 0. If all the $r_n \times r_n$ squares in Figure 1 are populated with nodes, then total number of squares N satisfies

$$C_1 \frac{n}{\log n} \le \frac{1}{(20r_n)^2} \le N \le \left(\frac{1}{r_n}\right)^2 \le C_2 \frac{n}{\log n}$$

for some constants $C_1, C_2 > 0$. Condition (1.7) is therefore satisfied and so the estimates for the expected length of T_l in Lemma 3 hold. From (2.36) we therefore have that

$$\mathbb{E}T_l \ge C_3 r_n \sqrt{\frac{n}{N}} \ge C_4 r_n \sqrt{\log n}$$

for some constants $C_3, C_4 > 0$. In other words, the expected total length of a cycle containing all the nodes of S_l is much larger than the intercity distance s_n . Therefore it is quite possible that the cycle C_{tot} locally crosses between two squares s_n apart multiple times.

We now allow s_n and r_n to be general as in the statement of the Theorem 1 and show that the length $TSPC_n$ of the minimum length cycle C_{tot} is well approximated by $\sum_{l=1}^{N} T_l$.

Lemma 7. The overall minimum length

$$TSPC_n \le (V_n + 2N(s_n + 8r_n))\mathbf{1}(U_{tot}(n)) + 5\sqrt{n}\mathbf{1}(U_{tot}^c(n)),$$
 (3.1)

where

$$V_n := \sum_{l=1}^{N} T_l, (3.2)$$

$$U_{tot} = U_{tot}(n) := \bigcap_{l=1}^{N} U_{l}$$
 (3.3)

and U_l is the event defined in (2.11). If the intercity distance $s_n > r_n \sqrt{2}$, then

$$TSPC_n \ge V_n.$$
 (3.4)

Proof of (3.1): Suppose that the event U_{tot} occurs and let C_l be minimum length cycle containing all the nodes in the square S_l , $1 \leq l \leq N$. Call the cycles $\{C_l\}_{1\leq l\leq N}$ as small cycles. We construct a big cycle containing all the n nodes by merging the small cycles C_l together iteratively, via a sequence of intermediate cycles $\{\mathcal{T}(i)\}_{1\leq i\leq N}$ as follows. Let $\mathcal{T}(1) = C_1$ so that the length of $\mathcal{T}(1)$ is

$$L(\mathcal{T}(1)) = L(\mathcal{C}_1) = T_1. \tag{3.5}$$

To proceed with the iteration, recall from Section 1 that the squares $\{S_l\}$ are well connected in the sense that there exists a square in $\{S_j\}_{2 \leq j \leq N}$ at a distance s_n from S_1 . Without loss of generality, we assume that S_i , $2 \leq i \leq N$ is at a distance s_n from some square $S_{q(i)} \in \{S_1, \ldots, S_{i-1}\}$.

Consider the small cycle C_2 containing all the nodes of S_2 . Remove any edge e_1 from the intermediate cycle $\mathcal{T}(1)$ and any edge e_2 from C_2 and add "cross edges" f_1 and f_2 connecting the endvertices of e_1 and e_2 . This is illustrated in Figure 3 where the edges $e_1 = ab$ and $e_2 = xy$ are replaced by the edges $f_1 = ax$ and $f_2 = by$.

The resulting intermediate cycle $\mathcal{T}(2)$ satisfies the following properties with i=2:

(f1) The cycle $\mathcal{T}(i)$ contains all the edges of the small cycles $\{C_j\}_{1 \leq j \leq i}$ not removed so far in the iteration process.

(f2) The length

$$L(\mathcal{T}(i)) \le L(\mathcal{T}(i-1)) + 2(s_n + 8r_n) \le \sum_{l=1}^{i} T_l + 2(i-1)(s_n + 8r_n).$$
 (3.6)

Property (f1) is true by construction and property (f2) is true since the length of each added edge f_i , i = 1, 2 is no more than $s_n + 8r_n$, the sum of the distance between the squares S_1 and S_2 and the total perimeter of S_1 and S_2 .

Consider now a general iteration step $i \geq 3$ where we need to merge the intermediate cycle $\mathcal{T}(i)$ with the small cycle \mathcal{C}_{i+1} containing all the nodes in the square S_{i+1} . Recall that the square S_{i+1} is at a distance of s_n from some square $S_{q(i)} \in \{S_1, \ldots, S_{i-1}\}$.

Since the event U_{tot} occurs, each square S_l , $1 \le l \le N$ contains at least

$$\frac{\eta_1 n}{2N} \ge \frac{\eta_1 M}{2} \log n \ge 8,$$



Figure 3: Merging cycles $\mathcal{T}(1) = \mathcal{C}_1 = acba$ and $\mathcal{C}_2 = xzyx$.

nodes of $\{X_k\}$ for all large n, by (2.14). In particular, $S_{q(i)}$ also contains at least 8 nodes and so the small cycle $C_{q(i)}$ contains at least 8 edges.

The square S_{i+1} is at a distance of s_n from $S_{q(i)}$ and so there are at most three squares in $\{S_j\}_{1\leq j\leq i-1}$ at a distance of s_n from $S_{q(i)}$. This means at most three edges have been removed from the small cycle $C_{q(i)}$ in the iteration process so far and so by property (f1), at least one edge $e_{q(i)}$ of $C_{q(i)}$ is still present in the intermediate cycle $\mathcal{T}(i)$.

Remove $e_{q(i)}$ and an edge from C_{i+1} and add cross edges as before to get the new cycle $\mathcal{T}(i+1)$. Arguing as above, the new intermediate cycle $\mathcal{T}(i+1)$ also satisfies properties (f1) - (f2). Performing the above process for a total of N-1 iterations, we finally obtain a big cycle C_{fin} containing all the nodes $\{X_i\}_{1\leq i\leq n}$, whose length satisfies

$$L(\mathcal{C}_{fin}) \le \sum_{i=1}^{N} T_i + 2(N-1)(s_n + 8r_n). \tag{3.7}$$

Since the overall minimum length $TSPC_n \leq L(\mathcal{C}_{fin})$ we obtain the upper bound (3.1) when $U_{tot}(n)$ occurs.

