On random symmetric travelling salesman problems

Alan Frieze
Department of Mathematical Sciences,
Carnegie Mellon University.
alan@random.math.cmu.edu*

September 30, 2003

Abstract

Let the edges of the complete graph K_n be assigned independent uniform $[0,1]$ random edge weights. Let Z_{TSP} and Z_{2FAC} be the weights of the minimum length travelling salesman tour and minimum weight 2-factor respectively. We show that **whp** $|Z_{TSP} - Z_{2FAC}| = o(1)$. The proof is via by the analysis of a polynomial time algorithm that finds a tour only a little longer than Z_{2FAC}.

1 Introduction

The starting point of this line of research is the foundational paper of Karp [10] in 1979. Karp considered the following problem: The arcs of the complete digraph D_n on vertex set $[n]$ are given independent uniform $[0,1]$ random edge weights. Z_{TSP} denotes the weight of the minimum length (directed) travelling salesman tour and Z_{AP} denotes the minimum weight of an assignment for the associated $n \times n$ matrix $\mathbf{M}$ of costs. Karp proved, via the analysis of an $O(n^3)$ time algorithm, that **whp** $|Z_{TSP} - Z_{AP}| = o(1)$. (By "with high probability" (**whp**) we mean "with probability 1-o(1) as $n \to \infty$".) This gave good theoretical backing for the empirical observation (see e.g. Balas and Toth [3]) that the assignment problem provides a good lower bound for use in branch and bound algorithms. A sequence of papers, Karp and Steele [11], Dyer and Frieze [5] and Frieze and Sorkin [9] tightened this result. In particular, [9] proves that

$$|Z_{TSP} - Z_{AP}| = O\left(\frac{(\ln n)^2}{n}\right) \qquad \textbf{whp}.$$

It is natural to try to prove a similar result for symmetric problems. Here the edges of the complete graph K_n are assigned independent uniform $[0,1]$ random edge weights. Up to now there has been almost no progress analysing this probabilistic model of the TSP. Let $\mathbf{M} = \mathbf{M}(i,j)$ once again denote the matrix of lengths. Here of course $\mathbf{M}$ is *symmetric* i.e. $\mathbf{M}(i,j) = \mathbf{M}(j,i)$ for all $i,j \in [n]$. Let $Z_{TSP} = Z_{TSP}(\mathbf{M})$ denote the length of the shortest Hamilton cycle. It is unlikely that solving the assignment problem for $\mathbf{M}$ will yield a good approximation to Z_{TSP} since its solution $(i,\pi(i)), i \in [n]$ will likely contain many instances i,j where $\pi(i) = j$ and $\pi(j) = i$. It is perhaps worth conjecturing that $Z_{TSP} - Z_{AP} = \Omega(1)$ **whp**. It seems more sensible to start with the solution of the minimum weight 2-factor problem. A 2-factor is a

*Supported in part by NSF grant CCR-9818411.

subgraph of K_n in which every vertex has degree 2 and so is a collection of vertex disjoint cycles which cover all vertices in $[n]$. A minimum weight 2-factor can be constructed in polynomial time. It is a classical problem in matching theory, see e.g. Lovász and Plummer [12]. Let $Z_{2FAC} = Z_{2FAC}(\mathbf{M})$ be the weight of the minimum weight 2-factor. A tour i.e. a cycle with n edges is a 2-factor and so clearly $z_{2FAC} \leq z_{TSP}$. Our main result is:

Theorem 1

$$z_{TSP} - z_{2FAC} = o(1) \qquad \textbf{whp}. \tag{1}$$

Furthermore, **whp** *a tour of length $z_{2FAC} + o(1)$ can be constructed by a polynomial time algorithm.*

Now $z_{TSP} \geq z_{MST}$, the weight of the minimum spanning tree, and **whp** $z_{MST} \geq \zeta(3) - o(1) \sim 1.202\ldots$, Frieze [7]. Thus the tour produced by our algorithm is *asymptotically optimal* i.e. **whp** the ratio of the tour produced by the algorithm and the optimum is $1 + o(1)$.

It as well to point out now what makes the symmetric case seemingly more difficult than the asymmetric case studied in [10], [11], [5] and [9]. The solution to the assignment problem can be represented as a permutation $(i, \pi^*(i)), i \in [n]$. It is straightforward to show that the distribution of π^* is *uniform* over the set of possible permutations. As such, the number of cycles of π^* is $O(\ln n)$ **whp**. This is a great help in the analysis. Following Karp, one tries to merge cycles together until one has a tour. Each merger has to be shown to entail a small cost **whp**. One can in fact show in both the symmetric and asymmetric cases that each merger costs at most $O\left(\frac{\ln n}{n}\right)$ **whp** giving an increased cost of $O\left(\frac{(\ln n)^2}{n}\right)$ overall in the asymmetric case. When we come the symmetric case things are not so nice. A random 2-factor will also have $O(\ln n)$ cycles **whp** but it is not at all clear that the minimum weight 2-factor has a uniform distribution. If it did, then we could replace the o(1) in (1) by $O\left(\frac{(\ln n)^2}{n}\right)$, and it would be easier to do because of the symmetry. Thus one of our problems has been to put a high probability bound on the number of cycles in the minimum weight 2-factor. We have not been too successful. We have only managed a "miserable" $O\left(\frac{n}{\ln n}\right)$ which is just on the borderline of being useful. (This number seems to be grossly large when compared to the average number of cycles in a random 2-factor.) Any fewer and the paper would be much shorter. We will however have to content ourselves with what we have and make the best of it.

Structure of the paper: We first show that $Z_{TSP} \leq 6$ **whp**, which we then use to show that **whp** the optimum two factor has $O(n/\ln n)$ cycles **whp**. We then show that **whp** the longest edge in the minimum 2-FACTOR $F^* = F^*(\mathbf{M})$ is of length $O\left(\frac{\ln n}{n}\right)$. This is achieved by showing that **whp** one can find short alternating paths which eliminate the need for longer edges.

In Section 3 we conditon on the cycle structure of F^* and describe a model for use given the cycle structure of F^*. In Section 4 we will describe our algorithm for finding a tour and prove that it is asymptotically optimal. It consists of two Phases.

In the first phase we try to merge small cycles into a long path using short edges. We start it by deleting an arbitrary edge of the shortest cycle to obtain a path P and a collection of cycles $\mathcal{C}$. Then we repeatedly, choose the shortest *acceptable* edge e joining an endpoint of P to one of the cycles in $C \in \mathcal{C}$. We remove an edge of C which is incident with e and thereby grow P and reduce the number of cycles in $\mathcal{C}$ by one. We continue this process until the number of cycles in $\mathcal{C}$ is $o\left(\frac{n}{\ln n}\right)$. The total increase in cost is kept to $o(1)$ in expectation.

At the end of the first phase we will **whp** have one long path and $o\left(\frac{n}{\ln n}\right)$ cycles. Then, using an *extension-rotation* type of algorithm, we merge these cycles together at a cost of $O\left(\frac{\ln n}{n}\right)$ per cycle, a total of $o(1)$ extra cost in all.

It is the first phase which is the most complicated to analyse and if we could prove that F^* had $o\left(\frac{\ln n}{n}\right)$ cycles **whp**, then we would not need it.

