Approximation Hardness of TSP with Bounded Metrics (Revised Version)

Lars Engebretsen[1,⋆] and Marek Karpinski[2,⋆⋆]

[1] MIT Laboratory for Computer Science
200 Technology Square, NE43-367
Cambridge, Massachusetts 02139-3594
E-mail: enge@mit.edu

[2] Department of Computer Science
University of Bonn
53117 Bonn
E-mail: marek@cs.uni-bonn.de

Abstract. The general asymmetric TSP with triangle inequality is known to be approximable only to within an $O(\log n)$ factor, and is also known to be approximable within a constant factor as soon as the metric is bounded. In this paper we study the asymmetric and symmetric TSP problems with bounded metrics and prove approximation lower bounds of 131/130 and 174/173, respectively, for these problems, improving over the previous best lower bounds of 2805/2804 and 3813/3812 by an order of magnitude. Our bound 174/173 for the symmetric TSP with bounded metric is also the currently best known approximation lower bound for the general metric symmetric TSP problem.

We prove also approximation lower bounds of 321/320 and 743/742 for the asymmetric and symmetric TSP with distances one and two, improving over the previous best lower bounds of 2805/2804 and 5381/5380.

Key words. Approximation Ratios; Lower Bounds; Metric TSP; Bounded Metric.

⋆Supported by the Marcus Wallenberg foundation.
⋆⋆Supported in part by DFG grant, DIMACS, and IST grant 14036 (RAND-APX).

1 Introduction

A common special case of the Traveling Salesman Problem (TSP) is the metric TSP, where the distances between the cities satisfy the triangle inequality. The decision version of this special case was shown to be **NP**-complete by Karp [11], which means that we have little hope of computing exact solutions in polynomial time. Christofides [7] has constructed an elegant algorithm approximating the metric TSP within 3/2, i.e., an algorithm that always produces a tour whose weight is at most a factor 3/2 from the weight of the optimal tour. For the case when the distance function may be asymmetric, the best known algorithm approximates the solution within $O(\log n)$, where n is the number of cities [9], although a constant factor approximation algorithm has recently been conjectured [6]. As for lower bounds, Papadimitriou and Yannakakis [14] have shown that there exists some constant, see also [1], such that it is **NP**-hard to approximate the TSP where the distances are constrained to be either one or two—note that such a distance function always satisfies the triangle inequality—within that constant. This lower bound was improved by Engebretsen [8] to $2805/2804 - \epsilon$ for the asymmetric and $5381/5380 - \epsilon$ for the symmetric, respectively, TSP with distances one and two. Böckenhauer et. al [4, 5] considered the symmetric TSP with distances one, two and three, and were able to prove a lower bound of $3813/3812 - \epsilon$. (For a discussion of bounded metric TSP, see also Trevisan [15].) It appears that the metric TSP lacks the good definability properties which were needed (so far) for proving strong nonapproximability results. Therefore, any new insights into explicit lower bounds here seem to be of a considerable interest.

Papadimitriou and Vempala [12] recently announced lower bounds of $42/41 - \epsilon$ and $129/128 - \epsilon$, respectively, for the asymmetric and symmetric versions, respectively, of the TSP with graph metric, but left the question of the approximability for the case with bounded metric open. However, their proof contained an error influencing the explicit constants. A corrected proof with the new constants of $98/97 - \epsilon$ and $234/233 - \epsilon$, respectively, has been communicated to us by Papadimitriou and Vempala [13]. Apart from being an interesting question on its own, it is conceivable that the special cases with bounded metric are easier to approximate than the cases when the distance between two points can grow with the number of cities in the instance. Indeed, the asymmetric TSP with distances bounded by B can be approximated within B by just picking any tour as the solution and the asymmetric TSP with distances one and two can be approximated within 4/3 [3]. The symmetric version of the latter problem can be approximated within 7/6 [14].

In this paper, we consider the case when the metric contains only integer distances between one and six and prove a lower bound of $131/130 - \epsilon$ for the asymmetric case and $174/173 - \epsilon$ for the symmetric case. This is an improvement of an order of magnitude compared to the previous best known bounds of $2805/2804-\epsilon$ and $3813/3812-\epsilon$ for this case, respectively [4, 5, 8]. Our bound for the symmetric case is currently the best known bound for the general metric TSP, improving over the recent bound of Papadimitriou and Vempala [13]. We also prove that it is **NP**-hard to approximate the asymmetric TSP with distances one and two within $321/320 - \epsilon$, for any constant $\epsilon > 0$. For the symmetric version of the latter problem we show a lower bound of $743/742-\epsilon$. The previously best known bounds for this case are $2805/2804 - \epsilon$ and $5381/5380 - \epsilon$, respectively [8]. Our proofs depend on explicit reductions from certain bounded dependency instances of linear equations satisfiability. The main idea is to construct certain uniform circles of equation gadgets and, in the second part, certain combined hybrid circle constructions.

Definition 1.1. *The* Asymmetric Traveling Salesman Problem (ATSP) *is the following minimization problem: Given a collection of cities and a matrix whose entries are interpreted as the distance from a city to another, find the shortest tour starting and ending in the same city and visiting every city exactly once.*

Definition 1.2. (1,B)-ATSP *is the special case of ATSP where the entries in the distance matrix obey the triangle inequality and the off-diagonal entries in the distance matrix are integers between 1 and B.* (1,B)-TSP *is the special case of (1,B)-ATSP where the distance matrix is symmetric.*

2 The hardness of (1,B)-ATSP

We reduce, similarly to Papadimitriou and Vempala [12], from Håstad's lower bound for E3-Lin mod 2 [10]. In fact, our gadgets for the (1,B)-ATSP case are syntactically identical to those of Papadimitriou and Vempala [12] but we use a slightly different accounting method. The construction consists of a circle of *equation gadgets* testing odd parity. This is no restriction since we can easily transform a test for even parity into a test for odd parity by flipping a literal. Three of the edges in the equation gadget correspond to the variables involved in the parity check. These edges are in fact gadgets, so called *edge gadgets*, themselves. Edge gadgets from different equation gadgets are connected to ensure consistency among the edges representing a literal. This requires the number of negative occurrences of a variable to be equal to the number of positive occurrences. This is no restriction since we can duplicate every equation a constant number of times and flip literals to reach this property.

Definition 2.1. E3-Lin mod 2 *is the following problem: Given an instance of n variables and m equations over $\boldsymbol{Z}_2$ with exactly three unknowns in each equation, find an assignment to the variables that satisfies as many equations as possible.*

Theorem 2.1 [10]. *There exists instances of E3-Lin mod 2 with $2m$ equations such that, for any constant $\epsilon > 0$, it is **NP**-hard to decide if at most ϵm or at least $(1-\epsilon)m$ equations are left unsatisfied by the optimal assignment. Each variable in the instance occurs a constant number of times, half of them negated and half of them unnegated.*

We describe our instance of (1,B)-ATSP by constructing a weighted directed graph and then let the (1,B)-ATSP instance have the nodes of this graph as cities. The distance between two cities u and v is the (1,B)-ATSP instance is then defined to be $\min\{B, \ell(u,v)\}$, where $\ell(u,v)$ be the length of the shortest path from u to v in the graph.

2.1 The gadgets

The gadgets are parametrized by the parameters a, b and d; they will be specified later. The equation gadget for equations of the form $x+y+z=0$ is shown in Fig. 1. The key property of this gadget is that there is a Hamiltonian path through the gadget only if zero or two of the ticked edges are traversed. To form the circle of equation gadgets, vertex A in one gadget coincides with vertex B in another gadget.

