On Labeled Traveling Salesman Problems

Basile Couëtoux[1], Laurent Gourvès[1], Jérôme Monnot[1], and 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 and show that it 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 not $O(r^{1-\epsilon})$-approximable. For fixed constant r we analyze a polynomial-time $(r+H_r)/2$-approximation algorithm (H_r is the r-th harmonic number), and prove **APX**-hardness. Analysis of the studied algorithms is shown to be tight.

1 Introduction

We consider labeled versions of the Traveling Salesman Problem (TSP), defined upon a complete graph K_n of n vertices along with 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) $|\mathcal{L}(T)|$, where $\mathcal{L}(T) = \{\mathcal{L}(e) :\ e \in T\}$. We refer to the corresponding problems with MaxLTSP and MinLTSP respectively. The *color frequency* of a MinLTSP instance is the maximum number of equi-colored edges. We use MinLTSP$_{(r)}$ to refer to the class of MinLTSP instances with fixed color frequency r.

Labeled network optimization over colored graphs has seen extended study [17, 18, 3, 5, 12, 2, 4, 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

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

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. It was shown in [4] that both MaxLTSP and MinLTSP are **NP**-hard.

Contribution We present approximation algorithms and hardness results for MaxLTSP and MinLTSP. In section 2 we provide a $\frac{1}{2}$-approximation local improvement algorithm for the MaxLTSP problem and show that the problem is **APX**-hard. In section 3 we show that MinLTSP is not approximable within a factor $n^{1-\epsilon}$ for every $\epsilon > 0$ or within a factor $O(r^{1-\epsilon})$ when color frequency r is an increasing function of n (paragraph 3.1). For the case of *fixed constant* r, we analyze a simple greedy algorithm with approximation ratio $(r + H_r)/2$, where $H_r = \sum_{i=1}^{r} \frac{1}{i}$ is the r-th harmonic number (paragraph 3.2). For $r = 2$ MinLTSP$_{(2)}$ is shown to be **APX**-hard. We conclude with open problems.

Related Work Identification of conditions for the existence of single-colored or multi-colored cycles on colored graphs was first treated in [6]. A great amount of work that followed concerned identification of such conditions and bounds on the number of colors [4, 1, 7, 9]. The optimization problems that we consider here were shown to be **NP**-hard in [4]. To the best of our knowledge no further theoretical development prior our work exists with respect to MaxLTSP and MinLTSP. An experimental study of MinLTSP appeared in [19]. *TSP under categorization* [17, 18] generalizes several TSP problems, and is also a weighted generalization of MinLTSP. For metric edge weights and at most q colors appearing in the graph a $2q$ approximation is achieved in [17, 18].

The recent literature on labeled network optimization problems includes several interesting results from both perspectives of hardness and approximation algorithms. 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. **NP**-hardness of finding paths with the fewest different colors was shown in [4]. The labeled minimum spanning tree problem was introduced in [5]. In [12] a greedy approximation algorithm is analyzed, and in [2] bounded color frequency is considered. The labeled perfect matching problems were studied in [14, 15], while Maffioli *et al.* worked on a labeled matroid problem [13]. Complexity of approximation of bottleneck labeled problems is studied in [11].

2 MaxLTSP: Constant factor Approximation

A simple greedy algorithm yields a 1/3 approximation of MaxLTSP (see full version).We analyze a $\frac{1}{2}$-approximation algorithm based on local search. The algorithm grows iteratively by local improvements a subset $S \subseteq E$ of edges, such that *(i)* each label of $\mathcal{L}(S)$ appears at most once in S and *(ii)* S does not induce vertices of degree three or more, or a cycle of length less than n. We call S a *labeled valid* subset of edges. Finding a labeled valid subset S of maximum size is clearly equivalent to MaxLTSP.

Given a labeled valid subset S of $(K_n, \mathcal{L})$, a *1-improvement* of S is a labeled valid subset $S \cup \{e_1\}$ where $e_1 \notin S$, whereas a *2-improvement* of S is a labeled valid subset $(S \setminus \{e\}) \cup \{e_1, e_2\}$ where $e \in S$ and $e_1, e_2 \notin S \setminus \{e\}$. An 1- or 2-improvement of S is a labeled valid subset S' such that $|S'| = |S| + 1$. An 1-improvement can be viewed as a particular 2-improvement but we separate the two cases for ease of presentation. The local improvement algorithm - denoted by locim - initializes $S = \emptyset$ and performs iteratively either an 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.