2 Preliminary Analysis

2.1 Upper bound on Z_{TSP}

The first thing we shall do is prove a high probability upper bound on z_{TSP}. It is quite weak and obtaining a more precise bound remains an interesting open problem. (The work of Aldous [1] combined with Karp [10] shows that for the asymmetric case, $z_{TSP} \sim \frac{\pi^2}{6}$ **whp**).

Lemma 1

$$z_{TSP} \leq 6 \qquad \textbf{whp}.$$

Proof We use an old idea of Walkup [13]. Let Z be a random variable on $[0,1]$ with $\mathbf{Pr}(Z \leq x) = 1 - \sqrt{1-x}$ for $x \in [0,1]$. Then if Z_1, Z_2 are independent copies of Z, $\min\{Z_1, Z_2\}$ is distributed as a uniform $[0,1]$ random variable. So now for $i < j$ let the edge lengths $\mathbf{M}(i,j)$ be generated by $\min\{Z_1(i,j), Z_1(j,i)\}$ where the set $Z_1(i,j)$, $1 \leq i \neq j \leq n$ are independent copies of Z.

For each integer $m \geq 1$ define a random digraph $\vec{\Gamma}_m$ with vertex set $[n]$ and a directed arc (x,y) if $Z_1(x,y)$ is one of the m smallest values $Z_1(x,j), j \neq x$. By ignoring orientation in $\vec{\Gamma}_m$ we obtain the random graph $\Gamma_m = G_{m-out}$. Cooper and Frieze [4] showed that Γ_4 is Hamiltonian **whp**. The expected value of the kth smallest of n independent copies of Z is $(1+o(1))\frac{2k}{n}$, for $k = O(1)$. Thus the expected length of an edge of Γ_4 is $(1+o(1))\frac{2+4+6+8}{4n} = (1+o(1))\frac{5}{n}$. If H is a randomly chosen Hamilton cycle of Γ_4 then the expected length of H is at most $5+o(1)$. It will be at most 6 **whp** since it is the sum of n independent bounded random variables. (The variables here are sums of the lengths of the zero, one or two H-edges which are *directed out* of each vertex. Its length bounds z_{TSP} from above. □

2.2 The number of cycles in F^*

We can use Lemma 1 to help bound the number of cycles in a minimum weight 2-factor.

Lemma 2 **Whp** *F^* consists of at most $\frac{3n}{\ln n}$ cycles.*

Proof Let Z denote the number of cycles in the minimum weight 2-factor. For $3n/\ln n \leq r \leq n/3$, we write

$$\mathbf{Pr}(Z = r) \leq \mathbf{Pr}(Z_{TSP} > 6) + \mathbf{Pr}(\exists \text{ 2-factor with } r \text{ cycles and total weight at most } 6). \quad (2)$$

The number $A_{n,r}$ of 2-factors in K_n with r cycles is given by

$$A_{n,r} = \sum_{k_1+\cdots+k_r=n} \frac{1}{r!}\binom{n}{k_1,\ldots,k_r}\prod_{i=1}^{r}\frac{(k_i-1)!}{2}.$$

The weight of a fixed 2-factor is the sum of n independent uniform [0,1] random variables. The probability that that this sum is at most 6 is bounded by the volume of the simplex equal to the intersection of $\{x_1 + x_2 + \cdots + x_n \le 6\}$ and the positive orthant. The volume of this simplex is $\frac{6^n}{n!}$. Thus,

$$
\begin{aligned}
\mathbf{Pr}(Z \ge r) &\le o(1) + \sum_{r=3n/\ln n}^{n/3} A_{n,r} \frac{6^n}{n!} \\
&= o(1) + \sum_{r=3n/\ln n}^{n/3} \frac{6^n}{2^r r!} \sum_{k_1+\cdots+k_r=n} \prod_{i=1}^{r} \frac{1}{k_i} \\
&= o(1) + \sum_{r=3n/\ln n}^{n/3} \frac{6^n}{2^r r!} [x^n](-\ln(1-x))^r,
\end{aligned}
$$

where $[x^n]f(x)$ denotes the coefficient of x^n in the expansion of $f(x)$ around zero i.e. if $f(x) = \sum_{n\ge 0} f_n x^n$ then $[x^n]f(x) = f_n$. Note that $f_n \le f(\xi)/\xi^n$ for all $\xi > 0$. So

$$
\begin{aligned}
\mathbf{Pr}(Z \ge r) &\le o(1) + \sum_{r=3n/\ln n}^{n/3} \frac{6^n}{2^r r!} \frac{(-\ln(1-\xi))^r}{\xi^n} && \forall 0 < \xi < 1 \\
&= o(1) + \sum_{r=3n/\ln n}^{n/3} \frac{(12)^n (\ln 2)^r}{2^r r!} && \xi = 1/2 \\
&\le o(1) + \sum_{r=3n/\ln n}^{n/3} (12)^n \left(\frac{e}{2r}\right)^r \\
&\le o(1) + \sum_{r=3n/\ln n}^{n/3} e^{-n/2} && \text{for } n \text{ large} \\
&= o(1).
\end{aligned}
$$

□

2.3 The longest edge in F^*

We show that **whp** the longest edge in F^* is of length $O\left(\frac{\ln n}{n}\right)$. For this we define

$$\mathcal{T} = 20000$$

and let $\Gamma_{\mathcal{T}}$ be as defined in the proof of Lemma 1. The edges of $G_{\mathcal{T}}$ will tend to short, $O\left(\frac{1}{n}\right)$ in length. We show that **whp** $\Gamma_{\mathcal{T}}$ is a good expander. From this we will see that any long edge in a 2-factor, can be replaced using alternating paths with $O(\ln n)$ edges and total weight $O\left(\frac{\ln n}{n}\right)$.

For $S \subseteq [n]$ let

$$
\begin{aligned}
\vec{N}_{\mathcal{T}}(S) &= \{w \notin S : \exists v \in S \text{ such that } (v, w) \text{ is an arc of } \vec{\Gamma}_{\mathcal{T}}\}. \\
N_{\mathcal{T}}(S) &= \{w \notin S : \exists v \in S \text{ such that } (v, w) \text{ is an edge of } \Gamma_{\mathcal{T}}\}.
\end{aligned}
$$

Lemma 3 **Whp** $|\vec{N}_{\mathcal{T}}(S)| \ge 20|S|$ *for all* $S \subseteq [n]$, $|S| \le \lceil n/25 \rceil$.

Proof The arcs leaving S terminate in $S \cup \vec{N}_{\mathcal{T}}(S)$ and so

$$\begin{aligned}
\mathbf{Pr}(\exists S:\ |S| \leq \lceil n/25 \rceil,\ |\vec{N}_{\mathcal{T}}(S)| < 20|S|) &\leq \sum_{s=1}^{\lceil n/25 \rceil} \binom{n}{s}\binom{n}{20s}\left(\frac{\binom{21s}{\mathcal{T}}}{\binom{n}{\mathcal{T}}}\right)^s \\
&\leq \sum_{s=1}^{\lceil n/25 \rceil} \left(\frac{ne}{s}\right)^s \left(\frac{ne}{20s}\right)^{20s} \left(\frac{21s}{n}\right)^{\mathcal{T}s} \\
&= \sum_{s=1}^{\lceil n/25 \rceil} \left(\frac{e^{21}(21)^{\mathcal{T}} s^{\mathcal{T}-21}}{(20)^{20} n^{\mathcal{T}-21}}\right)^s \\
&= o(1).
\end{aligned}$$

□

We will need to be sure we can connect a pair of alternating paths. For this purpose we prove that **whp** every pair of large subsets of $\Gamma_{\mathcal{T}}$ are joined by many edges:

