# On Labeled Traveling Salesman Problems

Basile Couëtoux[1] Laurent Gourvès[1] Jérôme Monnot[1] Orestis A. Telelis[2*]

[1]CNRS UMR 7024 LAMSADE, Université de Paris-Dauphine, France

basile.couetoux@dauphine.fr

{laurent.gourves, monnot}@lamsade.dauphine.fr

[2]Department of Computer Science, University of Aarhus, Denmark

telelis@daimi.au.dk

**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 a simpler $\frac{1}{3}$-approximation algorithm. We show that the problem is **APX**-hard. For the minimization version, we show that it is not approximable within $n^{1-\epsilon}$ for every $\epsilon > 0$. 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 $\mathcal{L} : E(K_n) \to \{c_1, \ldots, c_q\}$. 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 equi-colored edges or, equivalently, the maximum number of appearances of any color in the graph. For the class of MinLTSP instances with specified color frequency $r$, we use 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 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 in [2] for both MaxLTSP and MinLTSP.

*The Center for Algorithmic Game Theory, funded by the Carlsberg Foundation, Denmark

**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. We also analyze a significantly simpler greedy algorithm that achieves $\frac{1}{3}$ factor approximation. We show that analysis of both algorithms is tight, by devising non-trivial worst-case examples. 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; we show that it cannot be approximated within a factor strictly less than $n^{1-\epsilon}$ for every $\epsilon > 0$. 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 every $\epsilon > 0$. Therefore we turn our attention to the case of constant color frequency instances, and find that a simple greedy algorithm achieves an approximation factor of $\frac{r+H_r}{2}$, where $H_r = \sum_{i=1}^{r} \frac{1}{i}$ is the $r$-th harmonic number. We illustrate tightness of analysis of the greedy algorithm by a far from trivial worst-case example. Finally we prove that MINLTSP$_{(2)}$ is **APX**-complete. We conclude with open problems concerning both versions, minimization and maximization.

The paper is organized as follows. In the next section (2) we discuss related work with respect to combinatorial optimization problems on colored graphs. Sections 3 and 4 are devoted to the study of MAXLTSP and MINLTSP respectively. We analyze approximation algorithms for MAXLTSP in paragraphs 3.1 and 3.2, and settle the problem's complexity in 3.3. For MINLTSP we study the problem's hardness of approximation in 4.1. For constant color frequency we analyze a greedy approximation algorithm and prove **APX**-hardness in 4.2. For the latter greedy algorithm we develop our argument for tightness of its analysis in 4.3. We conclude the paper by mentioning issues that remain open. A preliminary version of our results appeared in [5].

## 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. 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]; 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] similar problems to this are studied; 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 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 [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].

**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]. The problem was shown to be **NP**-complete in [4], even for complete graphs. The authors presented an exact and two heuristic algorithms. In [12] a greedy approximation algorithm is analyzed, that achieves $O(\ln n)$ approximation. Bounded color frequency $r$ for the Minimum Label Spanning tree is considered in [3]; 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] the authors 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] it is shown 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. **APX**-completeness of the minimization version is shown to persist in the case of complete bipartite graphs for any fixed color frequency $r \geq 6$. In absence of information with respect to color frequency the minimization problem is not approximable with $(\frac{1}{2} - \epsilon) \ln n$ for any $\epsilon > 0$, while a simple greedy algorithm achieves $\frac{H_r + r}{2}$ approximation for fixed color frequency $r$. Maffioli *et al.* present results on a labeled matroid problem [13]. Complexity of approximation of bottleneck labeled problems is studied in [11]. 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.

# 3 MaxLTSP: Constant factor Approximation

In the following paragraphs we analyze two approximation algorithms for MaxLTSP. Although the first yields factor $\frac{1}{2}$ approximation and is based on local improvements, the second is simpler and achieves a $\frac{1}{3}$ factor. Subsequently we prove **APX**-hardness of the problem.

## 3.1 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 at most 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. 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 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 $l$.

