

## Ola Svensson, Jakub Tarnawski and [László A. Végh](#) Constant factor approximation for ATSP with two edge weights

**Article (Accepted version)  
(Refereed)**

**Original citation:**

Svensson, Ola, Tarnawski, Jakub and Végh, László A. (2017). *Constant factor approximation for ATSP with two edge weights*. [Mathematical Programming](#). ISSN 0025-5610.

DOI: [10.1007/s10107-017-1195-7](https://doi.org/10.1007/s10107-017-1195-7)

© 2017 [Springer-Verlag GmbH](#) Germany and Mathematical Optimization Society

This version available at: <http://eprints.lse.ac.uk/85331/>

Available in LSE Research Online: November 2017

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

# Constant Factor Approximation for ATSP with Two Edge Weights

Ola Svensson\*  
ola.svensson@epfl.ch

Jakub Tarnawski†  
jakub.tarnawski@epfl.ch

László A. Végh‡  
l.vegh@lse.ac.uk

## Abstract

We give a constant factor approximation algorithm for the Asymmetric Traveling Salesman Problem on shortest path metrics of directed graphs with two different edge weights. For the case of unit edge weights, the first constant factor approximation was given recently by Svensson. This was accomplished by introducing an easier problem called Local-Connectivity ATSP and showing that a good solution to this problem can be used to obtain a constant factor approximation for ATSP. In this paper, we solve Local-Connectivity ATSP for two different edge weights. The solution is based on a flow decomposition theorem for solutions of the Held-Karp relaxation, which may be of independent interest.

## 1 Introduction

The traveling salesman problem — one of finding the shortest tour of  $n$  cities — is one of the most classical optimization problems. Its definition dates back to the 19th century and since then a large body of work has been devoted to designing “good” algorithms using heuristics, mathematical programming techniques, and approximation algorithms. The focus of this work is on approximation algorithms. A natural and necessary assumption in this line of work that we also make throughout this paper is that the distances satisfy the triangle inequality: for any triple  $i, j, k$  of cities, we have  $d(i, j) + d(j, k) \geq d(i, k)$  where  $d(\cdot, \cdot)$  denotes the pairwise distances between cities. In other words, it is not more expensive to take the direct path compared to a path that makes a detour.

With this assumption, the approximability of TSP turns out to be a very delicate question that has attracted significant research efforts. Specifically, one of the first approximation algorithms (Christofides’ heuristic [Chr76]) was designed for the *symmetric* traveling salesman problem (STSP) where we assume symmetric distances ( $d(i, j) = d(j, i)$ ). Several works (see e.g. [FGM82, AGM<sup>+</sup>10, GS11, AG15, Sve15]) have addressed the more general *asymmetric* traveling salesman problem (ATSP) where we make no such assumption.

However, there are still large gaps in our understanding of both STSP and ATSP. In fact, for STSP, the best approximation algorithm remains Christofides’  $3/2$ -approximation algorithm from the 70’s [Chr76]. For the harder ATSP, the state of the art is a  $\mathcal{O}(\log n / \log \log n)$ -approximation algorithm by Asadpour et al. [AGM<sup>+</sup>10] and a recent  $\mathcal{O}(\text{poly } \log \log n)$ -estimation algorithm<sup>1</sup> by Anari and Oveis Gharan [AG15]. On the negative side, the best inapproximability results only say that STSP and ATSP are hard to approximate within factors  $123/122$

---

\*École Polytechnique Fédérale de Lausanne. Supported by ERC Starting Grant 335288-OptApprox.

†École Polytechnique Fédérale de Lausanne.

‡London School of Economics. Supported by EPSRC First Grant EP/M02797X/1.

<sup>1</sup>An estimation algorithm is a polynomial-time algorithm for approximating/estimating the optimal value without necessarily finding a solution to the problem.

and  $75/74$ , respectively [KLS15]. Closing these gaps is a major open problem in the field of approximation algorithms (see e.g. “Problem 1” and “Problem 2” in the list of open problems in the recent book by Williamson and Shmoys [WS11]). What is perhaps even more intriguing about these questions is that we expect that a standard linear programming (LP) relaxation, often referred to as the Held-Karp relaxation, already gives better guarantees. Indeed, it is conjectured to give a guarantee of  $4/3$  for STSP and a guarantee of  $\mathcal{O}(1)$  (or even 2) for ATSP.