Lemma 4 **Whp** *$\Gamma_{\mathcal{T}}$ contains at least $2n$ edges joining S and T, for all $S, T \subseteq [n]$, $|S|, |T| \geq n/50$ and $S \cap T = \emptyset$.*

Proof In the construction of $\Gamma_{\mathcal{T}}$ each vertex makes $\mathcal{T}$ distinct choices. Suppose we instead consider the graph $\tilde{\Gamma}_{\mathcal{T}}$ where the $\mathcal{T}$ choices at each vertex are made with replacement and so are independent. We can couple the construction of $\Gamma_{\mathcal{T}}, \tilde{\Gamma}_{\mathcal{T}}$ so that, after coalescing parallel edges, the edge-set of $\Gamma_{\mathcal{T}}$ is a super-set of $\tilde{\Gamma}_{\mathcal{T}}$. Now the expected size of the difference between the two graphs is at most $n\binom{\mathcal{T}}{2}\frac{1}{n-1} = O(1)$ and so **whp** this difference is $O(\ln n)$. It will therefore be justifiable to continue the rest of the proof of this lemma in terms of $\tilde{\Gamma}_{\mathcal{T}}$.

For fixed disjoint sets S, T of size $n/50$, let X be the number of edges joining S and T in $\tilde{\Gamma}_{\mathcal{T}}$. Each $v \in S \cup T$ makes $\mathcal{T}$ choices, and this gives $\mathcal{T}n/25$ independent choices overall. The probability that a choice for $v \in S$ is in T is at least $1/50$ and vice-versa. Thus X dominates the binomial $B\left(800n, \frac{1}{50}\right)$. Thus, using the following (Chernoff) tail bound for the binomial,

$$\mathbf{Pr}(B(N,p) \leq (1-\epsilon)Np) \leq e^{-\epsilon^2 Np/2}$$

we see that (with $N = 800n$ and $\epsilon = 5/16$),

$$\mathbf{Pr}(X \leq 3n) \leq e^{-3n}.$$

So the probability there exists a pair S, T for which $X \leq 3n$ is at most

$$\binom{n}{n/50}^2 e^{-3n} = o(1).$$

(We should take account of multiple edges (x, y) where x chooses y and vice-versa. But **whp** there are $O(\ln n)$ of these). □

No two vertices have many common neighbours in $\Gamma_{\mathcal{T}}$.

Lemma 5 **Whp** *every pair of vertices x, y have at most 2 common neighbours in $\Gamma_{\mathcal{T}}$.*

Proof The probability that there is a pair of vertices x, y with 3 common neighbours in $\Gamma_{\mathcal{T}}$ is at most

$$\binom{n}{2}\binom{n}{3}\left(\frac{2\mathcal{T}}{n-1}\right)^6 = o(1)$$

since the probability that a pair of vertices are adjacent in $\Gamma_{\mathcal{T}}$ is $\frac{2\mathcal{T}}{n-1} - \frac{\mathcal{T}^2}{(n-1)^2}$ and this probability is not increased if we are given the existence of other edges. $\Box$

Now let F be an arbitrary 2-factor. We consider alternating paths in $\Gamma_{\mathcal{T}}$ with respect to F i.e. paths of the form $x_0, x_1, \ldots, x_k$ where the edges $(x_{2i}, x_{2i+1}) \notin F$ for $0 \leq i \leq \lceil k/2 \rceil$ and the edges $(x_{2i-1}, x_{2i}) \in F$ for $1 \leq i \leq \lfloor k/2 \rfloor$.

Lemma 6 Whp *for every 2-factor F and for every pair of vertices x, y there is an odd length alternating path from x to y of length at most $2i_0 + 1$ where $i_0 = 1 + \lceil \log_2(n/(50(\mathcal{T} - 4)) \rceil$.*

Proof Assume that the conditions of Lemmas 3, 4 and 5 hold.

Fix x, y, F and arbitrarily orient the cycles of F to obtain $\vec{F}$. For a vertex z let $\nu(z)$ be defined by $(z, \nu(z))$ is an arc of $\vec{F}$.

We define a collection of sets $S_0 = \{x\}$, $T_0 = \{y\}$, S_i, T_i, $i = 1, 2, \ldots, i_0$ where

(i) The collection of sets $S_i, \nu(S_i), T_i, \nu(T_i)$, $i = 1, 2, \ldots, i_0$ are pair-wise disjoint, even for different i.

(ii) S_i is reachable from x by an alternating path of length $2i - 1$ (number of edges), $1 \leq i \leq i_0$.

(iii) T_i is reachable from y by an alternating path of length $2i - 1$, $1 \leq i \leq i_0$.

(iv) $|S_1| = |T_1| = \mathcal{T} - 4$.

(v) $|S_{i+1}| = 2|S_i|$ and $|T_{i+1}| = 2|T_i|$ for $1 \leq i < i_0$.

Since (Lemma 5) $|N_{\mathcal{T}}(x) \cap N_{\mathcal{T}}(y)| \leq 2$ we can choose $\tilde{S}_1 \subseteq N_{\mathcal{T}}(x)$, $\tilde{T}_1 \subseteq N_{\mathcal{T}}(y)$ such that $|\tilde{S}_1| = |\tilde{T}_1| = \mathcal{T} - 2$ and $\tilde{S}_1, \tilde{T}_1$ are disjoint. At most 2 of the $x, \tilde{S}_1$ edges are in F and so we can choose $\mathcal{T} - 4$ members of $\tilde{S}_1$ to make S_1 and similarly we can choose $\mathcal{T} - 4$ members of $\tilde{T}_1$ to make T_1.

Now suppose that we have chosen S_i, T_i for some $1 \leq i < i_0$ and that $s = |S_i| = |T_i|$. Let $A_0 = N_{\mathcal{T}}(S_i), B_0 = N_{\mathcal{T}}(T_i)$ and $C = N_{\mathcal{T}}(S_i \cup T_i) \subseteq A_0 \cup B_0$. Since $|A_0|, |B_0| \geq 20s$ and $|C| \geq 40s$ we can choose $A_1 \subseteq A_0, B_1 \subseteq B_0$ such that $A_1 \cap B_1 = \emptyset$ and $|A_1| = |B_1| = 20s$. By deleting $2s$ vertices from A_1 we obtain a set $A_2 \subseteq A_0$ of size $18s$ such that all $S_i : A_2$ edges are not in F. Define B_2 analogously.

Next let $A_3 = \nu(A_2)$ and $B_3 = \nu(B_2)$. At this point the only possible intersections among A_2, A_3, B_2, B_3 are between A_3 and B_2 or between A_2 and B_3. Now choose $6s$ vertices $A_4 \subseteq A_3$ and let $A_5 = \nu^{-1}(A_4)$. Next choose $6s$ vertices B_4 from $B_3 \setminus (\nu(A_4) \cup A_5)$. By so doing we have $A_4, A_5, B_4, B_5 = \nu^{-1}(B_4)$ disjoint.

Since

$$\left| \bigcup_{j=0}^{i-1} (S_j \cup \nu^{-1}(S_j) \cup T_j \cup \nu^{-1}(T_j)) \right| \leq 4s$$

we can find $S_{i+1} \subseteq A_4, T_{i+1} \subseteq B_4$ so that (i) above holds and so complete the inductive step.

