Labeled Traveling Salesman Problems: Complexity and Approximation *

Basile Couëtoux[1] Laurent Gourvès[1] Jérôme Monnot[1] Orestis A. Telelis[2†]
[1]CNRS FRE 3234 LAMSADE, Université de Paris-Dauphine, France
basile.couetoux@dauphine.fr
{laurent.gourves, monnot}@lamsade.dauphine.fr
[2]PNA1, Centrum Wiskunde & Informatica Amsterdam, The Netherlands
telelis@cwi.nl

Abstract

We consider labeled Traveling Salesman Problems, defined upon a complete graph of n vertices with colored edges. The objective is to find a tour of maximum or minimum number of colors. We derive results regarding hardness of approximation and analyze approximation algorithms, for both versions of the problem. For the maximization version we give a $\frac{1}{2}$-approximation algorithm based on local improvements and show that the problem is **APX**-hard. For the minimization version, we show that it is not approximable within $n^{1-\epsilon}$ for any fixed $\epsilon > 0$. When every color appears in the graph at most r times and r is an increasing function of n, the problem is shown not to be approximable within factor $O(r^{1-\epsilon})$. For fixed constant r we analyze a polynomial-time $(r + H_r)/2$-approximation algorithm, where H_r is the r-th harmonic number, and prove **APX**-hardness for $r = 2$. For all of the analyzed algorithms we exhibit tightness of their analysis by provision of appropriate worst-case instances.

1 Introduction

We study labeled versions of the Traveling Salesman Problem (TSP). The problems are defined upon a complete graph K_n of n vertices, associated to an edge-labeling (or coloring) function $\mathcal{L} : E(K_n) \rightarrow \{c_1, \ldots, c_q\}$. The objective is to find a hamiltonian tour T of K_n optimizing (either maximizing or minimizing) the number of distinct labels used $|\mathcal{L}(T)|$, where $\mathcal{L}(T) = \{\mathcal{L}(e) : e \in T\}$. We refer to the corresponding problems with MaxLTSP and MinLTSP respectively. We also consider the case of an additional input parameter for MinLTSP, that we refer to as *color frequency*. The *color frequency* of a MinLTSP instance is the maximum number of appearances of any color in the graph. For the class of MinLTSP instances with specified color frequency r, we use $\text{MinLTSP}_{(r)}$.

Labeled network optimization over colored graphs has seen extensive study [17, 18, 1, 4, 12, 3, 2, 14, 10, 11, 15]. Minimization of used colors models naturally the need for using links with common properties, whereas the maximization case can be viewed as a maximum covering problem with a certain network structure (in our case such a structure is a hamiltonian cycle). If for example every color represents a technology consulted by a different vendor, then we wish to use as few colors as

*A preliminary version of our results appeared in [5].

†Part of this work was done while the author was a member of The Center for Algorithmic Game Theory, at the Computer Science Department of Aarhus University, Denmark.

possible, so as to diminish incompatibilities among different technologies. For the maximization case, consider the situation of designing a metropolitan peripheral ring road, where every color represents a different suburban area that a certain link would traverse. In order to maximize the number of suburban areas that such a peripheral ring covers, we seek a tour of a maximum number of colors. To the best of our knowledge, the only result known for labeled traveling salesman problems prior to ours is **NP**-hardness, shown by Broersma, Li and Woeginger in [2] for both MaxLTSP and MinLTSP.

1.1 Contribution

We present approximation algorithms and complexity results for MaxLTSP and MinLTSP. For MaxLTSP in particular, we analyze a $\frac{1}{2}$-approximation algorithm, that is based on local improvements and show that the analysis is tight. With respect to complexity we show that MaxLTSP is **APX**-hard, by an appropriate approximation-preserving reduction. This, along with our approximability results yields that the problem is complete for **APX**.

The MinLTSP problem is significantly harder in terms of approximability; we show that, unless $\mathbf{P} = \mathbf{NP}$, it cannot be approximated within a factor strictly less than $n^{1-\epsilon}$ for any fixed $\epsilon > 0$. When the color frequency r is specified as an increasing function of the number of vertices n, the problem is not approximable within a factor less than $O(r^{1-\epsilon})$ for any fixed $\epsilon > 0$; as we discuss later, any feasible tour is trivially r-factor approximate, thus the latter result rules out asymptotically non-trivial approximation factors even when $r = o(n)$. For the case of color frequency $r = 2$ we prove **APX**-completeness. Then we turn our attention to the case of constant color frequency instances and find that a simple greedy algorithm achieves an approximation factor of $\frac{r+H_r}{2}$ in time $O(n^3)$, where $H_r = \sum_{i=1}^{r} \frac{1}{i}$ is the r-th harmonic number. The complexity of the algorithm is however exponentially dependent on r. We illustrate tightness of analysis of the greedy algorithm by a worst-case example.

Organization. The paper is organized as follows. In Section 1.2 we discuss related work with respect to combinatorial optimization problems on colored graphs. Sections 2 and 3 are devoted to the study of MaxLTSP and MinLTSP respectively. We analyze an approximation algorithm for MaxLTSP in Subsection 2.1 and settle the problem's complexity in 2.2. For MinLTSP we study the problem's hardness of approximation in Subsection 3.1. For constant color frequency we analyze a greedy approximation algorithm and prove **APX**-hardness in Subsection 3.2. For the latter greedy algorithm we develop our argument for tightness of its analysis in 3.3.

1.2 Related Work

Multi/Mono-Chromatic Cycles and Paths Erdős, Nešetřil and Rödl [6] first mentioned a problem with respect to the conditions that a complete colored graph needs to satisfy, so as to contain *heterochromatic* Hamilton cycles, that is cycles that do not contain the same color twice. It was shown in [6] that constant color frequency r guarantees existence of such cycles for large graphs. Hahn and Thomassen [9] identified a similar but improved bound for the existence of a heterochromatic Hamilton cycle, namely that $n \geq cr^3$ suffices for some constant c and any color frequency r. This problem was further studied in [7] by Frieze and Reed; the authors showed that, if the edges of a complete graph are colored so that every color appears at most $r = \frac{n}{A \ln n}$ times for some large constant A, then a heterochromatic Hamilton cycle exists. In [2], Broersma, Li and Woeginger study similar problems to this; in particular the authors provide sufficient conditions for the existence of long monochromatic/heterochromatic paths and cycles. Furthermore they prove

NP-hardness of the problem of finding a long path/cycle of a minimum number of colors and provide exponential time exact and heuristic algorithms.

Traveling Salesman The only work that we are aware of dealing with polynomial-time approximation and hardness of Hamilton tours of few or many colors are the works of Punnen [17, 18]. The *TSP under categorization* problem studied in [17, 18] generalizes several traveling salesman problems, and is also a weighted generalization of MinLTSP as well; each edge is associated to a (metric) weight and a color simultaneously, and optimization of the sum of maximum weights of equi-colored edges of the Hamilton tour is sought for. If at most q colors appear in the graph, a $2q$ approximation algorithm is shown. The MinLTSP has also been experimentally investigated in [19] by Xiong, Golden and Wasil.

Labeled Spanning Trees and Paths The recent literature on labeled/colored network optimization problems includes several interesting results from both perspectives of hardness and approximation algorithms. The *Minimum Label Spanning Tree* problem is perhaps the most well explored [4, 12, 3, 10]. Chang and Leu showed that the problem is **NP**-complete in [4], even for complete graphs. The authors presented an (exponential time) exact and two heuristic algorithms. In [12] Krumke and Wirth analyze a greedy approximation algorithm, that achieves $O(\ln n)$ approximation. Bounded color frequency r for the Minimum Label Spanning tree is considered in [3] by Brüggermann, Monnot and Woeginger; the authors show that the problem is polynomial-time solvable for $r = 2$ and **APX**-complete for any fixed $r \geq 3$. They also show that local search can yield a factor of $\frac{r}{2}$ approximation. In [10] Hassin, Monnot and Segev investigate weighted generalizations of labeled minimum spanning tree and shortest paths problems, where each label is also associated with a positive weight and the objective generalizes to minimization of the weighted sum of different labels used. They analyze approximation algorithms and prove inapproximability results for both problems. In particular, they give a H_{n-1} approximation algorithm for the minimum weighted label spanning tree problem and a $H_r - \frac{1}{6}$ approximation algorithm for the case of given color frequency r and unweighted labels. For the minimum weighted label path a factor $O(\sqrt{n})$ approximation algorithm is given. For the case of fixed color frequency $r = O(1)$ the problem is shown to admit constant factor approximation. The minimum weighted label path problem is shown not to admit a polylogarithmic factor approximation unless $\mathbf{P} = \mathbf{NP}$.