We introduce further notations. Given $e \in S$, let $\ell(e)$ be the edge of S^* with the same label if such an edge exists. Formally, $\ell : S \to S^* \cup \{\perp\}$ is defined by:

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

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

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

$N(e)$ is partitionned into $N_1(e)$ 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 $k \neq j$. Either S has two edges incident to i, or $S \cup \{e^*\}$ contains a cycle passing through e and e^*.

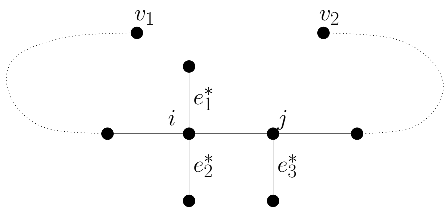
Property 1 holds at the end of the algorithm since otherwise $S \cup \{e^*\}$ would be an 1-improvement of S.

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^*.

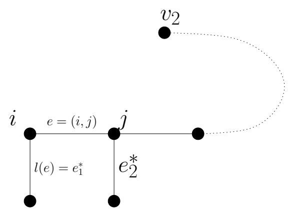
Property 2 holds at the end of the algorithm since otherwise $(S \setminus \{e\}) \cup \{e_1^*, e_2^*\}$ would be a 2-improvement of S. In order to prove the $\frac{1}{2}$ approximation factor for locim we use charging/discharging arguments based on the following function $g : S \to \mathbb{R}$:

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

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



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



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

Fig. 1: Cases studied in proof of proposition 1

Proposition 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 an 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 w.l.o.g., 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. This cycle must be $(P_e \setminus \{e\}) \cup \{e_1^*, e_3^*\}$ where P_e is the path containing e in S (see Fig. 1a: $e_1^* = [i, v_2]$ and $e_3^* = [j, v_1]$. Note that $e_2^* \neq [i, v_1]$ since $e_2^* \in N_1(e)$). Then $(S \setminus \{e\}) \cup \{e_2^*, e_3^*\}$ would be a 2-improvement of S, a contradiction.

- If $\ell(e) = \perp$ or $|N^{-1}(\ell(e))| \geq 2$, we deduce from $|N_1(e)| \leq 2$ that $g(e) \leq 2$.
- If $\ell(e) \neq \perp$ and $|N^{-1}(\ell(e))| = 1$, then $|N_1(e)| \leq 1$. Otherwise, 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. Hence, $|N_1(e)| \leq 1$ 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 $|S^*|$.

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

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

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

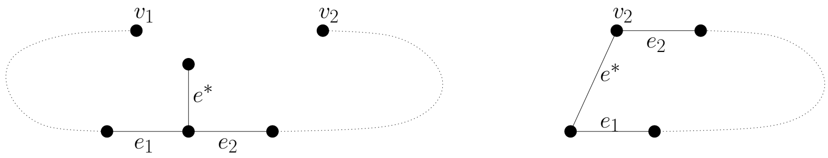


Fig. 2: The case where $N^{-1}(e^*) = \{e_1, e_2\}$.

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 $\ell(e_0) = e^*$. We distinguish two possibilities:

- $e^* \in N(e_0)$: 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:

- $N^{-1}(e^*) = \{e_1, e_2\}$: 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 propositions 1 and 2, we have $2|S| \geq \sum_{e \in S} g(e) \geq |S^*|$. Fig. 3 gives an example with approximation ratio $\frac{6}{10}$ achieved by LOCIM. This example can be generalized to asymptotic $\frac{1}{2}$ (to appear in the full version). □

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

Proof. (Sketch) We construct an L-reduction from the *maximum hamiltonian path problem on graphs with distances 1 and 2* (complete proof appears in the full version). □

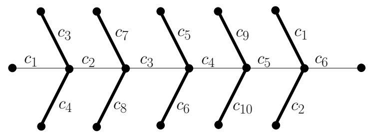


Fig. 3: A critical instance: undrawn edges have label c_1. LOCIM returns the horizontal path (colors c_1 to c_6). An optimum contains the other edges, using colors c_1 to c_{10}.

3 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.