Now $|S_{i_0}|, |T_{i_0}| \geq n/50$ and so (Lemma 4) there are at least $2n$ edges joining these 2 sets, at least one of which is not in F. This proves the existence of the required alternating path. $\Box$

We now define the *effect* of an alternating path P w.r.t. F^* to be the difference between the sum of the lengths of the edges of P which are not in F^* and the sum of the lengths of the edges of P which are in F^*.

Lemma 7 *There exists an absolute constant $A_1 > 0$ such that* **whp** *for every 2-factor F and for every pair of vertices x, y there is an odd length alternating path from x to y of effect at most $\frac{A_1 \ln n}{n}$.*

Proof We will need the following inequality, Lemma 4.2(b) of Frieze and Grimmett [8].

Suppose that $k_1 + k_2 + \cdots + k_s \leq a$, and $Y_1, Y_2, \ldots, Y_s$ are independent random variables with Y_i distributed as the k_ith minimum of N independent uniform [0,1] random variables. If $\mu > 1$ then

$$\mathbf{Pr}\left(Y_1 + \cdots + Y_s \geq \frac{\mu a}{N+1}\right) \leq e^{a(1+\ln\mu-\mu)}. \tag{3}$$

Let

$$W_1 = \max\left\{\sum_{i=0}^{k} \mathbf{M}(x_{2i}, x_{2i+1}) - \sum_{i=1}^{k} \mathbf{M}(x_{2i-1}, x_{2i})\right\}, \tag{4}$$

where the maximum is over sequences $x_0, x_1, \ldots, x_{2k+1}$ where $k \leq i_0$ and $\mathbf{M}(x_{2i}, x_{2i+1}), i = 0, 1, \ldots, k$ is stochastically dominated by the ρ_ith smallest ($\rho_i \leq \mathcal{T}$) of $n-3$ independent copies of the random variable Z of Lemma 1. Here we can use the fact that in our definition of an alternating path the weight of (x_{2i}, x_{2i+1}) is stochastically dominated by the minimum of the (at least) $n-3$ $Z_1(x_{2i}, y)$'s corresponding to edges not in F. Notice that this quantity is independent of the length of $(x_{2i+1}, x_{2i+2}) \in F$. Because we are taking the maximum over all possible sequences, the remaining edge lengths are uniform and all lengths can be taken as independent. Because $1 - \sqrt{1-x} \geq x/2$ for $x \in [0,1]$ we see that Z is stochastically dominated by $2U$ where U is uniform $[0,1]$.

We estimate the probability that W_1 is large. Indeed, for any $\zeta > 0$ we have

$$\mathbf{Pr}\left(W_1 \geq \zeta\frac{\ln n}{n}\right) \leq \sum_{k=1}^{i_0} n^{2k+2}\frac{1}{(n-3)^{k+1}} \times$$
$$\int_{y=0}^{\infty}\left[\frac{1}{(k-1)!}\left(\frac{y\ln n}{n}\right)^{k-1} \sum_{\rho_1+\cdots+\rho_{k+1}\leq(k+1)\mathcal{T}} q(\rho_1, \ldots, \rho_{k+1}; \zeta + y)\right] dy$$

where

$$q(\rho_1, \ldots, \rho_{k+1}; \eta) = \mathbf{Pr}\left(X_1 + \cdots + X_{k+1} \geq \eta\frac{\ln n}{n}\right),$$

$X_1, \ldots, X_{k+1}$ are independent and X_j is distributed as twice the ρ_jth minimum of $n-3$ independent copies of Z.

Explanation: We have at most n^{2k+2} choices for the sequence $x_0, x_1, \ldots, x_{2k+1}$. The term $\frac{1}{(k-1)!}\left(\frac{y\ln n}{n}\right)^{k-1} dy$ bounds the probability that the sum of k independent uniforms, $\mathbf{M}(x_1, x_2) + \cdots + \mathbf{M}(x_{2k-1}, x_{2k})$, is in $\frac{\ln n}{n}[y, y+dy]$. (We approximate this probability by the area of the simplex face $\left\{y_1 + y_2 + \cdots + y_k = \frac{y\ln n}{n},\ y_1, y_2, \ldots, y_k \geq 0\right\}$ multiplied by dy.) We then integrate over y. $\frac{1}{n-3}$ bounds the probability that (x_{2i}, x_{2i+1}) is the ρ_ith shortest (in terms of Z_1) edge leaving x_{2i}, and these events are independent for $0 \leq i \leq k-1$. The final summation bounds the probability that the associated edge lengths sum to at least $\frac{(\zeta+y)\ln n}{n}$.

It follows from (3) with $N = n-3, a = (k+1)\mathcal{T}$ and $\mu = \frac{n-2}{n}\frac{(\zeta/2+y)\ln n}{a}$, that if ζ is sufficiently large then for all $y \geq 0$, we have $\mu \geq 3(1 + \ln\mu)$ and so

$$q(\rho_1, \ldots, \rho_k; \zeta + y) \leq e^{-2a\mu/3} \leq n^{-(\zeta+y)/3}.$$

Thenumber of choices for $\rho_1, \rho_2, \ldots, \rho_{k+1}$ is $\binom{(k+1)\mathcal{T}}{k+1}$ and so

$$\begin{aligned}
\mathbf{Pr}\left(W_1 \geq \zeta \frac{\ln n}{n}\right) &\leq 2n^{2-\zeta/3} \sum_{k=1}^{i_0} \frac{(\ln n)^{k-1}}{(k-1)!} \binom{(k+1)\mathcal{T}}{k+1} \int_{y=0}^{\infty} y^{k-1} n^{-y/3} dy \\
&= 2n^{2-\zeta/3} \sum_{k=1}^{i_0} \frac{(\ln n)^{k-1}}{(k-1)!} \binom{(k+1)\mathcal{T}}{k+1} \frac{3^k (k-1)!}{(\ln n)^k} \\
&\leq 2n^{2-\zeta/3} \sum_{k=1}^{i_0} (3\mathcal{T}e)^k \\
&\leq 2n^{2-\zeta/3} (3\mathcal{T}e)^{i_0+1} \\
&= o(1).
\end{aligned}$$

for sufficiently large ζ. □

The following lemma is almost immediate:

Lemma 8 **Whp** *F^* contains no edge longer than $A_1 \frac{\ln n}{n}$.*

Proof Suppose that F^* contains an edge $e = (x, y)$ of length greater than $A_1 \frac{\ln n}{n}$. Construct the alternating P path from x to y promised by Lemma 7. By removing e and the F^* edges of P from F^* and replacing them with the non-F^* edges of P we obtain a 2-factor of lower weight. □

Note that **whp** F^* contains an edge of length $\geq \frac{\ln n}{n}$. The distribution of the subgraph induced by edges of length $\leq p$ is the random graph $G_{n,p}$ for any $p \in [0, 1]$ and we need $p \geq \frac{\ln n}{n}$ in order that $\delta(G_{n,p}) \geq 2$ **whp**.

We use the notation
$$p_0 = A_1 \frac{\ln n}{n}$$
for the remainder of the paper.

As consequence of Lemma 8, we see that **whp** F^* does not contain many very short cycles.