Labeled Matchings Labeled perfect matching problems were studied in [14, 15]. In [14] Monnot shows that both the minimum and maximum label perfect matching problem is **APX**-complete even in 2-regular bipartite graphs for any fixed color frequency $r \geq 2$. The maximization version is approximable within a factor of 0.7846 in 2-regular bipartite graphs. **APX**-completeness of the minimization version is shown to persist in the case of complete bipartite graphs for any fixed color frequency $r \geq 6$. The minimization problem is not approximable with $(\frac{1}{2} - \epsilon) \ln n$ for any fixed $\epsilon > 0$, while a simple greedy algorithm achieves $\frac{H_r + r}{2}$ approximation for fixed color frequency r. Maffioli, Rizzi and Benati present results on a labeled matroid problem [13]. Complexity of approximation of bottleneck labeled problems is studied in [11] by Hassin, Monnot and Segev. In such problems each color is associated to a weight and the target is maximization of the minimum or minimization of the maximum weight color used. The authors derive hardness results and approximation algorithms for labeled paths, spanning trees, and perfect matchings.

2 MaxLTSP: Constant factor Approximation

In the following subsections we analyze an approximation algorithm for MaxLTSP, that is based on local improvements and yields $\frac{1}{2}$ approximation. We only comment on an obvious greedy heuristic that achieves $\frac{1}{3}$ approximation; we do not provide its analysis but only a tight example for this heuristic. Subsequently we prove **APX**-hardness of the problem.

2.1 Local Improvements for $\frac{1}{2}$-approximation

The algorithm grows iteratively by local improvements a subset $S \subseteq E$ of edges, that satisfies the following properties:

1. Each label of $\mathcal{L}(S)$ appears exactly once in S.

2. S does not induce vertices of degree three or more, or a cycle of length less than n.

We call the set S a *labeled valid* subset of edges. Finding a labeled valid subset S of *maximum* size is clearly equivalent to MaxLTSP: once it has been found, it can be completed into a feasible Hamilton tour by insertion of appropriately connecting edges, regardless of their label/color. Notice that this augmentation will not increase the objective function. We define two kinds of improvements that the local improvement algorithm performs on the current labeled valid subset S:

- A *1-improvement* of S is a labeled valid subset $S \cup \{e_1\}$, where $e_1 \notin S$.
- A *2-improvement* of S is a labeled valid subset $(S \setminus \{e\}) \cup \{e_1, e_2\}$, where $e \in S$ and $e_1, e_2 \notin S \setminus \{e\}$.

Clearly, a 1- or 2-improvement of S is a labeled valid subset S' such that $|S'| = |S| + 1$. A 1-improvement can be viewed as a particular case of 2-improvement, but we separate the two cases for ease of presentation. The local improvement algorithm - henceforth referred to as LOCIM - initializes $S = \emptyset$ and performs iteratively either a 1- or a 2-improvement on the current S, as long as such an improvement exists. This algorithm works clearly in polynomial-time. We are going to prove the following performance guarantee:

Theorem 1 LOCIM *is a* $1/2$*-approximation algorithm and this ratio is tight.*

We denote by S the solution returned by LOCIM and by S^* an optimal solution, i.e. a *maximum* labeled valid subset of edges. Given $e \in S$, we define $\ell(e)$ to be the edge of S^* with the same label, if such an edge exists. Formally, $\ell : S \to S^* \cup \{\perp\}$ is defined as:

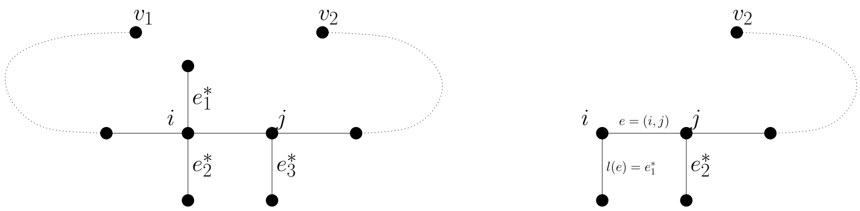
$$\ell(e) = \begin{cases} \perp & \text{if } \mathcal{L}(e) \notin \mathcal{L}(S^*), \\ e^* \in S^* & \text{such that } \mathcal{L}(e^*) = \mathcal{L}(e) \text{ otherwise.} \end{cases}$$

For $e = (i, j) \in S$, let $N(e)$ be the edges of S^* incident to i or j:

$$N(e) = \{(k, l) \in S^* \mid \{k, l\} \cap \{i, j\} \neq \emptyset\}.$$

Define a partition of $N(e)$ into two subsets, $N_1(e)$ and $N_0(e)$, as follows: $e^* \in N_1(e)$ iff $(S \setminus \{e\}) \cup \{e^*\}$ is a labeled valid subset, and $N_0(e) = N(e) \setminus N_1(e)$. In particular, $N_0(e)$ contains the edges $e^* \in S^*$ of $N(e)$ such that $(S \setminus \{e\}) \cup \{e^*\}$ is not labeled valid subset. Finally, for $e^* = (k, l) \in S^*$, let $N^{-1}(e^*)$ be the edges of S incident to k or l:

$$N^{-1}(e^*) = \{(i, j) \in S \mid \{k, l\} \cap \{i, j\} \neq \emptyset\}.$$



(a) $|N_1(e)| \geq 3$

(b) $|N^{-1}(\ell(e))| = 1$, $\ell(e) \in N_1(e)$

Figure 1: Cases studied in proof of Lemma 1.

Property 1 *Let $e = (i, j) \in S$ and $e^* = (i, k) \in N_1(e)$ with $k \neq j$, $e^* \neq \ell(e)$. Either S has two edges incident to i, or $S \cup \{e^*\}$ contains a cycle passing through e and e^*.*

Property 1 holds at the end of the algorithm, because otherwise $S \cup \{e^*\}$ would be a 1-improvement of S.

Property 2 *Let $e = (i, j) \in S$ and $e_1^*, e_2^* \in N_1(e)$. Either both e_1^* and e_2^* are adjacent to i (or to j) or there is a cycle in $S \cup \{e_1^*, e_2^*\}$ passing through e_1^*, e_2^*.*

Recall that in this Property e_1^* and e_2^* have different colors, because they belong to S^*, the maximum valid subset of edges. Property 2 holds at the end of the algorithm since otherwise $(S \setminus \{e\}) \cup \{e_1^*, e_2^*\}$ would be a 2-improvement of S. In order to prove the $\frac{1}{2}$ approximation factor for LOCIM we use charging/discharging arguments based on the following function $g : S \to \mathbb{R}$:

$$g(e) = \begin{cases} |N_0(e)|/4 + |N_1(e)|/2 + 1 - |N^{-1}(\ell(e))|/4 & \text{if } \ell(e) \neq \perp, \\ |N_0(e)|/4 + |N_1(e)|/2 & \text{otherwise.} \end{cases}$$

For simplicity the proof of the 1/2-approximation is cut into two Lemmas.

Lemma 1 *For every edge $e \in S$, $g(e) \leq 2$.*

Proof. Let $e = (i, j)$ be an edge of S. We study two cases, when $e \in S \cap S^*$ and when $e \in S \setminus S^*$. If $e \in S \cap S^*$ then $\ell(e) = e$. Observe that $|N^{-1}(e)| \geq |N_1(e)|$, since otherwise a 1- or 2-improvement would be possible. Since $|N(e)| = |N_0(e)| + |N_1(e)| \leq 4$ we obtain $g(e) \leq (|N_0(e)| + |N_1(e)|)/4 + 1 \leq 2$.