The ticked edges in Fig. 1 are gadgets themselves. This gadget is shown in Fig. 2. Each of the bridges is shared between two different edge gadgets, one corresponding to a positive occurrence of the literal and one corresponding to a negative occurrence. The precise coupling is provided by a perfect matching in a d-regular bipartite multigraph $(V_1 \cup V_2, E)$ on $2k$ vertices with the following property: For any partition of V_1 into subsets S_1, U_1 and T_1 and any partition of V_2 into subsets S_2, U_2 and T_2 such that there are no edges from T_1 to T_2 and no edges from U_1 to U_2,

$$(|S_1| + |S_2|) \min\{a/2, b, (a+b)/2 - 1\} \geq$$

$$\min\{k, |U_1| + |T_2| + |S_1| + |S_2|, |U_2| + |T_1| + |S_1| + |S_2|\}.$$

The purpose of this construction is to ensure that it is always optimal for the tour to traverse the graph in such a way that all variables are given consistent values. The edge gadget gives an assignment to an occurrence of a variable by the way it is traversed.

Definition 2.2. *We call an edge gadget where all bridges are traversed from left to right in Fig. 2* traversed *and an edge gadget where all bridges are traversed from right to left* untraversed. *All other edge gadgets are called* semitraversed.

2.2 Proof of correctness

If we assume that the tour behaves nicely, i.e., that the edge gadgets are either traversed or untraversed, it is straightforward to establish a correspondence between the length of the tour and the number of unsatisfied equations.

Lemma 2.1. *The only way to traverse the equation gadget in Fig. 1 with a tour of length 4—if the edge gadgets count as length one for the moment—is to traverse an odd number of edge gadgets. All other locally optimal traversals have length 5.*

Proof. It is easy to see that any tour traversing two ticked edges and leaving the third one untraversed has length 4. Any tour traversing one ticked edge and leaving the other two ticked edges untraversed has length at least 5. Strictly speaking, it is impossible to have three traversals since this does not result in a tour. However, we can regard the case when the tour leaves the edge gadget by jumping directly to the exit node of the equation gadget as a tour with three traversals; such a tour gives a cost of 5. ∎

Lemma 2.2. *In addition to the length* 1 *attributed to the edge gadget above, the length of a tour traversing an edge gadget in the intended way is $d(a+b)$.*

Proof. Each bridge has length a, and every bridge must have one of the incoming edge traversed. Thus, the total cost is $d(a+b)$. ∎

Lemma 2.3. *Suppose that there are $2m$ equations in the E3-Lin mod 2 instance. If the tour is shaped in the intended way, i.e., every edge gadget is either traversed or untraversed, the length of the tour is $3md(a+b)+4m+u$, where u is the number of unsatisfied equations resulting from the assignment represented by the tour.*

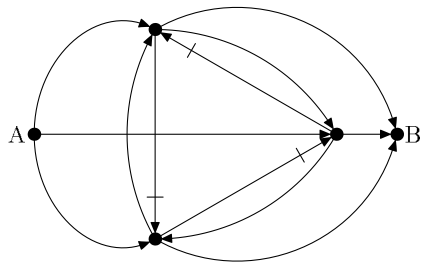

Figure 1. The gadget for equations of the form $x + y + z = 0$. There is a Hamiltonian path from A to B only if zero or two of the ticked edges, which are actually gadgets themselves (Fig. 2), are traversed. The non-ticked edges have weight 1.

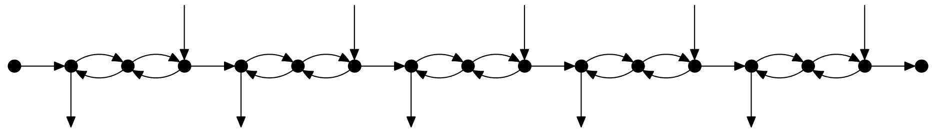

Figure 2. The edge gadget consists of d bridges. Each of the bridges are shared between two different edge gadgets and consist of a undirected edges of weight 1 each. The rightmost directed edge above has weight 1, the directed edges entering a bridge have weight b.

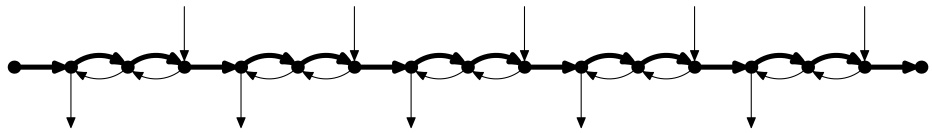

Figure 3. A traversed edge gadget represents the value 1.

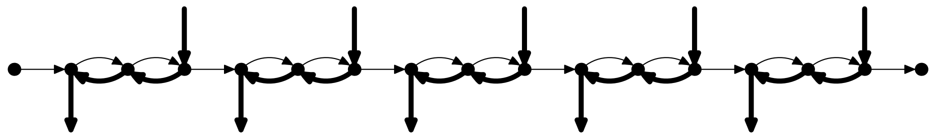

Figure 4. An untraversed edge gadget represents the value 0.

Proof. The length of the tour on an edge gadgets is $d(a+b)$. There are three edge gadgets corresponding to every equation and every bridge in the edge gadget is shared between two equation gadgets. Thus, the length of the tour on the edge gadgets is $2m \cdot 3d(a+b)/2 = 3md(a+b)$ The length of the tour on an equation gadget is 4 if the equation is satisfied and 5 otherwise. Thus, the total length is $3md(a+b) + 4m + u$. ∎

The main challenge now is to prove that the above correspondence between the length of the optimum tour and the number of unsatisfied equation holds also when we drop the assumption that the tour is shaped in the intended way. Our proof uses the following technical lemma (we provide a proof in the appendix):

Lemma A.1. *For every large enough constant k, there exists an 7-regular bipartite multigraph on $2k$ vertices such that for every partition of the left vertices into sets T_1, U_1 and S_1 and every partition of the right vertices into sets T_2, U_2 and S_2 such that there are no edges from T_1 to T_2, and there are no edges from U_1 to U_2,*

$$2(|S_1| + |S_2|) \geq \min\{k, |U_1| + |T_2| + |S_1| + |S_2|, |U_2| + |T_1| + |S_1| + |S_2|\}$$

with equality only if $S_1 = S_2 = U_1 = T_2 = \emptyset$ or $S_1 = S_2 = T_1 = U_2 = \emptyset$.

Given the above lemma, the following sequence of lemmas give a lower bound on the extra cost, not counting the "normal" cost of $d(a+b)$ per edge gadget and 4 per equation gadget, that results from a non-standard behavior of the tour. We have already seen that an unsatisfied equation adds an extra cost of 1. Edge gadgets that are either traversed or untraversed do not add any extra cost. Note that traversed edge gadgets never can share the same bridge, neither can untraversed edge gadgets. We now give a lower bound on the additional length of the tour due to semitraversed edge gadgets.

Lemma 2.4. *Suppose that $B \geq \max\{a+b, 3b\}$. Then every semitraversed edge gadget adds an extra cost of at least $\min\{a/2, b, (a+b)/2 - 1\}$ to the length of the tour.*

Proof. A bridge is said to have an undefined traversal if the tour does not traverse it in the intended way, i.e., the restriction of the tour to the bridge does not result in a simple path from one end of the bridge to the other. There are two reasons for an edge gadget to be classified as semitraversed. Either there is a bridge that has an undefined traversal, or the direction in which the bridges are traversed changes without any bridge having an undefined traversal.

In the former case there is an extra cost of at least $a - 2 + b$ which is shared between the two edge gadgets that cross at the bridge with an undefined traversal.