An equivalent formulation of STSP and ATSP from a more graph-theoretic point of view is the following. For STSP, we are given a weighted undirected graph  $G = (V, E, w)$  where  $w : E \rightarrow \mathbb{R}_+$  and we wish to find a multisubset  $F$  of edges of minimum total weight such that  $(V, F)$  is connected and Eulerian. Recall that an undirected graph is Eulerian if every vertex has even degree. We also remark that we use the term multisubset as the solution  $F$  may use the same edge several times. An intuitive point of view on this definition is that  $G$  represents a road network, and a solution is a tour that visits each vertex at least once (and may use a single edge/road several times). The definition of ATSP is similar, with the differences that the input graph is directed and the output is Eulerian in the directed sense: the in-degree of each vertex equals its out-degree. Having defined the traveling salesman problem in this way, there are several natural special cases to consider. For example, what if  $G$  is planar? Or, what if all the edges/roads have the same length, i.e., if  $G$  is unweighted?

For planar graphs, we have much better algorithms than in general. Grigni, Koutsoupias and Papadimitriou [GKP95] first obtained a polynomial-time approximation scheme for STSP restricted to unweighted planar graphs, which was later generalized to edge-weighted planar graphs by Arora et al. [AGK<sup>+</sup>98]. More recently, ATSP on planar graphs (and more generally bounded genus graphs) was shown to admit constant factor approximation algorithms (first by Oveis Gharan and Saberi [GS11] and later by Erickson and Sidiropoulos [ES14] who improved the dependency on the genus).

In contrast to planar graphs, STSP and ATSP remain APX-hard for unweighted graphs (ones where all edges have identical weight) and, until recently, there were no better algorithms for these cases. Then, in a recent series of papers, the approximation guarantee of  $3/2$  was finally improved for STSP restricted to unweighted graphs. Specifically, Oveis Gharan, Saberi and Singh [GSS11] first gave an approximation guarantee of  $1.5 - \epsilon$ ; Mömke and Svensson [MS11] proposed a different approach yielding a 1.461-approximation guarantee; Mucha [Muc12] gave a tighter analysis of this algorithm; and Sebő and Vygen [SV14] significantly developed the approach to give the currently best approximation guarantee of 1.4. Similarly, for ATSP, it was only very recently that the restriction to unweighted graphs could be leveraged: the first constant approximation guarantee for unweighted graphs was given by Svensson [Sve15]. In this paper we make progress towards the general problem by taking the logical next step and addressing a simple case left unresolved by [Sve15]: graphs with two different edge weights.

**Theorem 1.1.** *There is an  $\mathcal{O}(1)$ -approximation algorithm for ATSP on graphs with two different edge weights.*

The paper [Sve15] introduces an “easier” problem named Local-Connectivity ATSP, where one needs to find an Eulerian multiset of edges crossing only sets in a given partition rather than all possible sets (see next section for definitions). It is shown that an “ $\alpha$ -light” algorithm to this problem yields a  $(9 + \epsilon)\alpha$ -factor approximation for ATSP. For unweighted graphs (and slightly more generally, for node-induced weight functions<sup>2</sup>) it is fairly easy to obtain a 3-light algorithm for Local-Connectivity ATSP; the difficult part in [Sve15] is the black-box reduction of ATSP to this problem. Note that [Sve15] easily gives an  $\mathcal{O}(w_{\max}/w_{\min})$ -approximation algorithm in

---

<sup>2</sup>For ATSP, we can think of a node-weighted graph as an edge-weighted graph where the weight of an edge  $(u, v)$  equals the node weight of  $u$ .

general if we take  $w_{\max}$  and  $w_{\min}$  to denote the largest and smallest edge weight, respectively. However, obtaining a constant factor approximation even for two different weights requires substantial further work.

In Local-Connectivity ATSP we need a lower bound function  $\text{lb} : V \rightarrow \mathbb{R}_+$  on the vertices. The natural choice for node-induced weights is  $\text{lb}(v) = \sum_{e \in \delta^+(v)} w(e)x_e^*$ . With this weight function, every vertex is able to “pay” for the incident edges in the Eulerian subgraph we are looking for. This choice of  $\text{lb}$  does not seem to work for more general weight functions, and we need to define  $\text{lb}$  more “globally”, using a new flow theorem for Eulerian graphs (Theorem 2.1). In Section 1.2, after the preliminaries, we give a more detailed overview of these techniques and the proof of the theorem. Our argument is somewhat technical, but it demonstrates the potential of the Local-Connectivity ATSP problem as a tool for attacking general ATSP.