Suppose now that $e \in S \setminus S^*$. Let us first show that $|N_1(e)| \leq 2$. By contradiction, suppose that $\{e_1^*, e_2^*, e_3^*\} \subseteq N_1(e)$ and without loss of generality, assume that e_1^* and e_2^* are incident to i (see Fig. 1a for an illustration). The pairs e_1^*, e_3^* and e_2^*, e_3^* cannot be simultaneously adjacent since otherwise $\{e_1^*, e_2^*, e_3^*\}$ would form a triangle. Then e_1^*, e_3^* is a matching. Property 2 implies that $(S \setminus \{e\}) \cup \{e_1^*, e_3^*\}$ contains a cycle. If P_e is the path containing e in S, this cycle must be $(P_e \setminus \{e\}) \cup \{e_1^*, e_3^*\}$ (see Fig. 1a: $e_1^* = (i, v_2)$ and $e_3^* = (j, v_1)$; note that $e_2^* \neq (i, v_1)$ because $e_2^* \in N_1(e)$). Then $(S \setminus \{e\}) \cup \{e_2^*, e_3^*\}$ would be a 2-improvement of S, a contradiction. Thus $|N_1(e)| \leq 2$. For proving $g(e) \leq 2$ we consider the following cases, and make use of $|N(e)| = |N_0(e)| + |N_1(e)| \leq 4$.

- If $\ell(e) = \perp$ or $|N^{-1}(\ell(e))| \geq 2$, by $|N_1(e)| \leq 2$ we deduce that $g(e) \leq 2$.
- If $\ell(e) \neq \perp$ and $|N^{-1}(\ell(e))| = 1$, then it must be $|N_1(e)| \leq 1$. If not, let $\{e_1^*, e_2^*\} \subseteq N_1(e)$. We have $\ell(e) \neq e_1^*$ and $\ell(e) \neq e_2^*$ since otherwise $(S \setminus \{e\}) \cup \{e_1^*, e_2^*\}$ is a 2-improvement of S, see Fig. 1b for an illustration. In this case, we deduce that $(S \setminus \{e\}) \cup \{\ell(e), e_2^*\}$ or $(S \setminus \{e\}) \cup \{\ell(e), e_1^*\}$ is a 2-improvement of S, a contradiction. Thus $|N_1(e)| \leq 1$ and $g(e) \leq 2$.

- If $\ell(e) \neq \perp$ and $|N^{-1}(\ell(e))| = 0$, then $|N_1(e)| = 0$. Hence, $g(e) \leq 2$.

□

We apply a discharging method to establish a relationship between g and $|S^*|$.

Lemma 2 $\sum_{e \in S} g(e) \geq |S^*|$.

Proof. Let $f : S \times S^* \rightarrow \mathbb{R}$ be defined as:

$$f(e, e^*) = \begin{cases} 1/4 & \text{if } e^* \in N_0(e) \text{ and } \ell(e) \neq e^*, \\ 1/2 & \text{if } e^* \in N_1(e) \text{ and } \ell(e) \neq e^*, \\ 1 - |N^{-1}(e^*)|/4 & \text{if } e^* \notin N(e) \text{ and } \ell(e) = e^*, \\ 5/4 - |N^{-1}(e^*)|/4 & \text{if } e^* \in N_0(e) \text{ and } \ell(e) = e^*, \\ 3/2 - |N^{-1}(e^*)|/4 & \text{if } e^* \in N_1(e) \text{ and } \ell(e) = e^*, \\ 0 & \text{otherwise.} \end{cases}$$

For all $e \in S$ it is $\sum_{\{e^* \in S^*\}} f(e, e^*) = g(e)$. Because of the following:

$$\sum_{e \in S} g(e) = \sum_{e^* \in S^*} \sum_{e \in S} f(e, e^*),$$

it is enough to show that $\sum_{\{e \in S\}} f(e, e^*) \geq 1$ for all $e^* \in S^*$. For an edge $e^* \in S^*$, we study two cases: $\mathcal{L}(e^*) \in \mathcal{L}(S)$ and $\mathcal{L}(e^*) \notin \mathcal{L}(S)$. If $\mathcal{L}(e^*) \in \mathcal{L}(S)$ then there is $e_0 \in S$ such that $\ell(e_0) = e^*$. One of the two following cases occurs:

- $e^* \in N(e_0)$: it is possible that $e_0 = e^*$ if $e^* \in N_1(e_0)$. Then:

$$\sum_{e \in S} f(e, e^*) \geq f(e_0, e^*) + \sum_{e \in N^{-1}(e^*) \setminus \{e_0\}} f(e, e^*) \geq \frac{5}{4} - \frac{|N^{-1}(e^*)|}{4} + \frac{|N^{-1}(e^*)| - 1}{4} = 1.$$

- $e^* \notin N(e_0)$. Then:

$$\sum_{e \in S} f(e, e^*) \geq f(e_0, e^*) + \sum_{e \in N^{-1}(e^*)} f(e, e^*) \geq 1 - \frac{|N^{-1}(e^*)|}{4} + \frac{|N^{-1}(e^*)|}{4} = 1.$$

Now consider $\mathcal{L}(e^*) \notin \mathcal{L}(S)$. Then $|N^{-1}(e^*)| \geq 2$, otherwise $S \cup \{e^*\}$ would be an 1-improvement. We examine the following situations (exactly one of them occurs):

- $N^{-1}(e^*) = \{e_1, e_2\}$: By Property 1 e_1 and e_2 are adjacent, or there is a cycle passing through e^*, e_1 and e_2. In this case $e^* \in N_1(e_1)$ and $e^* \in N_1(e_2)$ (see Fig. 2). Thus:

$$\sum_{\{e \in S\}} f(e, e^*) \geq f(e_1, e^*) + f(e_2, e^*) = \frac{1}{2} + \frac{1}{2} = 1.$$

- $N^{-1}(e^*) = \{e_1, e_2, e_3\}$: Then, $e^* \in N_1(e_1) \cup N_1(e_2)$ where e_1 and e_2 are assumed adjacent. In the worst case e_3 is the ending edge of a path in S containing both e_1 and e_2. Assuming that e_2 is between e_1 and e_3 in this path we obtain $e^* \in N_1(e_2)$. In conclusion, we deduce:

$$\sum_{\{e \in S\}} f(e, e^*) \geq \sum_{i=1}^{3} f(e_i, e^*) \geq \frac{1}{2} + 2 \cdot \frac{1}{4} = 1.$$

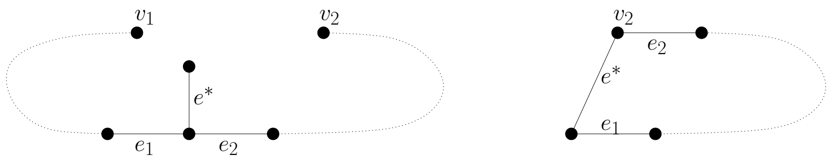


Figure 2: The case where $N^{-1}(e^*) = \{e_1, e_2\}$.

- $N^{-1}(e^*) = \{e_1, e_2, e_3, e_4\}$. Then:

$$\sum_{\{e \in S\}} f(e, e^*) \geq \sum_{i=1}^{4} f(e_i, e^*) \geq 4 \cdot \frac{1}{4} = 1.$$

□

By Lemmas 1 and 2, we have $2|S| \geq \sum_{e \in S} g(e) \geq |S^*|$.

Tightness of Analysis of LOCIM We describe a parameterized instance which shows that the analysis of LOCIM is assymptotically tight. Given an integer $l \geq 2$, the complete graph is composed of $6l - 1$ vertices $\{v_0, \ldots, v_{2l}\} \cup \{v'_1, \ldots, v'_{2l-1}\} \cup \{v''_1, \ldots, v''_{2l-1}\}$. The edges are labeled as follows (see Fig. 3 for an illustration).

- For $i = 1, \ldots, 2l - 2$: $\mathcal{L}(v'_i, v_i) = c_{i+2}$ if i is even, $\mathcal{L}(v'_i, v_i) = c^*_{i+2}$ if i is odd.
- For $i = 1, \ldots, 2l - 2$: $\mathcal{L}(v''_i, v_i) = c_{i+3}$ if i is even, $\mathcal{L}(v''_i, v_i) = c^*_{i+3}$ if i is odd.
- For $i = 0, \ldots, 2l - 1$: $\mathcal{L}(v_i, v_{i+1}) = c_{i+1}$.
- $\mathcal{L}(v'_{2l-1}, v_{2l-1}) = c_1$, $\mathcal{L}(v''_{2l-1}, v_{2l-1}) = c_2$, and the other edges have label c_1.