If the edge gadget has no bridges with an undefined traversal, it can still be semitraversed. The direction in which the bridges are traversed then changes between two bridges. There are two cases. Either the tour switches from traversing an edge gadget representing an occurrence of x to traversing an edge gadget representing an occurrence of $\bar{x}$—in this case the extra cost is a—or, the tour switches from traversing an edge gadget representing an occurrence of x to traversing another edge gadget representing an occurrence of x—in this case the extra cost is $2b$. In both of the above cases, the extra cost is shared evenly between the two semitraversed edge gadgets involved. ∎

Lemma 2.5. *For $a = 4$, $b = 2$ and $d = 7$, there exists a coupling of the equation gadgets with the property that it can never be advantageous to have inconsistently traversed equation gadgets.*

Proof. Repeat the following argument for every variable x:

Let k be the number of occurrences of x (and also the number of occurrences of $\bar{x}$). Pick a bipartite multigraph on $2k$ vertices such that for every partition of the left vertices into sets T_1, U_1 and S_1 and every partition of the right vertices into sets T_2, U_2 and S_2 such that there are no edges from T_1 to T_2, and there are no edges from U_1 to U_2,

$$2(|S_1| + |S_2|) \geq \min\{k, |U_1| + |T_2| + |S_1| + |S_2|, |U_2| + |T_1| + |S_1| + |S_2|\}$$

with equality only if $S_1 = S_2 = U_1 = T_2 = \emptyset$ or $S_1 = S_2 = T_1 = U_2 = \emptyset$. We know by Lemma A.1 that such a graph exists—since the graph has constant size, we can try all possible graphs in constant time.

Put occurrences of x at one side and occurrences of $\bar{x}$ on the other side of the bipartite graph. Each vertex in the graph can be labeled as T, U or S, depending on whether it is traversed, untraversed or semitraversed. Let T_1 be the set of traversed positive occurrences and T_2 be the set of traversed negative occurrences. Define U_1, U_2, S_1, and S_2 similarly. We can assume that $|U_1| + |T_2| \leq |U_2| + |T_1|$—otherwise we just change the indexing convention.

We now consider a modified tour where the positive occurrences are traversed and the negative occurrences are untraversed. This decreases the cost of tour by at least $2(|S_1| + |S_2|)$ and increases it by $\min\{k, |S_1| + |S_2| + |U_1| + |T_2|\}$. But the bipartite graph has the property that

$$2(|S_1| + |S_2|) \geq \min\{k, |U_1| + |T_2| + |S_1| + |S_2|\}$$

which implies that the cost of tour decreases by this transformation. Thus, we can assume that x is given a consistent assignment by the tour. ∎

Theorem 2.2. *For any constant $\epsilon > 0$, it is **NP**-hard to approximate (1,6)-ATSP within $131/130 - \epsilon$.*

Proof. Given an instance of E3-Lin mod 2 with $2m$ equations where every variable occurs a constant number of times, we construct the corresponding instance of (1,6)-ATSP with $a = 4$, $b = 2$ and $d = 7$. This can be done in polynomial time. By the above lemma, we can assume that all edge gadgets are traversed consistently in this instance. The assignment obtained from this traversal satisfies $2m - u$ equations if the length of the tour is $3md(a+b) + 4m + u$. If we could decide if the length of the optimum tour is at most $(3d(a+b)+4+\epsilon_1)m$ or at least $(3d(a+b)+5-\epsilon_2)m$, we could decide if at most $\epsilon_1 m$ or at least $(1-\epsilon_2)m$ of the equations are left unsatisfied by the corresponding assignment. But to decide this is **NP**-hard by Theorem 2.1. Therefore it is **NP**-hard to approximate (1,6)-ATSP within

$$\frac{3d(a+b)+5-\epsilon_2}{3d(a+b)+4+\epsilon_1} \geq \frac{131}{130} - \epsilon.$$ ∎

3 The hardness of (1,B)-TSP

To adapt the construction from the previous section for the symmetric case we need to change some of the gadgets. Most changes in the equation gadgets are minor—the main change being that we test odd instead of even parity for equations with three variables (Fig. 8). There is a more substantial change in the edge gadget; it is changed according to Fig. 9.

If we assume that the tour behaves nicely, it is straightforward to prove a correspondence between the length of a tour and the number of equations left unsatisfied by the corresponding assignment.

Lemma 3.1. *The only way to traverse the equation gadget in Fig. 8 with a tour of length 5—if the edge gadgets count as length one for the moment—is to traverse an odd number of edge gadgets. All other locally optimal traversals have length 6.*

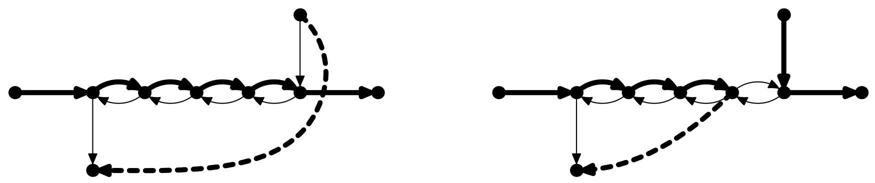

Figure 5. We can assume that traversals shown in the left figure above never occur since they can be transformed into the traversal shown in the right figure without increasing the length of the tour. A bridge with such a traversal gives an extra cost of $(a+b)/2-1$. The total length of the part of the tour that traverses the bridge and its two incoming edges in the right figure above is $2a-2-2b$; that gives an extra cost of $a-2-b$ which is shared evenly between the two semitraversed edge gadgets.

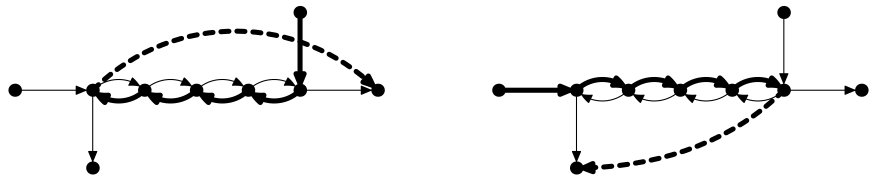

Figure 6. Switching from traversing an edge gadget representing an occurrence of x to traversing an edge gadget representing an occurrence of $\bar{x}$ gives an extra cost of at least $a/2$. The dashed edges above has length $a+b$; that gives an extra cost of a which is then shared evenly among the two semitraversed edge gadgets.

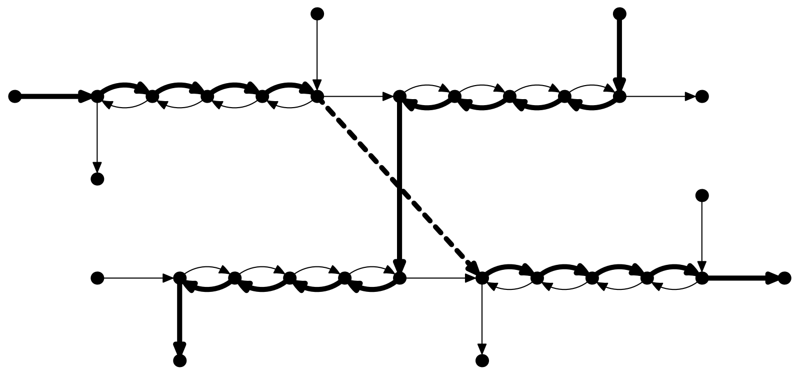

Figure 7. Switching from traversing an edge gadget representing an occurrence of x to traversing another edge gadget representing an occurrence of x gives an extra cost of at least b. The dotted edge above has length $3b$; that gives an extra cost of $2b$ which is then shared evenly among the two semitraversed edge gadgets.

