Parameterized TSP: Beating the Average

Gregory Gutin
Royal Holloway, University of London
e-mail: gutin@cs.rhul.ac.uk

Viresh Patel
University of Amsterdam
e-mail: viresh.s.patel@gmail.com

December 8, 2015

Abstract

In the Travelling Salesman Problem (TSP), we are given a complete graph K_n together with an integer weighting w on the edges of K_n, and we are asked to find a Hamilton cycle of K_n of minimum weight. Let $h(w)$ denote the average weight of a Hamilton cycle of K_n for the weighting w. Vizing (1973) asked whether there is a polynomial-time algorithm which always finds a Hamilton cycle of weight at most $h(w)$. He answered this question in the affirmative and subsequently Rublineckii (1973) and others described several other TSP heuristics satisfying this property. In this paper, we prove a considerable generalisation of Vizing's result: for each fixed k, we give an algorithm that decides whether, for any input edge weighting w of K_n, there is a Hamilton cycle of K_n of weight at most $h(w) - k$ (and constructs such a cycle if it exists). For k fixed, the running time of the algorithm is polynomial in n, where the degree of the polynomial does not depend on k (i.e. the generalised Vizing problem is fixed-parameter tractable with respect to the parameter k).

1 Introduction

The Travelling Salesman Problem (TSP) is one of the most well-known and widely studied combinatorial optimisation problems. In this problem, we are given an n-vertex complete graph K_n with weights on its edges and we are required to find a Hamilton cycle in K_n of minimum total weight. In its full generality, TSP is not only NP-hard, but also NP-hard to approximate to within any constant factor. Therefore there has been much attention in developing approximation algorithms for restricted instances of TSP. In this paper, we consider general TSP, but rather than seeking a Hamilton cycle of minimum weight, we seek a Hamilton cycle that beats the average weight of all Hamilton cycles by some given value.

Let us fix some notation in order to state our result. We use $\mathbb{Z}$, $\mathbb{N}$ and $\mathbb{Q}$ to denote the integers, the positive integers and the rational numbers respectively. As usual $V(G)$ and $E(G)$ denote the vertex and edge sets of a graph G. Let w be an integer edge weighting of K_n, i.e. $w : E(K_n) \to \mathbb{Z}$, and let G be a subgraph of K_n. We write

$$w(G) := \sum_{e \in E(G)} w(e) \qquad \text{and} \qquad w[G] := \sum_{e \in E(G)} |w(e)|$$

and we define the *density* $d = d(w)$ of w to be the average weight of an edge, i.e. $d := w(K_n)/\binom{n}{2}$. Note that $\mathbb{E}(w(\tilde{H})) = dn$, where $\tilde{H}$ is a uniformly random Hamilton cycle of K_n, and so there always exists a Hamilton cycle H^* satisfying $w(H^*) \le dn$.

Vizing [18] asked whether there is a polynomial-time algorithm which, given an integer edge weighting w of K_n, always finds a Hamilton cycle H^* of K_n satisfying $w(H) \le dn$. He answered this question in the affirmative and subsequently Rublineckii [17] described several other TSP heuristics satisfying this property. Such TSP heuristics including more recent ones are given in [10]. A natural question extending Vizing's question is the following: for each fixed k, is there a polynomial-time algorithm which, given w, determines if there exists a Hamilton cycle H^* satisfying $w(H^*) \le dn - k$? We give an affirmative answer to this question.

Theorem 1.1 *There exists an algorithm which, given (n, w, k) as input, where $n, k \in \mathbb{N}$ and $w : E(K_n) \to \mathbb{Z}$, determines whether there exists a Hamilton cycle H^* of K_n satisfying*

$$w(H^*) \le dn - k$$

(and outputs such a Hamilton cycle if it exists) in time $O(k^3)! + O(k^3 n) + O(n^7) = f(k)n^{O(1)}$.

Note that our algorithm includes arithmetic operations which are assumed to take time $O(1)$ and so our running times here and throughout are stated in the *strong* sense (see e.g. [8]). To obtain the running time in the *weak* sense, one simply multiplies by $\log M$, where $M := \max_{e \in E(K_n)} |w(e)|$ for the input instance (n, w, k).

Theorem 1.1 immediately implies that the following NP-hard problem (which is essentially TSP) is fixed-parameter tractable[1] when parameterised by k (and since this is our goal, we make no attempt to optimise the running time in Theorem 1.1).

Travelling Salesman Problem Below Average (TSP_{BA})

Instance: (n, w, k), where $n, k \in \mathbb{N}$ and $w : E(K_n) \to \mathbb{Z}$.
Question: Is there a Hamilton cycle H^* of K_n satisfying $w(H^*) \leq dn - k$?

Theorem 1.1 is proved by applying a combination of probabilistic, combinatorial, and algorithmic techniques, some of which are inspired by [12]. The key step to proving Theorem 1.1 is Theorem 1.2 below, which characterises those weightings in which all Hamilton cycles have weight close to the average. We believe this result will have further applications.

Theorem 1.2 *For any $n, k \in \mathbb{N}$ satisfying $n > 50000(k + 1)$, and $w : E(K_n) \to \mathbb{Z}$, in time $O(n^7)$ we can find either*

(a) *A Hamilton cycle H^* of K_n satisfying $w(H^*) < dn - k$, or*

(b) *A weighting $w^* : E(K_n) \to \mathbb{Z}$ and $\alpha \in \mathbb{Z}$ satisfying $w^*(H) = w(H) + \alpha$ for all Hamilton cycles H of K_n and $w^*[K_n] \leq 5000kn$.*

Note that since k is the parameter, it can be viewed as a constant, and so we may assume that $n \geq g(k)$ for any function g.

Related work The problem we consider in this paper falls into a class of problems introduced by Mahajan, Raman, and Sikdar [13]. The general framework is the following. Consider a combinatorial optimisation problem in which one is seeking a feasible solution of minimum (or maximum) *value* and suppose further that one is always guaranteed to find a feasible solution whose *value* is at most (or at least) some non-trivial (often tight) bound b (e.g. in our case, for any instance of TSP, one can always find a Hamilton cycle of weight at most dn where d is the average weight of an edge of K_n). One can then consider the problem, parameterised by k, of determining whether there exists a feasible solution of value at most $b - k$ (or at least $b + k$). A variety of techniques combining tools from linear algebra, the probabilistic method, harmonic analysis, combinatorics and graph theory have been applied to such

[1] For recent introductions to parameterised algorithms and complexity, see monographs [3, 4].

problems; see [9] for a survey. Here, we mention progress on only a small sample of such problems.

For the Maximum r-Satisfiability Problem, given a multiset of m clauses of size r, a straigtforward probabilistic argument shows that there exists a truth assignment satisfying at least $(2^r - 1)m/2^r$ clauses and this is tight. Alon et al. [1] showed that one can decide in time $O(m) + 2^{O(k^2)}$ if there is a truth assignment satisfying at least $((2^r - 1)m + k)/2^r$ clauses, where they used a combination of probabilistic, combinatorial and Harmonic analysis tools.

For the Max-Cut problem, the Edwards-Erdös bound [6, 7] states that every connected graph on n vertices and m edges has a cut of size at least $\frac{m}{2} + \frac{n-1}{4}$ and this is tight. Crowston, Jones, and Mnich [2] showed that it is fixed-parameter tractable to decide whether a given graph on n vertices and m edges has a cut of size at least $\frac{m}{2} + \frac{n-1}{4} + k$. This was later extended by Mnich et al. [16] to so-called λ-*extendible properties*; as special cases of their result, they could extend the Max-Cut result above to the Max q-Colourable Subgraph problem and the Oriented Max Acyclic Digraph problem.

Organisation The rest of the paper is organised as follows. In the next section we set out the notation we use throughout. Section 3 gives a brief discussion of some the ideas that are used to prove Theorem 1.1 and Theorem 1.2. In Section 4, we show how standard derandomisation techniques can be applied to the Travelling Salesman Problem in preparation for Sections 5 and 6. Section 5 is dedicated to the proof of Theorem 1.2 and this is used in Section 6 to prove Theorem 1.1.

2 Notation and Terminology

In this section, for convenience, we collect some notation and terminology (mostly standard) that we shall use throughout.

As mentioned earlier we use $\mathbb{Z}$, $\mathbb{N}$ and $\mathbb{Q}$ to denote the integers, the positive integers and the rational numbers respectively. We write $\mathbb{N}_0$ for $\mathbb{N} \cup \{0\}$. We use the symbols $\mathbb{P}$ and $\mathbb{E}$ to denote probabilities and expectations respectively.

Let G be a graph. The vertex set and edge set of G are denoted by $V(G)$ and $E(G)$ respectively and we write $e(G)$ for the number of edges in G. A graph F is a subgraph of G written $F \subseteq G$ if $V(F) \subseteq V(G)$ and $E(F) \subseteq E(G)$. We say F is a spanning subgraph of G if $V(F) = V(G)$ and $E(F) \subseteq E(G)$.