Lemma 9 **Whp** *F^* contains at most $n^{3/4}$ cycles with fewer than $\frac{\ln n}{2 \ln \ln n}$ edges.*

Proof Let Z_k denote the number of cycles of K_n, with k or fewer edges, all of whose edges are of length at most p_0. Then

$$\mathbf{E}(Z_k) = \sum_{l=3}^{k} \binom{n}{l} \frac{(l-1)!}{2} p_0^l \leq \sum_{l=3}^{k} \frac{(A_1 \ln n)^l}{2l} \leq (A_1 \ln n)^k \leq n^{1/2+o(1)}$$

if $k \leq \frac{\ln n}{2 \ln \ln n}$. Now use the Markov inequality. □

2.4 Long and short edges

We can divide the edges (x, y) of K_n into *long*, length $\mathbf{M}(x, y) \geq p_0$ and *short* edges. From the previous section we see that **whp** it is enough to find a minimum weight 2-factor in the graph induced by the short edges. If for each short edge (x, y) we generate an extra parallel edge with length uniform in the range $[p_0, 1]$ then we can consider that we start with G_{n,p_0}, with edge weights uniform in $[0, p_0]$ plus an independent K_n with edge weights uniform in $[p_0, 1]$ where we always use the shortest edge between a pair of vertices x, y.

We further divide the long edges into *very long*, length $\geq 2p_0$ and *medium* length edges. Thus we will obtain another *Red* copy of G_{n,p_0} with weights in the range $[p_0, 2p_0]$ and a *Blue* copy of K_n with edge lengths in the range $[2p_0, 1]$.

It is important to realise that when we say we use an edge of a particular graph, say a Red edge, we are really just upper bounding the length of the edge in the original K_n.

3 A Conditional Model

For a permutation π of $[n]$ and matrix of weights $\mathbf{M}$ we define $\mathbf{M}_\pi$ by $\mathbf{M}_\pi(i,j) = \mathbf{M}(\pi(i), \pi(j))$. Clearly $\mathbf{M}$ and $\mathbf{M}_\pi$ have the same distribution. So for any 2-factor F,

$$\mathbf{Pr}(F^*(\mathbf{M}) = F) = \mathbf{Pr}(F^*(\mathbf{M}_\pi) = F).$$

But $F^*(\mathbf{M}_\pi) = F$ iff $F^*(\mathbf{M}) = F \circ \pi$ where (i,j) is an edge of $F \circ \pi$ iff $(\pi(i), \pi(j))$ is an edge of F. So

$$\mathbf{Pr}(F^* = F) = \mathbf{Pr}(F^* = F \circ \pi).$$

Now as π ranges over the $n!$ permutations of $[n]$, $F \circ \pi$ ranges over all 2-factors having the same cycle structure as F – cycle $i_1, i_2, \ldots, i_l, i_1$ of F is mapped to cycle $\pi^{-1}(i_1), \pi^{-1}(i_2), \ldots, \pi^{-1}(i_l)$, $\pi^{-1}(i_1)$ of $F \circ \pi$. By symmetry each of these 2-factors appears the same number of times.

For a sequence $\mathbf{k} = 3 \leq k_1 \leq k_2 \leq \cdots \leq k_m$ we let $\Omega_{\mathbf{k}}$ denote the set of 2-factors with these cycle sizes. If we compute F^* by first choosing a random permutation π, then computing $F^*(\mathbf{M}_\pi)$ and then taking $F^*(\mathbf{M}_\pi) \circ \pi$ then we see that:

$$\text{Given } F^* \in \Omega_{\mathbf{k}},\ F^* \text{ is a uniform random member of } \Omega_{\mathbf{k}}. \tag{5}$$

So we will now fix the cycle sizes $\mathbf{k}$ and assume that the conditions of Lemmas 2, 9 hold. We will run our proposed algorithm under the assumption that we know $\mathbf{A} = \mathbf{M}_\pi$ and $\tilde{F} = F^*(\mathbf{M}_\pi)$ but that π is a random permutation that we will *expose* as necessary. More precisely we assume that

(i) $F^*(\mathbf{A}) \in \Omega_{\mathbf{k}}$.

(ii) The conditions of Lemma 10 below are satisfied.

(iii) We have the graph decomposition of Section 2.4.

It will help to imagine that we have m cycles $C_1, C_2, \ldots, C_m$ where $|C_i| = k_i$, $i = 1, 2, \ldots, m$. We can imagine these as being drawn in a plane. The vertices of these cycles are $X = \{x_1, x_2, \ldots, x_n\}$. We will assume that these cycles have an (arbitrary) orientation. Then for each $x \in X$ there is a predecessor $\nu(x)$ on the same cycle as x. As we go we expose a random mapping f from $[n] \to X$ and then $\pi = f^{-1}\nu f$. If we establish that $f(i) = x_k$ and $f(j) = x_l$ then we will also establish the length of the edge (x_k, x_l) as $\mathbf{A}(i,j)$.

The vertices of X and $[n]$ are divided into *exposed* and *unexposed*. $v \in [n]$ is exposed iff $f(v)$ has been determined and $x \in X$ is exposed iff $f^{-1}(x)$ has been determined.

4 The algorithm

We break our algorithm into 2 phases: A *Greedy Phase* and an *Extension-Rotation Phase*.

4.1 The Greedy Phase

We start by deleting an edge of C_1. This leaves a path P_0. In general, we have a path P, with endpoints a_0, a_1. Initially $P = P_0$. We further have a collection of cycles $\mathcal{C} = \mathcal{C}_S \cup \mathcal{C}_L$ where $\mathcal{C}_S = \{C \in \mathcal{C} : |C| \leq (\log n)^2\}$ and $\mathcal{C}_L = \{C \in \mathcal{C} : |C| > (\log n)^2\}$. Initially $\mathcal{C} = C_2, C_3, \ldots, C_m$. We define the set of vertices $R = \bigcup_{C \in \mathcal{C}_S} C$. At each iteration we find a short edge e from a_1 to a vertex x in a cycle $C \in \mathcal{C}$. Then we delete an edge of C incident with x. This lengthens P and reduces the number of cycles in $\mathcal{C}$ by one.

All of the vertices of P will be exposed. Most of R will be unexposed. We end the Greedy Phase when $|R|$ first drops below $\frac{n}{\sqrt{\ln n}}$.

U_R denotes the set of unexposed vertices in R and U_n denotes the unexposed vertices of $[n]$. We never allow the number of exposed vertices in R to reach more than $6\frac{n}{\ln n}$. We terminate the algorithm and fail if we expose this number.

A general step of this phase involves the following substeps:

(S1) Determine $f^{-1}(a_1)$ by a random choice from U_n.

(S2) Determine the shortest *acceptable* (defined below) edge (a_1, x) from a_1 to a vertex ξ of $\mathcal{C}_S$ for which $\nu(\xi)$ is unexposed. Assume that x lies in cycle C. Delete the edge $(x' = \nu(x), x)$ from C to create a path Q. Now replace P by $P + Q$ and delete C from $\mathcal{C}$.

Thus each step reduces the number of cycles left by one, at a cost of less than the length of the edge (a_1, x).

In step **S2** above, a_1 is replaced by x'. So we want x' to be unexposed and to have many unexposed vertices which are close to it. The following values are used in the definition of acceptable. The justification for choosing these values comes from Lemma 10 below.

$$\begin{array}{lll} \omega_1 = 2(\ln n)^{1/4} & \omega_2 = (\ln n)^{1/3} & \epsilon = (\ln n)^{-1/5} \\ \omega_1' = 2(\ln n)^5 & \omega_2' = (\ln n)^7 & \epsilon' = (\ln n)^{-1} \end{array}$$

$\xi \in R$ is *good* if

(i) ξ is unexposed.