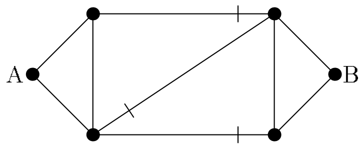

Figure 8. The gadget for equations of the form $x + y + z = 1$. There is a Hamiltonian path from A to B only if one or three of the ticked edges are traversed.

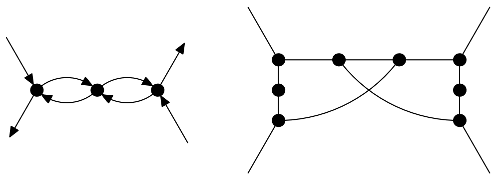

Figure 9. To transform the edge gadget from Fig. 2 into a gadget that can be used in the symmetric case, all occurrences of the structure to the left above are replaced with the structure to the right above.

Proof. It is easy to see that any tour traversing either one or three of the ticked edges and leaving the third one untraversed has length 5. Any tour traversing zero or two ticked edges end up on the wrong side of the gadget and needs an extra cost of at least one to get back to the other side. ∎

Lemma 3.2. *In addition to the length 1 attributed to the edge gadget above, the length of a tour traversing an edge gadget in the intended way is* 56.

Proof. The total cost is $7 \cdot (7 + 1) = 56$. ∎

Lemma 3.3. *Suppose that there are* $2m$ *equations in the E3-Lin instance. If the tour is shaped in the intended way, i.e., every edge gadget is either traversed or untraversed, the length of the tour is* $173m + u$, *where* u *is the number of unsatisfied equations resulting from the assignment represented by the tour.*

Proof. The length of the tour on the edge gadgets is 64. There are three edge gadgets corresponding to every equation and every bridge in the edge gadget is shared between two equation gadgets. Thus, the length of the tour on the edge gadgets is $2m \cdot 3 \cdot 56/2 = 168m$. The length of the tour in the equation gadgets is 5 if the equation is satisfied and 6 otherwise. Thus, the total length is $173m + u$. ∎

In the same way as in the asymmetric case, it can now be shown that the tour can be assumed to behave in the intended way. This gives the following lemma (we omit the proof):

Lemma 3.4. *Suppose that $B \geq 6$. Then every semitraversed edge gadget adds an extra cost of at least 2 to the length of the tour.*

There exists a coupling of the edge gadgets with the property that there can never be advantageous to have inconsistently traversed edge gadgets.

Given the above lemma, the main theorem follows in exactly the same way as in the asymmetric case.

Theorem 3.1. *For any constant $\epsilon > 0$, it is **NP**-hard to approximate (1,6)-TSP within $174/173 - \epsilon$.*

Proof. Given an instance of E3-Lin mod 2 with $2m$ equations where every variable occurs a constant number of times, we construct the corresponding instance of (1,6)-TSP. This can be done in polynomial time. By the above lemma, we can assume that all edge gadgets are traversed consistently in this instance. The assignment obtained from this traversal satisfies $2m - u$ equations if the length of the tour is $173m + u$. If we could decide if the length of the optimum tour is at most $(173 + \epsilon_1)m$ or at least $(174 - \epsilon_2)m$, we could decide if at most $\epsilon_1 m$ or at most $(1 - \epsilon_2)m$ of the equations are let unsatisfied by the corresponding assignment. But to decide this is **NP**-hard by Theorem 2.1. ∎

4 The hardness of (1,2)-ATSP

To prove a lower bound for (1,2)-ATSP we apply the construction used by Berman and Karpinski [2], a reduction from systems of linear equations mod 2 with exactly three unknowns in each equation to a problem called *Hybrid*, to prove hardness results for instances of several combinatorial optimization problems where the number of occurrences of every variable is bounded by some constant.

Definition 4.1. Hybrid *is the following maximization problem: Given a system of linear equations mod 2 containing n variables, m_2 equations with exactly two unknowns, and m_3 equations exactly with three unknowns, find an assignment to the variables that satisfies as many equations as possible.*

Theorem 4.1 [2]. *There exists instances of Hybrid with 42ν variables, 60ν equations with two variables, and 2ν equations with three variables such that:*

1. *Each variable occurs exactly three times.*

2. *For any constant $\epsilon > 0$, it is **NP**-hard to decide if at most $\epsilon\nu$ or at least $(1-\epsilon)\nu$ equations are left unsatisfied.*

Since we adopt the construction of Berman and Karpinski [2], we can partly rely on their main technical lemmas, which simplifies our proof of correctness.

On a high level, the (1,2)-ATSP instance in our reduction consists of a circle formed by *equation gadgets* representing equations of the form $x+y+z=0$ and $x+y=1$. These gadgets are coupled in a way ensuring that the three occurrences of a variable are given consistent values. In fact, the instances of Hybrid produced by the Berman-Karpinski construction have a very special structure. Every variable occurs in at least two equations with two unknowns, and those equations are all equivalences, i.e., equations of the form $x+y=0$. Since our gadget for equations with two unknowns tests odd parity, we have to rewrite those equations as $x+\bar{y}=1$ instead. Similarly, the equations of the form $x+y+z=1$ must be rewritten with one variable negated since our gadgets for equations with three unknowns only test even parity. Turning to the coupling needed to ensure consistency, we have three occurrences of every variable. Since we do not have any gadgets testing odd parity for three variables or even parity for two variables, we may have to negate some of the occurrences. We now argue that there are either one or two negated occurrences of every variable. The Hybrid instance produced by the Berman-Karpinski construction can be viewed as a collection of wheels where the nodes correspond to variables and edges to equations. The edges within a wheel all represent equations with two unknowns, while the equations with three unknowns are represented by hyperedges connecting three different wheels. Figure 10 gives an example of one such wheel. The equations corresponding to the edges forming the perimeter of the wheel can be written as $x_1+\bar{x}_2=1$, $x_2+\bar{x}_3=1$, $\ldots$, $x_{k-1}+\bar{x}_k=1$, and $x_k+\bar{x}_1=1$, which implies that there is at least one negated and at least one unnegated occurrence of each variable.

Corollary 4.1. *There are instances of Hybrid with 42ν variables, 60ν equations of the form $x+\bar{y}=1 \bmod 2$, and 2ν equations of the form $x+y+z=0 \bmod 2$ or $x+y+\bar{z}=0 \bmod 2$ such that:*

1. *Each variable occurs exactly three times.*

2. *There is at least one positive and at least one negative occurrence of each variable.*

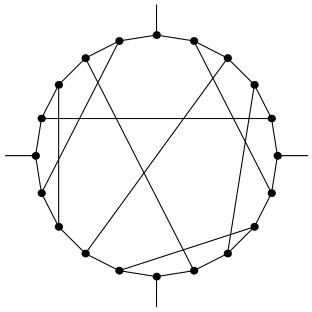

Figure 10. The Hybrid instance produced by the Berman-Karpinski construction can be viewed as a collection of wheels where the nodes correspond to variables and edges to equations.

3. *For any constant $\epsilon > 0$, it is **NP**-hard to decide if at most $\epsilon\nu$ or at least $(1-\epsilon)\nu$ equations are left unsatisfied.*

To prove our hardness result for (1,2)-ATSP, we reduce instances of Hybrid of the form described in Corollary 4.1 to instances of (1,2)-ATSP and prove that, given a tour in the (1,2)-ATSP instance, it is possible to construct an assignment to the variables in the original Hybrid instance with the property that the number of unsatisfied equations in the Hybrid instance is related to the length of the tour in the (1,2)-ATSP instance.