For $X \subseteq V(G)$, we write $X^{(2)}$ for the set of all edges ab such that $a, b \in X$

and $a \neq b$. We write $G[X]$ for the graph induced by G on X and $G - X$ for the graph obtained from G by deleting all vertices in X, i.e. $G - X := G[V(G) \setminus X]$. For $S \subseteq E(G)$, $G - S$ is the graph obtained from G by deleting all the edges in S, i.e. $G - S := (V(G), E(G) \setminus S)$. If $S \subset V(G)^{(2)}$ then we write $G \cup S := (V(G), E(G) \cup S)$ (and we write $G \cup e$ rather than $G \cup \{e\}$ if $S = \{e\}$). For disjoint subsets A, B of $V(G)$, we write $G[A, B]$ for the graph with vertex set $A \cup B$ and edge set $\{e = ab \in E(G) \mid a \in A, b \in B\}$. For a vertex $v \in V(G)$, $N_G(v)$ denotes the set of neighbours of v in G and $d_G(v) := |N_G(v)|$ denotes the degree of v. The maximum and minimum degree of G is denoted by $\delta(G)$ and $\Delta(G)$, respectively.

A path $P = v_1 v_2 \cdots v_k$ is the graph with vertices $v_1 \ldots, v_k$ and edges $v_i v_{i+1}$ for $i = 1, \ldots, k-1$. For a path P, we sometimes write $v_1 P v_k$ for the same path to indicate that v_1 and v_k are its end-vertices and we say $v_2, \ldots, v_{k-1}$ are the internal vertices of P. The notation extends in the natural way for concatenated paths so that if $v_1 P v_k$ and $w_1 Q w_\ell$ are disjoint paths and $x_1, \ldots, x_t$ are distinct vertices, then $v_1 P v_k x_1 \cdots x_t w_1 Q w_\ell$ is the path $v_1 \cdots v_k x_1 \cdots x_t w_1 \cdots w_\ell$. A cycle $C = v_1 \cdots v_k v_1$ is the graph with vertices $v_1 \ldots, v_k$ and edges $v_1 v_k$ and $v_i v_{i+1}$ for $i = 1, \ldots, k-1$. We call it a k-cycle if it has k vertices. A cycle that is a spanning subgraph of a graph G is called a Hamilton cycle of G. As before, we can write a cycle as a concatenation of paths. A matching of G is a subgraph of G in which every vertex has degree 1; a perfect matching of G is a spanning matching of G. The complete graph on n vertices is denoted by K_n.

Repeating notation from the introduction, recall that an instance of TSP_{BA} consists of a triple (n, w, k), where $n, k \in \mathbb{N}$ and $w : E(K_n) \to \mathbb{Z}$. We sometimes drop the parameter k (when it is not relevant) and refer to instances (n, w). For a subgraph G of K_n, we write

$$w(G) := \sum_{e \in E(G)} w(e) \qquad \text{and} \qquad w[G] := \sum_{e \in E(G)} |w(e)|$$

and we define the *density* $d = d(n, w)$ of (n, w) to be the average weight of an edge, i.e. $d := w(K_n)/\binom{n}{2}$.

Often, we work with functions $\phi : E(K_n) \to \mathbb{Q}$, and we shall tacitly extend these to subsets $S \subseteq E(K_n)$ by taking $\phi(S) := \sum_{e \in S} \phi(e)$. Throughout, addition of functions always refers to pointwise addition.

3 Overview

We remark at the outset that the discussion in this section is not required to understand the sections that follow; some definitions will be repeated later.

Structural result The key step for the algorithm of Theorem 1.1 is the structural result, Theorem 1.2. In order to explain the idea behind its proof, let us recast Theorem 1.2 in the language of norms.

We say two instances (n, w) and (n, w') of $\mathrm{TSP_{BA}}$ are *equivalent*, written $(n, w) \sim (n, w')$, if there exists some $\alpha \in \mathbb{Z}$ such that $w'(H) = w(H) + \alpha$ for every Hamilton cycle H of K_n. We define

$$\|(n, w)\|_{1/\sim} := \min\{w'[K_n] : (n, w') \sim (n, w)\}.$$

For an instance (n, w) of density d, if k^* is such that $dn - k^*$ is the weight of a minimum weight Hamilton cycle of K_n, we define $\|(n, w)\|_{\mathrm{HC}} := k^*$. Then the main substance of Theorem 1.2 is that the following inequality holds:

$$\|(n, w)\|_{1/\sim} \leq 4000n\|(n, w)\|_{\mathrm{HC}}. \tag{1}$$

This is proved by considering a third parameter $\|(n, w)\|_{4-\mathrm{cyc}}$ that is easily computed by examining the 4-cycles of K_n. This parameter is introduced in Section 5, and in the same section we prove the two inequalities

$$\|(n, w)\|_{1/\sim} \leq \frac{4000}{n^2}\|(n, w)\|_{4-\mathrm{cyc}} \tag{2}$$

$$\frac{1}{n^3}\|(n, w)\|_{4-\mathrm{cyc}} \leq \|(n, w)\|_{\mathrm{HC}}, \tag{3}$$

which together prove (1).

We make some further remarks. Note that for fixed n, the set of instances (n, w) with the obvious notions of addition and scalar multiplication is the vector space $\mathbb{R}^{E(K_n)}$. One can show that $\sim$ is an equivalence relation and the equivalence classes are translates of Z, the set of instances equivalent to the all-zero weighting, which turns out to be an $(n-1)$-dimensional subspace of $\mathbb{R}^{E(K_n)}$. Furthermore, each of $\|\cdot\|_{1/\sim}$, $\|\cdot\|_{4-\mathrm{cyc}}$, and $\|\cdot\|_{\mathrm{HC}}$ are seminorms on $\mathbb{R}^{E(K_n)}$ and norms on the quotient space $\mathbb{R}^{E(K_n)}/Z$. In particular, Z is precisely the set of instances (n, w) in which all Hamilton cycles have the same weight dn, where d is the density of (n, w) (see e.g. [17] for a characterisation).

Algorithmic result Once we have established the structural result Theorem 1.2, we can construct the algorithm of Theorem 1.1. Given (n, w, k), if Theorem 1.2 does not already give us the desired Hamilton cycle of weight at most $dn - k$, then we can find an equivalent instance (n, w', k) with $w'[K_n] \leq 5000kn$. If we can find a large matching in the graph of edges assigned a negative weight by w' then we can extend it (in a random way) to a desired Hamilton cycle of low weight. If such a matching does not exist, then one can easily conclude that all edges assigned a negative weight by w'

are incident with only a small set (of size depending on k but independent of n) of vertices. It turns out that, with this additional structure, one can in fact find the minimum weight Hamilton cycle for (n, w') (and hence the minimum weight Hamilton cycle for (n, w)) in FPT-time, i.e. time $f(k)n^{O(1)}$, where f is a function of k only.

4 Derandomisation

One can often convert a probabilistic proof for the existence of a certain object into an algorithmic proof using the well-known method of conditional expectation (see e.g. [14], [15]). In this section we use this method to give an algorithmic proof for the existence of Hamilton cycles with certain properties that we will need later.

We denote by $\mathcal{H}_n$ the set of all Hamilton cycles of the complete graph K_n. For G any subgraph (or set of edges) of K_n, let $\mathcal{H}_n^G := \{H \in \mathcal{H}_n \mid G \subseteq H\}$. In general we shall denote by $\tilde{H}$ a uniformly random element of $\mathcal{H}_n$, and by $\tilde{H}^G$ a uniformly random element of $\mathcal{H}_n^G$.

We say a graph $G \subseteq K_n$ is a *partial Hamilton cycle* of K_n if G is a spanning subgraph of some $H \in \mathcal{H}_n$; thus G is either a Hamilton cycle or the union of vertex disjoint paths (where we allow a path to be a singleton vertex). A path consisting of a single vertex is called a *trivial path*, and a path on several vertices is called a *non-trivial path*. We shall use the following simple fact several times: if G is a partial Hamilton cycle with r non-trivial paths and s trivial paths then

$$|\mathcal{H}_n^G| = 2^{r-1}(r+s-1)!. \tag{4}$$

For G a partial Hamilton cycle of K_n, we denote by $J(G)$ the set of edges in K_n which join two paths of G together into a single path. If G is a Hamilton path, $J(G)$ is defined to be the unique edge between the two ends of the path.

Lemma 4.1 *Suppose we have a function $X : \mathcal{H}_n \to \mathbb{Q}$ and for every partial Hamilton cycle G of K_n, assume we can compute $\mathbb{E}(X(\tilde{H}^G))$ in time $f(n)$. Then for any given partial Hamilton cycle G^*, we can find in time $O(n^3 f(n))$ a Hamilton cycle $H^* \in \mathcal{H}_n^{G^*}$ such that $X(H^*) \leq \mathbb{E}(X(\tilde{H}^{G^*}))$.*

Proof From the law of total expectation, we have

$$\mathbb{E}(X(\tilde{H}^G)) = \sum_{e \in J(G)} \mathbb{P}(e \in \tilde{H}^G)\, \mathbb{E}(X(\tilde{H}^{G \cup e}))$$

and so we know there exists some $e \in J(G)$ such that $\mathbb{E}(X(\tilde{H}^{G \cup e})) \leq \mathbb{E}(X(\tilde{H}^{G}))$.