Let $S = \{(v_i, v_{i+1}) \mid i = 0, \ldots, 2l - 1\}$. We first show that S can be returned by LOCIM. Since adding an edge with label in $\{c^*_1, \ldots, c^*_{2l}\}$ would induce a node with degree 3, no 1-improvement of S is possible. A 2-improvement consists in removing an edge of S and insert two edges with new labels. Suppose that we remove (v_i, v_{i+1}) for some $i \in \{1, \ldots, 2l - 1\}$. Since $\mathcal{L}(v_i, v_{i+1}) = c_i$, we must add two edges with labels in $NEW = \{c^*_1, \ldots, c^*_{2l}\} \cup \{c_i\}$. If i is even (resp. odd) then two edges having their label in NEW are adjacent to v_{i+1} (resp. v_i) whereas the label of the edges adjacent to v_i (resp. v_{i+1}) are already used in S. Thus, no 2-improvement is possible if we remove (v_i, v_{i+1}) where $i \in \{1, \ldots, 2l - 1\}$. If we remove (v_0, v_1) (resp. (v_{2l-1}, v_{2l})) then the label of every edge adjacent to v_0 and v_1 (resp. v_{2l-1} and v_{2l}) are already used in S. Thus, no 2-improvement is possible if we remove one of these edges.

As a consequence, no local improvement is possible and LOCIM can return S. By definition, $|S| = 2l + 1$. Take $S^* = \{(v'_i, v_i) \mid i = 1, \ldots, 2l - 1\} \cup \{(v''_i, v_i) \mid i = 1, \ldots, 2l - 1\}$; then $|S^*| = 4l - 2$, and the approximation ratio tends towards 1/2 when l tends towards $+\infty$.

There are ways to obtain inferior, yet constant, approximation factors for MAXLTSP. An example is the following greedy heuristic: start from an arbitrary vertex $x = v_0$ and grow a prospective hamiltonian path by visiting a neighbor y of x such that edge (x, y) is labeled with a so far unused color, if possible. It is quite straightforward to prove a tight $\frac{1}{3}$ approximation factor for this algorithm. The analysis amounts to comparing the algorithm's output against an optimum tour, by comparing the algorithm's steps against an imaginary optimum algorithm when they both start at the same vertex.

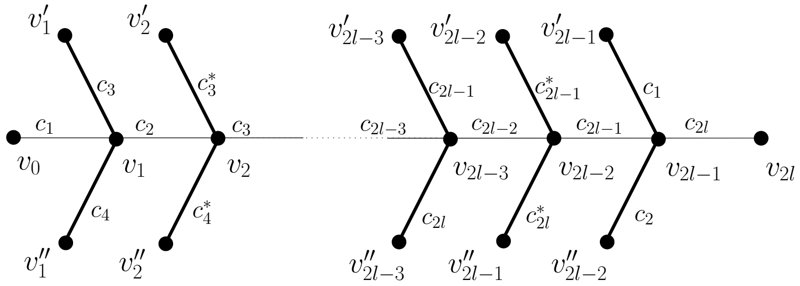


Figure 3: A tight instance for LOCIM.

2.2 Complexity of Approximation

The previous subsections established approximability of MAXLTSP within constant factor. We prove additionally the following result, which entirely establishes the complexity of the problem.

Theorem 2 MAXLTSP *is* **APX**-*hard.*

Proof. We carry out an L-reduction from the *maximum hamiltonian path problem on graphs with distances 1 and 2* (MAXHPP$_{1,2}$), which involves finding the "longest" hamiltonian path of the complete input graph with edge distances 1 and 2. The MAXHPP$_{1,2}$ is easily shown to be **APX**-complete, by a simple reduction from MINHPP$_{1,2}$. The latter is known to be **MaxSNP**-hard and, therefore, **APX**-hard, by the classic work of Papadimitriou and Yannakakis [16]. For an instance of MINHPP$_{1,2}$ one can simply reverse all edge weights of 1 to 2 and of 2 to 1. Then any path in the new instance has weight $w(P) = 3(n-1) - c(P)$, where $c(P)$ is the cost of P in the MINHPP$_{1,2}$ instance. Furthermore, and optimum path in one instance is also optimum in the other. Then it is easy to see that any α-approximation algorithm for MAXHPP$_{1,2}$ translates to a $(3-2\alpha)$-approximation algorithm for MINHPP$_{1,2}$.

Given an instance $I = (G, d)$ with $d : E(G) \to \{1, 2\}$ on n vertices of MAXHPP$_{1,2}$, we construct an instance $I' = (G', \mathcal{L})$ of MAXLTSP as follows. G' is a complete graph with vertex set $V' = V(G) \cup \{v_0\}$ where v_0 is a new node. The labeling function is defined as $\mathcal{L}(e) = c_e$ if $e \in E(G)$ and $d(e) = 2$, and $\mathcal{L}(e) = c_0$ otherwise. Given a feasible solution (hamiltonian path) P to I with total length $d(P) = \sum_{e \in P} d(e)$, we can derive a tour T' for I' using exactly $d(P) - n + 2$ labels, just by linking both endpoints of P to v_0. Thus:

$$|\mathcal{L}(T')| = d(P) - n + 2. \tag{1}$$

Conversely, given a feasible solution (hamiltonian tour) T' to I', using $|\mathcal{L}(T')|$ labels, we can derive a hamiltonian path for I of length $|\mathcal{L}(T')| + n - 2$ by simply removing the two edges incident to v_0. Hence:

$$d(P) = |\mathcal{L}(T')| + n - 2. \tag{2}$$

We denote by OPT the cost of an optimal solution to MAXHPP$_{1,2}$ and by OPT' the number of labels used by an optimal solution to MAXLTSP. It follows from equalities (1) and (2) that $OPT - d(P) = OPT' - |\mathcal{L}(T')|$.

Since every edge incident to v_0 in G' has label c_0, we know that the optimal tour like any other tour uses at most n labels. Hence, $OPT' \leq n$. Since every edge of G has weight 1 or 2, we deduce that the optimum solution to I, like any other hamiltonian cycle, has total weight at least $n-1$. Hence, $OPT \geq n - 1$. In conclusion, $OPT' \leq \frac{3}{2} OPT$ for $n \geq 3$ which concludes the proof. □

Corollary 1 MaxLTSP *is* **APX***-complete.*

3 MinLTSP: Hardness and Approximation

We show that the MinLTSP is generally inapproximable within $O(n^{1-\epsilon})$ for any fixed $\epsilon > 0$, unless $\mathbf{P} = \mathbf{NP}$. Notice that, given a bounded color frequency r, the number of colors appearing in an optimum tour of a MinLTSP$_{(r)}$ instance are at least n/r; hence, any Hamilton tour is at most r-approximate to the optimum. We show that this is asymptotically the best possible factor, when r is an increasing function of n, even if it is of $o(n)$ magnitude; i.e., that MinLTSP$_{(r)}$ is not $O(r^{1-\epsilon})$ approximable for any fixed $\epsilon > 0$. Thus, asymptotically non-trivial approximation factors are ruled out for any non-constant bound on color frequency r. We focus subsequently on fixed color frequency r, and show that a simple greedy algorithm exhibits a tight non-trivial approximation ratio equal to $(r + H_r)/2$, where H_r is the harmonic number of order r. Finally we consider the simple case of $r = 2$, for which the algorithm's approximation ratio becomes $\frac{7}{4}$, and show that MinLTSP$_{(2)}$ is **APX**-hard.

3.1 Hardness of MinLTSP

Without restrictions on color frequency, any algorithm for MinLTSP will trivially achieve an approximation factor of n. We show that this ratio is essentially optimal, unless **P**=**NP**, by reduction from the hamiltonian $s-t$-path problem which is defined as follows: given a graph $G = (V, E)$ with two specified vertices $s, t \in V$, decide whether G has a hamiltonian path from s to t. See [8] (problem [GT39]) for this problem's **NP**-completeness. The restriction of the hamiltonian $s-t$-path problem on graphs where vertices s, t are of degree 1 remains **NP**-complete. In the following let $OPT(\cdot)$ be the optimum solution value to some problem instance.