To describe a (1,2)-TSP instance, it is enough to specify the edges of weight one. We do this by constructing a graph G and then let the (1,2)-TSP instance have the nodes of G as cities. The distance between two cities u and v is defined to be one if (u, v) is an edge in G and two otherwise. To compute the weight of a tour, it is enough to study the parts of the tour traversing edges of G. In the asymmetric case G is a directed graph.

Definition 4.2. *We call a node where the tour leaves or enters G an* endpoint. *A node with the property that the tour both enters and leaves G in that particular node is called a* double endpoint *and counts as two endpoints.*

If c is the number of cities and $2e$ is the total number of endpoints, the weight of the tour is $c+e$ since every edge of weight two corresponds to two endpoints.

4.1 The gadgets

The equation gadget for equations of the form $x + y + z = 0$ is shown in Fig. 1—the same gadget as in the (1,B) case. However, the ticked edges

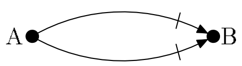

Figure 11. The gadget for equations of the form $x + y = 1$. There is a Hamiltonian path from A to B only if one of the ticked edges is traversed.

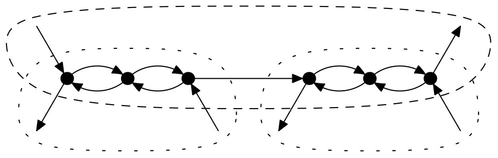

Figure 12. The gadget ensuring consistency for a variable. If there are two positive occurrences of the variable, the ticked edges corresponding to those occurrences are represented by the parts enclosed in the dotted curves and the ticked edge corresponding to the negative occurrence is represented by the part enclosed in the dashed curve. If there are two negative occurrences, the rôles are reversed.

now represent a different structure.

The equation gadget for equations of the form $x + y = 1$ is shown in Fig. 11. The key property of this gadget is that there is a Hamiltonian path through the gadget only if one of the ticked edges is traversed.

The ticked edges in the equation gadgets are syntactic sugar for a construction ensuring consistency among the three occurrences of each variable. As we noted above, either one or two of the occurrences of a variable are negated. The construction in Fig. 12 ensures that the occurrences are given consistent values, i.e., that either $x = 0$ and $\bar{x} = 1$, or $x = 1$ and $\bar{x} = 0$. If there is one negated occurrence of a variable, the upper part of the gadget connects with that occurrence and the lower part connects with the two unnegated occurrences. If there are two negated occurrences, the situation is reversed.

4.2 Proof of correctness

We want to prove that every unsatisfied equation has an extra cost of one associated with it. At first, it would seem that this is very easy—the gadget in Fig. 1 is traversed by a path of length four if the equation is satisfied and a path of length at least five otherwise; the gadget in Fig. 11 is traversed by a path of length one if the equation is satisfied and a path of length at least two otherwise; and the gadget in Fig. 12 ensures consistency and is traversed by a tour of length six, not counting the edges that were accounted for above. Unfortunately, things are more complicated than this. Due to the consistency gadgets, the tour can leave a ticked edge when it is half-way through it, which forces us to be more careful in our analysis.

We count the number of *endpoints* that occur within the gadgets; each endpoint gives an extra cost of one half. We say that an occurrence of a literal is *traversed* if both of its connected edges are traversed, *untraversed* if none of its connecting edges are traversed, and *semitraversed* otherwise. To construct an assignment to the literals, we use the convention that a literal is true if it is either traversed or semitraversed. We need to show that there are two endpoints in gadgets that are traversed in such a way that the corresponding assignment to the literals makes the equation unsatisfied. The following lemmas are easy, but tedious, to verify by case analysis:

Lemma 4.1. *It is locally optimal to traverse both* bridges*, i.e., both pairs of undirected edges, in the consistency gadget.*

Proof. By case analysis. ∎

Lemma 4.2. *Every semitraversed occurrence introduces at least one endpoint.*

Proof. By case analysis on traversed connection edges. ∎

Lemma 4.3. *It is always possible to change a semitraversed occurrence into a traversed one without introducing any endpoints in the consistency gadget.*

Proof. By case analysis on traversed connection edges. ∎

Given the above lemmas, the following two lemmas prove the properties we need regarding the equation gadgets.

Lemma 4.4. *A "satisfying traversal" of the gadget in Fig. 11 has length* 1*, all other locally optimal traversals have length at least* 2*, i.e., contain at least two endpoints within the gadget.*

Proof. If one of the ticked edges is traversed and the other is untraversed, the gadget is traversed by a tour of length 1. It is suboptimal to have one semitraversed and one untraversed edge, in this case it is possible to shorten the tour by transforming the semitraversed edge into a traversed one.

Two untraversed edges give a total cost of at least 2. It is impossible to have either two traversed edges or one traversed and one semitraversed ticked edge, since that gives a traversal which is not a tour. Two semitraversed edges give an extra cost of 1/2 each, giving a total cost of at least 2. ∎

Lemma 4.5. *A "satisfying traversal" of the gadget in Fig. 1 has length* 4*, all other locally optimal traversals have length at least* 5*, i.e., contain at least two endpoints within the gadget.*

Proof. It is easy to see that any tour traversing two ticked edges and leaving the third one untraversed has length 4. The case with two semitraversed occurrences and one untraversed is suboptimal since a shorter tour can be produced in this way: Make the semitraversed occurrences traversed and then adjust the tour on the non-ticked edges to get a tour of length 4. Similarly, the case with one traversed and one semitraversed occurrence can be transformed into two semitraversed occurrences.

Any tour traversing one ticked edge and leaving the other two ticked edges untraversed has length at least 5. A tour semitraversing one ticked edge and leaving the other ticked edges untraversed can be transformed into a tour with one traversal and two non-traversals. It is impossible to have three traversals since this does not result in a tour. The case with two traversals and one semitraversal gives a cost of 5, and so does case with one traversal and two semitraverals, since each semitraversal has an extra cost of $1/2$ associated with it. ∎

When the above lemmas have been proven, we only need to prove that the gadget we use for consistency actually implements consistency.

Lemma 4.6. *The gadget in Fig. 12 ensures consistency and is traversed by a tour of length* 6*, not counting the edges or endpoints that were accounted for in the above lemmas.*

Proof. If there are no semitraversed occurrences, the gadget implements consistency correctly.

Suppose that the upper occurrence in Fig. 12 is semitraversed in such a way that the leftmost connecting edge is traversed but the rightmost is not. Then it is possible to have the lower left occurrence untraversed and the lower right occurrence traversed. Since a semitraversed occurrence is always part of an unsatisfied equation gadget, the following procedure produces a tour with equal cost: Make the upper occurrence untraversed and the lower left occurrence traversed. This makes the equation gadget that the upper occurrence is connected to satisfied and may make the equation gadget that the lower left occurrence is connected to unsatisfied.

Suppose that the lower left occurrence in Fig. 12 is semitraversed in such a way that the leftmost connecting edge is traversed but the rightmost is not. Then it is possible to have the lower right occurrence untraversed and the upper occurrence semitraversed. Since a semitraversed occurrence is always part of an unsatisfied equation gadget, the following procedure produces a tour with equal cost: Make the upper occurrence untraversed and the lower right occurrence traversed. This makes the equation gadget that the upper occurrence is connected to satisfied and may make the equation gadget that the lower right occurrence is connected to unsatisfied.

With similar arguments it can be shown that the lemma holds for all other possible cases. ∎

By combining the above lemmas, we have shown the following connection between the length of an optimum tour and the number of unsatisfied equations in the corresponding instance of Hybrid.