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

**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 $e_1^*$, $e_2^*$.*

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 $g : S \to \mathbb{R}$:

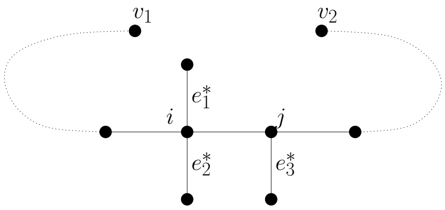
$$g(e) = \left\{ \begin{array}{ll} |N_0(e)|/4 + |N_1(e)|/2 + 1 - |N^{-1}(\ell(e))|/4 & \text{if } \ell(e) \neq \perp \\ |N_0(e)|/4 + |N_1(e)|/2 & \text{otherwise} \end{array} \right\}$$

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

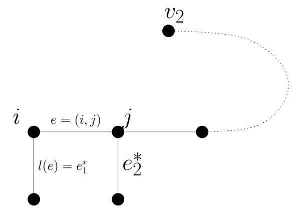
**Lemma 1** *$\forall e \in S$, $g(e) \leq 2$.*

**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



(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.

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, $g(e) \leq 2$.

$\square$

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 $f : S \times S^* \to \mathbb{R}$ 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:

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

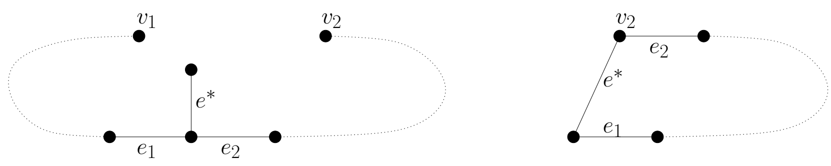



Figure 2: The case where $N^{-1}(e^*) = \{e_1, e_2\}$.

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

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

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

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

□

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

**Proof.** 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 $i = 0, \ldots, 2l-1$: $\mathcal{L}([v_i, v_{i+1}]) = c_{i+1}$.

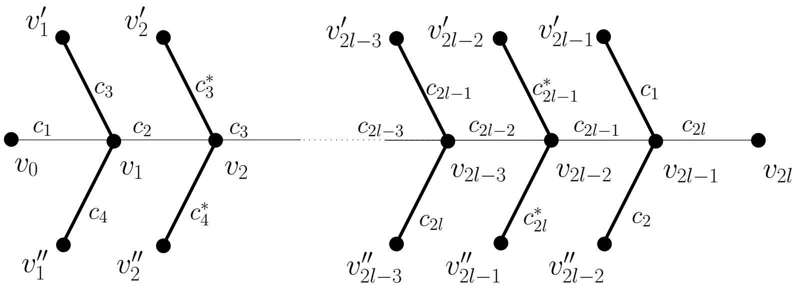



Figure 3: A critical instance for LOCIM.

- $\mathcal{L}([v'_{2l-1}, v_{2l-1}]) = c_1$, $\mathcal{L}([v''_{2l-1}, v_{2l-1}]) = c_2$, and the other edges have label $c_1$.

Let $S = \{[v_i, v_{i+1}] \mid i = 0, \dots, 2l-1\}$ and $S^* = \{[v'_i, v_i] \mid i = 1, \dots, 2l-1\} \cup \{[v''_i, v_i] \mid i = 1, \dots, 2l-1\}$. We first show that $S$ can be returned by LOCIM.

Since adding an edge with label in $\{c_1^*, \dots, c_{2l}^*\}$ 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 $i \in \{1, \dots, 2l-1\}$. Since $\mathcal{L}([v_i, v_{i+1}]) = c_i$, we must add two edges with labels in $NEW = \{c_1^*, \dots, c_{2l}^*\} \cup \{c_i\}$. 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 $i \in \{1, \dots, 2l-1\}$. 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 $S$. Since $|S| = 2l+1$ and $|S^*| = 4l-2$, the approximation ratio tends towards $1/2$ when $l$ tends towards $+\infty$.