If the event $U_{tot}(n)$ does not occur, then we use the strips estimate (2.2) with a = n and b = 1 to get that the minimum length cycle $TSPC_n$ has a total length of at most $5\sqrt{n}$.

Proof of (3.4): For illustration we consider the case of two squares first. Let $Q_1 = (v_1, \ldots, v_{k_1}, v_1 =: v_{k_1+1})$ be the minimum length cycle containing all the nodes in S_1 and let $Q_2 = (u_1, \ldots, u_{k_2}, u_1 =: u_{k_2+1})$ be minimum length cycle containing all the nodes in S_2 . If C_{tot} is the minimum length cycle containing all the nodes $\{v_i\} \cup \{u_i\}$, then

$$L\left(\mathcal{C}_{tot}\right) \ge L\left(\mathcal{Q}_{1}\right) + L\left(\mathcal{Q}_{2}\right) \tag{3.8}$$

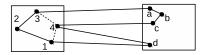


Figure 4: Replace cycle $C_{tot} = 123abc4d1$ with the cycle $C_1 = 12341$.

where $L(Q_j)$, j = 1, 2 is length of the cycle Q_j as defined in (1.4). Proof of (3.8): For a node $v \in \{v_j\} \cup \{u_j\}$, let $l(v, C_{tot})$ be the sum of length of the edges containing the node v in the cycle C_{tot} . Using (1.4)

$$L\left(\mathcal{C}_{tot}\right) = I_1 + I_2,\tag{3.9}$$

where

$$I_1 = \frac{1}{2} \sum_{j=1}^{k_1} l(v_j, \mathcal{C}_{tot}) \text{ and } I_2 = \frac{1}{2} \sum_{j=1}^{k_2} l(u_j, \mathcal{C}_{tot}).$$
 (3.10)

To estimate I_1 , assume without loss of generality that the cycle C_{tot} is of the form

$$C_{tot} = (v_1, \mathcal{E}_1, v_2, \mathcal{E}_2, \dots, \mathcal{E}_{k_1 - 1}, v_{k_1}, \mathcal{E}_{k_1}, v_1 = v_{k_1 + 1}), \tag{3.11}$$

where each \mathcal{E}_j is either empty or is a path containing only nodes of $\{u_j\}$. For $1 \leq j \leq k_1$, replace the subpath \mathcal{E}_j of \mathcal{C}_{tot} with the edge (v_j, v_{j+1}) . Let \mathcal{C}_1 be the resulting cycle as shown in Figure 4, where v_i is denoted by i for $1 \leq i \leq 4$.

For any fixed $1 \leq j \leq k_1$ the sum length of the edges containing v_j as an endvertex is less in the new cycle C_1 than in the original cycle C_{tot} i.e.,

$$l(v_j, \mathcal{C}_1) \le l(v_j, \mathcal{C}_{tot}) \tag{3.12}$$

To see (3.12) is true, let e_1 and e_2 be the edges of \mathcal{C}_{tot} containing v_j as an endvertex in the original cycle \mathcal{C}_{tot} . Using the representation of \mathcal{C}_{tot} in (3.11), we assume that the other endvertex of e_1 is either v_{j-1} or a node in $\{u_k\}$. If v_{j-1} is the other endvertex of e_1 , then e_1 is also present in the new cycle \mathcal{C}_1 . Else the length of e_1 is at least $s_n > r_n \sqrt{2}$ and e_1 is replaced by the edge $f_1 = (v_{j-1}, v_j)$ in \mathcal{C}_1 . The length of f_1 is at most $r_n \sqrt{2}$ since both endvertices of f_1 lie within the $r_n \times r_n$ square S_1 . A similar argument holds for the edge e_2 and so (3.12) is true.

Using (3.12) in (3.10), we have

$$I_1 \ge \frac{1}{2} \sum_{j=1}^{k_1} l(v_j, \mathcal{C}_1) = L(\mathcal{C}_1) \ge L(\mathcal{Q}_1),$$
 (3.13)

since Q_1 is the minimum length cycle containing all the nodes $\{v_j\}$. An analogous argument obtains that $I_2 \geq L(Q_2)$ and so from (3.9), we get (3.8). The argument for the general case is analogous.

We use Lemma 7 to prove Theorem 1. From Lemma 7, we have that the overall minimum length $TSPC_n$ is bounded above and below by the sum of the local minimum lengths $\sum_{l=1}^{N} T_l$. From the bounds on $\mathbb{E}T_l$ in (2.36) of Lemma 3, we have that $\sum_{l=1}^{N} \mathbb{E}T_l$ is of the order of $Nr_n\sqrt{\frac{n}{N}} = r_n\sqrt{nN} = b_n$ as defined in (1.6). We therefore study the convergence of $\frac{TSPC_n}{b_n}$. We henceforth fix M > 0 large so that (2.32) of Lemma 4 holds.