Theorem 3 *For any fixed $\epsilon > 0$,* MinLTSP *is not $n^{1-\epsilon}$-approximable unless* ***P*=*NP****, where n is the number of vertices.*

Proof. Fix $\epsilon > 0$ and let $I = (G, s, t)$ be an instance of the hamiltonian $s-t$-path problem on a graph $G = (V, E)$ with two specified vertices $s, t \in V$ having degree 1 in G. Let $p = \lceil \frac{1}{\epsilon} \rceil - 1$. We construct the following instance $I' = (G', \mathcal{L})$ of MinLTSP: take a graph G' consisting of n^p copies of G, where the i-th copy is denoted by $G_i = (V_i, E_i)$ and s_i, t_i are the corresponding copies of vertices s, t. Set $\mathcal{L}(e) = c_0$ for every $e \in \cup_{i=1}^{n^p} E_i$, $\mathcal{L}(t_i, s_{i+1}) = c_0$ for all $i = 1, \ldots, n^p - 1$, and $\mathcal{L}(t_{n^p}, s_1) = c_0$. Complete this graph by taking a new color per remaining edge. Because ϵ is a fixed constant, this construction can obviously be done in polynomial time and the resulting graph has n^{p+1} vertices.

If G has a hamiltonian $s-t$-path, then $OPT(I') = 1$. Otherwise, G has no hamiltonian path for any pair of vertices, since vertices $s, t \in V$ have a degree 1 in G. Hence $OPT(I') \geq n^p + 1$, because for each copy G_i either the restriction of an optimal tour T^* (with value $OPT(I')$) in copy G_i is a hamiltonian path, and T^* uses a new color (distinct of c_0) or T^* uses at least two new colors linking G_i to the other copies. Since $|V(K_{n^{p+1}})| = n^{p+1}$, we deduce that it is **NP**-complete to distinguish between $OPT(I') = 1$ and $OPT(I') \geq |V(K_{n^{p+1}})|^{1-\frac{1}{p+1}} + 1 > |V(K_{n^{p+1}})|^{1-\epsilon}$. □

The hamiltonian $s-t$-path problem is also **NP**-complete in graphs of maximum degree 3 (problem [GT39] in [8]). Applying the reduction given in Theorem 3 to this restriction, we deduce that the color frequency r of I' is upper bounded by $(\frac{3n+2}{2})n^p = O(n^{p+1})$. Thus, when r increases with n we obtain:

Corollary 2 *There exists constant $c > 0$ such that for all $\epsilon > 0$,* MinLTSP *is not $cr^{1-\epsilon}$-approximable where r is the color frequency, unless* ***P=NP****.*

Theorem 4 MinLTSP$_{(2)}$ *is* ***APX****-complete.*

Proof. Consider the mininimum cost hamiltonian path problem, on a complete graph with edge costs 1 and 2 (MinHPP$_{1,2}$ - [ND22] in [8]). We are going to prove that a ρ-approximation for MinLTSP$_{(2)}$ can be polynomially transformed into a $(\rho+\epsilon)$-approximation for MinHPP$_{1,2}$, for any fixed $\epsilon > 0$. Since the traveling salesman problem with distances 1 and 2 (MinTSP$_{1,2}$) is **APX**-hard [16], then MinHPP$_{1,2}$ is also **APX**-hard) and we conclude that MinLTSP$_{(2)}$ is **APX**-hard. Moreover, MinLTSP$_{(2)}$ belongs to **APX** because any feasible tour is 2-approximate.

Let I be an instance of MinHPP$_{1,2}$, with $V(K_n) = \{v_1, \dots, v_n\}$, and $d : E(K_n) \to \{1, 2\}$. We construct an instance I' of MinLTSP$_{(2)}$ on K_{2n} as follows. The vertex set of K_{2n} is $V(K_{2n}) = \{v_1, \dots, v_n\} \cup \{v'_1, \dots, v'_n\}$. For every edge $e = (x, y) \in E(K_n)$ with $d(x, y) = 1$ we define two edges $(x, y), (x', y') \in E(K_{2n})$ with the same color $\mathcal{L}((x, y)) = \mathcal{L}(x', y') = c_e$. We complete the coloring of K_{2n} by adding a new color for each of the remaining edges of K_{2n}.

Let P^* be an optimum hamiltonian path (with endpoints s and t) of K_n with cost $OPT(I)$. Build a tour T' of K_{2n} by taking P^*, the edges (s, s'), (t, t') and a copy of P^* on vertices $\{v'_1, \dots, v'_n\}$. Then $|\mathcal{L}(T')| = OPT(I) + 2$, and:

$$OPT(I') \leq OPT(I) + 2. \tag{3}$$

Now let T' be a feasible solution of I'. Assume that n_2 colors appear twice in T' (thus $2n - 2n_2$ colors appear once in T'). In K_n, the set of edges with these colors corresponds to a collection of disjoint paths $P_1, \dots, P_k$ with edges of distance 1. Then, by adding exactly $n - 1 - n_2$ edges we obtain a hamiltonian path P of K_n with cost at most:

$$d(P) \leq |\mathcal{L}(T')| - 2. \tag{4}$$

where $d(P) = \sum_{e \in P} d(e)$. Using inequalities (3) and (4), we deduce $OPT(I') = OPT(I) + 2$. Now, if T is a ρ-approximation for MinLTSP$_{(2)}$, we deduce $d(P) \leq \rho OPT(I) + 2(\rho - 1) \leq (\rho + \epsilon) OPT(I)$ when n is large enough. □

3.2 The Case of Fixed Color Frequency

In this section we improve over the trivial r factor approximation of MinLTSP$_{(r)}$, when the color frequency is upper bounded by a constant. We describe and analyze a greedy approximation algorithm (referred to as *Greedy Tour*) for the MinLTSP$_{(r)}$, for fixed $r = O(1)$. In the description of the algorithm *Greedy Tour* we use the notion of a *valid* subset of edges, which do not induce vertices of degree three or more and do not induce a cycle of length less than n. This definition of a valid subset of edges differs from the one already used in subsection 2.1 in that edges of the same color may appear in the subset. For the analysis and the description of the algorithm we use the notation $\mathcal{L}^{-1}(c)$, where c is a label; this stands for the subset of the graph's edges that are labeled with c.

The algorithm augments iteratively a valid subset of edges by a chosen subset E', until a feasible tour of the input graph is formed. It initializes the set of colors K and iteratively identifies the color that offers the largest valid set of edges with respect to the current (partial) tour T; it adds this set to the tour and eliminates the selected color from the current set of colors. Greedy Tour terminates with a complete tour T, because the input graph is complete and, therefore, every partial tour can be augmented to a complete Hamilton tour. Thus, until T is a feasible solution there will always be valid subset of edges for the algorithm to pick. For constant $r \geq 1$, Greedy Tour is of polynomially bounded complexity proportional to $O(n^3)$; choosing the maximum subset of valid edges for any color within the main loop is of $O(2^r) = O(1)$ time, and there are $O(\frac{n^2}{r}) = O(n^2)$ colors to choose from at any iteration. Clearly a tour is completed within at most $O(n)$ iterations, because at each iteration at least one edge is introduced to T. We introduce some definitions and notations that we use in the analysis of Greedy Tour. Let T^* denote an optimum tour and T be a tour produced by Greedy Tour.

Definition 1 ***(Blocks)*** *For $j = 1, \ldots, r$, the* j-block *with respect to the execution of Greedy Tour is the subset of iterations during which at least j edges of the same color were added to the partial tour, i.e. $|E'| \geq j$. Define T_j to be the subset of edges selected by Greedy Tour during the j-block and $V_j = V(T_j)$ be the set of vertices that are endpoints of edges in T_j.*

Definition 2 ***(Color Degree)*** *For a color $c \in \mathcal{L}(T^*)$ in the optimum tour, define its color degree $f_j(c)$ with respect to V_j (vertex set corresponding to the j-block of Greedy Tour) to be the sum of degrees of vertices in V_j, in the graph $G_c = (V, \mathcal{L}^{-1}(c) \cap T^*)$. That is, $f_j(c) = \sum_{v \in V_j} d_{G_c}(v)$.*

Let us clarify this definition: $f_j(c)$ equals at most 2 times the number of edges of color c, that belong in the optimum tour T^* and each such edge has at least one end-vertex in V_j; i.e. at least one of its end-vertices was picked during an iteration of Greedy Tour that inserted at least j edges of the same color. For each value of color frequency $j \in \{2, \ldots, r\}$ we denote by $\mathcal{L}_j(T^*)$ the set of colors that appear at least j times in T^*: $\mathcal{L}_j(T^*) = \{c \in \mathcal{L}(T^*) : |\mathcal{L}^{-1}(c) \cap T^*| \geq j\}$. T_j may generally contain $k \geq 0$ paths with distinct end-points; in case $k = 0$, T_j is a tour. End-vertices of distinct paths in T_j have degree one in T_j. Let p denote the number of those end-vertices of T_j that are adjacent to two edges of $T^* \setminus T_j$ each, with colors in $\mathcal{L}_j(T^*)$. We denote them by $v_1, \ldots, v_p \in V_j$ and call them *black terminals*. Each such vertex v_i, $i = 1, \ldots, p$, has been "collected" in V_j by Greedy Tour, during some iteration where the algorithm picked at least j edges of the same color; furhtermore there are at most two distinct colors each labeling at least j edges of T^* each, so that v_i is incident to two of these edges (that do not belong in T_j). In this case Greedy Tour has clearly "missed" the optimum structure during the j-block. By $q \geq 0$ we denote the number of path end-vertices of T_j that are adjacent to one edge of $T^* \setminus T_j$ with color in $\mathcal{L}_j(T^*)$. Then it must be $p + q \leq 2k$. Finally, vertices in $V_j \setminus \{v_1, \ldots, v_p\}$ are referred to as *white terminals* and the ones in $V \setminus V_j$ we call *optional* (see Fig. 4 for an illustration).