Finally, let us remark that both STSP [PY93, BK06] and ATSP [Blä04] have been studied in the case when all distances are either 1 or 2. That restriction is very different from our setting, as in those cases the input graph is complete. In particular, it is trivial to get a 2-approximation algorithm there, whereas in our setting – where the input graph is *not* complete – a constant factor approximation guarantee already requires non-trivial algorithms. (In our setting, we can still think about the metric completion, but it will usually have more than two different edge weights.)

## 1.1 Notation and preliminaries

We consider an edge-weighted directed graph  $G = (V, E, w)$  with  $w : E \rightarrow \mathbb{R}_+$ . For a vertex subset  $S \subseteq V$  we let  $\delta^+(S) = \{(u, v) \in E : u \in S, v \in V \setminus S\}$  and  $\delta^-(S) = \{(u, v) \in E : u \in V \setminus S, v \in S\}$  denote the sets of outgoing and incoming edges, respectively. For two vertex subsets  $X, Y \subseteq V$ , we let  $\delta(X, Y) = \{(u, v) \in E : u \in X \setminus Y, v \in Y \setminus X\}$ . For a subset of edges  $E' \subseteq E$ , we use  $\delta_{E'}^+(S) = \delta^+(S) \cap E'$  and  $\delta_{E'}^-(S) = \delta^-(S) \cap E'$ . We also let  $\mathcal{C}(E') = (\tilde{G}_1, \dots, \tilde{G}_k)$  denote the set of weakly connected components of the graph  $(V, E')$ ; the vertex set  $V$  will always be clear from the context. For a directed graph  $\tilde{G}$  we use  $V(\tilde{G})$  to denote its vertex set and  $E(\tilde{G})$  the edge set. For brevity, we denote the singleton set  $\{v\}$  by  $v$  (e.g.  $\delta^+(v) = \delta^+(\{v\})$ ), and we use the notation  $x(F) = \sum_{e \in F} x_e$  for a subset  $F \subseteq E$  of edges and a vector  $x \in \mathbb{R}^E$ . For a multiset  $F$ , we have  $\mathbb{1}_F$  denote the indicator vector of  $F$ , which has a coordinate for each edge  $e$  with value equal to the number of copies of  $e$  in  $F$ . For the case of two edge weights, we use  $0 \leq w_0 < w_1$  to denote the two possible values, and partition  $E = E_0 \cup E_1$  so that  $w(e) = w_0$  if  $e \in E_0$  and  $w(e) = w_1$  if  $e \in E_1$ . We will refer to edges in  $E_0$  and  $E_1$  as cheap and expensive edges, respectively.

We define ATSP as the problem of finding a connected Eulerian subgraph of minimum weight. As already mentioned in the introduction, this definition is equivalent to that of visiting each city exactly once (in the metric completion) since we assume the triangle inequality. The formal definition is as follows.

### ATSP

*Given:* An edge-weighted (strongly connected) digraph  $G = (V, E, w)$ .

*Find:* A multisubset  $F$  of  $E$  of minimum total weight  $w(F) = \sum_{e \in F} w(e)$  such that  $(V, F)$  is Eulerian and connected.

**Held-Karp Relaxation.** The Held-Karp relaxation has a variable  $x_e \geq 0$  for every edge in  $G$ . The intended meaning is that  $x_e$  should equal the number of times  $e$  is used in the solution. The relaxation  $\text{LP}(G)$  is defined as follows:

$$\begin{aligned} & \text{minimize} && \sum_{e \in E} w(e)x_e \\ & \text{subject to} && x(\delta^+(v)) = x(\delta^-(v)) \quad v \in V, \\ & && x(\delta^+(S)) \geq 1 \quad \emptyset \neq S \subsetneq V, \\ & && x \geq 0. \end{aligned} \tag{LP}(G)$$

The first set of constraints says that the in-degree should equal the out-degree for each vertex, i.e., the solution should be Eulerian. The second set of constraints enforces that the solution is connected; they are sometimes referred to as subtour elimination constraints. Finally, we remark that although the Held-Karp relaxation has exponentially many constraints, it is well-known that we can solve it in polynomial time either by using the ellipsoid method with a separation oracle or by formulating an equivalent compact (polynomial-size) linear program. We will use  $x^*$  to denote an optimal solution to  $\text{LP}(G)$  of value  $\text{OPT}$ , which is a lower bound on the value of an optimal solution to ATSP on  $G$ .