We construct H^* by adding edges one at a time to G^* as follows. Assume G^* has q edges for some $q \geq 0$ and set $G_q := G^*$. Assume we have constructed a partial Hamilton cycle $G_{q'} \supseteq G_q$ with $q' \geq q$ edges satisfying $\mathbb{E}(X(\tilde{H}^{G_{q'}})) \leq \mathbb{E}(X(\tilde{H}^{G_q}))$. For each $e \in J(G_{q'})$ we compute $\mathbb{E}(X(\tilde{H}^{G_{q'} \cup e}))$ and determine an edge e^* for which $\mathbb{E}(X(\tilde{H}^{G_{q'} \cup e^*})) \leq \mathbb{E}(X(\tilde{H}^{G_{q'}}))$. This can be done in time $O(f(n)n^2)$. We set $G_{q'+1} := G_{q'} \cup e^*$. After at most n iterations of this process, we obtain a Hamilton cycle H^* satisfying the desired condition. The running time is therefore bounded by $O(f(n)n^3)$. □

Lemma 4.2 *Given any instance (n, w) and any partial Hamilton cycle G of K_n, we can find in time $O(n^5)$ a Hamilton cycle $H^* \in \mathcal{H}_n^G$ such that $w(H^*) \leq \mathbb{E}(w(\tilde{H}^G))$.*

Proof Apply the previous lemma. We can compute $\mathbb{E}(w(\tilde{H}^G))$ in time $O(n^2)$. Indeed, note that

$$\mathbb{E}(w(\tilde{H}^G)) = w(G) + \sum_{e \in J(G)} \mathbb{P}(e \in \tilde{H}^G) w(e)$$

and using (4),

$$\mathbb{P}(e \in \tilde{H}^G) = \frac{|\mathcal{H}_n^{G \cup e}|}{|\mathcal{H}_n^G|} = \frac{2^{r'-1}(r' + s' - 1)!}{2^{r-1}(r + s - 1)!},$$

where r, s are the numbers of non-trivial and trivial paths in G and r', s' are the numbers of non-trival and trivial paths in $G \cup e$. In fact it is not hard to see that if $e \in J(G)$, then $r + s = r' + s' + 1$, and so

$$\mathbb{P}(e \in \tilde{H}^G) = \frac{2^{r'-r}}{r + s - 1}. \tag{5}$$

Thus since $|J(G)| = O(n^2)$, we can compute $\mathbb{E}(w(\tilde{H}^G))$ in time $O(n^2)$ as required. □

5 The structural result

Our aim in this section is to prove Theorem 1.2. In this section, an instance refers to a pair (n, w) where $w : E(K_n) \to \mathbb{Z}$. Let us denote the set of 4-cycles of K_n by $\mathcal{C}_n$. Given an instance (n, w) and a 4-cycle $C = v_1v_2v_3v_4v_1 \in \mathcal{C}_n$, we define the *balance* of C (with respect to (n, w)) to be

$$\mathrm{bal}_{(n,w)}(C) = \mathrm{bal}(C) := |w(v_1v_2) + w(v_3v_4) - w(v_2v_3) - w(v_1v_4)|.$$

We say that C is *balanced* if $\mathrm{bal}(C) = 0$; otherwise we say C is *unbalanced*. For a set $\mathcal{A} \subseteq \mathcal{C}_n$, we define

$$\mathrm{bal}(\mathcal{A}) := \sum_{C \in \mathcal{A}} \mathrm{bal}(C),$$

and we set

$$\|(n, w)\|_{4-\mathrm{cyc}} := \mathrm{bal}(\mathcal{C}_n) = \sum_{C \in \mathcal{C}_n} \mathrm{bal}(C).$$

Our first lemma gives a polynomial-time witness to the inequality (3).

Lemma 5.1 *For a given instance (n, w) and $k \in \mathbb{Q}$, suppose $\|(n, w)\|_{4-\mathrm{cyc}} \geq kn^3$. Then there exists a Hamilton cycle H satisfying $w(H) < dn - k$. Furthermore, we can find such a Hamilton cycle in time $O(n^7)$.*

Before we can prove this lemma, we require several preliminary results. Suppose $H = v_1 v_2 \cdots v_n v_1$ is a Hamilton cycle of K_n. Consider two edges $e_1 = v_a v_b$ and $e_2 = v_x v_y$ of K_n, where without loss of generality $a < b$, $x < y$, and $a < x$. We say e_1 and e_2 are *crossing (relative to H)* if $a < x < b < y$; otherwise we say e_1 and e_2 are non-crossing. This is just the natural notion of crossing in a planar drawing of H, e_1, and e_2.

Proposition 5.2 *Suppose $H = v_1 v_2 \cdots v_n v_1$ is a Hamilton cycle of K_n and suppose we have a weight function $t : E(K_n) \to \mathbb{N}_0$. Then we can find $T^* \subseteq E(K_n)$ such that every pair of edges in T^* is non-crossing and $t(T^*) \geq t(K_n)/(2n)$. Furthermore, we can find T^* in time $O(n^2)$.*

Proof Let $Q = \{3, 4, \ldots, 2n\}$, and for each $q \in Q$, let $E_q = \{v_i v_j \in E(K_n) \mid i + j = q\}$. Observe that the sets $(E_q)_{q \in Q}$ partition $E(K_n)$ and that for each fixed $q \in Q$ each pair of edges in E_q is non-crossing. Thus since

$$t(K_n) = \sum_{q \in Q} t(E_q),$$

there is some $q^* \in Q$ for which $t(E_{q^*}) \geq t(K_n)/|Q| \geq t(K_n)/2n$. Set $T^* = E_{q^*}$.

We can determine E_q and $t(E_q)$ in time $O(n^2)$ and so we can determine T^* in time $O(n^2)$. □

We introduce some more terminology. Let (n, w) be an instance and let H be a Hamilton cycle of K_n. We say a 4-cycle $C = v_1 v_2 v_3 v_4 v_1$ of K_n is *embedded* in H if either $v_1 v_2, v_3 v_4 \in E(H)$ or $v_2 v_3, v_4 v_1 \in E(H)$; note that if three edges of C are in H, then C is necessarily embedded in H. We say C is *correctly embedded* in H if $E(H) \triangle E(C)$ forms a Hamilton cycle[2]; note that this happens if and only if C is embedded in H, exactly two edges of C belong to H, and the other two edges of C are crossing relative to H. If C is an unbalanced 4-cycle and is embedded in H, we say it is *heavily embedded in* H if the edges of the heavier perfect matching of C appear in H. We write $\mathcal{C}_n^*(H)$ for the set of 4-cycles heavily and correctly embedded in H. We define

$$q(H) := \sum_{C \in \mathcal{C}_n^*(H)} \mathrm{bal}(C).$$

The next lemma shows how a Hamilton cycle H can be locally improved using 4-cycles heavily and correctly embedded in it.

Lemma 5.3 *Let (n, w) be an instance. Given a Hamilton cycle H of K_n, there exists another Hamilton cycle H' such that $w(H') \leq w(H) - (q(H)/2n)$, and we can find it in time $O(n^4)$.*

Proof Let $e_1, \ldots, e_n$ be the edges of H in order. Suppose C_1 and C_2 are 4-cycles that are correctly embedded in H, where e_a, e_b with $a < b$ are the edges of C_1 in H and e_x, e_y with $x < y$ are the edges of C_2 in H, and without loss of generality, assume $a < x$. We say that C_1 and C_2 are *crossing (relative to H)* if $a < x < b < y$. If C_1 and C_2 are not crossing relative to H, we find that $E(H) \triangle E(C_1) \triangle E(C_2)$ forms a Hamilton cycle. More generally, one can check that if $C_1, \ldots, C_r$ are all 4-cycles correctly embedded in H and no pair of these 4-cycles are crossing relative to H, then $E(H) \triangle E(C_1) \triangle \cdots \triangle E(C_r)$ forms a Hamilton cycle.

Consider an auxiliary complete graph K_n° with vertex set $e_1, \ldots, e_n$ and an auxiliary Hamilton cycle $H^\circ = e_1 e_2 \cdots e_n e_1$. For each edge $e_x e_y$ of $E(K_n^\circ) \setminus E(H^\circ)$ let C_{xy} be the unique 4-cycle of K_n that is correctly embedded in H and satisfies $E(H) \cap E(C) = \{e_x, e_y\}$. This correspondence is clearly bijective and crossing edges of K_n° (relative to H°) correspond to crossing 4-cycles of K_n (relative to H). Define $t : E(K_n^\circ) \to \mathbb{N}_0$ by

$$t(e_x e_y) = \begin{cases} \mathrm{bal}(C_{xy}) & \text{if } C_{xy} \text{ is heavily embedded in } H; \\ 0 & \text{otherwise,} \end{cases}$$

and note $t(K_n^\circ) = q(H)$. Applying Proposition 5.2, we can find a set of edges $T^* \subseteq E(K_n^\circ)$ which are pairwise non-crossing relative to H° and satisfying $t(T^*) > q(H)/2n$. By removing from T^* any edges assigned weight 0 by t,

[2] $E(H) \Delta E(C)$ is often called a 2-Opt move in TSP experimental studies [11]

we may further assume that every edge in T^* is assigned a positive weight by t. The edges of T^* then correspond to 4-cycles $C_1, \ldots, C_r$ of K_n that are heavily and correctly embedded in H such that no pair of these 4-cycles are crossing (relative to H) and where $\mathrm{bal}(C_1) + \cdots + \mathrm{bal}(C_r) \geq q(H)/2n$. The time needed to find $C_1, \ldots, C_r$ is $O(n^2)$.