Proof of (1.8) in Theorem 1: From the upper and lower bounds (3.1) and (3.4) in Lemma 7, we have that

$$\frac{1}{b_n}(V_n - \mathbb{E}V_n) - \Delta_n \le \frac{1}{b_n}(TSP_n - \mathbb{E}TSP_n) \le \frac{1}{b_n}(V_n - \mathbb{E}V_n) + \Delta_n \quad (3.14)$$

where $V_n = \sum_{l=1}^{N} T_l$ is as defined (3.2) and

$$\Delta_n = \frac{2(N-1)(s_n + 8r_n)}{b_n} \mathbf{1}(U_{tot}(n)) + \frac{5\sqrt{n}}{b_n} \mathbf{1}(U_{tot}^c(n)).$$

The variance of V_n satisfies

$$var(V_n) \le C \frac{r_n^2 n^2}{N} = Cb_n^2 \left(\frac{n}{N^2}\right)$$
(3.15)

for some constant C > 0 and all n large and since $\frac{n}{N^2} \longrightarrow 0$ (see (1.7)), we get that

$$\frac{1}{b_n} \left(V_n - \mathbb{E}V_n \right) \longrightarrow 0 \text{ in probability} \tag{3.16}$$

as $n \to \infty$. Also

$$\Delta_n \longrightarrow 0 \text{ a.s.}$$
 (3.17)

as $n \to \infty$. This proves (1.8) and we prove (3.15) and (3.17) separately below.

Proof of (3.15): Write

$$var(V_n) = \sum_{l} var(T_l) + \sum_{l_1, l_2} cov(T_{l_1}, T_{l_2})$$

$$\leq \sum_{l} \mathbb{E}T_l^2 + \sum_{l_1, l_2} cov(T_{l_1}, T_{l_2}), \qquad (3.18)$$

where $cov(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$. Using (2.36) of Lemma 3 to estimate $\mathbb{E}T_l^2$ we get

$$\sum_{l=1}^{N} \mathbb{E}T_{l}^{2} \le NC_{1} \left(r_{n} \sqrt{\frac{n}{N}} \right)^{2} = C_{1} r_{n}^{2} n \tag{3.19}$$

for some constant $C_1 > 0$. Similarly using estimate (2.32) of Lemma 4 for the covariance, we get

$$\sum_{l_1, l_2} cov(T_{l_1}, T_{l_2}) \le N^2 \left(C_2 \frac{r_n^2 n^2}{N^3} \right) = C_2 \frac{r_n^2 n^2}{N}. \tag{3.20}$$

for some constants C > 0. Substituting (3.19) and (3.20) into (3.18), we get

$$var(V_n) \le C_1 r_n^2 n + C_2 \frac{r_n^2 n^2}{N} = \frac{r_n^2 n^2}{N} \left(C_1 \frac{N}{n} + C_2 \right).$$

Since $\frac{N}{n} \leq \frac{1}{M \log n} \leq 1$ for all n large (see (2.14)), we get that $var(V_n) \leq C_3 \frac{r_n^2 n^2}{N}$ for some positive constant C_3 and for all n large.

Proof of (3.17): From (3.15) and the fact that $r_n < r_n \sqrt{2} < s_n$ (see statement of the Theorem), we get

$$0 \le \Delta_n \le \frac{18Ns_n}{b_n} + \frac{5\sqrt{n}}{b_n} \mathbf{1}(U_{tot}^c(n))$$
 (3.21)

and so

$$0 \le \limsup_{n} \Delta_n \le \limsup_{n} \frac{5\sqrt{n}}{b_n} \mathbf{1}(U_{tot}^c(n)), \tag{3.22}$$

since $\frac{Ns_n}{b_n} \longrightarrow 0$ as $n \to \infty$ by the statement of the Theorem. From the estimate for the event U_l in (2.11),

$$\mathbb{P}(U_{tot}^c(n)) \le \sum_{l=1}^N \mathbb{P}(U_l^c) \le N \exp\left(-C\frac{n}{N}\right), \tag{3.23}$$

for some constant C > 0. Using the fact that $\frac{n}{N} \geq M \log n$ (see (2.14)), we get

$$\mathbb{P}(U_{tot}^c(n)) \le \frac{n}{M \log n} \frac{1}{n^{MC}} \le \frac{1}{n^2},\tag{3.24}$$

provided M > 0 is large. Fixing such an M, we have from Borell-Cantelli lemma that $\mathbb{P}(\limsup_n U^c_{tot}(n)) = 0$ and so a.s. $\mathbf{1}(U^c_{tot}(n)) = 0$ for all large n. From (3.22), we therefore get (3.17).

Proof of (1.9) in Theorem 1: Recalling that $V_n = \sum_{i=1}^N T_i$ from (3.2), we use Lemma 7 to get

$$\mathbb{E}V_n \le \mathbb{E}TSPC_n \le \mathbb{E}V_n + b_n \mathbb{E}\Delta_n, \tag{3.25}$$

where Δ_n satisfies (see (3.21))

$$\mathbb{E}\Delta_n \le \frac{18Ns_n}{b_n} + \frac{5\sqrt{n}}{b_n} \mathbb{P}(U_{tot}^c(n)) \le 18 + \frac{5\sqrt{n}}{b_n} \mathbb{P}(U_{tot}^c(n)), \tag{3.26}$$

since $\frac{Ns_n}{b_n} \longrightarrow 0$ as $n \to \infty$ (see statement of the Theorem). Using (3.24) for estimating the probability of the event U_{tot} we get

$$\sqrt{n}\mathbb{P}(U_{tot}^c(n)) \le \frac{\sqrt{n}}{n^2} \le \sqrt{\frac{M\log n}{n}} \le r_n$$
(3.27)

for all n large, where the final inequality is true by the condition for r_n in (1.7).

On the other hand $b_n = r_n \sqrt{nN} \ge r_n$ and so we get from (3.27) that

$$\frac{5\sqrt{n}}{b_n} \mathbb{P}(U_{tot}^c(n)) \le 5 \tag{3.28}$$

and using (3.28) in (3.26) we get $\mathbb{E}\Delta_n \leq 23$ and so from (3.25),

$$\mathbb{E}V_n \le \mathbb{E}TSPC_n \le \mathbb{E}V_n + 23b_n. \tag{3.29}$$

To estimate $\mathbb{E}V_n$ use the bounds for $\mathbb{E}T_l$ in (2.36) of Lemma 3 to get

$$C_1 b_n = N\left(C_1 r_n \sqrt{\frac{n}{N}}\right) \le \mathbb{E}V_n \le N\left(C_2 r_n \sqrt{\frac{n}{N}}\right) = C_2 b_n$$
 (3.30)

for some constants $C_1, C_2 > 0$. From (3.30) and (3.29), we get the bounds for $\mathbb{E}TSPC_n$ in (1.9).