(ii) Let the unexposed vertices of R be enumerated as $x_1, x_2, \ldots$ where $\mathbf{A}(\xi, x_i) \leq \mathbf{A}(\xi, x_{i+1})$ for $i \geq 1$. Also, let $Y_1 = \{\xi_1, \xi_2, \ldots, \xi_{\omega_1}\}$ and let $Y_1' = \{\xi_1, \xi_2, \ldots, \xi_{\omega_1'}\}$

(a) $\mathbf{A}(\xi, \xi_{\omega_1}) \leq \frac{\omega_2}{|R|}$.

(b) $\mathbf{A}(\xi, \xi_{\omega_1'}) \leq \frac{\omega_2'}{|R|}$.

(iii) Let the cycle containing v be C. Then either

(a) $|R \setminus C| \leq \frac{n}{\sqrt{\ln n}}$ or

(b) $|R \setminus C| > \frac{n}{\sqrt{\ln n}}$, and $|f(Y_1) \cap C| \leq |Y_1|/2$ and $|f(Y_1') \cap C| \leq |Y_1'|/2$

We define the edge (a_1, x) to be *acceptable* if x' is good.

We check now that the search for x can actually be done without exposing too many vertices: We know $\tilde{a} = f^{-1}(a_1)$. We go through $\xi \in f^{-1}(R)$ in increasing order of $\mathbf{A}(\tilde{a}, \xi)$. If ξ is exposed then we go on to the next ξ. If ξ is unexposed then we choose $\eta = f(\xi)$ randomly from U_R. Let

$\eta' = \nu(\eta)$. If η' has been exposed, we go onto the next ξ. Otherwise we expose η' by randomly choosing $\nu = f^{-1}(\eta')$ from U_n. We then check to see whether or not η' is good. We go through $\xi \in f^{-1}(R)$ in increasing order of $\mathbf{A}(\nu, \xi)$ and we examine the first ω_1' unexposed ξ and see whether conditions (ii), (iii) of goodness are satisfied. We do not expose these ξ unless η' passes this test and we take $x' = \eta'$.

We need to be sure that in Step **S2** we are likely to find a short acceptable choice of edge. Before considering the expected length of the accepted edge, we mention what are the only possibilities: Recall that C is the cycle containing x.

(A) $|R \setminus C| \le \frac{n}{\sqrt{\ln n}}$.

(B) $|R \setminus C| > \frac{n}{\sqrt{\ln n}}$ and $|f(Y_1) \cap C| \le \omega_1/2$ and $|f(Y_1') \cap C| \le \omega_1'/2$.

(C) $|R \setminus C| > \frac{n}{\sqrt{\ln n}}$ and $|f(Y_1) \cap C| > \omega_1/2$

(D) $|R \setminus C| > \frac{n}{\sqrt{\ln n}}$ and $|f(Y_1') \cap C| > \omega_1'/2$.

In Case **A** we terminate the Greedy Phase and begin the Extension-Rotation phase. Otherwise we can expose the vertices of C and determine which of Cases **B**–**D** we are in. Now,

$$\mathbf{Pr}(\mathbf{C} \text{ or } \mathbf{D}) \le \binom{\omega_1}{\omega_1/2}\left(\frac{(\ln n)^{5/2}}{n}\right)^{\omega_1/2} + \binom{\omega_1'}{\omega_1'/2}\left(\frac{(\ln n)^{5/2}}{n}\right)^{\omega_1'/2} \le n^{-10}. \tag{6}$$

Explanation If $S \subseteq Y_1$ then $\mathbf{Pr}(f(S) \subseteq C) = \binom{|C|}{|S|} / \binom{|R|}{|S|} \le \left(\frac{(\ln n)^{5/2}}{n}\right)^{|S|}$.

The probability in (6) is small enough that we can afford to fail if either Case **C** or **D** happens. We assume therefore that we do not come across these cases. They would cause trouble, because when we extend P by adding C, we would find that many of the unexposed vertices close to x' are on the new path and are therefore unusable.

We next need to estimate the expected length of the edge (a_1, x) in Case **B**.

Lemma 10 *Suppose the following holds as $n \to \infty$,*

$$\alpha_0 \to 0, \qquad \frac{\alpha\beta_2}{\ln\ln n} \to \infty, \qquad \frac{\beta_2}{\beta_1 \ln \beta_2} \to \infty \qquad \beta_2 \le (\log n)^{10}.$$

Then **whp**, *for every $K \subseteq [n]$, $k = |K| \ge n/\sqrt{\ln n}$ and $L \subseteq K$, $|L| \le \alpha_0 k$ there are at most αk vertices v for which $|\{w \in K \setminus L : \ \mathbf{A}(v, w) \le \frac{\beta_2}{k}\}| < \beta_1$ i.e. whose β_1th closest neighbour in $K \setminus L$ is at $\mathbf{A}$ distance $\ge \frac{\beta_2}{k}$.*

Proof Fix $S, L \subseteq K \subseteq [n]$ with $|K| = k \ge n/\sqrt{\ln n}$ and $|L| = \alpha_0 k$ and $|S| = \alpha k$. The probability that for each $v \in S$, $|\{w \in K \setminus L : \ \mathbf{A}(v, w) \le \frac{\beta_2}{k}\}| < \beta_1$ is at most

$$\left(\sum_{i=0}^{\beta_1} \binom{k(1-\alpha_0)}{i} \left(1 - \sqrt{1 - \frac{\beta_2}{k}}\right)^i \left(1 - \frac{\beta_2}{k}\right)^{(k(1-\alpha_0)-i)/2}\right)^{\alpha k}. \tag{7}$$

Explanation: We can express $\mathbf{A}(x, y)$ as $\min\{Z_{\mathbf{A}}(x, y), Z_{\mathbf{A}}(y, x)\}$ where the $Z_{\mathbf{A}}(x, y)$ are independent copies of Z. For $x \in S$, the term $\left(1 - \sqrt{1 - \frac{\beta_2}{k}}\right)^i \left(1 - \frac{\beta_2}{k}\right)^{(k(1-\alpha_0)-i)/2}$ is the probability that exactly i of the quantities $Z_{\mathbf{A}}(x, y), y \in K \setminus L$ are at most $\frac{\beta_2}{k}$. The expression in (7) is then the probability that for each $x \in S$, at most β_1 of the quantities $Z_{\mathbf{A}}(x,y), y \in K \setminus L$ are at most $\frac{\beta_2}{k}$. This event is implied by the event that for each $x \in S$, at most β_1 of the quantities $\mathbf{A}(x,y), y \in K \setminus L$ are at most $\frac{\beta_2}{k}$.