We construct a sequence of Hamilton cycles $H_0, \ldots, H_r$, where $H_0 = H$ and $E(H_i) = E(H_{i-1}) \triangle E(C_i)$ (which takes time $O(r) = O(n)$). We have $w(H_i) = w(H_{i-1}) - \mathrm{bal}(C_i)$ since C_i is heavily embedded in H_{i-1}. Hence setting $H' = H_r$, we have

$$w(H') = w(H) - \mathrm{bal}(C_1) - \cdots - \mathrm{bal}(C_r) \leq w(H) - (q(H)/2n),$$

and the time required to construct H' is $O(n^2)$. □

The previous lemma motivates our interest in the function $q : \mathcal{H}_n \to \mathbb{Z}$. We shall later apply Lemma 4.1 to q and for this we require the following straightforward proposition. We spell out the details for completeness.

Proposition 5.4 *Let (n, w) be an instance and let G be a partial Hamilton cycle of K_n. Then we can compute $\mathbb{E}(q(\tilde{H}^G))$ in time $O(n^4)$ (recalling that $\tilde{H}^G$ is a uniformly random Hamilton cycle containing the edges of G).*

Proof Note that

$$\mathbb{E}(q(\tilde{H}^G)) = \sum_{C \in \mathcal{C}_n} \mathbb{P}(C \in \mathcal{C}_n^*(\tilde{H}^G))\mathrm{bal}(C),$$

and so it is sufficient to show how to compute $\mathbb{P}(C \in \mathcal{C}_n^*(\tilde{H}^G))$ in time $O(1)$ for each $C \in \mathcal{C}_n$. This is intuitively clear, but slightly tedious to explain.

Clearly if C is balanced then $\mathbb{P}(C \in \mathcal{C}_n^*(\tilde{H}^G)) = 0$. So assume $C = v_1v_2v_3v_4v_1 \in \mathcal{C}_n$ is unbalanced and that $e_1 = v_1v_2$ and $e_2 = v_3v_4$ are the edges of the heavier perfect matching of C. Then (using (4)) the probability $p(C, G)$ that C is heavily embedded in $\tilde{H}^G$ is

$$p(C, G) = \mathbb{P}(e_1, e_2 \in \tilde{H}^G) = \frac{|\mathcal{H}_n^{G \cup \{e_1, e_2\}}|}{|\mathcal{H}_n^G|} = \frac{2^{r'-1}(r' + s' - 1)!}{2^{r-1}(r + s - 1)!},$$

where r and s are the numbers of non-trivial and trivial paths in G and r' and s' are the numbers of non-trivial and trivial paths in $G \cup \{e_1, e_2\}$.

If e_1 and e_2 appear on different non-trivial paths of $G' := G \cup \{e_1, e_2\}$, say $a_1P_1b_1$ and $a_2P_2b_2$, then $\mathbb{P}(C \in \mathcal{C}_n^*(\tilde{H}^G)) = \frac{1}{2}p(C, G)$. To see this, we note that a Hamilton cycle $H \in \mathcal{H}_n^{G'}$ can take one of two forms depending on relative orientations of P_1 and P_2 on H. In one of these forms C is correctly embedded in H and in the other it is not. Thus C is correctly embedded in exactly half the Hamilton cycles of $\mathcal{H}_n^{G'}$.

If e_1 and e_2 appear on the same non-trivial path of G' then either $\mathbb{P}(C \in \mathcal{C}_n^*(\tilde{H}^G)) = p(C, G)$ or $\mathbb{P}(C \in \mathcal{C}_n^*(\tilde{H}^G)) = 0$ depending on the order of the vertices $v_1, \ldots, v_4$ on the path. Hence we have shown how to compute $\mathbb{P}(C \in \mathcal{C}_n^*(\tilde{H}^G))$ in time $O(1)$. $\square$

We are now ready to prove Lemma 5.1.

Proof (of Lemma 5.1) Let $\tilde{H}$ be a uniformly random Hamilton cycle of K_n and suppose $C = v_1v_2v_3v_4v_1$ is an unbalanced 4-cycle, with $e_1 = v_1v_2$ and $e_2 = v_3v_4$ the edges of its heavier perfect matching. Then the probability that C is heavily embedded in $\tilde{H}$ is given by

$$\mathbb{P}(e_1, e_2 \in \tilde{H}) = \frac{|\mathcal{H}_n^{\{e_1,e_2\}}|}{|\mathcal{H}_n|} = \frac{2(n-3)!}{(n-1)!/2} = \frac{4}{(n-1)(n-2)} \geq \frac{4}{n^2}.$$

Now (as in the proof of Proposition 5.4) it is not too hard to see that the probability that C is heavily and correctly embedded in $\tilde{H}$ is $\frac{1}{2}\mathbb{P}(e_1, e_2 \in \tilde{H}) \geq \frac{2}{n^2}$. Thus we have that

$$\mathbb{E}(q(\tilde{H})) = \sum_{C \in \mathcal{C}_n} \mathbb{P}(C \in \mathcal{C}_n^*(\tilde{H}))\mathrm{bal}(C) \geq \frac{2}{n^2}\|(n, w)\|_{4-\mathrm{cyc}} > 2kn.$$

and so

$$\mathbb{E}(w(\tilde{H}) - (q(\tilde{H})/2n)) < dn - k.$$

Thus there exists a Hamilton cycle H^* satisfying $w(H^*) - (q(H^*)/2n) < dn - k$. Furthermore we can find this Hamilton cycle in time $O(n^7)$. Indeed, by Lemma 4.1, it is sufficient to check that we can compute $\mathbb{E}(w(\tilde{H}^G) - (q(\tilde{H}^G)/2n)) = \mathbb{E}(w(\tilde{H}^G)) - \mathbb{E}(q(\tilde{H}^G))/2n$ in time $O(n^4)$ for every partial Hamilton cycle G. This holds since we can compute $\mathbb{E}(w(\tilde{H}^G))$ in $O(n^2)$ time by the proof of Lemma 4.2 and we can compute $\mathbb{E}(q(\tilde{H}^G))$ in time $O(n^4)$ by Proposition 5.4.

Finally, we apply Lemma 5.3 to H^* to obtain a Hamilton cycle H satisfying $w(H) \leq w(H^*) - (q(H^*)/2n) < dn - k$ as required. This takes time $O(n^4)$, so the total running time is $O(n^7)$. $\square$

Our next task is to prove (2) and give a polynomial-time witness for the inequality. Recall that two instances (n, w) and (n, w') are equivalent if there exists some $\alpha \in \mathbb{Z}$ such that $w'(H) = w(H) + \alpha$ for all Hamilton cycles of K_n.

Lemma 5.5 *Suppose* (n, w) *is an instance and* $0 \le k \in \mathbb{Q}$ *with* $n > 50000(k+1)$. *If* $\|(n, w)\|_{4-\mathrm{cyc}} \le kn^3$ *then there exists an equivalent instance* (n, w^*) *satisfying* $w^*[K_n] \le 5000kn$. *Moreover, we can determine* (n, w^*) *in time* $O(kn^5)$.

After proving two preliminary results, we will prove Lemma 5.8, which states that if $\|(n, w)\|_{4-\mathrm{cyc}}$ is small then there is an equivalent instance (n, w^*) in which most edges of K_n are assigned weight 0. From this, we shall deduce Lemma 5.5.

Lemma 5.6 *Suppose* (n, w) *is an instance and* $0 \le k \in \mathbb{Q}$ *with* $n > 15$ *such that* $\|(n, w)\|_{4-\mathrm{cyc}} \le kn^3$. *Then there exists a set* S *of at most* $72kn$ *edges of* K_n *such that every* 4*-cycle of* $K_n - S$ *is balanced. Moreover, we can find* S *in time* $O(kn^5)$.

Proof Note first that if $\|(n, w)\|_{4-\mathrm{cyc}} \le kn^3$ then there are at most kn^3 unbalanced 4-cycles.

Suppose $C = v_1v_2v_3v_4v_1$ is an unbalanced 4-cycle of K_n and let ab be any edge of K_n that is vertex disjoint from C. Then one of the following 4-cycles is unbalanced: av_1v_2ba, bv_2v_3ab, av_3v_4ba, bv_4v_1ab. Indeed, if not then we have

$$\begin{aligned} w(ab) &= w(av_1) + w(bv_2) - w(v_1v_2) \\ &= w(bv_2) + w(av_3) - w(v_2v_3) \\ &= w(av_3) + w(bv_4) - w(v_3v_4) \\ &= w(bv_4) + w(av_1) - w(v_4v_1). \end{aligned}$$

But then $w(v_1v_2) + w(v_3v_4) = w(v_2v_3) + w(v_1v_4)$ and so C is a balanced 4-cycle, a contradiction. Hence we see that for every $ab \in E(K_n - V(C))$, there exists an unbalanced 4-cycle C' which contains ab and shares an edge with C.

We construct edge sets $E_0 \subseteq E_1 \subseteq \cdots \subseteq E_r \subseteq E(K_n)$ and multisets of 4-cycles $\mathcal{M}_0 \subseteq \mathcal{M}_1 \subseteq \cdots \subseteq \mathcal{M}_r \subseteq \mathcal{C}_n$ as follows. Set $E_0 = \emptyset$ and $\mathcal{M}_0 = \emptyset$. Given E_i, if $K_n - E_i$ has no unbalanced 4-cycles then set $r = i$ and stop. Otherwise pick any unbalanced 4-cycle C_i of $K_n - E_i$ and set $E_{i+1} = E_i \cup E(C_i)$. We call C_i the ith base cycle and note that the base cycles are edge-disjoint. For each edge $ab \in E(K_n - V(C_i))$, we know there exists at least one unbalanced 4-cycle of K_n that contains ab and shares an edge with C_i; pick one arbitrarily and call it C_i^{ab}. We obtain $\mathcal{M}_{i+1}$ by adding C_i^{ab} to M_i for all $ab \in E(K_n - V(C_i))$ and keep track of multiple occurrences.