Theorem 4.2. *Suppose that we are given an arbitrary instance of Hybrid with n variables, m_2 equations of the form $x+\bar{y} = 1 \bmod 2$, and m_3 equations of the form $x + y + z = 0 \bmod 2$ or $x + y + \bar{z} = 0 \bmod 2$ such that:*

1. *Each variable occurs exactly three times.*
2. *There is at least one positive and at least one negative occurrence of each variable.*

Then we can construct an instance of (1,2)-ATSP with the property that a tour of length $6n + m_2 + 4m_3 + u$ corresponds to an assignment satisfying all but u of the equations in the Hybrid instance.

Corollary 4.2. *For any constant $\epsilon > 0$, it is **NP**-hard to approximate (1,2)-ATSP within $321/320 - \epsilon$.*

Proof. We connect Theorem 4.2 with Corollary 4.1 and obtain an instance of (1,2)-ATSP with the property that a tour of length

$$6n + m_2 + 4m_3 + u = 6 \cdot 42\nu + 60\nu + 4 \cdot 2\nu + u = 320\nu + u$$

corresponds to an assignment satisfying all but u of the equations in the Hybrid instance. Since, for any constant $\epsilon' > 0$, it is **NP**-hard to distinguish the cases $u \leq \epsilon'$ and $u \geq 1 - \epsilon'$, it is **NP**-hard to approximate (1,2)-ATSP within $321/320 - \epsilon$ for any constant $\epsilon > 0$. ∎

5 The hardness of (1,2)-TSP

It is possible to adapt the above construction for (1,2)-ATSP to prove a lower bound also for (1,2)-TSP. The equation gadget for equations containing three variables is changed in the same way as in the $(1,B)$ case, the consistency gadget is change in a similar way.

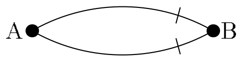

Figure 13. The gadget for equations of the form $x + y = 1$. There is a Hamiltonian path from A to B only if one of the ticked edges is traversed.

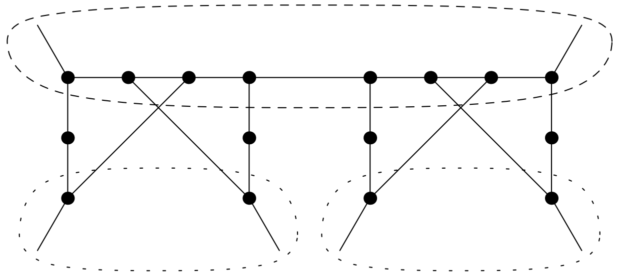

Figure 14. The gadget ensuring consistency for a variable. If there are two positive occurrences of the variable, the ticked edges corresponding to those occurrences are represented by the parts enclosed in the dotted curves and the ticked edge corresponding to the negative occurrence is represented by the part enclosed in the dashed curve. If there are two negative occurrences, the rôles are reversed.

5.1 The gadgets

The equation gadget for equations of the form $x+y=1$ is shown in Fig. 13. The key property of this gadget is that there is a Hamiltonian path through the gadget only if one of the ticked edges is traversed.

The equation gadget for equations of the form $x+y+z=1$ is shown in Fig. 8—the same gadget as in the $(1,B)$ case.

The ticked edges in the equation gadgets are syntactic sugar for a construction ensuring consistency among the three occurrences of each variable. As we noted above, either one or two of the occurrences of a variable are negated. The construction in Fig. 14 ensures that the occurrences are given consistent values, i.e., that either $x = 0$ and $\bar{x} = 1$, or $x = 1$ and $\bar{x} = 0$. If there is one negated occurrence of a variable, the upper part of the gadget connects with that occurrence and the lower part connects with the two unnegated occurrences. If there are two negated occurrences, the situation is reversed.

5.2 Proof of correctness

In the same way as in the asymmetric case, it can be shown that the tour can be assumed to behave in the intended way. When this result is combined with the lower bound on the approximability of Hybrid, we obtain the following theorem:

Theorem 5.1. *Suppose that we are given an instance of Hybrid with n variables, m_2 equations of the form $x + \bar{y} = 1 \bmod 2$, and m_3 equations of the form $x + y + z = 0 \bmod 2$ or $x + y + \bar{z} = 0 \bmod 2$ such that:*

1. *Each variable occurs exactly three times.*
2. *There is at least one positive and at least one negative occurrence of each variable.*

Then we can construct an instance of (1,2)-TSP with the property that a tour of length $16n + m_2 + 5m_3 + u$ corresponds to an assignment satisfying all but u of the equations in the Hybrid instance.

Theorem 5.2. *For any constant $\epsilon > 0$, it is **NP**-hard to approximate (1,2)-TSP within $743/742 - \epsilon$.*

Proof. We connect Theorem 5.1 with Corollary 4.1 and obtain an instance of (1,2)-TSP with the property that a tour of length

$$16n + m_2 + 5m_3 + u = 16 \cdot 42nu + 60\nu + 5 \cdot 2\nu + u = 742\nu + u$$

corresponds to an assignment satisfying all but u of the equations in the Hybrid instance. Since, for any constant $\epsilon' > 0$, it is **NP**-hard to distinguish the cases $u \leq \epsilon'$ and $u \geq 1 - \epsilon'$, it is **NP**-hard to approximate (1,2)-TSP within $743/742 - \epsilon$ for any constant $\epsilon > 0$. ∎

6 Conclusions

It should be possible to improve the reduction by eliminating the vertices that connect the equation gadgets for $x + y + z = \{0, 1\}$ with each other. This reduces the cost of those equation gadgets by one, which improves our bounds—but only by a miniscule amount. The big bottleneck, especially in the (1,2) case, is the consistency gadgets. If, for the asymmetric case, we were able to decrease the cost of them to four instead of six, we would improve the bound to $237/236 - \epsilon$; if we could decrease the cost to three, the bound would become $195/194 - \epsilon$. We conjecture that some improvement for the $(1, 2)$ case is still possible along these lines.

Acknowledgments

We thank Santosh Vempala for many clarifying discussions on the subject of this paper.

A The bipartite graph

This section is devoted to the proof of the following technical lemma:

Lemma A.1. *For every large enough constant k, there exists an 7-regular bipartite multigraph on $2k$ vertices such that for every partition of the left vertices into sets T_1, U_1 and S_1 and every partition of the right vertices into sets T_2, U_2 and S_2 such that there are no edges from T_1 to T_2, and there are no edges from U_1 to U_2,*

$$2(|S_1|+|S_2|) \geq \min\{k, |U_1|+|T_2|+|S_1|+|S_2|, |U_2|+|T_1|+|S_1|+|S_2|\} \quad (1)$$

with equality only if $S_1 = S_2 = U_1 = T_2 = \emptyset$ or $S_1 = S_2 = T_1 = U_2 = \emptyset$.

The proof uses the same main idea as the proof of a similar expansion theorem communicated to us by Papadimitriou and Vempala [13]. In particular, it uses a lemma that bounds the size of neighbor sets in 7-regular bipartite graphs.