### 3.2 Greedy $\frac{1}{3}$-approximation

In this paragraph we analyze a greedy heuristic that is simpler and faster than algorithm LOCIM, with only moderately worse approximation performance. We refer to it as *Labeled Neighbor* (LN). Starting from an arbitrary vertex $x = v_0$, LN grows 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. If no such edge is incident to $x$, LN selects $y$ arbitrarily. Once a hamiltonian path is constructed, its endpoints are linked to yield a tour (see Algorithm 1 for a formal description). When the partial solution is a path of length $\ell$, growing it requires $n - \ell - 1$ operations. Thus LN runs in time $\mathcal{O}(n^2)$.

**Theorem 2** LN *is* $1/3$*-approximate and the result is asymptotically tight.*

**Proof.** Let $T$ be the tour returned byLN and $T^*$ be an optimum tour. Let $C$ and $C^*$ denote the set of colors used by $T$ and $T^*$ respectively. For every vertex $x$ we define $succ(x)$ to be the vertex visited by LN after $x$. We choose a direction for $T^*$ and define accordingly $succ_*(x)$. By $succ^i$ and $succ_*^i$ we denote composition of $succ/succ_*$ $i$ times. By construction, $T$ will consist of:

$$T = \{[v_0, succ(v_0)], \dots, [succ^{n-2}(v_0), succ^{n-1}(v_0)], [succ^{n-1}(v_0), succ^n(v_0)]\}$$

We assume that $v_0 = succ^0(v_0)$ and $v_0 = succ^n(v_0)$. Similarly, the following $n$ edges are contained in $T^*$:

$$T^* = \{[v_0, succ_*(v_0)], \dots, [succ_*^{n-2}(v_0), succ_*^{n-1}(v_0)], [succ_*^{n-1}(v_0), succ_*^n(v_0)]\}$$

**Algorithm 1**: LN

```
Input: A graph G = (V, E) and a labelling function ℒ
Output: A tour T
choose v0 ∈ V arbitrarily;
set p ← v0; T ← ∅; K ← ∅, VISITED ← {v0} ;
while VISITED ≠ V do
    if ∃v ∈ V \ VISITED such that ℒ([p, v]) ∉ K then
        K ← K ∪ {ℒ([p, v])} ;
    else
        choose v ∈ V \ VISITED arbitrarily ;
    end
    T ← T ∪ {[p, v]}; VISITED ← VISITED ∪ {v}; p ← v;
end
return T ∪ {[p, v0]} ;
```

For each $c \in C$ define $d(c) = \min\{i \mid \mathcal{L}([succ^i(v_0), succ^{i+1}(v_0)]) = c\}$, to be the position of the current vertex when LN "discovered" color $c$ for the first time. Accordingly for $c \in C^*$ define $d^*(c) = \min\{i \mid \mathcal{L}([succ_*^i(v_0), succ_*^{i+1}(v_0)]) = c\}$. Let $V' = \{succ^{d(c)}(v_0) \mid c \in C\}$ and $V^* = \{succ^{d^*(c)}(v_0) \mid c \in C^*\}$. By definition, we have $|C| = |V'|$ (resp., $|C^*| = |V^*|$) since each color in $C$ (resp., $C^*$) has its corresponding vertex.

Let $V_1^* = \{v \in V^* \backslash V' \mid \mathcal{L}(v, succ_*(v)) \in C\}$ and $V_2^* = \{v \in V^* \backslash V' \mid \mathcal{L}(v, succ_*(v)) \notin C\}$. We have $|V_1^*| \leq |V'|$ because $|V_1^*| \leq |C| = |V'|$. We also have $|V_2^*| \leq |V'|$ because $v \in V_2^* \Rightarrow succ_*(v) \in V'$. Indeed, if the algorithm inserts $v \in V_2^*$ before $succ_*(v)$ without taking the edge $[v, succ_*(v)]$ (because $\mathcal{L}([v, succ_*(v)]) \in C^* \setminus C$), this means that color $\mathcal{L}([v, succ(v)])$ was new when $[v, succ(v)]$ was added. Then, $v \in V'$ and $v \notin V_2^*$, contradiction. If the algorithm inserts $v \in V_2^*$ after $succ_*(v)$ without taking the edge $[v, succ_*(v)]$, this means that color $\mathcal{L}([succ_*(v), succ(succ_*(v))])$ was new when $[succ_*(v), succ(succ_*(v))]$ was added. Then $succ_*(v) \in V'$. By adding inequalities $|V_1^*| \leq |V'|$, $|V_2^*| \leq |V'|$ and $|V^* \cap V'| \leq |V'|$, we obtain:

$$|V_1^*| + |V_2^*| + |V^* \cap V'| \leq 3|V'| \tag{1}$$

Since $V_1^*$, $V_2^*$ and $V^* \cap V'$ form a partition of $V^*$, (1) becomes $|V^*| \leq 3|V'|$. We replace $|V^*|$ by $|C^*|$ (resp., $|V'|$ by $|C|$) and the result follows. □

**Tightness of Analysis** We turn to the (asymptotic) tightness of the analysis by considering a family of instances depending on a parameter $i$ (an integer). For a fixed $i$, the graph has $3i$ vertices $\{v_1, \ldots, v_i\} \cup \{v'_1, \ldots, v'_i\} \cup \{v''_1, \ldots, v''_i\}$ and uses $3i$ labels $\{c_1, \ldots, c_i\} \cup \{c'_1, \ldots, c'_i\} \cup \{c''_1, \ldots, c''_i\}$. For $i = 2, \ldots, i$ we set $\mathcal{L}([v_{i-1}, v_i]) = c_i$. For $i = 1, \ldots, i$ we set $\mathcal{L}([v_i, v'_i]) = c'_i$ and $\mathcal{L}([v_i, v''_i]) = c''_i$. Every edge whose label is not defined above has label $c_1$. See Figure 4 for an illustration of the graph. The tour $T = v_1\, v_2\, v_3\, \ldots\, v_i\, v'_i\, v'_{i-1} v'_{i-2}\, \ldots\, v'_1\, v''_i\, v''_{i-1}\, \ldots\, v''_1$ uses $i+2$ labels $\{c_1, c_2, \ldots, c_i\} \cup \{c''_1, c'_i\}$. It is not difficult to see that $T$ can be returned by LN. The (optimal) tour $T^* = v'_1\, v_1\, v''_1\, v'_2\, v_2\, v''_2\, \ldots\, v'_i\, v_i\, v''_i$ uses all $3i$ colors $\{c_1, c_2, \ldots, c_i\} \cup \{c'_1, c'_2, \ldots, c'_i\} \cup \{c''_1, c''_2, \ldots, c''_i\}$. Because $(i+2)/(3i)$ tends to $1/3$ as $i$ tends to $\infty$, the result follows.

### 3.3 Complexity of Approximation

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

**Theorem 3** 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, and is known to be **APX**-complete.

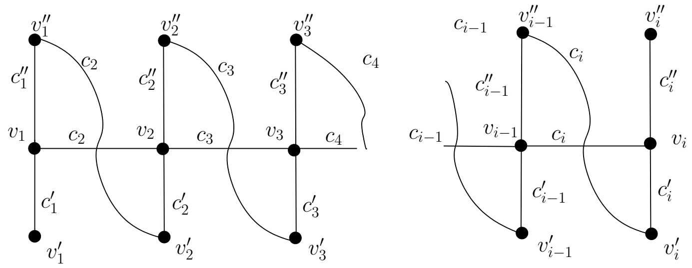



Figure 4: Asymptotically tight instance for MaxLTSP. Undrawn edges have label $c_1$.

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:

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

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:

$$d(P) = |\mathcal{L}(T')| + n - 2 \tag{3}$$

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 (2) and (3) 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. □

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

# 4 MinLTSP: Hardness and Approximation

We show that the MinLTSP is generally inapproximable, unless $\mathbf{P} = \mathbf{NP}$: MinLTSP$_{(r)}$ where $r$ is any increasing function of $n$ is not $r^{1-\epsilon}$ approximable for any $\epsilon > 0$. 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.