We consider a partition of $\mathcal{L}_j(T^*)$ into two subsets, $\mathcal{L}^*_{j,in}$ and $\mathcal{L}^*_{j,out}$. A color $c \in \mathcal{L}_j(T^*)$ belongs in $\mathcal{L}^*_{j,out}$ if it labels an edge that does not belong in T_j and is incident to a black terminal of V_j. Then $\mathcal{L}^*_{j,in} = \mathcal{L}_j(T^*) \setminus \mathcal{L}^*_{j,out}$. Notice that $\mathcal{L}_{j,out}$ contains exactly the colors of the optimum tour that *Greedy Tour* missed during the part of its execution that corresponds to the j-block.

Lemma 3 (Color Degree Lemma) *For any $j = 2, \ldots, r$ the following hold:*

(i) If $c \in \mathcal{L}^*_{j,in}$, then $f_j(c) \geq |\mathcal{L}^{-1}(c) \cap T^*| + 1 - j$.

(ii) $$\sum_{c \in \mathcal{L}^*_{j,out}} f_j(c) \geq \sum_{c \in \mathcal{L}^*_{j,out}} (|\mathcal{L}^{-1}(c) \cap T^*| + 1 - j) + p.$$

Proof. (i): Out of the $|\mathcal{L}^{-1}(c) \cap T^*| \geq j$ edges in T^* with color c, at most $j-1$ that are valid with respect to T_j may be missing from T_j (and possibly collected in T_{j-1}): if there are more than $j-1$, then they should have been collected by Greedy Tour in T_j. Then at least $|\mathcal{L}^{-1}(c) \cap T^*| - (j-1)$ edges of color c must have one endpoint in V_j and, by definition of $f_j(c)$, the result follows.

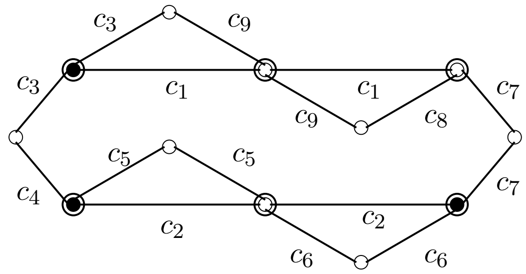


Figure 4: Graphical illustration of definitions: here $r = 2$ and let T_2 consist of the two horizontal paths of length 2 (with colors c_1 and c_2), and the optimum tour consist of the rest of the edges. Imagine all missing edges being labeled by a distinct color each. V_2 contains all the circled nodes; the rest are *optional*. The three black nodes are *black terminals* and the rest of the nodes of V_2 are white terminals.

(ii): First we note that for each color contained in $\mathcal{L}^*_{j,out}$ exactly one of the two edges of $T^* \setminus T_j$ that are incident to a black terminal in V_j belongs to the set of at most $j-1$ c-colored edges, that are valid with respect to T_j and were not collected in T_j. This is simply because the two edges to which the black terminal is incident are invalid with respect to T_j because the black terminal has already degree 1 in T_j. Using the same argument as in statement **(i)**, we have that at least $|\mathcal{L}^{-1}(c) \cap T^*| - (j-1)$ c-colored edges that are incident to at least one vertex of V_j. We can tighten this bound even further, by carefully counting into the *color degree* $f_j(c)$ the contribution of edge belonging to the set of at most $j-1$ valid ones: an edge incident to a black terminal is also incident to either an optional vertex, or another terminal (black or white). Take a black terminal $v_i \in V_j$ with two edges $(x, v_i), (v_i, y) \in T^* \setminus T_j$ incident to it, and consider the cases:

- If x is a white or black terminal: then the color degree must be increased by one, because (x, v_i) can be counted twice in the color degree. The same fact also holds for y, if it is a white or black terminal of V_j.
- If x and y are optional vertices: then the color degree must be increased by at least one, because each edge set $\{(x, v_i)\} \cup T_j$ or $\{(v_i, y)\} \cup T_j$ is valid (and the edge was subtracted from $|\mathcal{L}^{-1}(c) \cap T^*|$ with the at most $j-1$ valid ones). However, if both edges have the same color, the color degree only increases by one since the set $\{(x, v_i), (v_i, y)\} \cup T_j$ is not valid.

Therefore we have an increase of one in the color degree of some colors in $\mathcal{L}^*_{j,out}$ and, in fact, of at least p of them. Thus statement **(ii)** follows. □

The following Lemma is our main tool for proving the approximation ratio of *Greedy Tour*:

Lemma 4 *Let y^*_i and y_i be the number of colors appearing exactly i times in the optimum tour T^* and in solution T returned by Greedy Tour respectively. Then for any value of color frequency $j = 2, \ldots, r$:*

$$\sum_{i=j}^{r} (i+1-j) y^*_i \leq 2 \sum_{i=j}^{r} i \, y_i. \tag{5}$$

Proof. We prove the inequality by upper and lower bounding $F^*_j = \sum_{c \in \mathcal{L}_j(T^*)} f_j(c)$. A lower bound stems from Lemma 3:

$$\begin{aligned} F_j^* = \sum_{c\in\mathcal{L}_j(T^*)} f_j(c) &= \sum_{c\in\mathcal{L}^*_{j,in}} f_j(c) + \sum_{c\in\mathcal{L}^*_{j,out}} f_j(c) \\ &\geq \sum_{c\in\mathcal{L}^*_{j,in}} \big(|\mathcal{L}^{-1}(c)\cap T^*| + 1 - j\big) + \sum_{c\in\mathcal{L}^*_{j,out}} |\mathcal{L}^{-1}(c)\cap T^*| + p \\ &\geq \sum_{i=j}^{r}(i+1-j)y_i^* + p. \end{aligned} \tag{6}$$

Assume now that T_j consists of k disjoint paths. Then $|V_j| = \sum_{i=j}^r iy_i + k$. Furthermore, the number of *internal* vertices on all k paths of T_j is: $\sum_{i=j}^r iy_i - k$. Each internal vertex of V_j may contribute at most twice to F_j^*. Each black terminal of T_j, i.e. each vertex of $\{v_1,\ldots,v_p\}$, contributes exactly twice by definition; we remind the reader that every black terminal of V_j is incident to exactly two edges belonging in $T^*\setminus T_j$ each being labeled with a color that appears at least j times in (labels at least j edges of) T^*. If the number of path end-vertices in T_j that contribute exactly once to F_j^* is q, then $p+q\leq 2k$. Then by giving a contribution of at most two to internal vertices of paths, 2 for black terminals and 1 for the rest of the end-vertices, we obtain:

$$F_j^* \leq 2(\sum_{i=j}^r iy_i - k) + 2p + q \leq 2\sum_{i=j}^r iy_i + p. \tag{7}$$

The result follows by combination of (6) and (7). □

We prove the approximation ratio of Greedy Tour by using the latter Lemma.