Proof of (1.10) of Theorem 1: We consider Poissonization and recall the Poisson process \mathcal{P} on the squares $\{S_l\}_{1\leq l\leq N}$, defined on the probability space $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ (see paragraph prior to (2.33)). Analogous to $TSPC_n$ defined (1.5), let $TSPC_n^{(P)}$ denote the length of the minimum length cycle containing all the nodes of the Poisson process \mathcal{P} . Recall from (2.35) that $T_l^{(P)}$ denotes the length of the minimum length cycle containing all the nodes of \mathcal{P} in the square S_l .

Analogous to (3.4), we have that if the intercity distance $s_n > r_n \sqrt{2}$, then

$$TSPC_n^{(P)} \ge V_n^{(P)} = \sum_{l=1}^N T_l^{(P)}.$$
 (3.31)

Define the event

$$E_l^{(P)} = \left\{ T_l^{(P)} \ge \delta_4 r_n \sqrt{\frac{n}{N}} \right\},\,$$

where δ_4 is the constant in (2.37) of Lemma 5. Since the Poisson process is independent on disjoint sets, the events $E_l^{(P)}$ are independent and each occurs with probability at least δ_5 , by (2.37). If

$$F_{sum}^{(P)} := \sum_{l=1}^{N} \mathbf{1}(E_l^{(P)}) \tag{3.32}$$

then $\mathbb{E}_0\left(F_{sum}^{(P)}\right) \geq \delta_5 N$ and from the standard Chernoff bound estimate for sums of independent Bernoulli random variables (see Corollary A.1.14, pp. 312 of Alon and Spencer (2008)) we also have

$$\mathbb{P}_0\left(F_{sum}^{(P)} \ge C_1 N\right) \ge 1 - e^{-2C_2 N} \tag{3.33}$$

for some positive constants C_1 and C_2 . If $F_{sum}^{(P)} \geq C_1 N$, then by (3.32), the sum

$$\sum_{l=1}^{N} T_l^{(P)} \ge C_1 N \left(\delta_4 r_n \sqrt{\frac{n}{N}} \right) = C_3 b_n$$

for some constant $C_3 > 0$ and so from (3.31),

$$\mathbb{P}_0(TSPC_n^{(P)} \ge C_3 b_n) \ge 1 - e^{-2C_2 N} \tag{3.34}$$

for all n large.

To convert the probability estimates to the Binomial process, let

$$A_P = \{TSPC_n^{(P)} \ge C_3b_n\}, A = \{TSPC_n \ge C_3b_n\}$$

and use the dePoissonization formula

$$\mathbb{P}(A) \ge 1 - D\sqrt{n}\mathbb{P}(A_P^c) \tag{3.35}$$

for some constant D > 0 and (3.34) to get that

$$\mathbb{P}(TSPC_n \ge C_3 b_n) \ge 1 - D\sqrt{n}e^{-2C_2 N} = 1 - e^{-\alpha_N}, \tag{3.36}$$

where

$$\alpha_N = 2C_2N - \log D - \frac{1}{2}\log n \ge C_2N$$

for all n large, since $N \ge \sqrt{n}$ for all n large (see (2.14)). This proves (1.10) and it only remains to prove (3.35).

To prove (3.35), let N_P denote the random number of nodes of \mathcal{P} in all the squares $\bigcup_{j=1}^{N} S_j$ so that $\mathbb{E}_0 N_P = n$ and $\mathbb{P}_0(N_P = n) = e^{-n} \frac{n^n}{n!} \geq \frac{D_1}{\sqrt{n}}$ for some constant $D_1 > 0$, using the Stirling formula. Given $N_P = n$, the nodes of \mathcal{P} are i.i.d. with distribution g_N as defined in (1.3); i.e.,

$$\mathbb{P}_0(A_P^c|N_P=n) = \mathbb{P}(A^c)$$

and so

$$\mathbb{P}_0(A_P^c) \ge \mathbb{P}_0(A_P^c|N_P = n)\mathbb{P}_0(N_P = n) = \mathbb{P}(A^c)\mathbb{P}_0(N_P = n) \ge \mathbb{P}(A^c)\frac{D_1}{\sqrt{n}},$$
 proving (3.35).

Proof of (1.11) of Theorem 1: As in the proof of (1.10) above, we consider the Poisson process \mathcal{P} on the squares $\{S_l\}_{1\leq l\leq N}$ defined in the paragraph prior to (2.33). As before, let $TSPC_n^{(P)}$ denote the length of the minimum length cycle containing all the nodes of the Poisson process \mathcal{P} . Recall from (2.35) that $T_l^{(P)}$ denotes the length of the minimum length cycle containing all the nodes of \mathcal{P} in the square S_l .

Analogous to (3.1), we have

$$TSPC_n^{(P)} \le \left(V_n^{(P)} + 2(N-1)(s_n + 8r_n)\right) \mathbf{1}(U_{tot}^{(P)}(n)) + 5\sqrt{n}\mathbf{1}(U_{tot}^{(P)}(n))^c,$$
(3.37)

where

$$V_n^{(P)} := \sum_{l=1}^N T_l^{(P)}, \tag{3.38}$$

$$U_{tot}^{(P)} = U_{tot}^{(P)}(n) := \bigcap_{l=1}^{N} U_{l}^{(P)}$$
(3.39)

and $U_l^{(P)} = \{\frac{\eta_1 n}{2N} \leq N_l^{(P)} \leq \frac{2\eta_2 n}{N}\}$ is the event defined in (2.38). Recall that $N_l^{(P)}$ is the total number of nodes of \mathcal{P} inside the square S_l .