## 4.1 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 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 4** *For all $\varepsilon > 0$,* MinLTSP *is not $n^{1-\varepsilon}$-approximable unless **P=NP**, where $n$ is the number of vertices.*

**Proof.** Let $\varepsilon > 0$ 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 $p = \lceil \frac{1}{\varepsilon} \rceil - 1$. 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 $i = 1, \dots, n^p - 1$, and $\mathcal{L}([t_{n^p}, s_1]) = c_0$. Complete this graph by taking a new color per remaining edge. 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 $OPT(I') \geq |V(K_{n^{p+1}})|^{1-\frac{1}{p+1}} + 1 > |V(K_{n^{p+1}})|^{1-\varepsilon}$. □

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 4 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:

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

## 4.2 The Case of Fixed Color Frequency

We describe and analyze a greedy approximation algorithm (referred to as *Greedy Tour* - algorithm 2) 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 also do not induce a cycle of length less than $n$. 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. We remind the reader that *validity* of an edge set excludes vertices of induced degree more than 2 and cycles of length less than $n$. For constant $r \geq 1$ Greedy Tour is of polynomially bounded complexity proportional to $O(n^{r+1})$. 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 $j = 1, \dots, r$, the* $j$-block *with respect to the execution of Greedy Tour is the subset of iterations during which it was $|E'| \geq j$. Let $T_j$ 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 $c \in \mathcal{L}(T^*)$ define its color degree $f_j(c)$ in $V_j$ to be $f_j(c) = \sum_{v \in V_j} d_{G_c}(v)$, where $G_c = (V, \mathcal{L}^{-1}(c) \cap T^*)$ and $d_{G_c}(v)$ is the degree of $v$ in graph $G_c$.*

**Algorithm 2**: Greedy Tour

> Let $T \leftarrow \emptyset$;
> Let $K \leftarrow \{c_1, \dots, c_q\}$;
> **while** $T$ *is not a tour* **do**
>     Consider $c_j \in K$ maximizing $|E'|$ such that $E' \subseteq \mathcal{L}^{-1}(c_j)$ and $T \cup E'$ is valid;
>     $T \leftarrow T \cup E'$;
>     $K \leftarrow K \setminus \{c_j\}$;
> **end**
> return $T$ ;

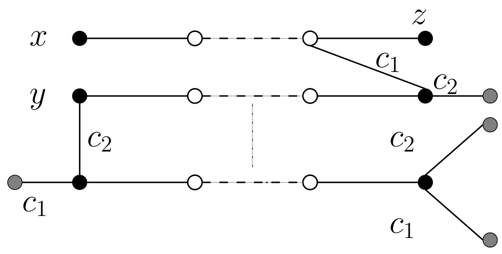



Figure 5: Graphical illustration of definitions: if $c_1, c_2 \in \mathcal{L}_j(T^*)$, apart from vertices $x, y, z$, the remaining endpoints of paths are *black terminals*. Inner vertices are *white terminals* (drawn white), while vertices outside the paths are *optional vertices*.

For $j \in \{2, \dots, r\}$ let $\mathcal{L}_j(T^*)$ be the set of colors that appear at least $j$ times in $T^*$: $\mathcal{L}_j(T^*) = \{c \in \mathcal{L}(T^*) : |\mathcal{L}^{-1}(c) \cap T^*| \geq j\}$. In general $T_j$ contains $k \geq 0$ paths (in case $k = 0$, $T_j$ is a tour). We consider $p$ vertices $\{v_1, \dots, v_p\} \subseteq V_j$ of degree 1 in $T_j$ (i.e. they are endpoints of paths), such that each such vertex is adjacent to two edges of $T^*$ that have colors in $\mathcal{L}_j(T^*)$. We refer to vertices of $\{v_1, \dots v_p\}$ as *black terminals*. We refer to vertices in $V_j \setminus \{v_1, \dots, v_p\}$ as *white terminals* and to vertices in $V \setminus V_j$ as *optional* (see Fig. 5 for an illustration). We also assume the existence of $q \geq 0$ path endpoints of $T_j$ adjacent to one edge of $T^*$ with color in $\mathcal{L}_j(T^*)$. Clearly $p + q \leq 2k$.