Lemma A.2. *For every large enough constant k, there exists a 7-regular bipartite multigraph on $2k$ vertices such that every subset U of vertices contained entirely in V_1 or V_2 has a set $N(U)$ of neighbors satisfying the following constraints:*

$$\begin{aligned}
&0 < |U| \leq k/10 \implies |N(U)| > 29|U|/10, && (2)\\
&k/10 \leq |U| \leq 3k/10 \implies |N(U)| > 13k/100 + 8|U|/5, && (3)\\
&3k/10 \leq |U| \leq 39k/100 \implies |N(U)| > 31k/100 + |U|, && (4)\\
&39k/100 \leq |U| \leq 62k/100 \implies |N(U)| > k/2 + |U|/2, && (5)\\
&62k/100 \leq |U| < k \implies |N(U)| > \max\{85k/100, k/2 + |U|/2\}. && (6)
\end{aligned}$$

Proof. We select a d-regular bipartite graph on $2k$ vertices by selecting d perfect matchings independently and uniformly at random. Let $A_{U,N}$ be the event that the set U has neighbors only inside the set N and let Ω be the subset of $\{0, 1, 2, \ldots, k\} \times \{0, 1, 2, \ldots, k\}$ such that if $(a, b) \in \Omega$,

$$\begin{aligned}
&0 < a \leq k/10 \implies b \leq 29a/10,\\
&k/10 \leq a \leq 3k/10 \implies b \leq 13k/100 + 8a/5,\\
&3k/10 \leq a \leq 39k/100 \implies b \leq 31k/100 + a,\\
&39k/100 \leq a \leq 62k/100 \implies b \leq k/2 + a/2,\\
&62k/100 \leq a < k \implies b \leq \max\{85/100, k/2 + a/2\}.
\end{aligned}$$

Denote the vertex set of the bipartite graph by $V_1 \cup V_2$. We need to prove that

$$\Pr\left[\bigcup_{i=1}^{2} \bigcup_{(a,b)\in\Omega} \bigcup_{\substack{U\subset V_i\\ |U|=a}} \bigcup_{\substack{N\subset V_{2-i}\\ |N|=b}} A_{U,N}\right] < 1$$

and we do this by using the union bound, i.e., we prove that

$$\sum_{i=1}^{2} \sum_{(a,b)\in\Omega} \sum_{\substack{U\in V_i\\ |U|=a}} \sum_{\substack{N\in V_{2-i}\\ |N|=b}} \Pr[A_{U,N}] < 1.$$

First note that $\Pr[A_{U,N}] = 0$ when $|U| > |N|$, therefore it suffices to consider only (a,b) such that $a \leq b$. Since Ω contains less than k^2 pairs and

$$\Pr[A_{U,N}] = \binom{d|N|}{d|U|} \frac{(d|U|)!(dk - d|U|)!}{(dk)!} = \frac{\binom{d|N|}{d|U|}}{\binom{dk}{d|U|}}$$

when $a \leq b$ it is enough to prove that

$$2k^2 \max_{\substack{(a,b)\in\Omega\\ a\leq b}} \binom{k}{a}\binom{k}{b}\frac{\binom{db}{da}}{\binom{dk}{da}} = \max_{\substack{(a,b)\in\Omega\\ a\leq b}} P(a,b) < 1.$$

We prove this inequality by case analysis. When a can be written as αk where $10^{-5} \leq \alpha \leq 1 - 10^{-5}$ we expand the above expression using Stirling's formula. The case when a, and therefore also b, is very close to either 0 or k is dealt with separately.

Case I: $10^{-5} \leq a/k \leq 1 - 10^{-5}$. By Stirling's formula

$$\binom{k}{\alpha k} = (\alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)})^k \operatorname{poly}(k).$$

Now write $a = \alpha k$ and $b = \beta k$ and apply Stirling's formula to the expression we want to bound. This gives us the equality

$$P(\alpha k, \beta k) = \left(\frac{(1-\alpha)^{(d-1)(1-\alpha)}\beta^{(d-1)\beta}}{\alpha^\alpha(1-\beta)^{(1-\beta)}(\beta-\alpha)^{d(\beta-\alpha)}}\right)^k \operatorname{poly}(k)$$

This expression is certainly strictly less than 1 for all (α,β) such that $\alpha \leq \beta$ and $(\alpha k, \beta k) \in \Omega$ as soon as

$$Q(\alpha,\beta) = \frac{(1-\alpha)^{(d-1)(1-\alpha)}\beta^{(d-1)\beta}}{\alpha^\alpha(1-\beta)^{(1-\beta)}(\beta-\alpha)^{d(\beta-\alpha)}} < 1$$

for all (α,β) such that $\alpha \leq \beta$ and $(\alpha k, \beta k) \in \Omega$ and k is large enough. The validity of the latter inequality can then be verified numerically.

Case II: $0 < a/k \leq 10^{-5}$. For every fixed a in that range and every b such that $a \leq b \leq 10a$, $P(a,b)$ is increasing with b. Therefore it suffices to prove that $P(a,b) < 1$ when $b = 3a$; that implies (2). Let us first note that

$$P(1,3) = 2k^2 \binom{k}{1}\binom{k}{3}\frac{\binom{21}{7}}{\binom{7k}{7}} < \frac{240k^6}{(k-1)^7},$$

therefore $P(1,3) < 1$ when $k > 250$. We now show that $P(a,3a)/P(a+1,3(a+1)) > 1$ when $0 < a/k < 10^{-5}$ and $k > 10^5$, thereby establishing that (2) holds in that region. Since

$$P(a,3a) = 2k^2 \binom{k}{a}\binom{k}{3a}\frac{\binom{21a}{7a}}{\binom{7k}{7a}}$$

we need to bound quotients of the following forms:

$$\binom{k}{a}\Big/\binom{k}{a+1} = \frac{a+1}{k-1},$$

$$\binom{k}{3a}\Big/\binom{k}{3a+3} = \frac{(3a+3)!(k-3a-3)!}{(3a)!(k-3a)!} > \left(\frac{3a+1}{k-3a}\right)^3,$$

$$\binom{7k}{7a+7}\Big/\binom{7k}{7a} = \frac{(7a)!(k-7a)!}{(7a+7)!(k-7a-7)!} > \left(\frac{k-a-1}{a+1}\right)^7,$$

$$\binom{21a}{7a}\Big/\binom{21a+21}{7a+7} = \frac{(21a)!(7a+7)!(14a+14)!}{(21a+21)!(7a)!(14a)!} > \frac{1}{54^7}.$$

The above bounds imply that when $0 < a \leq \delta k$, where $\delta = 10^{-5}$ and $k > 10^5$,

$$\begin{aligned}
\frac{P(a,3a)}{P(a+1,3(a+1))} &> \frac{a+1}{k-1}\left(\frac{3a+1}{k-3a}\right)^3\left(\frac{a+1}{k-a-1}\right)^7\frac{1}{54^7} \\
&> \frac{(k-\delta k-1)^7}{k^4(\delta k+1)^3 54^7} \\
&= \frac{(1-\delta-1/k)^7}{(\delta+1/k)^3 54^7} \\
&> \frac{1-14\cdot 10^{-5}}{8\cdot 10^{-15}\cdot 54^7} \\
&> \frac{9900}{8\cdot 10^{-10}\cdot 54^7} \\
&> \frac{12\cdot 10^{12}}{54^7} \\
&> 1.
\end{aligned}$$

Case III: $1 - 10^{-5} \le a/k < 1$. Note that $P(a,b) = P(k-b, k-a)$ since

$$\binom{d(k-a)}{d(k-b)} \Big/ \binom{dk}{d(k-b)} = \frac{(dk-da)!}{(dk-db)!(db-da)!} \cdot \frac{(dk-db)!(db)!}{(dk)!}$$
$$= \frac{(dk-da)!(db)!}{(dk)!(db-da)!}$$
$$= \frac{(dk-da)!(da)!}{(dk)!} \cdot \frac{(db)!}{(db-da)!(da)!}$$
$$= \binom{db}{da} \Big/ \binom{dk}{da}.$$

Therefore, (6) in the region $a \ge (1 - 10^{-5})k$ follows directly from Case II.