**Local-Connectivity ATSP.** The Local-Connectivity ATSP problem can be seen as a two-stage procedure. In the first stage, the input is an edge-weighted digraph  $G = (V, E, w)$  and the output is a “lower bound” function  $\text{lb} : V \rightarrow \mathbb{R}_+$  on the vertices such that  $\text{lb}(V) \leq \text{OPT}$ . In the second stage, the input is a partition of the vertices, and the output is an Eulerian multisubset of edges which crosses each set in the partition and where the ratio of weight to  $\text{lb}$  of every connected component is as small as possible. We now give the formal description of the second stage, assuming the  $\text{lb}$  function is already computed.

### Local-Connectivity ATSP

*Given:* An edge-weighted digraph  $G = (V, E, w)$ , a function  $\text{lb} : V \rightarrow \mathbb{R}_+$  with  $\text{lb}(V) \leq \text{OPT}$ , and a partitioning  $V = V_1 \cup V_2 \cup \dots \cup V_k$  of the vertices.

*Find:* A Eulerian multisubset  $F$  of  $E$  such that

$$|\delta_F^+(V_i)| \geq 1 \quad \text{for } i = 1, 2, \dots, k \quad \text{and} \quad \max_{\tilde{G} \in \mathcal{C}(F)} \frac{w(\tilde{G})}{\text{lb}(\tilde{G})} \text{ is minimized.}$$

Here we used the notation that for a connected component  $\tilde{G}$  of  $(V, F)$ ,  $w(\tilde{G}) = \sum_{e \in E(\tilde{G})} w(e)$  (summation over the edges) and  $\text{lb}(\tilde{G}) = \sum_{v \in V(\tilde{G})} \text{lb}(v)$  (summation over the vertices). We say that an algorithm for Local-Connectivity ATSP is  $\alpha$ -light on  $G$  if it is guaranteed, for any partition, to find a solution  $F$  such that for every component  $\tilde{G} \in \mathcal{C}(F)$ ,

$$\frac{w(\tilde{G})}{\text{lb}(\tilde{G})} \leq \alpha.$$

In [Sve15],  $\text{lb}$  is defined as  $\text{lb}(v) = \sum_{e \in \delta^+(v)} w(e)x_e^*$ ; note that  $\text{lb}(V) = \text{OPT}$  in this case. We remark that we use the “ $\alpha$ -light” terminology to avoid any ambiguities with the concept

of approximation algorithms (an  $\alpha$ -light algorithm does not compare its solution to an optimal solution to the given instance of Local-Connectivity ATSP).

Perhaps the main difficulty of ATSP is to satisfy the connectivity requirement, i.e., to select an Eulerian subset  $F$  of edges which connects the whole graph. Local-Connectivity ATSP relaxes this condition – we only need to find an Eulerian set  $F$  that crosses the  $k$  cuts defined by the partition. This makes it intuitively an “easier” problem than ATSP. Indeed, an  $\alpha$ -approximation algorithm for ATSP (with respect to the Held-Karp relaxation) is trivially an  $\alpha$ -light algorithm for Local-Connectivity ATSP for an arbitrary lb function with  $\text{lb}(V) = \text{OPT}$ : just return the same Eulerian subset  $F$  as the algorithm for ATSP; since the set  $F$  connects the graph, we have  $\max_{\tilde{G} \in \mathcal{C}(F)} w(\tilde{G})/\text{lb}(\tilde{G}) = w(F)/\text{lb}(V) \leq \alpha$ . Perhaps more surprisingly, the main technical theorem of [Sve15] shows that the two problems are equivalent up to small constant factors.