We consider a partition of $\mathcal{L}_j(T^*)$: $\mathcal{L}^*_{j,in}$ and $\mathcal{L}^*_{j,out}$. A color $c \in \mathcal{L}_j(T^*)$ belongs in $\mathcal{L}^*_{j,out}$ if there is an edge with this color incident to a black terminal of $V_j$. Then $\mathcal{L}^*_{j,in} = \mathcal{L}_j(T^*) \setminus \mathcal{L}^*_{j,out}$.

**Lemma 3 (Color Degree Lemma)** *For any $j = 2, \dots, r$ 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):** Except of the $|\mathcal{L}^{-1}(c) \cap T^*| \geq j$ edges of color $c$ in $T^*$, at most $j - 1$ valid ones (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 the result follows.

**(ii):** First we note an important fact for each color $c \in \mathcal{L}^*_{j,out}$: exactly one of the two edges incident to a black terminal (suppose one with color $c$) belongs to the set of at most $j - 1$ valid $c$-colored edges, that were not collected 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 $V_j$.

The fact that we mentioned can help us tighten this bound even further, by counting to the color degree the contribution of one 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 a terminal (black or white). Take one black terminal $v_i$ of the two edges $[x, v_i]$, $[v_i, y]$ of $T^*$ 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 this edge can be counted twice in the color degree. The same fact also holds for $y$.
- 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 was subtracted from $|\mathcal{L}^{-1}(c) \cap T^*|$ with the at most $j-1$ valid ones). However, if the both edges have the same color, the color degree only increases by one unit 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 $p$ of them at least. Thus statement **(ii)** follows. □

Let $y_i^*$ and $y_i$ be the number of colors appearing exactly $i$ times in $T^*$ and $T$ respectively. Then we show that:

**Lemma 4** *For $j = 2, \ldots, r$:* $\sum_{i=j}^{r} (i + 1 - j) y_i^* \leq \sum_{i=j}^{r} 2i\, y_i$.

**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:

$$F_j^* \geq \sum_{i=j}^{r} (i + 1 - j) y_i^* + p \tag{4}$$

Assume now that $T_j$ consists of $k$ disjoint paths. Then $|V_j| = \sum_{i=j}^{r} i y_i + k$ and the number of internal vertices on all $k$ paths of $T_j$ is: $\sum_{i=j}^{r} i y_i - k$. Each internal vertex of $V_j$ may contribute at most twice to $F_j^*$. Furthermore, each black terminal of $T_j$, i.e. each vertex of $\{v_1, \ldots, v_p\}$, also contributes twice by definition. Assume that there are $q$ endpoints of paths in $T_j$, each contributing once to $F_j^*$. Clearly $p + q \leq 2k$. Then:

$$F_j^* \leq 2(\sum_{i=j}^{r} i y_i - k) + 2p + q \leq \sum_{i=j}^{r} i 2 y_i + p \tag{5}$$

The result follows by combination of (4) and (5). □

We prove the approximation ratio of Greedy Tour by using Lemma 4:

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

**Proof.** By summing up inequality of Lemma 4 with coefficient $\frac{1}{2(j-1)j}$ for $j = 2, \ldots, r$, we obtain:

$$\sum_{j=2}^{r} \sum_{i=j}^{r} \frac{i + 1 - j}{2j(j-1)} y_i^* \leq \sum_{j=2}^{r} \sum_{i=j}^{r} \frac{i}{j(j-1)} y_i \tag{6}$$

For the right-hand part of inequality (6) we have:

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

For the left-hand part of inequality (6) we obtain:

$$\sum_{j=2}^{r}\sum_{i=j}^{r}\frac{i+1-j}{2j(j-1)}\,y_i^* = \sum_{i=2}^{r}\frac{y_i^*}{2}\sum_{j=2}^{i}\frac{i+1-j}{j(j-1)} \tag{7}$$

But we also have:

$$\sum_{j=2}^{i}\frac{i+1-j}{j(j-1)} = \sum_{j=2}^{i}(\frac{i-(j-1)}{j-1}-\frac{i-j}{j}) - (H_i-1) = i - H_i \tag{8}$$

where $H_i = \sum_{k=1}^{i}\frac{1}{k}$. Therefore relation (7) becomes by (8):

$$\sum_{j=2}^{r}\sum_{i=j}^{r}\frac{i+1-j}{2j(j-1)}\,y_i^* = \sum_{i=2}^{r}\frac{i-H_i}{2}y_i^* \tag{9}$$