Note that a 4-cycle can only appear in $\mathcal{M}_r$ at most 4 times since each time is appears, it shares an edge with a different base cycle and the base cycles are edge-disjoint. Also the size of $\mathcal{M}_r$ (with multiplicities) is exactly $r\binom{n-4}{2}$, and since all 4-cycles in $\mathcal{M}_r$ are unbalanced and appear at most 4 times, we have $\frac{r}{4}\binom{n-4}{2} \leq kn^3$, whence

$$r \leq \frac{8kn^3}{(n-4)(n-5)} \leq 18kn,$$

for our choice of n.

Thus $|E_r| = 4r \leq 72kn$ and $K_n - E_r$ has no unbalanced 4-cycles, so we set $S = E_r$.

The proof is constructive: it takes time $O(n^4)$ to search for and remove an unbalanced 4-cycle and we do this at most $18kn$ times, so the total running time is $O(kn^5)$. □

The following proposition is straightforward, but we spell out the details for completeness.

Proposition 5.7 *Suppose G is a graph on n vertices and at least $\binom{n}{2} - t$ edges for some $0 \leq t \leq \binom{n}{2}$. Then G has a connected component with at least $\binom{n}{2} - 2t$ edges.*

Proof Let $A \subseteq V(G)$ be a connected component of G with the maximum number of vertices. Then it must be the case that $|A| \geq n - 1 - (2t/n)$; if not then $\Delta(G) < n - 1 - (2t/n)$, which implies $e(G) \leq n\Delta(G)/2 < \binom{n}{2} - t$, a contradiction. Hence

$$e(G[A]) \geq e(G) - \binom{n-|A|}{2} \geq \binom{n}{2} - t - \binom{1+(2t/n)}{2} \geq \binom{n}{2} - 2t,$$

where the last inequality follows using that $\binom{n}{2} \geq t$. □

Next, we describe simple linear operations one can apply to an instance to obtain an equivalent instance. For a vertex v of K_n we write $I_v : E(K_n) \to \{0, 1\}$ for the edge weighting of K_n where $I_v(e) = 1$ if v is an end-vertex of e, and $I_v(e) = 0$ otherwise. Observe that if (n, w) is an instance and $w' = w + \lambda I_v$ for some $\lambda \in \mathbb{Z}$, then $w'(H) = w(H) + 2\lambda$ for all Hamilton cycles H of K_n and both w and w' have the same balanced 4-cycles. More generally, given an integer λ_v for each $v \in V(K_n)$, if

$$w' = w + \sum_{v \in V(K_n)} \lambda_v I_v \tag{6}$$

then $w'(H) = w(H) + 2\sum_v \lambda_v$ for all Hamilton cycles H of K_n and so (n, w) and (n, w') are equivalent instances. Note further that, as before, (n, w) and (n, w') have the same balanced 4-cycles.

Lemma 5.8 *Suppose (n, w) is an instance with $k \in \mathbb{Q}$ and $n > 288k + 1$. If $\|(n, w)\|_{4-\mathrm{cyc}} \leq kn^3$ then there exists an equivalent instance (n, w^*), where $w^*(e) = 0$ for all but at most $144k(n + 1)$ edges $e \in E(K_n)$. Moreover, we can determine (n, w^*) in time $O(kn^5)$.*

Proof From Lemma 5.6, in time $O(kn^5)$, we can find $S \subseteq E(K_n)$ of size at most $72kn$ such that all 4-cycles in $K_n - S$ are balanced. Write $G := K_n - S$. Let v_0 be a vertex of maximum degree in G (so $d_G(v_0) \geq n - 1 - 144k$). Consider the weighting of K_n given by

$$w' = w - \sum_{u \in N_G(v_0)} w(uv_0) I_u.$$

Since w' takes the form of (6), we see (n, w) is equivalent to (n, w').

We show that w' assigns the same weight, α say, to all but at most $144kn + (n - 1)$ edges $e \in E(K_n)$. It is clear that $w'(uv_0) = 0$ for every u adjacent to v_0 in G. Now for any vertex $a \in V(G) \setminus \{v_0\} = V(K_n) \setminus \{v_0\}$, all edges incident to a in G (except possibly av_0) have the same weight: indeed, if $ax, ay \in E(G)$ with $x, y \neq v_0$, then axv_0ya forms a balanced 4-cycle and so we must have that $w'(ax) = w'(v_0y) + w'(ax) = w'(v_0x) + w'(ay) = w'(ay)$. This implies that w' assigns the same weight to every connected component of $G - \{v_0\}$: given two edges in the same component, there is a path from one edge to the other and each pair of incident edges have the same weight, so all edges on the path have the same weight.

Since $G - \{v_0\}$ is an $(n - 1)$-vertex graph with at least $\binom{n-1}{2} - 72kn$ edges (and where $\binom{n-1}{2} \geq 72kn$ by our choice of n), Proposition 5.7 implies there is a component A of $G - \{v_0\}$ with at least $\binom{n-1}{2} - 144kn = \binom{n}{2} - 144kn - (n - 1)$ edges. Thus w' assigns the same weight, α say, to at least $e(G[A]) \geq \binom{n}{2} - 144kn - (n - 1)$ edges of K_n.

We set $w'' := w' - \alpha$ (i.e. we reduce w' by α for all edges of K_n). Thus (n, w'') is equivalent to (n, w') (and hence to (n, w)) where $w''(e) = 0$ for all but at most $144kn + (n - 1)$ edges $e \in E(K_n)$ and $w''(e) = -\alpha$ for all edges $e \in E(G)$ incident to v_0. Setting $w^* = w'' + \alpha I_{v_0}$, we have that (n, w^*) is equivalent to (n, w'') (and hence to (n, w)) with $w^*(e) = 0$ for all but at most $144kn + (n-1) - d_G(v_0) \leq 144k(n+1)$ edges $e \in E(K_n)$. So (n, w^*) satisfies the requirements of the lemma.

Finding w^* takes time $O(kn^5)$; the time is dominated by the time to find S. □

We are now ready to prove Lemma 5.5.

Proof (of Lemma 5.5) By Lemma 5.8, we can find in time $O(kn^5)$ an instance (n, w') equivalent to (n, w) such that $w'(e) = 0$ for all but at most $144k(n+1) \leq 200kn$ edges of K_n (for our choice of large n).

Let F be the spanning subgraph of K_n whose edge set consists of the edges $e \in E(K_n)$ for which $w'(e) \neq 0$ (so F may have isolated vertices). Define $X = \{v \in V(K_n) : d_F(v) \geq n/4\}$. Then $|X|n/4 \leq 2e(F) \leq 400kn$, and so $|X| \leq 1600k$. We write $\overline{X} := V(K_n) \setminus X$.

For each $x \in X$ define $\alpha_x \in \mathbb{Q}$ by

$$\alpha_x := \frac{1}{|\overline{X}|} \sum_{v \in \overline{X}} w'(xv) \qquad \text{and} \qquad w^* := w' - \sum_{x \in X} \lfloor \alpha_x \rfloor I_x.$$

Thus (n, w^*) is equivalent to (n, w') (since it takes the form of (6)) and hence to (n, w) and it is easy to check that we can compute (n, w^*) in $O(n^2)$ time. It remains for us to show that $w^*[K_n] \leq 5000kn$. We shall show that $w^*[K_n[\overline{X}]] \leq 2kn$, that $w^*[K_n[X, \overline{X}]] \leq 1616kn$ and that $w^*[K_n[X]] \leq 3208kn$, proving the lemma. In each case, we use averaging arguments to show that if $w^*[\cdot]$ is large then there is a set of 4-cycles with large balance.

Note that $w^*(e) = w'(e)$ for all $e \in \overline{X}^{(2)}$. Define F^* to be the spanning subgraph of K_n consisting of edges e for which $w^*(e) \neq 0$. For the rest of the proof, balance is with respect to (n, w^*), i.e. $\mathrm{bal}(\cdot) = \mathrm{bal}_{(n,w^*)}(\cdot)$.

Claim 1 We have $w^*[K_n[\overline{X}]] \leq 2kn$.

Proof (of Claim 1) Consider $e = xy \in E(F) \cap \overline{X}^{(2)} = E(F^*) \cap \overline{X}^{(2)}$ and define A_{xy} to be the set of all 4-cycles $uxyvu$ satisfying $u, v \in \overline{X}$ and $ux, vy, uv \notin E(F^*) \cap \overline{X}^{(2)}$, so $w^*(ux) = w^*(vy) = w^*(uv) = 0$. Note that for each $C \in A_{xy}$, $\mathrm{bal}(C) = |w^*(xy)|$.