Theorem 5 *For any fixed $r\geq 1$, Greedy Tour yields a $\frac{r+H_r}{2}$−approximation for* MinLTSP$_{(r)}$ *and the analysis is tight.*

Proof. As in the previous Lemma, y_i and y_i^* denote the number of colors that label exactly i edges in the solution returned by *Greedy Tour* and in the optimum tour respectively. By summing up inequality (5) with coefficient $\frac{1}{2(j-1)j}$ over color frequencies $j=2,\ldots,r$, we obtain:

$$\sum_{j=2}^r\sum_{i=j}^r \frac{i+1-j}{2j(j-1)}\,y_i^* \leq \sum_{j=2}^r\sum_{i=j}^r \frac{i}{j(j-1)}\,y_i \tag{8}$$

$$\begin{aligned} &= \sum_{i=2}^r i\,y_i \sum_{j=2}^i \frac{1}{j(j-1)} \\ &= \sum_{i=2}^r i\,y_i \sum_{j=2}^i (\frac{1}{j-1} - \frac{1}{j}) \\ &= \sum_{i=2}^r i\,y_i(1-\frac{1}{i}) = \sum_{i=2}^r (i-1)y_i. \end{aligned} \tag{9}$$

For the left-hand part of inequality (8) we obtain:

$$\sum_{j=2}^r\sum_{i=j}^r \frac{i+1-j}{2j(j-1)}\,y_i^* = \sum_{i=2}^r \frac{y_i^*}{2}\sum_{j=2}^i \frac{i+1-j}{j(j-1)}$$

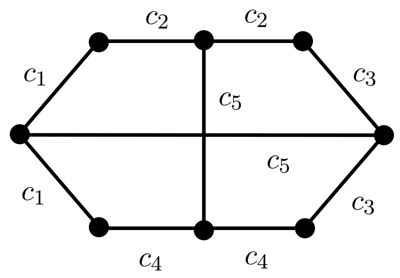


Figure 5: Only colors appearing twice are shown. The rest appear once.

$$= \sum_{i=2}^{r} \frac{y_i^*}{2} \left[\sum_{j=2}^{i} \left(\frac{i-(j-1)}{j-1} - \frac{i-j}{j} \right) - (H_i - 1) \right]$$

$$= \sum_{i=2}^{r} \frac{y_i^*}{2} (i - H_i), \tag{10}$$

where $H_i = \sum_{k=1}^{i} \frac{1}{k}$. By (8, 9) and (10), we obtain: $\sum_{i=2}^{r} \frac{i-H_i}{2} y_i^* \le \sum_{i=2}^{r} (i-1) y_i$, which becomes:

$$2 \sum_{i=1}^{r} (1 - y_i) + 2n \le \sum_{i=1}^{r} (H_i - i) y_i^* + 2n. \tag{11}$$

If APX and OPT denote the number of colors used by Greedy Tour and by the optimum solution respectively, then we use the following:

$$OPT = \sum_{i=1}^{r} y_i^*, \quad APX = \sum_{i=1}^{r} y_i, \quad \text{and} \quad \sum_{i=1}^{r} i y_i = \sum_{i=1}^{r} i y_i^* = n. \tag{12}$$

By (12) we can write $APX = n - \sum_{i=2}^{r} (i-1) y_i$ and then, by (11):

$$2APX \le 2n + \sum_{i=1}^{r} (H_i - i) y_i^* = 2 \sum_{i=1}^{r} i y_i^* + \sum_{i=1}^{r} (H_i - i) y_i^* = \sum_{i=1}^{r} (H_i + i) y_i^*.$$

The result follows by taking $H_i + i \le H_r + r$. Fig. 5 illustrates tightness for $r = 2$. Only colors appearing twice are drawn. The optimal tour uses colors c_1 to c_4, whereas Greedy Tour takes c_5 and completes the tour with 6 new colors appearing once. This yields factor $\frac{7}{4} = \frac{2+H_2}{2}$ approximation. A detailed example for $r \ge 3$ is given in the next subsection. □

3.3 Tightness of Analysis of Greedy Tour

We consider the case of fixed $r \ge 3$. Take a complete graph of $n = 2r(r!)$ vertices where $r! = 1 \cdot 2 \cdot \ldots \cdot r$. We define the following subsets of colors appearing in the graph:

1. **Colors appearing r times:** there are $2(r!) + (r-1)!$ such colors, each denoted by c_i^*, $i = 1, \ldots, 2(r!)$ and $c_{r,i}$, $i = 1, \ldots, (r-1)!$.

2. **Colors appearing j times:** for $j = 2, \ldots, r-1$ and $i = 1, \ldots, \frac{r!}{j}$ let color $c_{j,i}$ appear j times (there are $\frac{r!}{j}$ colors appearing j times).

3. **Colors appearing once:** there are $2(r!)^2 - 3(r!) - (r-1)(r!)$ such colors.

We will exhibit an instance of MinLTSP$_{(r)}$ for fixed $r \geq 3$ in which the optimal tour T^* uses colors c_i^* for $i = 1, \ldots, 2(r!)$ (i.e. exactly $2(r!)$ colors), and the tour constructed by Greedy Tour algorithm uses colors $c_{j,i}$ for $j = 2, \ldots, r$ and $i = 1, \ldots, \frac{r!}{j}$ and exactly $2r(r!) - (r-1)(r!)$ colors appearing once. Then the Greedy Tour solution value will be: $2r(r!) - (r-1)(r!) + \sum_{j=2}^{r} \frac{r!}{j} = 2(r!)(r - \frac{r-1}{2} + \frac{H_r - 1}{2}) = 2(r!)\frac{r+H_r}{2}$, i.e. exactly $(r + H_r)/2$ times the optimum value.

Let us explain how Greedy Tour constructs a feasible tour T, by concurrently deciding how edges of the considered colors are placed on the graph. In the beginning, during the r-block, Greedy Tour includes in T_r edges of colors $c_{r,i}$, $i = 1, \ldots, (r-1)!$ (each of these colors appears exactly r times in the graph). Edges of these colors ($(r-1)! \times r = r!$ in total) are arranged in such a way, that $r! - 1$ paths are formed: $r! - 2$ paths consisting of a single edge each, and one path consisting of 2 edges. We place edges of colors c_i^*, $i = 3, \ldots, 2(r!)$, in such a way that they are incident to vertices of these $r! - 1$ paths. More precisely, for each endpoint of the $r! - 1$ paths two edges with distinct colors c_i^*, c_j^* are incident to the endpoint. One edge of color c_1^* and one of color c_2^* are incident to the *middle vertex* of the length-2 path. Observe that by this construction we cannot take r times any color c_i^* in the r-block.

During the $(r-1)$-block we assume that Greedy Tour takes valid edges of colors $c_{r-1,i}$, $i = 1, \ldots, \frac{r!}{r-1}$, each color appearing $r-1$ times, so that in T_{r-1} the $r! - 1$ paths of T_r are connected into one long path with extreme edges of colors $c_{r-1,i}$. See Fig. 6 and 8 for an illustration.

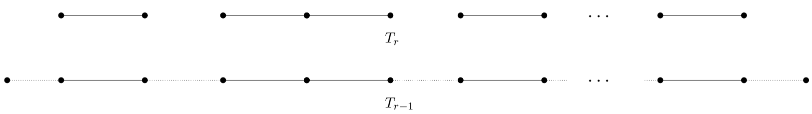


Figure 6: Construction of the r-block T_r and the $(r-1)$-block T_{r-1}.

Finally we let two edges of color c_1^* be incident to one endpoint of the path T_{r-1} and two edges of color c_2^* be incident to the other endpoint of T_{r-1}. Now notice that none of the colors c_i^* can be added $r-1$ times to T_{r-1}. See Fig. 7 for an illustration of how edges of T^* are incident to T_r and T_{r-1}.

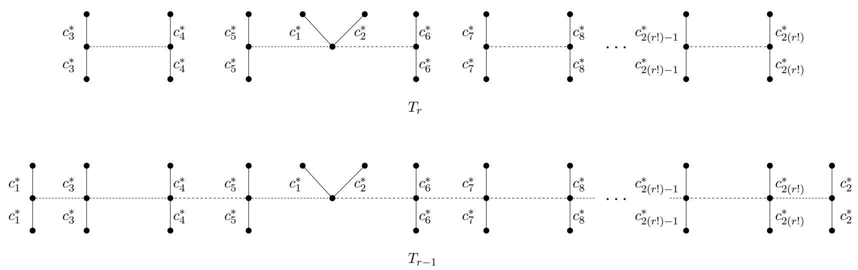


Figure 7: The colors of T^* adjacent to T_r and T_{r-1}.