We now bound

$$\begin{aligned}
&\sum_{i=0}^{\beta_1} \binom{k(1-\alpha_0)}{i} \left(1-\sqrt{1-\frac{\beta_2}{k}}\right)^i \left(1-\frac{\beta_2}{k}\right)^{(k(1-\alpha_0)-i)/2} \\
\leq &\sum_{i=0}^{\beta_1} \frac{k^i(1-\alpha_0)^i}{i!} \left(\frac{\beta_2}{k}\right)^i e^{-\beta_2(1-\alpha_0)/3} \\
\leq &((1-\alpha_0)\beta_2)^{\beta_1} e^{-\beta_2(1-\alpha_0)/3}.
\end{aligned}$$

Thus the probability that there exist K, L not satisfying the conditions of the lemma is at most

$$\begin{aligned}
&\binom{n}{k}\binom{k}{\alpha k}\binom{k}{\alpha_0 k}(((1-\alpha_0)\beta_2)^{\beta_1} e^{-\beta_2(1-\alpha_0)/3})^{\alpha k} \\
\leq &\left(\frac{ne}{k} \cdot \left(\frac{e}{\alpha}\right)^{\alpha} \cdot \left(\frac{e}{\alpha_0}\right)^{\alpha_0} \left(((1-\alpha_0)\beta_2)^{\beta_1} e^{-\beta_2(1-\alpha_0)/3}\right)^{\alpha}\right)^k \\
\leq &\left((\ln n)^{1/2} \cdot e \cdot e \cdot e^{-\beta_2 \alpha/4}\right)^k \\
= &\, o(1).
\end{aligned}$$

□

We know that a_1 was either obtained from C_1 or was a good vertex chosen in the previous step. In both cases there were at least ω_1 (resp. ω_1') edges of length $\leq \frac{\omega_2}{|R|}$ (resp. $\leq \frac{\omega_2'}{|R|}$) to unexposed members of R. Because cases **C**, **D** are ruled out, at most 1/2 of these are absorbed into P. Using Lemma 10 with $\alpha_0 = \frac{6}{\sqrt{\ln n}}$ (a bound on the proportion of exposed vertices in R, see (10) below) and $\alpha = \epsilon, \epsilon'$ we see that the expected length of the edge (a_1, x) is at most

$$\frac{\omega_2}{|R|} + (\alpha_0 + \epsilon + \epsilon')^{\omega_1/2} \frac{\omega_2'}{|R|} + (\alpha_0 + \epsilon + \epsilon')^{\omega_1'/2} \leq \frac{2(\ln n)^{1/3}}{|R|}. \tag{8}$$

Explanation Assume the condition of Lemma 10 hold with $\alpha_0, \alpha = \epsilon, \beta_1 = \omega_1, \beta_2 = \omega_2$ and with $\alpha_0, \alpha = \epsilon', \beta_1 = \omega_1', \beta_2 = \omega_2'$. Let $K = R$ and $L = R \setminus U_R$. $\alpha_0 + \epsilon + \epsilon'$ bounds the probability that a randomly chosen vertex x of R has $\nu(x) \notin U_R$ (prob. $\leq \alpha_0$) or $\nu(x)$ fails tests (iia) (prob. $\leq \epsilon$), (iib) (prob. $\leq \epsilon'$). of goodness.

Thus the number of vertices v_i exposed in the ith step before finding an acceptable edge, given the previous history is dominated by a geometric random variable with probability of success $p = 1 - (\alpha_0 + \epsilon + \epsilon')$. Thus for any $\lambda > 0$ such that $(1-p)e^{\lambda} < 1$ we have

$$\mathbf{E}(e^{\lambda v_i} \mid v_j, j < i) \leq 1 + \frac{e^{\lambda} - 1}{1 - (1-p)e^{\lambda}}. \tag{9}$$

Now let T_G denote the number of steps in the Greedy Phase. Now we know from Lemma 2 that **whp** $T \leq T_0 = \frac{3n}{\ln n}$. Defining $v_i = 0$, $T < i \leq T_0$ we see that (9) holds for $1 \leq i \leq T_0$. Now let $\Upsilon = \sum_{i=1}^{T_0} v_i$ be the total number of vertices exposed (**whp**). Then

$$\mathbf{Pr}(\Upsilon \geq 2T_0) \leq e^{-2\lambda T_0} \mathbf{E}(e^{\lambda \Upsilon}) \leq \exp\left\{-2T_0\lambda + T_0 \frac{e^{\lambda} - 1}{1 - (1-p)e^{\lambda}}\right\} = o(e^{-T_0/5}), \tag{10}$$

if we put $\lambda = 1$. Thus **whp** the number of exposed vertices in R is always at most $2T_0$.

This explains why we can take as our upper bound on $\frac{|R \setminus U_R|}{|R|}$, the value $\alpha_0 = \frac{6}{\sqrt{\ln n}}$. The probabilistic bound on the number of exposed vertices will hold throughout the Greedy Phase.

We now return to the main cost of the Greedy Phase (as defined in (8)).

We remind the reader that the lengths of the cycles in the optimal 2-factor F^* are $k_1, k_2, \ldots, k_m$ where $m \leq 3\frac{n}{\ln n}$ and $k_1 = \min_i k_i$.

We can take $k_1 \leq (\ln n)^2$ for otherwise $m \leq n/(\ln n)^2$ and we can dispense with a Greedy Phase and just use the Extension-Rotation Phase. In latter this Phase **whp** we remove a cycle at the cost of $O\left(\frac{\ln n}{n}\right)$ per cycle. Thus in the case of $k_1 \geq (\ln n)^2$ we can **whp** find a tour of length $O((\ln n)^{-1})$ more than the length of F^*.

We now wish to bound the expected sum of the lengths of the edges added. Suppose that we have re-ordered the cycles so that they are absorbed into P in order 1,2,.... Assume that $\mathcal{C}_L = \{C_{\rho+1}, C_{\rho+2}, \ldots, C_m\}$ and let $K_L = \sum_{i=\rho+1}^m k_i$.

Let $L_i = n - k_1 - k_2 - \cdots - k_i - K_L = k_{i+1} + k_{i+2} + \cdots + k_\rho$, $m' = \min\{i : L_i \leq \frac{n}{\sqrt{\ln n}}\}$. We note that the Greedy Phase is only concerned with the first $m' - 1$ cycles. Next let $I = \{i : k_i \leq L = \frac{\ln n}{2 \ln \ln n}\}$. We can assume (Lemma 9) that $|I| \leq n^{3/4}$. Then let

$$S(m') = \sum_{i=1}^{m'-1} \frac{(\ln n)^{1/3}}{L_i}.$$

It will suffice to show that $S(m') = o(1)$. For then (8) will imply that the expected weight of the edges added in the Greedy Phase is $o(1)$ and so is $o(1)$ **whp** by the Markov inequality. Let $J = \lceil (L_1 - n/\sqrt{\ln n})/L \rceil$. Then

$$\sum_{i=1}^{m'-1} \frac{1}{L_i} \leq |I| \frac{\sqrt{\ln n}}{n} + \sum_{j=0}^{J} \frac{1}{L_1 - jL}$$

since $|L_i| \geq \frac{\sqrt{\ln n}}{n}$ for $i \in I$ and if $j, j' \in [m'-1] \setminus I = \{j_1, j_2, \ldots, \}$ we have $\lceil \frac{L_1 - L_j}{L} \rceil \neq \lceil \frac{L_1 - L_{j'}}{L} \rceil$ and $L_j \geq L_1 - JL$.

$$\tag{11}$$

$$\begin{aligned} &\leq O(n^{-1/4}\sqrt{\ln n}) + \int_{x=0}^{J} \frac{dx}{L_1 - xL} \\ &= O(n^{-1/4}\sqrt{\ln n}) + \frac{1}{L} \ln\left(\frac{L_1}{L_1 - JL}\right). \end{aligned} \tag{12}$$

Now

$$L_1 - JL \geq \frac{n}{2\sqrt{\ln n}}.$$

Thus we have

$$\frac{1}{L} \ln\left(\frac{L_1}{L_1 - JL}\right) \leq \frac{2(\ln \ln n)^2}{\ln n}.$$

Plugging this into (12) we see that

$$S(m') = O\left(\frac{(\ln \ln n)^2}{\ln n}\right)$$

and we are done with the Greedy Phase.