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 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 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 $G = (V, E)$ be an instance of the hamiltonian $s - t$-path problem on a graph 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 of MinLTSP: take a graph 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

Algorithm 1: Greedy Tour

Let $T \leftarrow \emptyset$;
Let $K \leftarrow \{c_1, \ldots, c_q\}$;
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 ;

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]). Thus, 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 grows with n we obtain:

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

3.2 The Case of Fixed Color Frequency

We describe and analyze a greedy approximation algorithm (referred to as *Greedy Tour* - algorithm 1) 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 set of edges that is valid with respect to the current (partial) tour T and adds it to the tour, while also eliminating the selected color from the current set of colors. 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 $j = 1, \ldots, r$, 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.*

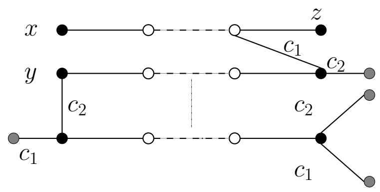


Fig. 4: Graphical illustration of definitions: if $c_1, c_2 \in \mathcal{L}_j(T^*)$, apart from vertices x, y, z, the remaining endpoints of paths are *black terminals*. Inner vertices are *white terminals* (drawn white), while vertices outside the paths are *optional vertices*.

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. 4 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 1 (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) $\displaystyle\sum_{c \in \mathcal{L}^*_{j,out}} f_j(c) \geq \sum_{c \in \mathcal{L}^*_{j,out}} (|\mathcal{L}^{-1}(c) \cap T^*| + 1 - j) + p.$

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

Proof. We prove the inequality by upper and lower bounding the quantity $F_j^* = \sum_{c \in \mathcal{L}_j(T^*)} f_j(c)$. A lower bound stems from Lemma 1:

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

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:

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

The result follows by combination of (1) and (2). □

We prove the approximation ratio of Greedy Tour by using Lemma 2:

Theorem 4. *For any $r \geq 1$ fixed, Greedy tour gives a $\frac{r+H_r}{2}$−approximation for* MinLTSP$_{(r)}$ *and the analysis is tight.*

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

$$\sum_{j=2}^{r} \sum_{i=j}^{r} \frac{i+1-j}{2j(j-1)} y_i^* \leq \sum_{j=2}^{r} \sum_{i=j}^{r} \frac{i}{j(j-1)} y_i \tag{3}$$

For the right-hand part of inequality (3) 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 (3) we have:

$$\sum_{j=2}^{r}\sum_{i=j}^{r}\frac{i+1-j}{2j(j-1)}\,y_i^* = \sum_{i=2}^{r}\frac{y_i^*}{2}\sum_{j=2}^{i}\frac{i+1-j}{j(j-1)} \tag{4}$$

But we also have:

$$\sum_{j=2}^{i}\frac{i+1-j}{j(j-1)} = \sum_{j=2}^{i}(\frac{i-(j-1)}{j-1}-\frac{i-j}{j}) - (H_i-1) = i - H_i \tag{5}$$

where $H_i = \sum_{k=1}^{i}\frac{1}{k}$. Therefore relation (4) becomes by (5):

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

By plugging the right-hand equality and (6) into inequality (3), we obtain:

$$\sum_{i=2}^{r}\frac{i-H_i}{2}y_i^* \leq \sum_{i=2}^{r}(i-1)y_i \tag{7}$$

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

$$OPT = \sum_{i=1}^{r} y_i^*,\;\; APX = \sum_{i=1}^{r} y_i,\;\; \text{and}\;\; \sum_{i=1}^{r} iy_i = \sum_{i=1}^{r} iy_i^* = n \tag{8}$$

where $n = |T| = |T^*|$ is the number of vertices of the graph. By (8) we can write $APX = n - \sum_{i=2}^{r}(i-1)y_i$, and using inequality (7), we deduce:

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

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

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

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 $\frac{7}{4} = \frac{2+H_2}{2}$

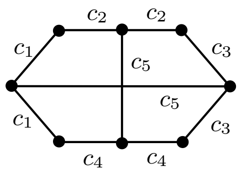


Fig. 5: Only colors appearing twice are represented. The others appears once.

approximation. A detailed example for $r \geq 3$ is given in the full version of the paper. $\square$