There are at least $|\overline{X}| - n/4$ choices of $u \in \overline{X}$ such that $ux \notin E(F^*)$ and similarly $|\overline{X}| - n/4 - 1$ choices of $v \in \overline{X}$ such that $vy \notin E(F^*)$ and $v \neq u$. Amongst these choices of u, v, the number of possible choices where $uv \in E(F^*) \cap \overline{X}^{(2)}$ is at most $|E(F) \cap \overline{X}^{(2)}| \leq e(F) \leq 200kn$. Hence

$$|A_{xy}| \geq (|\overline{X}| - n/4 - 1)^2 - 200kn \geq (3n/4 - 1600k - 1)^2 - 200kn \geq n^2/2$$

where the last inequality follows by our choice of sufficiently large n. Now we have

$$kn^3 \geq \|(n, w^*)\|_{4-\mathrm{cyc}} \geq \mathrm{bal}\left(\bigcup_{xy \in E(F^*) \cap X^{(2)}} A_{xy} \right) \geq \sum_{xy \in E(F^*) \cap \overline{X}^{(2)}} |w^*(xy)||A_{xy}|$$
$$\geq (n^2/2) w^*[K_n[\overline{X}]],$$

and so $w^*[K_n[\overline{X}]] \leq 2kn$. □

Claim 2 We have $w^*[K_n[X,\overline{X}]] \leq 416kn$.

Proof (of Claim 2) For each $x \in X$ and $ab \in \overline{X}^{(2)}$ with $a \neq b$, let $A_{x,ab}$ be the set of 4-cycles $C = axbca$ with $c \in \overline{X}$ satisfying $w^*(ac) = w^*(bc) = 0$ (i.e. $ac, bc \notin E(F^*) \cap \overline{X}^{(2)} = E(F) \cap \overline{X}^{(2)}$). The number of choices for such a vertex c is thus at most

$$|\overline{X}| - 2 - d_F(a) - d_F(b) \geq n - 2 - 1600k - n/4 - n/4 \geq n/4$$

where the last inequality follows by our choice of sufficiently large n. Thus we have $|A_{x,ab}| \geq n/4$ for all $x \in X$ and $ab \in \overline{X}^{(2)}$ with $a \neq b$. Note also that, for all $C \in A_{x,ab}$, we have $\mathrm{bal}(C) = |w^*(ax) - w^*(bx)|$.

Fix $x \in X$, let $\beta_x := |\overline{X}|^{-1} \sum_{v \in \overline{X}} w^*(vx)$, and note that $\beta_x = \alpha_x - \lfloor \alpha_x \rfloor \in [0,1]$. Write $S^+ := \{v \in \overline{X} : w^*(vx) \geq \beta_x\}$ and $S^- := \{v \in \overline{X} : w^*(vx) \leq \beta_x\}$. Note for later that, since β_x is the average of $(w^*(vx))_{v \in \overline{X}}$, we have

$$\sum_{v \in S^+} |w^*(vx) - \beta_x| = \sum_{v \in S^-} |w^*(vx) - \beta_x| = \frac{1}{2} \sum_{v \in \overline{X}} |w^*(vx) - \beta_x|. \tag{7}$$

We have

$$\begin{aligned}
\mathrm{bal}\left(\bigcup_{ab \in \overline{X}^{(2)}} A_{x,ab} \right) &\geq \sum_{a,b \in \overline{X}} |A_{x,ab}||w^*(ax) - w^*(bx)| \\
&\geq (n/4) \sum_{\substack{a \in S^+ \\ b \in S^-}} |w^*(ax) - w^*(bx)| \\
&= (n/4) \sum_{\substack{a \in S^+ \\ b \in S^-}} |w^*(ax) - \beta_x| + |w^*(bx) - \beta_x| \\
&\overset{(7)}{=} (n/4) \left(\frac{1}{2}(|S^-| + |S^+|) \sum_{v \in \overline{X}} |w^*(vx) - \beta_x| \right).
\end{aligned}$$

Using that $|S^+| + |S^-| \geq |\overline{X}|$ and $\beta_x \in [0,1]$, the last expression is bounded below by

$$(n|\overline{X}|/8) \sum_{v \in \overline{X}} (|w^*(vx)| - 1) \geq (n^2/16) \sum_{v \in \overline{X}} (|w^*(vx)| - 1)$$

for n sufficiently large. Finally, we have

$$\begin{aligned} kn^3 \geq \|(n,w)\|_{4-\text{cyc}} &\geq \text{bal}\Big(\bigcup_{\substack{x\in X\\ ab\in \overline{X}^{(2)}}} A_{x,ab}\Big) \geq (n^2/16)\sum_{\substack{x\in X\\ v\in \overline{X}}}(|w^*(vx)|-1) \\ &= (n^2/16)(w^*[K_n[X,\overline{X}]] - |X||\overline{X}|) \\ &\geq (n^2/16)(w^*[K_n[X,\overline{X}]] - 1600kn), \end{aligned}$$

from which we obtain that $w^*[K_n[X,\overline{X}]] \leq 1616kn$. □

Claim 3 We have $w^*[K_n[X]] \leq 3208kn$.

Proof (of Claim 3) For each $xy \in E(F^*)\cap X^{(2)}$, define A_{xy} to be the set of all 4-cycles of the form $C = xyuvx$, where $u, v \in \overline{X}$. For fixed $xy \in E(F^*)\cap X^{(2)}$, we have

$$\begin{aligned} \text{bal}(A_{xy}) &= \sum_{\substack{u,v\in\overline{X}\\ u\neq v}} |w^*(xy) + w^*(uv) - w^*(xv) - w^*(yu)| \\ &\geq \Big|\sum_{\substack{u,v\in\overline{X}\\ u\neq v}} w^*(xy) + w^*(uv) - w^*(xv) - w^*(yu)\Big| \\ &= \Big||\overline{X}|(|\overline{X}|-1)w^*(xy) \;+\; 2w^*(K_n[\overline{X}]) \;-\; (|\overline{X}|-1)(\beta_x+\beta_y)|\overline{X}|\Big| \\ &\geq |\overline{X}|(|\overline{X}|-1)|w^*(xy)| \;-\; 2w^*[K_n[\overline{X}]] \;-\; 2|\overline{X}|(|\overline{X}|-1), \end{aligned}$$

where $\beta_x, \beta_y \in [0,1]$ are as defined in the previous claim. For our choice of large n, we have $n/2 \leq |\overline{X}| \leq n$ and so $n^2/8 \leq |\overline{X}|(|\overline{X}|-1) \leq n^2$. By Claim 1, we have $w^*[K_n[\overline{X}]] \leq 2kn \leq n^2$. Putting this together, the final expression above is at most $\frac{1}{8}n^2|w^*(xy)| - 4n^2$. Finally

$$\begin{aligned} kn^3 \geq \|(n,w^*)\|_{4-\text{cyc}} &\geq \text{bal}(\bigcup_{xy\in X^{(2)}} A_{xy}) \geq \sum_{xy\in X^{(2)}} \frac{1}{8}n^2|w^*(xy)| - 4n^2 \\ &\geq (n^2/8)w^*[K_n[X]] - 4(1600k)^2n^2 \\ &\geq (n^2/8)w^*[K_n[X]] - 400kn^3, \end{aligned}$$

where the last inequality holds by the choice of large n. Rearranging shows that $w^*[K_n[X]] \leq 3208kn$. □

This completes the proof of the lemma. □

We can now combine Lemma 5.1 and Lemma 5.5 to prove the main result of this section, Theorem 1.2.

Proof (of Theorem 1.2) Given an instance (n, w) and $k \in \mathbb{N}$ with $n > 50000(k+1)$, if $\|(n, w)\|_{4-\text{cyc}} \geq kn^3$ then by Lemma 5.1, in time $O(n^7)$, we can find a Hamilton cycle H of K_n satisfying $w(H) < dn - k$. If $\|(n, w)\|_{4-\text{cyc}} \leq kn^3$ then in time $O(kn^5) = O(n^7)$, we can find an equivalent instance (n, w') satisfying $w'[K_n] \leq 5000kn$. □

6 Algorithms

In this section we prove Theorem 1.1. Let (n, w, k) be an instance of TSP_{BA} and let $X \subseteq V(K_n)$. Recall that for $X \subseteq V(K_n)$, we define $\overline{X} := V(K_n) \setminus X$. An (X)*-partial Hamilton cycle* of K_n is any spanning subgraph G of K_n that can be obtained by taking a Hamilton cycle of K_n and deleting all its edges that lie in $E(K_n[\overline{X}])$. The following proposition gives an equivalent characterisation of (X)-partial Hamilton cycles that will be more convenient to use.

Proposition 6.1 *Suppose $X \subseteq V(K_n)$ with $|X| < n/2$ and suppose G is a spanning subgraph of K_n. Then G is an (X)-partial Hamilton cycle of K_n if and only if*

(a) G consists of vertex disjoint paths $P_1, \ldots, P_r$ for some r (where we allow trivial paths);

(b) every $x \in X$ is an internal vertex of some P_i;

(c) no edge of $E(K_n[\overline{X}])$ is present in G.

Proof Since $|X| < n/2$ any Hamilton cycle of K_n must have at least one edge in $\overline{X}$. Thus if G is an (X)-partial Hamilton cycle of K_n it cannot be a Hamilton cycle so it is the vertex-disjoint union of paths and (a) holds. The endpoints of these paths must be in $\overline{X}$, so (b) holds, and trivially (c) holds.

If G is a spanning subgraph of K_n satisfying (a), (b), and (c) then G consists of vertex-disjoint paths with endpoints in $\overline{X}$ but no edges in $\overline{X}$. Thus we can use edges in $\overline{X}$ to connect G into a Hamilton cycle H. Now we see that G can be obtained by deleting the edges of H in $\overline{X}$, showing that G is an (X)-partial Hamilton cycle. □

Next we give an efficient algorithm for finding an (X)-partial Hamilton cycle of minimum weight when the size of X is fixed.