4.2 Final Extension-Rotation Phase

We enter this phase with a path P and, most importantly, only $m'' = m - m' + |\mathcal{C}_L| = O\left(\frac{n \ln \ln n}{(\ln n)^{3/2}}\right)$ cycles. We will absorb each cycle into P at an expected cost of $O\left(\frac{\ln n}{n}\right)$ and with the same cost turn the final Hamilton path into a tour and so complete the proof of Theorem 1.

We will use rotations and $\Gamma_{\mathcal{T}}$ of Section 2.3 for this task. At a general stage we have, as usual, a path P plus a collection of vertex disjoint cycles $\mathcal{C}$ which cover a set of vertices R. Let the endpoints of P be a, b.

If a (or b) has a $\Gamma_{\mathcal{T}}$-neighbour x in R then we replace P by $P + C - (x, x')$ where C is the cycle containing C and x' is a neighbour of x on C. We do not need to be concerned anymore with good or bad vertices. We call this operation, *extending* P and for every path obtained by rotation, we also see if an extension is possible. So, in the discussion of rotations below, **assume that no extension is possible for any path produced**.

For a path $P = (a = a_0, a_1, \ldots, a_h)$ and edge $a_h a_i$ we say that the path $P' = a_0, a_1, ...a_i, a_h, a_{h-1}, \ldots, a_{i+1}$ is obtained from P by a rotation with a_0 as fixed endpoint. For a vertex $v \in P$ let $\rho(v)$ be the minimum number of rotations, with a fixed, needed to construct a path with v as an endpoint. $\rho(v) = \infty$ if it is not possible to construct such a path. Then let

$$S(P, a, t) = \{v \in P : \ \rho(v) = t\}.$$

It follows from Lemma 3 that

$$|S(P, a, t)| \leq n/25 \text{ implies } |S(P, a, t+1)| \geq 9|S(P, a, t)| \tag{13}$$

Indeed, assuming no extensions are possible, and proceeding inductively,

$$|S(P, a, t+1)| \geq \frac{1}{2}\left(|N_{\mathcal{T}}(S(P, a, t))| - \sum_{\tau=0}^{t} |S(P, a, \tau)|\right) \geq \frac{1}{2}\left(20 - \frac{9}{8}\right)|S(P, a, t)|.$$

It follows that for some $t^* \leq \log_9 n/25$ we find that $|S(P, a, t^* + 1)| \geq 9n/25$. Let $END(a) = S(P, a, t^* + 1)$.

Now we have for each $v \in END(a)$ a path P_v with endpoints a, v which goes through all vertices of P (unless we have found an extension). For each such v construct the set $END(v) = S(P_v, v, t_v)$, $t_v \leq 1 + \log_9 n/25$ for which $|END(v)| \geq 9n/25$. Putting $END = \{a\} \cup END(a)$ we see that we have created a collection of sets $END(v), v \in END$, each of size $\geq 9n/25$ with the property that

> $v \in END$, $w \in END(v)$ implies that there is a path $P[v, w]$ with endpoints v, w going through all vertices of P and such that $P[v, w]$ differs from P in at most $2(1 + \log_9 n/25)$ edges (14)

Now we can use the Red copy of G_{n,p_0} (see Section 2.4) to find a Red edge joining some $v \in END$ to $w \in END(v)$. **Whp** we only need to check $O(\ln n)$ such pairs v, w for each cycle before finding a red edge, $O(n)$ pairs altogether. Thus we will **whp** find a Red edge each time we need to.

Once we have turned P into a cycle C, we can use the fact that **whp** Γ_m is connected for $m \geq 2$, Fenner and Frieze [6], to assert the existence of a $\Gamma_{\mathcal{T}}$ edge joining C to $\mathcal{C}$. In the event that $\mathcal{C}$ is empty, we have finished. Otherwise, we choose an edge (x, y) with $x \in C$ and $y \in C' \in \mathcal{C}$. We then remove an edge adjacent to x from C and an edge adjacent to y from C'. This gives us a path P, from which to continue with the process of reducing the number of cycles.

The cost of the added Red edges is $O(m'' p_0) = o(1)$. We can use (3) and (14) to see that **whp** the total weight of $\Gamma_{\mathcal{T}}$ edges used in this phase is $O\left(m'' \frac{\ln n}{n}\right) = o(1)$.

4.3 Running Time

We now summarize the running time of our algorithm. It takes $O(n^3)$ time to find the minimum weight 2-factor and as we will see, this dominates the rest of the algorithm. Each iteration of the Greedy Phase requires a search for an acceptable edge and this takes $O(n)$ time and so the Greedy Phase requires $O(n^2)$ time. We finally consider the Extension-Rotation phase. For a path P, it needs $O(n^2)$ rotations to create the sets END and $END(v)$ for $v \in END$). Angluin and Valiant [2] describe a data structure which allows a rotation to be done in $O(\ln n)$ time. Thus the total time for the Extension-Rotation phase is $O\left(\frac{n}{(\ln n)^{3/2}} n^2 \ln n\right) = o(n^3)$.

The proof of Theorem 1 is now complete.

References

[1] D. Aldous, The $\zeta(2)$ Limit in the Random Assignment Problem. *Random Structures and Algorithms* 18 (2001) 381-418.

[2] D. Angluin and L.G. Valiant, Fast Probabilistic Algorithms for Hamiltonian Circuits and Matchings, *Journal of Computer and System Sciences* 18 (1979) 155-193.

[3] E. Balas and P. Toth, Branch and bound methods, in *The traveling salesman problem: a guided tour of combinatorial optimization, E.L.Lawler, J.K.Lenstra, A.H.G.Rinnooy Kan and D.B.Shmoys Eds.*, Wiley, 1985.

[4] C.Cooper and A.M.Frieze, Hamilton cycles in random graphs and digraphs, *Random Structures and Algorithms* 16 (2000) 369-401.

[5] M.E. Dyer and A.M. Frieze, On patching algorithms for random asymmetric travelling salesman problems, *Mathematical Programming* 46 (1990) 361-378.

[6] T.I. Fenner and A.M.Frieze, On the connectivity of random m-orientable graphs and digraphs, *Combinatorica* 2, 347-359

[7] A.M. Frieze, On the value of a random minimum spanning tree problem, *Discrete Applied Mathematics* 10 (1985) 47-56.

[8] A.M. Frieze and G.R. Grimmett, *The shortest path problem for graphs with random arc-lengths*, Discrete Applied Mathematics 10 (1985) 57–77.

[9] A.M. Frieze and G. Sorkin, The probabilistic relationship between the assignment and asymmetric traveling salesman problems, *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms, Baltimore MD* (2001) 652-660

[10] R.M. Karp, A patching algorithm for the non-symmetric traveling salesman problem, *SIAM Journal on Computing* 8 (1979) 561-573.

[11] R.M. Karp and J.M. Steele, Probabilistic analysis of heuristics, in *The traveling salesman problem: a guided tour of combinatorial optimization, E.L.Lawler, J.K.Lenstra, A.H.G.Rinnooy Kan and D.B.Shmoys Eds.*, Wiley, 1985.

[12] L. Lovász and M.D. Plummer, Matching Theory, North Holland, 1986.

[13] D.W. Walkup, On the expected value of a random asignment problem, *SIAM Journal on Computing* 8 (1979) 440-442.