By plugging the right-hand equality and (9) into inequality (6), we obtain:

$$\sum_{i=2}^{r}\frac{i-H_i}{2}y_i^* \le \sum_{i=2}^{r}(i-1)y_i \tag{10}$$

Denote by $APX$ and $OPT$ the number of colors used by Greedy Tour and by the optimum solution respectively. Then:

$$OPT = \sum_{i=1}^{r} y_i^*,\;\; APX = \sum_{i=1}^{r} y_i,\;\; \text{and}\;\; \sum_{i=1}^{r} iy_i = \sum_{i=1}^{r} iy_i^* = n \tag{11}$$

where $n = |T| = |T^*|$ is the number of vertices of the graph. By (11) we can write $APX = n - \sum_{i=2}^{r}(i-1)y_i$, and using inequality (10), we deduce:

$$APX \le \sum_{i=1}^{r} iy_i^* - \sum_{i=2}^{r}\frac{i-H_i}{2}y_i^* = \sum_{i=1}^{r}\frac{i+H_i}{2}y_i^*$$

Finally, since $i + H_i \le r + H_r$ when $i \le r$, we obtain:

$$APX \le \frac{r+H_r}{2}\sum_{i=1}^{r} y_i^* = \frac{r+H_r}{2}OPT$$

Fig. 6 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 paragraph. □

We show next that MinLTSP$_{(2)}$ proves as hard to approximate as the min-cost hamiltonian path on a complete graph with edge costs 1 and 2 (MinHPP$_{1,2}$ - [ND22] in [8]).

**Theorem 6** *A $\rho$-approximation for* MinLTSP$_{(2)}$ *can be polynomially transformed into a $(\rho+\varepsilon)$-approximation for* MinHPP$_{1,2}$*, for all $\varepsilon > 0$.*

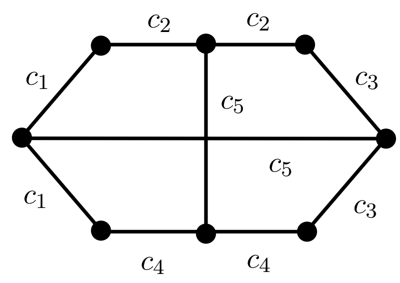



Figure 6: Only colors appearing twice are shown. The rest appear once.

**Proof.** Let $I$ be an instance of MINHPP$_{1,2}$, with $V(K_n) = \{v_1, \ldots, v_n\}$, and $d : E(K_n) \rightarrow \{1, 2\}$. We construct an instance $I'$ of MINLTSP$_{(2)}$ on $K_{2n}$ as follows. The vertex set of $K_{2n}$ is $V(K_{2n}) = \{v_1, \ldots, v_n\} \cup \{v'_1, \ldots, v'_n\}$. 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 rest of the edges $K_{2n}$.

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 $[x, x']$, $[y, y']$ and a copy of $P^*$ on vertices $\{v'_1, \ldots, v'_n\}$. Then $|\mathcal{L}(T')| = OPT(I) + 2$, and:

$$OPT(I') \leq OPT(I) + 2 \tag{12}$$

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 $P_1, \ldots, P_k$ 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:

$$d(P) \leq |\mathcal{L}(T')| - 2 \tag{13}$$

where $d(P) = \sum_{e \in P} d(e)$. Using inequalities (12) and (13), we deduce $OPT(I') = OPT(I) + 2$. Now, if $T$ is a $\rho$-approximation for MINLTSP$_{(2)}$, we deduce $d(P) \leq \rho OPT(I) + 2(\rho - 1) \leq (\rho + \varepsilon) OPT(I)$ when $n$ is large enough. □

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), we conclude by Theorem 6 that MINLTSP$_{(2)}$ is **APX**-hard. Moreover, MINLTSP$_{(2)}$ belongs to **APX** because any feasible tour is 2-approximate.