Suppose now that the event $U_{tot}^{(P)}(n)$ occurs so that

$$TSPC_n^{(P)} \le V_n^{(P)} + 2(N-1)(s_n + 8r_n) = \sum_{l=1}^{N} T_l^{(P)} + 2(N-1)(s_n + 8r_n).$$
 (3.40)

Since $U_l^{(P)} \supseteq U_{tot}^{(P)}$ occurs for every $1 \le l \le N$, we use the strips estimate (2.2) with $a = \frac{2\eta_2 n}{N}$ and $b = r_n$ to get that the corresponding minimum length $T_l^{(P)} \le 5b\sqrt{a} \le Cr_n\sqrt{\frac{n}{N}}$ for some constant C > 0 and for every $1 \le l \le N$. Thus

$$V_n^{(P)} = \left(\sum_{l=1}^N T_l^{(P)}\right) \le Cb_n$$

and from (3.40) we therefore get

$$TSPC_n^{(P)} \le Cb_n + 2(N-1)(s_n + 8r_n) \le Cb_n + 18Ns_n \le (C+1)b_n, (3.41)$$

for all n large. The second inequality in (3.41) is true since $r_n < r_n \sqrt{2} < s_n$. The final inequality in (3.41) is true since $\frac{Ns_n}{b_n} \longrightarrow 0$ and so $\frac{Ns_n}{b_n} \le \frac{1}{18}$ for all n large.

Summarizing, we have that if the event $U_{tot}^{(P)}$ occurs, then the overall minimum length $TSPC_n^{(P)} \leq C_1b_n$ for some constant $C_1 > 0$. To evaluate $\mathbb{P}(U_{tot}^{(P)})$, use the estimate (2.39) for the event $U_l^{(P)}$ to get

$$\mathbb{P}_0(U_{tot}^{(P)}) \ge 1 - N \exp\left(-2C\frac{n}{N}\right) \tag{3.42}$$

for some constant C > 0. Thus

$$\mathbb{P}_0\left(TSPC_n^{(P)} \le C_1 b_n\right) \ge \mathbb{P}(U_{tot}^{(P)}) \ge 1 - N \exp\left(-2C\frac{n}{N}\right). \tag{3.43}$$

To convert the probabilities to the Binomial process, we again use the dePoissonization formula (3.35) to get that

$$\mathbb{P}\left(TSPC_n \le C_1 b_n\right) \ge 1 - DN\sqrt{n} \exp\left(-2C\frac{n}{N}\right) = 1 - e^{-\delta_N}, \quad (3.44)$$

where D > 0 is as in (3.35) and

$$\delta_N = 2C \frac{n}{N} - \log D - \log N - \frac{1}{2} \log n.$$
 (3.45)

Since $\frac{n}{N} \ge M \log n$ for all n large (see (2.14)), we get

$$\log D + \log N + \frac{1}{2}\log n \le \log D + \log\left(\frac{n}{M\log n}\right) + \frac{1}{2}\log n \le 2\log n \le C\frac{n}{N},$$

provided M > 0 is large. Fixing such an M we get that $\delta_N \geq C \frac{n}{N}$ and so (1.11) follows from (3.44).

4 Proof of Theorem 2

We need preliminary estimates regarding the change in length of the minimum length cycle upon adding or deleting a single node.

Let X_1, \ldots, X_{n+1} be n+1 random nodes distributed according to the density f in the unit square S. For $1 \le j \le n+1$, let \mathcal{D}_j denote the minimum length cycle containing all the nodes $\{X_k\}_{1 \le k \ne j \le n+1}$ with length

$$L(\mathcal{D}_j) = TSP(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_{n+1}; S), \tag{4.1}$$

where TSP(.;.) is as defined in (2.1). For future use, we estimate lengths of edges in \mathcal{D}_{j} .

Divide the unit square S into $2Aw_n \times 2Aw_n$ squares $\{W_i^{(1)}\}_{1 \leq i \leq N_W}$, each of side length $2Aw_n$ where

$$\frac{1}{n^{1/6}} \le w_n := \frac{1 + c_n}{n^{1/6}} \le \frac{2}{n^{1/6}}, \quad A := \left(\frac{3}{\epsilon_1}\right)^{\frac{1}{3}} \tag{4.2}$$

and $\epsilon_1 > 0$ is as in (1.1). The term $c_n \in (0,1)$ is chosen such that $\frac{1}{2Aw_n}$ is an integer for all n large. For $1 \le i \le N_W$, let $W_i^{(2)}$ be the bigger square

with same centre as $W_i^{(1)}$ but with side length $4Aw_n$. For $1 \leq j \leq n+1$ and $1 \leq i \leq N_W$, let $F_j(i)$ be the event there exists an edge $e_j(i) \in \mathcal{D}_j$ with both endvertices in the bigger square $W_i^{(2)}$ and let

$$F_{tot}(n+1) := \bigcap_{j=1}^{n+1} \bigcap_{i=1}^{N_W} F_j(i). \tag{4.3}$$

The following Lemma is used in the proof of Theorem 2.

Lemma 8. We have that

$$\mathbb{P}\left(F_{tot}(n+1)\right) \ge 1 - \exp\left(-Cn^{2/3}\right) \tag{4.4}$$

for some constant C > 0 and for all n > 3.