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

Theorem 5. *A ρ-approximation for* MinLTSP$_{(2)}$ *can be polynomially transformed into a $(\rho + \varepsilon)$-approximation for* MinHPP$_{1,2}$*, for all $\varepsilon > 0$.*

Proof. Let I be an instance of MinHPP$_{1,2}$, with $V(K_n) = \{v_1, \ldots, v_n\}$, 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 $V(K_{2n}) = \{v_1, \ldots, v_n\} \cup \{v'_1, \ldots, v'_n\}$. 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 $OPT(I)$. We build a tour T' of K_{2n} by taking P^*, the edges $[x, x']$, $[y, y']$ and a copy of P^* on vertices $\{v'_1, \ldots, v'_n\}$. We obtain $|\mathcal{L}(T')| = OPT(I) + 2$, and deduce:

$$OPT(I') \leq OPT(I) + 2 \tag{9}$$

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:

$$d(P) \leq |\mathcal{L}(T')| - 2 \tag{10}$$

where $d(P) = \sum_{e \in P} d(e)$. Using inequalities (9) and (10), we deduce $OPT(I') = OPT(I) + 2$. Now, if T is a ρ-approximation for MinLTSP$_{(2)}$, we deduce $d(P) \leq \rho\, OPT(I) + 2(\rho - 1) \leq (\rho + \varepsilon) OPT(I)$ when n is large enough. $\square$

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 5 that MinLTSP$_{(2)}$ is **APX**-hard. Moreover, MinLTSP$_{(2)}$ belongs to **APX** because any feasible tour is 2-approximate.

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

4 Open questions and future work

Can we provide a better approximation algorithm for $\textsc{MinLTSP}_{(r)}$, when r is a fixed small constant (e.g. $r = 2$)? Concerning $\textsc{MaxLTSP}$, local search using $k-$improvements for fixed $k \geq 3$ could exhibit better performance but its analysis appears quite non-trivial. It would be also interesting to explore the complexity of $\textsc{MaxLTSP}_{(r)}$ with bounded color frequency r.

References

1. M. Albert, A. Frieze, and B. Reed. Multicoloured Hamilton Cycles. *The Electronic Journal of Combinatorics* (1995), **2**, #R10.
2. T. Brueggemann, J. Monnot and G. J. Woeginger. Local search for the minimum label spanning tree problem with bounded color classes, *Oper. Res. Lett.* 31(3):195-201, 2003.
3. 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.
4. H. Broersma, X. Li, G. J. Woeginger and S. Zhang. Paths and Cycles in Colored Graphs, *Australasian Journal on Combinatorics* 31:299-311, 2005.
5. R-S. Chang and S-J. Leu. The Minimum Labeling Spanning Trees, *Inf. Process. Lett.* 63(5):277-282, 1997.
6. P. Erdős, J. Nesetril, and V. Roedl. Some problems related to partitions of edges of a graph. *Graphs and Other Combinatorial Topics*, Teubner, Leipzig (1983), 54-63.
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* (1986), 62(1):29-33.
10. R. Hassin, J. Monnot and D. Segev. Approximation Algorithms and Hardness Results for Labeled Connectivity Problems", *J. Comb. Opt.* 14(4):437-453, 2007.
11. R. Hassin, J. Monnot and D. Segev. The Complexity of Bottleneck Labeled Graph Problems, *In Proc. of WG'07*, LNCS 4769, Springer, 2007.
12. S. O. Krumke and H.-C. Wirth. On the Minimum Label Spanning Tree Problem, *Inf. Process. Lett.* 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, *Inf. Process. Lett.* 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*, to appear, 2007.
16. C. H. Papadimitriou and M. Yannakakis. The traveling salesman problem with distances one and two, *Math. of Operations Research*, vol. 18:1 pp. 1-11, 1993.
17. A. P. Punnen. Traveling Salesman Problem under Categorization, *Oper. Res. Lett.* 12(2):89-95, 1992.
18. A. P. Punnen. Erratum: Traveling Salesman Problem under Categorization, *Oper. Res. Lett.* 14(2):121, 1993.
19. Y. Xiong, B. Golden, and E. Wasil. The Colorful Traveling Salesman Problem, *Extending the Horizons: Advances in Computing, Optimization, and Decision Technologies*, Springer, 115-123, 2007.