Lemma 6.2 *Let (n, w, k) be an instance of* $\mathrm{TSP_{BA}}$ *and let $X \subseteq V(K_n)$ with $t := |X|$. We can find an (X)-partial Hamilton cycle of K_n of minimum weight in time $O(t^3)! + O(t^3 n)$.*

Proof We may assume $t < n/10$; otherwise we can easily find an (X)-partial Hamilton cycle in the claimed time by examining all Hamilton cycles of K_n and checking their weight after deletion of edges in $\overline{X}$.

For each $a \in X$, let $y_1^a, y_2^a, \dots$ be an ordering of the vertices of $\overline{X}$ such that $w(ay_1^a) \le w(ay_2^a) \le \dots$. For each $ab \in X^{(2)}$, let $y_1^{ab}, y_2^{ab}, \dots$ be an ordering of the vertices of $\overline{X}$ such that $w(ay_1^{ab}) + w(by_1^{ab}) \le w(ay_2^{ab}) + w(by_2^{ab}) \le \dots$.

For each positive integer ℓ, define

$$Y_\ell^1(X) := \bigcup_{a \in X} \bigcup_{i=1}^{\ell} \{y_i^a\} \subseteq \overline{X} \quad \text{and} \quad Y_\ell^2(X) := \bigcup_{ab \in X^{(2)}} \bigcup_{i=1}^{\ell} \{y_i^{ab}\} \subseteq \overline{X}$$

and define

$$M_\ell^1(X) := \bigcup_{a \in X} \bigcup_{i=1}^{\ell} \{ay_i^a\} \quad \text{and} \quad M_\ell^2(X) := \bigcup_{ab \in X^{(2)}} \bigcup_{i=1}^{\ell} \{ay_i^{ab}, by_i^{ab}\}.$$

Finally, let $Y := Y_{2t+1}^1(X) \cup Y_{2t+1}^2(X)$ and let $M := M_{2t+1}^1(X) \cup M_{2t+1}^2(X)$.

We note for later that we can determine Y in time $O(t^3 n)$. Indeed, to determine $Y_{2t+1}^2(X)$ we compute the set $S_{ab} = \{y_i^{ab} : i \in [2t+1]\}$ for each $ab \in X^{(2)}$ and take the union. We compute S_{ab} by finding the $2t+1$ vertices $v \in \overline{X}$ that give the smallest values of $w(av) + w(bv)$. This takes time $O(tn)$ and we do this $|X^{(2)}| = O(t^2)$ times giving a running time of $O(t^3 n)$. Similarly we obtain $Y_{2t+1}^1(X)$, which in fact only takes time $O(t^2 n)$.

Claim 1 Suppose G is an (X)-partial Hamilton cycle of K_n of minimum weight. Subject to this, assume further that $|E_G(X, \overline{X}) \setminus M|$ is minimised. Then $E_G(X, \overline{X}) \subseteq M$. Note that in particular, this implies that the non-trivial paths of G all lie in $X \cup Y$.

Proof (of Claim 1) Suppose $rs \in E_G(X, \overline{X})$ with $r \in X$ and $s \in \overline{X}$ but $rs \notin M$. Since G is an (X)-partial Hamilton cycle, we have $|N_G(s)| = 1$ or 2 and $N_G(s) \subseteq X$ (since G has no edges in $E(K_n[\overline{X}])$. Let $N_G(s) = \{r, r'\}$, where $r = r'$ if $|N_G(s)| = 1$.

Let A be the set of all $x \in \overline{X}$ satisfying $|N_G(x)| > 0$; then $|A| \le 2|X| = 2t$ since $N_G(x) \subseteq X$ for all $x \in \overline{X}$ and $d_G(x) = 2$ for all $x \in X$. Note also that, for every $a \in \overline{X} \setminus A$, the graph G' obtained by deleting the edges $rs, r's$ and replacing them with the edges $ra, r'a$ is an (X)-partial Hamilton cycle.

Suppose $|N_G(s)| = 1$. Pick any $i \in [2t+1]$ such that $y_i^r \notin A$ (this is possible since $|A| \leq 2t$) and replace rs with ry_i^r in G to form G'. Then G' is an (X)-partial Hamilton cycle, and $w(G') \leq w(G)$ since $w(ry_i^r) \leq w(rs)$ (by the choice of y_i^r and the fact that $rs \notin M_{2t+1}^1(X) \subseteq M$). However $|E_{G'}(X, \overline{X}) \setminus M| < |E_G(X, \overline{X}) \setminus M|$ (since $rs \notin M$ and $ry_i^r \in M$), a contradiction to our choice of G.

Suppose $|N_G(s)| = 2$. Pick any $i \in [2t+1]$ such that $y_i^{rr'} \notin A$ (this is possible since $|A| \leq 2t$) and replace $rs, r's$ with $ry_i^{rr'}, r'y_i^{rr'}$ in G to form G'. Then G' is an (X)-partial Hamilton cycle, and $w(G') \leq w(G)$ since $w(ry_i^{rr'}) + w(r'y_i^{rr'}) \leq w(rs) + w(r's)$ (by the choice of $y_i^{rr'}$ and the fact that $rs \notin M_{2t+1}^2(t) \subseteq M$). However $|E_{G'}(X, \overline{X}) \setminus M| < |E_G(X, \overline{X}) \setminus M|$ (since $rs \notin M$ and $ry_i^{rr'}, r'y_i^{rr'} \in M$), a contradiction to our choice of G. □

By Claim 1, in order to find an (X)-partial Hamilton cycle of K_n of minimum weight, it is sufficient to find an (X)-partial Hamilton cycle of $K_n[X \cup Y]$ of minimum weight. We can do this in time $O(t^3)! + O(t^3 n)$ by brute force as follows.

We first determine Y in time $O(t^3 n)$. Next, we find all Hamilton cycles of $K_n[X \cup Y]$, which takes time $O(|X|+|Y|)! = O(t^3)!$. For each such Hamilton cycle, we delete its edges in $Y^{(2)}$ to obtain an (X)-partial Hamilton cycle H' and we determine $w(H')$ (this takes time $O(t^3)O(t^3)! = O(t^3)!$). We find the (X)-partial Hamilton cycle of $K_n[X \cup Y]$ of minimum weight, and its non-trivial paths form the non-trivial paths of an (X)-partial Hamilton cycle of K_n of minimum weight. □

We shall need the following (slightly generalised) algorithmic version of the well-known theorem of Dirac from graph theory; a proof can be found in [12].

Lemma 6.3 (Lemma 5.11 in [12]) *There exists an $O(n^3)$-time algorithm which, given an n-vertex graph G with $\delta(G) \geq n/2 + 3k/2$ and a matching $M = \{e_1, \dots, e_k\}$, outputs a Hamilton cycle H of G such that $M \subset E(H)$.*

Given an instance (n, w, k) of $\mathrm{TSP_{BA}}$, we define the subgraph K_w^+ (resp. K_w^-, K_w^0) of K_n to have vertex set $V(K_n)$ and to have edge set consisting of the edges assigned a positive (resp. negative, zero) weight by w.

Lemma 6.4 *Suppose that (n, w, k) is an instance of $\mathrm{TSP_{BA}}$ and $X \subseteq V(K_n)$ with $|X| = t$ satisfies the following properties:*

(a) $w(e) \geq 0$ *for all* $e \in E(K_n[\overline{X}])$*;*

(b) $\delta(K_w^0[\overline{X}]) \geq \frac{1}{2}|\overline{X}| + 4t$.

Then we can find a minimum weight Hamilton cycle H^* of K_n in time $O(t^3)! + O(t^3 n + n^3)$.

Proof In time $O(t^3)!+O(t^3 n)$, we can find an (X)-partial Hamilton cycle G^* of K_n of minimum weight. Let $P_1, \dots, P_r$ be the non-trivial paths of G^* and note that $r \le t$ since each path has an internal vertex in X. Let $a_i, b_i \in \overline{X}$ be the end-points of P_i and let $e_i = a_i b_i$.

Let X' be the set of internal vertices of $P_1, \dots, P_r$. Thus $X \subseteq X'$ and it is not hard to see that $|X'| \le 2|X| = 2t$ (since for any path, each internal vertex not in X has a unique predecessor in X). Thus

$$\begin{aligned}\delta(K_w^0[\overline{X'}]) \ge \delta(K_w^0[\overline{X}]) - |\overline{X} \setminus \overline{X'}| \ge \delta(K_w^0[\overline{X}]) - 2t \ge \frac{1}{2}|\overline{X}| + 4t - 2t \\ \ge \frac{1}{2}|\overline{X'}| + 2t \ge \frac{1}{2}|\overline{X'}| + \frac{3}{2}r.\end{aligned}$$

By Lemma 6.3, we can find, in time $O(n^3)$, a Hamilton cycle H of $K_w^0[\overline{X'}] \cup \{e_1, \dots, e_r\}$ that contains all the edges $e_1, \dots, e_r$. Replacing each edge e_i with the path P_i gives a Hamilton cycle H^* of K_n, which we claim has minimum weight.

Indeed, let H be any Hamilton cycle of K_n. Let G be obtained from H by deleting all the edges of H in $\overline{X}^{(2)}$; thus G is an (X)-partial Hamilton cycle, so $w(G^*) \le w(G)$. Note also by condition (a) that $w(H[\overline{X}]) \ge 0$, whereas by construction $w(H^*[\overline{X}]) = 0$. Therefore

$$w(H) = w(G) + w(H[\overline{X}]) \ge w(G^*) + w(H^*[\overline{X}]) = w(H^*)$$

showing that H^* is a minimum weight Hamilton cycle of K_n. □

Finally we can prove Theorem 1.1.