Proof of Lemma 8: We first perform some preliminary computations. Fix $1 \le j \le n+1$ and $1 \le i \le N_W$. Using (1.1) and the fact that $w_n \ge n^{-\frac{1}{6}}$ (see (4.2)), the average number of nodes of $\{X_k\}_{1 \le k \ne j \le n+1}$ in the square $W_i^{(1)}$ is

$$n \int_{W_{\epsilon}^{(1)}} f(x) dx \ge n\epsilon_1 (2Aw_n)^2 \ge 4A^2 \epsilon_1 n^{2/3},$$

where $\epsilon_1 > 0$ is as in (1.1). Let $Z_j(i)$ denote the event that the square $W_i^{(1)}$ contains at least $2A^2\epsilon_1n^{2/3}$ nodes of $\{X_k\}_{1\leq k\neq j\leq n+1}$. By standard Binomial estimates (see Corollary A.1.14, pp. 312 of Alon and Spencer (2008)) and the fact that $w_n^2 \geq n^{-\frac{1}{3}}$ (see (4.2)), we get

$$\mathbb{P}(Z_i(i)) \ge 1 - \exp(-C_1 n w_n^2) \ge 1 - \exp(-C_2 n^{2/3}) \tag{4.5}$$

for some positive constants C_1 and C_2 .

If

$$Z_{tot}(n+1) := \bigcap_{i=1}^{n+1} \bigcap_{i=1}^{N_W} Z_j(i), \tag{4.6}$$

then we have from (4.5) that

$$\mathbb{P}(Z_{tot}(n+1)) \ge 1 - (n+1)N_W \exp(-C_2 n^{2/3}). \tag{4.7}$$

The total number of squares is

$$N_W = \left(\frac{1}{2Aw_n}\right)^2 \ge Dn^{1/3} \tag{4.8}$$

for some constant D > 0 using $w_n \ge n^{-\frac{1}{6}}$ (see (4.2)) and so we get from (4.7) that

$$\mathbb{P}(Z_{tot}(n+1)) \ge 1 - \exp(-C_3 n^{2/3}) \tag{4.9}$$

for some constant $C_3 > 0$.

The estimate (4.9) and the following property imply Lemma 8.

(f1) If the event $Z_{tot}(n+1)$ occurs, then for every $1 \leq j \leq n+1$ and $1 \leq i \leq N_W$, there exists an edge $e_j(i) \in \mathcal{D}_j$ with both endvertices in the bigger square $W_i^{(2)}$.

Proof of (f1): Suppose $Z_{tot}(n+1)$ occurs and suppose that the node X_j is present in the square $W_i^{(1)}$. Let $\{Y_k\}_{1 \leq k \leq q} \subset \{X_k\}_{1 \leq k \neq j \leq n+1}$ be the other nodes present in the square $W_i^{(1)}$. Since the event $Z_j(i) \supseteq Z_{tot}(n+1)$ occurs,

$$q \ge 2A^2 \epsilon_1 n^{2/3}.\tag{4.10}$$

For $1 \leq k \leq q$, let $e_k(1)$ and $e_k(2)$ be the edges containing the node Y_k as an endvertex in the cycle \mathcal{D}_j . If no edge of \mathcal{D}_j has both its endvertices inside the bigger square $W_i^{(2)}$, then all the edges $\{e_k(1), e_k(2)\}_{1 \leq k \leq q}$ are distinct and each such edge has length at least Aw_n , since it must cross the annulus $W_i^{(2)} \setminus W_i^{(1)}$. Therefore if $l(Y_k, \mathcal{D}_j)$ is the sum of length of the edges containing Y_k as an endvertex in the cycle \mathcal{D}_j , then $l(Y_k, \mathcal{D}_j) \geq 2Aw_n$.

From (1.4) we therefore have that the total length of \mathcal{D}_i is

$$L(\mathcal{D}_j) \ge \frac{1}{2} \sum_{k=1}^{q} l(Y_k, \mathcal{D}_j) \ge q \cdot Aw_n \ge 2A^3 \epsilon_1 n^{2/3} w_n.$$
 (4.11)

Using the fact that $w_n \ge n^{-\frac{1}{6}}$ (see (4.2)) we then get that

$$L(\mathcal{D}_i) \ge 2A^3 \epsilon_1 \sqrt{n} \ge 6\sqrt{n},\tag{4.12}$$

by our choice of A in (4.2).

But using the strips estimate (2.2) with a = n and b = 1, we have that the length of the cycle \mathcal{D}_i is at most

$$L(\mathcal{D}_j) \le 5\sqrt{n}$$

and this contradicts (4.12).

The above Lemma allows us to estimate the variance of the length of the minimum length cycle.

Proof of 1.12 of Theorem 2: We use the martingale difference method and for $1 \le j \le n+1$, let

$$\mathcal{F}_j = \sigma\left(X_1, \dots, X_j\right)$$

denote the sigma field generated by the random variables X_1, \ldots, X_j . Defining the martingale difference

$$D_{j} = \mathbb{E}(TSP_{n+1}|\mathcal{F}_{j}) - \mathbb{E}(TSP_{n+1}|\mathcal{F}_{j-1}), \tag{4.13}$$

we have

$$TSP_{n+1} - \mathbb{E}TSP_{n+1} = \sum_{j=1}^{n+1} D_j$$

and so by the martingale property

$$var(TSP_{n+1}) = \mathbb{E}\left(\sum_{j=1}^{n+1} D_j\right)^2 = \sum_{j=1}^{n+1} \mathbb{E}D_j^2.$$
 (4.14)

There is a constant C > 0 such that

$$\max_{1 \le j \le n+1} \mathbb{E}D_j^2 \le \frac{C}{n^{1/3}} \tag{4.15}$$

for all $n \ge 1$ and this proves (1.12).

Proof of (4.15): We first rewrite D_j in a more convenient form. Let $\omega = (x_1, \ldots, x_{n+1})$ and $\omega' = (y_1, \ldots, y_{n+1})$ be two vectors in $(\mathbb{R}^2)^{n+1}$. Defining

$$\omega_j = (x_1, \dots, x_j, y_{j+1}, \dots, y_{n+1})$$

for $1 \le j \le n+1$ and using Fubini's theorem, we get

$$|D_j| = \left| \int (T(\omega_j) - T(\omega_{j-1})) f(y_j) \dots f(y_n) dy_j \dots dy_{n+1} \right| \le H_j$$

where

$$H_{j} := \int |T(\omega_{j}) - T(\omega_{j-1})| f(y_{j}) \dots f(y_{n}) dy_{j} \dots dy_{n+1}$$
 (4.16)

and $T(\omega_t), t = j, j - 1$ is the length of the minimum length cycle containing all the nodes in ω_t .