**Theorem 1.2 ([Sve15]).** *Let  $\mathcal{A}$  be an algorithm for Local-Connectivity ATSP. Consider an ATSP instance  $G = (V, E, w)$ , and let  $\text{OPT}$  denote the optimum value of the Held-Karp relaxation. If  $\mathcal{A}$  is  $\alpha$ -light on  $G$ , then there exists a tour of  $G$  with value at most  $5\alpha \text{OPT}$ . Moreover, for any  $\varepsilon > 0$ , a tour of value at most  $(9 + \varepsilon)\alpha \text{OPT}$  can be found in time polynomial in the number  $n = |V|$  of vertices, in  $1/\varepsilon$ , and in the running time of  $\mathcal{A}$ .*

In other words, the above theorem says that in order to approximate an ATSP instance  $G$ , it is sufficient to devise a polynomial-time algorithm to calculate a lower bound  $\text{lb}$  and a polynomial time algorithm for Local-Connectivity ATSP that is  $\mathcal{O}(1)$ -light on  $G$  with respect to this lb function. Our main result is proved using this framework.

## 1.2 Technical overview

**Singleton partition.** Let us start by outlining the fundamental ideas of our algorithm and comparing it to [Sve15] for the special case of Local-Connectivity ATSP when all partition classes  $V_i$  are singletons. For unit weights, the choice  $\text{lb}(v) = \sum_{e \in \delta^+(v)} w(e)x_e^* = x^*(\delta^+(v))$  in [Sve15] is a natural one: intuitively, every node is able to pay for its outgoing edges. We can thus immediately give an algorithm for this case: just select an arbitrary integral solution  $z$  to the circulation problem with node capacities  $1 \leq z(\delta^+(v)) \leq \lceil x^*(\delta^+(v)) \rceil$ . Then for any  $v$  we have  $z(\delta^+(v)) \leq x^*(\delta^+(v)) + 1 \leq 2x^*(\delta^+(v))$  and hence  $\sum_{e \in \delta^+(v)} w(e)z_e \leq 2\text{lb}(v)$ , showing that  $z$  is a 2-light solution.

The same choice of lb does not seem to work in the presence of two different edge costs. Consider a case when every expensive edge carries only a small fractional amount of flow. Then  $\sum_{e \in \delta^+(v)} w(e)x_e^*$  can be much smaller than the expensive edge cost  $w_1$ , and thus the vertex  $v$  would not be able to “afford” even a single outgoing expensive edge. To resolve this problem, we bundle small fractional amounts of expensive flow, channelling them to reach a small set of terminals. This is achieved via Theorem 2.1, a flow result which might be of independent interest. It shows that within the fractional Held-Karp solution  $x^*$ , we can send the flow from an arbitrary edge set  $E'$  to a sink set  $T$  with  $|T| \leq 8x^*(E')$ ; in fact,  $T$  can be any set minimal for inclusion such that it can receive the total flow from  $E'$ . We apply this theorem for  $E' = E_1$ , the set of expensive edges; let  $f$  be the flow from  $E_1$  to  $T$ , and call elements of  $T$  *terminals*. Now, whenever an expensive edge is used, we will “force” it to follow  $f$  to a terminal in  $T$ , where it can be paid for. Enforcement is technically done by splitting the vertices into two copies, one carrying the  $f$  flow and the other the rest. Thus we obtain the *split graph*  $G_{\text{sp}}$  and split fractional optimal solution  $x_{\text{sp}}^*$ .

The design of the split graph is such that every walk in it which starts with an expensive edge must proceed through cheap edges until it reaches a terminal before visiting another expensive

edge. In our terminology, expensive edges create “debt”, which must be paid off at a terminal. Starting from an expensive edge, the debt must be carried until a terminal is reached, and no further debt can be taken in the meantime. The bound on the number of terminals guarantees that we can assign a lower bound function  $\text{lb}$  with  $\text{lb}(V) \leq \text{OPT}$  such that (up to a constant factor) cheap edges are paid for locally, at their heads, whereas expensive edges are paid for at the terminals they are routed to. Such a splitting easily solves Local-Connectivity ATSP for the singleton partition: find an arbitrary integral circulation  $z_{\text{sp}}$  in the split graph with an upper bound  $z_{\text{sp}}(\delta^+(v)) \leq \lceil 2x_{\text{sp}}^*(\delta^+(v)) \rceil$  on every node, and a lower bound 1 on whichever copy of  $v$  transmits more flow. Note that  $2x_{\text{sp}}^*$  is a feasible fractional solution to this problem. We map  $z_{\text{sp}}$  back to an integral circulation  $z$  in the original graph by merging the split nodes, thus obtaining a constant-light solution.