Proof (of Theorem 1.1) We assume $n > ck$ for a sufficiently large constant c; otherwise we can find a Hamilton cycle of minimum weight by brute force within the claimed number of steps. We describe the steps of the algorithm and give the running time in brackets.

1. Given (n, w, k), by Theorem 1.2, either we can output a Hamilton cycle H^* satisfying $w(H^*) < dn - k$ (and we are done) or we can find an equivalent instance (n, w', k) where $w'[K_n] \le 5000kn$. (time $O(n^7)$)

2. Find a maximum matching Q of $K_{w'}^-$, the subgraph of K_n containing the edges assigned a negative weight by w'. Let q be the number of edges of Q and set $d' := w'(K_n)/\binom{n}{2}$. (time $O(n^3)$; see [5] for a polynomial-time maximum matching algorithm)

3. If $q > 10^5k$ then $\mathbb{E}(w'(\tilde{H}^Q)) \le d'n - k$ (see Claim A below) and so by Lemma 4.2, we can find a Hamilton cycle H^* of K_n in time $O(n^5)$ satisfying $w'(H^*) \le \mathbb{E}(w'(\tilde{H}^Q)) \le d'n - k$ and hence $w(H^*) \le dn - k$. We output H^* and stop. (time $O(n^5)$)

4. If $q \le 10^5k$, then construct the set $X \subseteq V(K_n)$ such that $X = V(Q) \cup \{v \in V(K_n) \mid d_{K^+_{w'}}(v) \ge \frac{1}{4}n\}$. (time $O(n^2)$)

5. Using the properties of X proved in Claim B below, we can apply Lemma 6.4 to $(n, w', k), X$ to find a Hamilton cycle of minimum weight in K_n (w.r.t. w'). If $w'(H^*) \le d'n - k$ then output H^* and note $w(H^*) \le dn - k$. Otherwise we conclude there is no Hamilton cycle beating the average by at least k. (time $O(k^3)! + O(k^3n + n^3)$)

Claim A If $q > 10^5k$ then $\mathbb{E}(w'(\tilde{H}^Q)) \le d'n - k$.

Proof (of Claim A) If $q > 10^5k$ we have

$$\begin{aligned}\mathbb{E}(w'(\tilde{H}^Q)) &= w'(Q) + \sum_{e \in J(Q)} \mathbb{P}(e \in \tilde{H}^Q) w'(e) \\ &\le -10^5k + (4/(n-2)) \sum_{e \in J(Q)} w'(e) \\ &\le -10^5k + (4/(n-2))5000kn \\ &\le -50000k,\end{aligned}$$

where we have used that $w'(Q) \le -q$ (since $Q \subseteq K^-_{w'}$) and $\mathbb{P}(e \in \tilde{H}^Q) \le 4/(n-2)$ (using (5)) for the first inequality and where we have used that $w'[K_n] \le 4000kn$ for the second inequality, and that n is large enough for the third inequality. Note also that

$$\begin{aligned}d'n - k = \binom{n}{2}^{-1} w'(K_n)n - k \ge -5000k\frac{2}{n-1} - k \ge -50000k \\ \ge \mathbb{E}(w'(\tilde{H}^Q)),\end{aligned}$$

where the we have used that n is large enough for the second inequality. This gives the desired result. □

Claim B For the set X defined in Step 4, we have

(a) $t := |X| \le 3 \cdot 10^5k$;

(b) $w'(e) \ge 0$ for all $e \in E(K_n[\overline{X}])$;

(c) $\delta(K^0_{w'}[\overline{X}]) \geq \frac{1}{2}|\overline{X}| + 4t$.

Proof (of Claim B) (a) We have $X = V(Q) \cup S$ where $S := \{v \in V(K_n) \mid d_{K^+_{w'}}(v) \geq \frac{1}{4}n\}$. We show that $|S| \leq 10^5 k$ and since $|V(Q)| \leq 2q \leq 2 \cdot 10^5 k$, we have $|X| \leq 3 \cdot 10^5 k$ as required.

Observe that

$$\frac{1}{4}n|S| \leq 2e(K^+_{w'}) \leq 10000kn,$$

which implies $|S| \leq 40000k \leq 10^5 k$, as required.

(b) follows from the construction of X, since X contains all vertices of a maximum matching from the graph of negatively weighted edges.

To prove (c) observe that for each $x \in \overline{X}$

$$\begin{aligned} d_{K^0_{w'}[\overline{X}]}(x) &= (n-1) - |X| - d_{K^+_{w'}[\overline{X}]}(x) - d_{K^-_{w'}[\overline{X}]}(x) \\ &\geq (n-1) - 3 \cdot 10^5 k - (n/4) - 0 \\ &\geq (3/5)n - 3 \cdot 10^5 k \\ &\geq (1/2)n + 4t, \end{aligned}$$

where the last inequality holds for n large enough. □

This completes the proof of the theorem. □

7 Concluding Remarks

We believe that an analogue of our result ought to hold for the Asymmetric Travelling Salesman Problem. Formally, let $\overleftrightarrow{K}_n$ denote the complete directed graph on n vertices, so that there are two edges, one in each direction, between each pair of vertices.

Asymmetric Travelling Salesman Problem Below Average (ATSP$_{\text{BA}}$)

Instance: (n, w, k), where $n, k \in \mathbb{N}$ and $w : E(\overleftrightarrow{K}_n) \to \mathbb{Z}$

Question: Is there a directed Hamilton cycle H^* of $\overleftrightarrow{K}_n$ satisfying $w(H^*) \leq dn - k$, (where d is the average weight of an edge of $\overleftrightarrow{K}_n$)?

Problem 7.1 *Is* ATSP$_{\text{BA}}$ *fixed parameter tractable when parameterised by k?*

We believe the answer to this question is yes. However, several methods of this paper do not generalise in a straightforward way to directed graphs, so we expect that several new ideas will be needed to solve the problem above.

Acknowledments

The authors are grateful to the referees for a number of useful suggestions leading to an improvement in presentation. GG's research was partially supported by Royal Society Wolfson Research Merit Award and VP was supported by the Queen Mary - Warwick Strategic Alliance.

References

[1] N. Alon, G. Gutin, E.J. Kim, S. Szeider, and A. Yeo, Solving MAX–SAT Above a Tight Lower Bound. *Algorithmica* **61** (2011), 638–655.

[2] R. Crowston, M. Jones, and M. Mnich, Max-cut parameterized above the Edwards-Erdös bound, *Proc. ICALP 2012, Part I, Lect. Notes Comput. Sci.* **7391** (2012), 242–253.

[3] M. Cygan, F. V. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh, *Parameterized Algorithms*, Springer, to appear.

[4] G. Downey and M. R. Fellows, *Fundamentals of Parameterized Complexity*, Springer, 2013.

[5] J. Edmonds, Maximum matching and a polyhedron with $0,1$-vertices, *Journal of Research of the National Bureau of Standards* **69B** (1965), 125–130.

[6] C. S. Edwards, Some extremal properties of bipartite subgraphs, *Canadian Journal of Mathematics* **25** (1973), 475–483.

[7] C. S. Edwards, An improved lower bound for the number of edges in a largest bipartite subgraph, In *Recent Advances in Graph Theory*, (1975), 167–181.

[8] M. Grötschel, L. Lovász, and A. Schrijver, Complexity, Oracles, and Numerical Computation, In *Geometric Algorithms and Combinatorial Optimization*, 21–45, (1988).

[9] G. Gutin and A. Yeo, Constraint satisfaction problems parameterized above or below tight bounds: a survey. In *Fellows Festschrift, Lect. Notes Comput. Sci.* **7370** (2012), 257–286.

[10] G. Gutin, A. Yeo, and A. Zverovitch, Exponential neighborhoods and domination analysis for the TSP. In *The Traveling Salesman Problem and its Variations* (G. Gutin and A. Punnen, eds.), Kluwer, 2002.

[11] D.S. Johnson and L.A. McGeoch, Experimental analysis of heuristics for the TSP, in *The Traveling Salesman Problem and Its Variations* (G. Gutin and A.P. Punnen, eds.), Kluwer, 2002.

[12] D. Kühn, D. Osthus and V. Patel, A domination algorithm for $\{0,1\}$-instances of the Travelling Salesman Problem, *Random Structures and Algorithms*, to appear; preprint in arXiv:1401.4931.

[13] M. Mahajan, V. Raman, and S. Sikdar, Parameterizing above or below guaranteed values, *J. Comput. System Sci.* **75** (2009), 137–153.

[14] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, CUP, 2005.

[15] R. Motwani and P. Raghavan, *Randomized Algorithms*, CUP, 1995.

[16] M. Mnich, G. Philip, S. Saurabh and O. Suchy, Beyond Max-Cut: λ-extendible properties parameterised above the Poljak-Turzík bound, *J. Comput. System Sci.* **80** (2014), 1384–1403.

[17] V.I. Rublineckii, Estimates of the accuracy of procedures in the Traveling Salesman Problem, *Numer. Math. Comput. Tech.* **4** (1973), 18–23 (in Russian).

[18] V.G. Vizing, Values of the target functional in a priority problem that are majorized by the mean value, *Kibernetika*, Kiev **5** (1973), 76–78 (in Russian).