Let $F_{tot}(n+1)$ be the event defined in (4.3) and write

$$H_j = I_1 + I_2, (4.17)$$

where

$$I_{1} = \int |T(\omega_{j}) - T(\omega_{j-1})| \mathbf{1}(\omega_{j} \in F_{tot}(n+1)) \mathbf{1}(\omega_{j-1} \in F_{tot}(n+1))$$
$$f(y_{j}) \dots f(y_{n+1}) dy_{j} \dots dy_{n+1}$$
(4.18)

and $I_2 = I_1 - H_j$.

There is a positive constant C > 0 such that

$$\mathbb{E}I_1^2 \le \frac{C}{n^{1/3}}$$
 and $\mathbb{E}I_2^2 \le \exp\left(-Cn^{2/3}\right)$ (4.19)

and so using $H_j^2 = (I_1 + I_2)^2 \le 2(I_1^2 + I_2^2)$ we get

$$\mathbb{E}H_j^2 \le 2\left(\frac{C}{n^{1/3}} + \exp\left(-Cn^{2/3}\right)\right) \le \frac{3C}{n^{1/3}}$$

for all n large. This proves (4.15).

We obtain the estimates for I_1 and I_2 in (4.19) separately below. Estimate for I_1 : Let \mathcal{D}_j be the minimum length cycle containing all the nodes $\{x_k\}_{1\leq k\leq j-1}\cup\{y_k\}_{j+1\leq k\leq n+1}$. If $L(\mathcal{D}_j)$ is the length of \mathcal{D}_j , then for $t\in\{j-1,j\}$

$$|T(\omega_t) - L(\mathcal{D}_t)|\mathbf{1}(\omega_t \in F_{tot}(n+1)) \le 4Aw_n\sqrt{2}$$
(4.20)

and so from (4.20), (4.18) and triangle inequality, we have

$$I_1 \le 8Aw_n\sqrt{2} \text{ and so } \mathbb{E}(I_1^2) \le C_1w_n^2 \le \frac{C_2}{n^{1/3}}$$
 (4.21)

for some constants C_1, C_2 since $w_n \leq \frac{2}{n^{1/6}}$ (see (4.2)).

Proof of (4.20): We prove for t=j and an analogous analysis holds for t=j-1. By monotonicity (2.2), we have that $T(\omega_j) \geq L(\mathcal{D}_j)$. Also, since $\omega_j \in F_{tot}(n+1)$, every square $W_k^{(2)}$, $1 \leq k \leq N_W$ of side length $4Aw_n$ defined prior to Lemma 8 contains an edge of \mathcal{D}_j . Suppose the "new" node x_j belongs to the square $W_i^{(1)}$. Since there is an edge $e \in \mathcal{D}_j$ having both its endnodes z_1, z_2 inside $W_i^{(1)}$, we remove e and add the edges (z_1, x_j) and (x_j, z_2) to form a

new cycle containing all the nodes of ω_j . The total length of the two edges added is at most $4Aw_n\sqrt{2}$. This implies that $T(\omega_j) \leq L(\mathcal{D}_j) + 4Aw_n\sqrt{2}$, proving (4.20).

Estimate for I_2 : Every edge within the unit square S has length at most $\sqrt{2}$ and any cycle containing all the n+1 nodes of ω_t has n+1 edges. Therefore $T(\omega_t) \leq (n+1)\sqrt{2}$ for $t \in \{j-1,j\}$. Thus from the definition of I_2 in (4.17), we get $I_2 \leq J_1 + J_2$, where

$$J_1 = (n+1)\sqrt{2} \int \mathbf{1}(\omega_j \notin F_{tot}(n+1)) f(y_j) \dots f(y_n) dy_j \dots dy_n$$

and

$$J_2 = (n+1)\sqrt{2} \int \mathbf{1}(\omega_{j-1} \notin F_{tot}(n+1)) f(y_j) \dots f(y_n) dy_j \dots dy_n.$$

Using Cauchy-Schwarz inequality,

$$J_1^2 = 2(n+1)^2 \left(\mathbb{E}(\mathbf{1}(F_{tot}^c(n+1))|\mathcal{F}_j) \right)^2 \le 2(n+1)^2 \mathbb{E}(\mathbf{1}(F_{tot}^c(n+1))|\mathcal{F}_j)$$

and similarly

$$J_2^2 \leq 2(n+1)^2 \mathbb{E} \left(\mathbf{1} (F_{tot}^c(n+1)) | \mathcal{F}_{j-1} \right).$$

Since $I_2^2 = (J_1 + J_2)^2 \le 2(J_1^2 + J_2^2)$ and $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_j)|\mathcal{F}_{j-1}) = \mathbb{E}(X|\mathcal{F}_{j-1})$, we get that

$$\mathbb{E}(I_2^2|\mathcal{F}_{i-1}) \le 4(n+1)^2 \mathbb{P}(F_{tot}^c(n+1)|\mathcal{F}_{i-1})$$

and therefore that

$$\mathbb{E}I_2^2 \le 4(n+1)^2 \mathbb{P}\left(F_{tot}^c(n+1)\right) \le 4(n+1)^2 \exp\left(-2Cn^{2/3}\right) \le \exp\left(-Cn^{2/3}\right)$$
(4.22)

for some constant C > 0 and for all n large, using (4.4).

Proof of (1.13) and (1.14) of Theorem 2: The upper bound for $\mathbb{E}TSP_n$ in (1.13) is obtained from the strips estimate (2.2) with a=n and b=1. This also proves (1.14).