Example $r = 3$. At this point we can illustrate the value of our construction by considering the case of $r = 3$: the path of T_2 is going to be completed into a tour by insertion of a batch of edges of distinct colors appearing only once. A tour consists of $2 \times 3 \times 3! = 36$ edges, and Greedy Tour has already picked (up to construction of T_2) $12 = 2 \times 3!$ edges for colors $c_{3,i}$ (for $i = 1, 2, 3$) and $c_{2,i}$ (for $i = 1, 2$) and needs to include exactly 24 more edges of distinct colors, while the optimum tour will contain $2 \times (3!) = 12$ colors. Thus it will be $|\mathcal{L}(T)| = 24 + 2 + 3 = 29$, whereas $|\mathcal{L}(T^*)| = 12$ and the ratio is $29/12 = (3 + H_3)/2$.

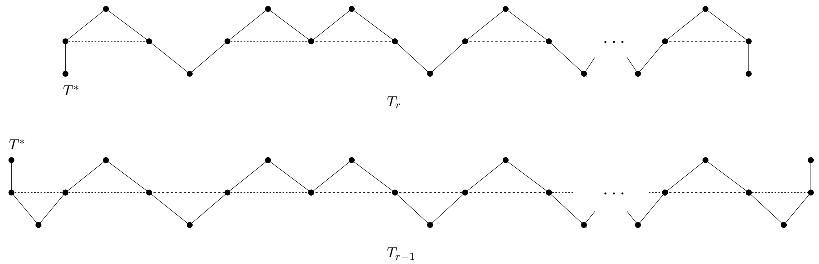


Figure 8: Construction of T^* from T_r and T_{r-1}.

Continuing, during by completion of the $(r-2)$-block Greedy Tour has added iteratively edges of colors $c_{r-2,i}$ by maintaining a path with T_{r-1} in such a way that each color added forms a path of length $r-2$ which is linked to an endpoint (by alternating the endpoints) of the path constructed previously. Thus, T_{r-2} is a path and $T_{r-2} \setminus T_{r-1}$ forms two paths, each using exactly $\frac{r!}{2(r-2)}$ colors of type $c_{r-2,i}$. To each internal vertex of the two paths of $T_{r-2} \setminus T_{r-1}$ the colors among $\{c_5^*, \dots, c_{2(r!)}^*\}$ are added in such a way that each of these $2(r!) - 4$ colors are counted once in total. It is possible because $|T_{r-2} \setminus T_{r-1}| = r!$ and there are 2 paths (so, $r! - 2$ internal vertices). Finally, color c_3^* is added twice to one endpoint of path T_{r-2} whereas color c_4^* is added twice to the other endpoint. Like previously, none of the colors of T^* can be added $r-2$ times.

In general, for each j-block, $j = 2, \dots, r-3$, Greedy Tour proceeds alike. The set $T_j \setminus T_{j+1}$ consists of 2 paths with $|T_j \setminus T_{j+1}| = r!$ edges in total. Edges of T^* with colors in $\{c_1^*, \dots, c_{2(r!)}^*\} \setminus \{c_{2r-2j-3}^*, \dots, c_{2r-2j+1}^*\}$ are made incident to internal vertices of the two paths $T_j \setminus T_{j+1}$, so that one edge per color is incident to $T_j \setminus T_{j+1}$. Two edges of color c_{2r-2j}^* are incident to one endpoint of the path T_j and two edges of color $c_{2r-2j+1}^*$ are incident to its other endpoint. Notice that this is possible because $r \geq 3$. Furthermore, by this pattern, for each path T_j, $j = 2, \dots, r-3$ no color c_i^* can be included j times. This way, Greedy Tour will have used, up to completion of the 2-block, $(r-1)(r!)$ edges for colors $c_{j,i}$ with $j = 2, \dots, r!$ and must use $2r(r!) - (r-1)(r!)$ new edges each having a distinct new color to complete the tour. Thus the value of the constructed tour will be $|\mathcal{L}(T)| = 2r(r!) - (r-1)(r!) + \sum_{j=2}^{r} \frac{r!}{j} = r(r!) + (r!)H_r$ as indicated previously.

In concluding our construction let us describe the structure of the optimal tour T^*. Edges of T^* incident to T_2 can be "patched" in pairs, in order to form a unique path of length $2(r-1)(r!) + 2$ (see Fig. 8 for an illustration of this construction from T_r and T_{r-1}). This path is completed into a tour by addition of $2(r!) - 2$ edges, one for each color in $\{c_1^*, \dots, c_{2(r!)}^*\} \setminus \{c_{2r-3}^*, c_{2r-4}^*\}$ (this is possible because $r \geq 3$). Then, each color in $\{c_1^*, \dots, c_{2(r!)}^*\}$ appears r times in T^* and we have $|\mathcal{L}(T^*)| = 2(r!)$.

References

[1] H. Broersma and X. Li. Spanning Trees with Many or Few Colors in Edge-Colored Graphs. *Discussiones Mathematicae Graph Theory*, 17(2):259–269, 1997.

[2] H. Broersma, X. Li, G. J. Woeginger, and S. Zhang. Paths and Cycles in Colored Graphs. *Australasian Journal on Combinatorics*, 31:299–311, 2005.

[3] T. Brüggemann, J. Monnot, and G. J. Woeginger. Local search for the minimum label spanning tree problem with bounded color classes. *Operations Research Letters*, 31(3):195–201, 2003.

[4] R.-S. Chang and S.-J. Leu. The minimum labeling spanning trees. *Information Processing Letters*, 63(5):277–282, 1997.

[5] B. Couëtoux, L. Gourvès, J. Monnot, and O. Telelis. On Labeled Traveling Salesman Problems. In *Proceedings of the International Symposium on Algorithms and Computation (ISAAC), Springer LNCS 5369*, pages 776–787, 2008.

[6] P. Erdős, J. Nešetřil, and V. Rödl. Some problems related to partitions of edges of a graph. In *Graphs and Other Combinatorial Topics: Proceedings of the 3rd Czechoslovak Symposium on Graph Theory (Teubner-Texte zur Mathematik 59)*, pages 54–63, 1983.

[7] A. M. Frieze and B. A. Reed. Polychromatic Hamilton Cycles. *Discrete Mathematics*, 118:69–74, 1993.

[8] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, New York, 1979.

[9] G. Hahn and C. Thomassen. Path and cycle sub-Ramsey numbers and an edge-coloring conjecture. *Discrete Mathematics*, 62(1):29–33, 1986.

[10] R. Hassin, J. Monnot, and D. Segev. Approximation Algorithms and Hardness Results for Labeled Connectivity Problems. *Journal of Combinatorial Optimization*, 14(4):437–453, 2007.

[11] R. Hassin, J. Monnot, and D. Segev. The Complexity of Bottleneck Labeled Graph Problems. In *Proceedings of the International Workshop on Graph-Theoretic Concepts in Computer Science (WG), Springer LNCS 4769*, pages 328–340, 2007.

[12] S. O. Krumke and H.-C. Wirth. On the Minimum Label Spanning Tree Problem. *Information Processing Letters*, 66(2):81–85, 1998.

[13] F. Maffioli, R. Rizzi, and S. Benati. Least and most colored bases. *Discrete Applied Mathematics*, 155(15):1958–1970, 2007.

[14] J. Monnot. The labeled perfect matching in bipartite graphs. *Information Processing Letters*, 96(3):81–88, 2005.

[15] J. Monnot. A note on the hardness results for the labeled perfect matching problems in bipartite graphs. *RAIRO-Operations Research*, 42(3):315–324, 2008.

[16] C. H. Papadimitriou and M. Yannakakis. The traveling salesman problem with distances one and two. *Mathematics of Operations Research*, 18(1):1–11, 1993.

[17] A. P. Punnen. Traveling Salesman Problem under Categorization. *Operations Research Letters*, 12(2):89–95, 1992.

[18] A. P. Punnen. Erratum: Traveling Salesman Problem under Categorization. *Operations Research Letters*, 14(2):121–121, 1993.

[19] Y. Xiong, B. Golden, and E. Wasil. The Colorful Traveling Salesman Problem. In K. B. Edward, A. Joseph, A. Mehrotra, and M. A. Trick, editors, *Extending the Horizons: Advances in Computing, Optimization, and Decision Technologies*, pages 115–123. Springer, 2006.