∎

Proof of Lemma A.1. We use the shorthands $|T_1| = t_1$, $|U_1| = u_1$, $|S_1| = s_1$, $|T_2| = t_2$, $|U_2| = u_2$, and $|S_2| = s_2$. We can assume without loss of generality that $u_1 \le u_2$. This implies that $u_1 < 7k/20$; otherwise $t_2 + s_2 \ge |N(U_1)| > 1 - u_1$ which is equivalent to $u_2 = k - t_2 - s_2 < u_1$, a contradiction. The proof now proceeds by case analysis on t_2 and u_1.

Case I a: $0 < t_2 \le 3k/10$ and $u_1 \le 3k/10$. Then $u_1 + s_1 \ge |N(T_2)| > 2t_2$ and $t_2 + s_2 \ge |N(U_1)| \ge 2u_1$ by (2) and (3)—we use non-strict inequality to also cover the case when $u_1 = 0$. Adding these two inequalities gives the inequality $s_1 + s_2 > t_2 + u_1$ which is equivalent to $2(s_1 + s_2) > u_1 + t_2 + s_1 + s_2$; therefore (1) holds.

Case I b: $0 < t_2 \le k/10$ and $u_1 \ge 3k/10$. Then $t_2 + s_2 \ge |N(U_1)| > 31k/100 + u_1$ by (4) which is equivalent to $s_2 > 31k/100 + u_1 - t_2 > k/2$; therefore (1) holds.

Case I c: $k/10 \le t_2 \le 3k/10$ and $u_1 \ge 3k/10$. Then $u_1 + s_1 \ge |N(T_2)| > 13k/100 + 8t_2/5$ by (3) and $t_2 + s_2 \ge |N(U_1)| > 31k/100 + u_1$ by (4). Adding these inequalities gives the inequality $s_1 + s_2 > 44k/100 + 3t_2/5 \ge k/2$; therefore (1) holds.

Case II a: $3k/10 \le t_2 \le 39k/100$ and $u_1 \le k/10$. Then $u_1 + s_1 \ge |N(T_2)| > 31k/100 + t_2$ by (4) which is equivalent to $s_1 > 31k/100 + t_2 - u_1 > k/2$, therefore (1) holds.

Case II b: $3k/10 \leq t_2 \leq 39k/100$ and $u_1 \geq k/10$. Then $u_1 + s_1 \geq |N(T_2)| > 31k/100 + t_2$ by (4) which is equivalent to $s_1 + s_2 > 31k/100 + t_2 + s_2 - u_1$, and $s_2 + t_2 \geq |N(U_1)| > 13k/100 + 8u_1/5$ by (3). Therefore $s_1 + s_2 > 31k/100 + t_2 + s_2 - u_1 > 44k/100 + 3u_1/5 > k/2$ and (1) holds.

Case III a: $39k/100 \leq t_2 < k$ and $u_1 \leq 3k/10$. Then $u_1 + s_1 \geq |N(T_2)| > k/2 + t_2/2$ by (5) and (6), which is equivalent to $s_1 + s_2 > k/2 + s_2 + t_2/2 - u_1$, and $t_2 + s_2 \geq |N(U_1)| > 2u_1$ by (2) and (3); therefore $s_1 + s_2 > k/2 + s_2 + t_2/2 - u_1 > k/2 + s_2 + t_2/2 - (s_2 + t_2)/2 \geq k/2$ and (1) holds.

Case III b: $39k/100 \leq t_2 \leq 62k/100$ and $u_1 \geq 3k/10$. Then $u_1 + s_1 \geq |N(T_2)| > k/2 + t_2/2$ by (5), which is equivalent to $s_1 + s_2 > k/2 + s_2 + t_2/2 - u_1$, and $t_2 + s_2 \geq |N(U_1)| > 31k/100 + u_1$ by (4); therefore $s_1 + s_2 > k/2 + s_2 + t_2/2 - u_1 > 81k/100 - t_2/2 \geq k/2$ and (1) holds.

Case III c: $62k/100 \leq t_2 < k$ and $u_1 \geq 3k/10$. Then $u_1 + s_1 \geq |N(T_2)| > 85k/100$ by (6), which is equivalent to $s_1 > 85k/100 - u_1 \geq k/2$ where the last inequality follows since $u_1 \leq 7k/20$; therefore (1) holds.

Case IV: $t_2 = 0$. Since $u_1 \leq u_2$ and $t_2 = 0$, $u_1 + t_2 \leq u_2 + t_1$, therefore it suffices to show that $2(s_1 + s_2) > \min\{u_1 + s_1 + s_2, k\}$. But this always holds if $u_1 > 0$ since then $s_2 \geq |N(U_1)| > u_1$. And if $u_1 = 0$, the inequality holds trivially as soon as either s_1 or s_2 are non-zero. Therefore, (1) holds when $t_2 = 0$.

Case V: $t_2 = k$. Since vertices in T_1 are not connected to vertices in T_2, $t_2 = k$ implies that $t_1 = 0$. Moreover, since $u_1 \leq u_2 = 0$, also $u_1 = 0$. Therefore, $s_1 = k$, which implies that (1) holds when $t_2 = k$. ∎

References

1. Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Márió Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.
2. Piotr Berman and Marek Karpinski. On some tighter inapproximability results. In *Proceedings of 26th International Colloquium on Automata, Languages and Programming*, volume 1644 of *Lecture Notes in Computer Science*, pages 200–209, 1999.
3. Markus Bläser and Bodo Siebert. Computing cycle covers without short cycles. In *Proceedings of the 9th Annual European Symposium on Algorithms*, 2001. To appear.
4. Hans-Joachim Böckenhauer, Juraj Hromkovič, Ralf Klasing, Sebastian Seibert, and Walter Unger. An improved lower bound on the approximability of metric TSP and approximation algorithms for the TSP with sharpened triangle inequality. In *Proceedings of the 17th Annual Symposium on Theoretical Aspects of Computer Science*, volume 1770 of *Lectures Notes in Computer Science*, pages 382–391, 2000.
5. Hans-Joachim Böckenhauer and Sebastian Seibert. Improved lower bounds on the approximability of the traveling salesman problem. *RAIRO Theoretical Informatics and Applications*, 34(3):213–255, 2000.
6. Robert D. Carr and Santosh Vempala. Towards a 4/3 approximation for the asymmetric traveling salesman problem. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 116–125, 2000.
7. Nicos Christofides. Worst-case analysis of a new heuristic for the traveling salesman problem. Technical Report CS-93-13, G.S.I.A., Carnegie Mellon University, Pittsburgh, 1976.
8. Lars Engebretsen. An explicit lower bound for TSP with distances one and two. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, volume 1563 of *Lecture Notes in Computer Science*, pages 373–382, 1999.
9. Alan Frieze, Giulia Galbiati, and Francesco Maffioli. On the worst-case performance of some algorithms for the asymmetric traveling salesman problem. *Networks*, 12(1):23–39, 1982.
10. Johan Håstad. Some optimal inapproximability results. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 1–10, 1997. Accepted for publication in *Journal of the ACM*.
11. Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, New York, 1972.
12. Christos H. Papadimitriou and Santosh Vempala. On the approximability of the traveling salesman problem. In *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing*, pages 126–133, 2000.
13. Christos H. Papadimitriou and Santosh Vempala. Personal communication, April 2001.
14. Christos H. Papadimitriou and Mihalis Yannakakis. The traveling salesman problem with distances one and two. *Mathematics of Operations Research*, 18(1):1–11, 1993.
15. Luca Trevisan. When Hamming meets Euclid: The approximability of geometric TSP and MST. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 21–29, 1997.