For the lower bound in (1.13), we argue as follows. For $1 \leq i \leq n$, let $d(X_i, \{X_j\}_{1 \leq j \neq i \leq n})$ denote the minimum distance of node X_i from all other nodes $\{X_j\}_{1 \leq j \neq i \leq n}$. The TSP length TSP_n then satisfies then $TSP_n \geq \sum_{i=1}^n d(X_i, \{X_j\}_{1 \leq j \neq i \leq n})$ and so

$$\mathbb{E}TSP_n \ge n\mathbb{E}d(X_1, \{X_j\}_{2 \le j \le n}). \tag{4.23}$$

Analogous to the proof of (2.17), we have that

$$\mathbb{E}d(X_1, \{X_j\}_{2 \le j \le n}) \ge \frac{C}{\sqrt{n}}$$

for some constant C > 0 not depending on the choice of i and so from (4.23) we get (1.15).

Proof of (1.15): Divide the unit square S into $r_n \times r_n$ squares $\{S_l\}_{1 \le l \le N}$ placed s_n apart as in Figure 1 with r_n and s_n as follows:

$$r_n^2 = \frac{M \log n + c_n}{n} \text{ and } s_n^2 = \frac{2M \log n + d_n}{n}$$
 (4.24)

where $c_n \in (0,1)$ and $d_n \in (4,5)$ are such that $\frac{1-r_n}{r_n+s_n}$ is an integer. With this choice of r_n and s_n , the number of $r_n \times r_n$ squares N and the scaling factor b_n defined in (1.6) satisfy

$$C_1 n r_n^2 \le C_2 \frac{n}{\log n} \le N \le C_3 \frac{n}{\log n} \le C_4 n r_n^2$$
 (4.25)

and

$$C_5\sqrt{n} \le b_n = r_n\sqrt{nN} \le C_6\sqrt{n} \tag{4.26}$$

for some positive constants $\{C_i\}_{1 \leq i \leq 6}$.

Let C_n denote the minimum length cycle containing all the nodes of $\{X_k\}_{1 \leq k \leq n}$ present in all the $r_n \times r_n$ squares $\{S_j\}_{1 \leq j \leq N}$. If $L(C_n)$ denotes the length of C_n , then by monotonicity (2.2) we have that

$$TSP_n \ge L(\mathcal{C}_n)$$
 (4.27)

and since the term $s_n > r_n \sqrt{2}$ strictly (see (4.24)), we have from (3.4) that

$$L(\mathcal{C}_n) \ge \sum_{l=1}^{N} T_l, \tag{4.28}$$

where $T_l, 1 \leq l \leq N$ is the minimum length cycle containing all the nodes of $\{X_k\}$ in the square S_l .

Estimates for $\mathbb{E}T_l$ in Lemma 3 and estimates for $\mathbb{E}_0T_l^{(P)}$, the Poissonized process, in Lemma 5 hold in this case as well. Moreover if M > 0 is large

in (4.24), then the covariance estimate in Lemma 4 holds as well. For illustration, we prove the lower bound for $\mathbb{E}T_l$ here. From (1.1), any node of $\{X_k\}_{1\leq k\leq n}$ is present in the square S_l with probability

$$D_1 \frac{n}{N} \le \epsilon_1 r_n^2 \le q_l = \int_{S_l} f(x) dx \le \epsilon_2 r_n^2 \le D_2 \frac{n}{N}$$

for some positive constants D_1 and D_2 , using (4.25). The estimates for q_l are analogous to the estimates for p_l in (2.5). Arguing as in the proof of (2.10) we then get that $\mathbb{E}T_l \geq Cr_n\sqrt{\frac{n}{N}}$.

Arguing as in the proof of (1.10), we get

$$\mathbb{P}\left(TSP_n \ge L(\mathcal{C}_n) \ge C_4 b_n\right) \ge 1 - e^{-C_5 N}$$

for some positive constants C_4, C_5 . Finally, using (4.25) and (4.26) to estimate b_n and N we get (1.15).

Acknowledgement

I thank Professors Rahul Roy and Federico Camia for crucial comments and for my fellowships.

References

- [1] N. Alon and J. Spencer. (2008). The probabilistic method. Wiley.
- [2] S. Arora. (1998). PTAS for Euclidean Traveling Salesman and Other Geometric Problems. *Journal of the ACM*, **45**, pp. 753–782.
- [3] J. Beardwood, J. H. Halton and J. M. Hammersley. (1959). The shortest path through many points. *Proceedings Cambridge Philosophical Society*, **55**, pp. 299–327.
- [4] W. Cook. (2011). In pursuit of the Traveling Salesman: Mathematics at the Limits of Computation. Princeton University Press.
- [5] G. Gutin and A. P. Punnen. (2006). The traveling salesman problem and its variations. Springer.

- [6] M. Karpinski, M. Lampis and R. Schmied (2015). New Inapproximability bounds for TSP. Journal of Computer and System Sciences, 81, 1665-1677.
- [7] C-M. Pintea, P. C. Pop and C. Chira. (2017) The generalized traveling salesman problem solved with ant algorithms. *Complex Adaptive Systems Modeling*, **5**, available at https://doi.org/10.1186/s402940017-0048-9.
- [8] L. V. Snyder and M. S. Daskin. (2006). A random-key genetic algorithm for the generalized traveling salesman problem. *European Journal of Operational Research*, **174**, 38–53.
- [9] J. M. Steele. (1981). Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Annals of Probability*, **9**, pp. 365–376.
- [10] J. M. Steele. (1993). Probability and Problems in Euclidean Combinatorial Optimization. *Statistical Science*, **8**, pp. 48–56.
- [11] S. Steinerberger. (2015). New bounds for the Traveling Salesman constant. Advances in Applied Probability, 47, pp. 27–36.
- [12] V. V. Vazirani. (2001). Approximation algorithms. Springer Verlag.