**Corrolary 3** MINLTSP$_{(2)}$ *is* ***APX****-complete.*

## 4.3 Tightness of Analysis of Greedy Tour

We consider the case of fixed $r \geq 3$. 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. 7 and 9 for an illustration.

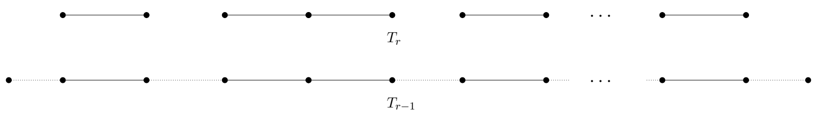



Figure 7: Construction of the $r$-block $T_r$ and the $(r-1)$-block $T_{r-1}$.

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. 8 for an illustration of how edges of $T^*$ are incident to $T_r$ and $T_{r-1}$.

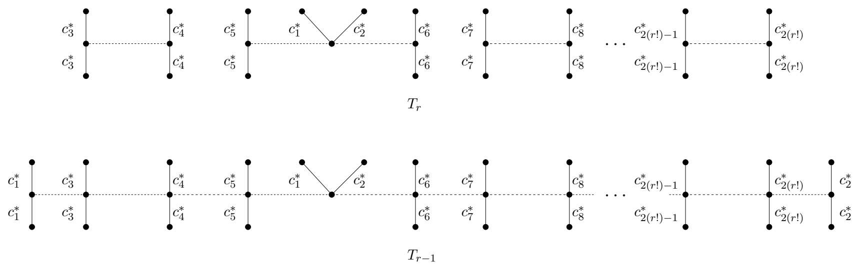



Figure 8: The colors of $T^*$ adjacent to $T_r$ and $T_{r-1}$.

**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 $29/12 = (3 + H_3)/2$.

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 $T_{r-2} \setminus T_{r-1}$

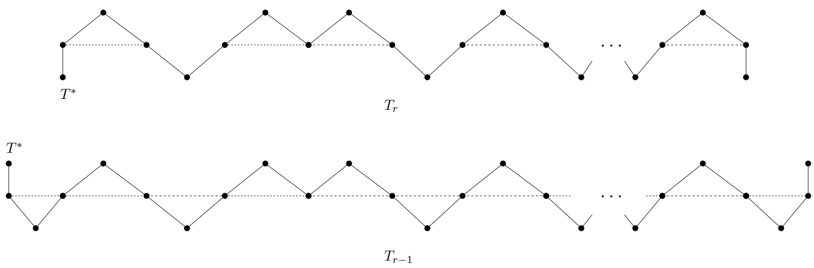



Figure 9: Construction of $T^*$ from $T_r$ and $T_{r-1}$.

the colors among $\{c_5^*, \ldots, c_{2(r!)}^*\}$ 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, $j = 2, \ldots, r-3$, 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 $\{c_1^*, \ldots, c_{2(r!)}^*\} \setminus \{c_{2r-2j-3}^*, \ldots, c_{2r-2j+1}^*\}$ 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 $T_j$, $j = 2, \ldots, r-3$ 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 $j = 2, \ldots, r!$ 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. 9 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 $\{c_1^*, \ldots, c_{2(r!)}^*\} \setminus \{c_{2r-3}^*, c_{2r-4}^*\}$ (this is possible because $r \geq 3$). Then, each color in $\{c_1^*, \ldots, c_{2(r!)}^*\}$ appears $r$ times in $T^*$ and we have $|\mathcal{L}(T^*)| = 2(r!)$.

# 5 Open Questions

Is there a better approximation algorithm for MinLTSP$_{(r)}$, when $r$ is a fixed small constant (e.g. $r = 2$)? For MaxLTSP, using $k-$improvements for fixed $k \geq 3$ could yield better performance but analysis appears quite non-trivial. It is also interesting to study MaxLTSP$_{(r)}$ with bounded color frequency $r$.

# 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*, pages 54–63. Teubner-Texte zur Mathematik 59, 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, 1993.

[19] Y. Xiong, B. Golden, and E. Wasil. The Colorful Traveling Salesman Problem. In *Extending the Horizons: Advances in Computing, Optimization, and Decision Technologies*, pages 115–123. Springer, 2007.