**Arbitrary partitions.** Let us now turn to the general case of Local-Connectivity ATSP, where the input is an arbitrary partition  $V = V_1 \cup \dots \cup V_k$ . For unit weights this is solved in [Sve15] via an integer circulation problem on a modified graph. Namely, an auxiliary node  $A_i$  is added to represent each partition class  $V_i$ , and one unit of in- and outgoing flow from  $V_i$  is rerouted through  $A_i$ . In the circulation problem, we require exactly one in- and one outgoing edge incident to  $A_i$  to be selected. When we map the solution back to the original graph, there will be one incoming and one outgoing arc from every set  $V_i$  (thus satisfying the connectivity requirement) whose endpoints inside  $V_i$  violate the Eulerian condition. In [Sve15] every  $V_i$  is assumed to be strongly connected, and therefore we can “patch up” the circulation by connecting the loose endpoints by an arbitrary path inside  $V_i$ . This argument easily gives a 3-light solution.

Let us observe that the strong connectivity assumption is in fact not needed for the result in [Sve15]. Indeed, given a component  $V_i$  which is not strongly connected, consider its decomposition into strongly connected (sub)components, and pick a  $U_i \subseteq V_i$  which is a sink (i.e. it has no edges outgoing to  $V_i \setminus U_i$ ). We proceed by rerouting 1 unit of flow through a new auxiliary vertex just as in that algorithm, but we do this for  $U_i$  instead. This guarantees that  $U_i$  has at least one outgoing edge in our solution, and that edge must leave  $V_i$  as well.

Our result for two different edge weights takes this observation as the starting point, but the argument is much more complicated. We will find an integer circulation in a graph based on the split graph  $G_{\text{sp}}$ , and for every  $1 \leq i \leq k$ , there will be an auxiliary vertex  $A_i$  representing a certain subset  $U_i \subseteq V_i$ . These sets  $U_i$  will be obtained as sink components in certain auxiliary graphs we construct inside each  $V_i$ . This construction is presented in Section 3.2; we provide a roadmap to the construction at the beginning of that section.

## 2 The Flow Theorem

In this section we prove our main flow decomposition result. As indicated in Section 1.2, we will use it to channel the flow from the expensive edges  $E_1$  to a small set of terminals  $T$  (where  $|T| \leq 8w(E_1)$ ). We will use the theorem stated below by moving the tail of every edge in  $E_1$  to a new vertex  $s$ . If  $w(E_1) \geq 1$ , then the constraints of the Held-Karp relaxation guarantee condition (1). The details of the reduction are given in Lemma 3.4.

**Theorem 2.1.** *Let  $D = (V \cup \{s\}, E)$  be a directed graph, let  $c : E \rightarrow \mathbb{R}_+$  be a nonnegative capacity vector, and let  $s$  be a source node with no incoming edges, i.e.,  $\delta^-(s) = \emptyset$ . Assume that for all  $\emptyset \neq S \subseteq V$  we have*

$$c(\delta^-(S)) \geq \max\{1, c(\delta^+(S))\}. \quad (1)$$

Consider a set  $T \subseteq V$  such that there exists a flow  $f \leq c$  of value  $c(\delta^+(s))$  from the source  $s$  to the sink set  $T$ , and  $T$  is minimal subject to this property.<sup>3</sup> Then  $|T| \leq 8c(\delta^+(s))$ .

The proof of this theorem can be skipped on first reading, as the algorithm in Section 3 only uses it in a black-box manner.

*Proof.* Fix a minimal set  $T$  and denote  $k = |T|$ . Our goal is to prove that  $k \leq 8c(\delta^+(s))$ . We know that there exists a flow of value  $c(\delta^+(s))$  from  $s$  to  $T$ . For any such flow  $f$  we define its *imbalance sequence* to be the sequence of values  $z(t) = f(\delta^-(t)) - f(\delta^+(t)) \in \mathbb{R}_+$  for all  $t \in T$  sorted in non-increasing order. We select the flow  $f$  which maximizes the imbalance sequence (lexicographically). We write  $T = \{t_1, \dots, t_k\}$  so that  $z(t_1) \geq z(t_2) \geq \dots \geq z(t_k)$ ; denote  $z_i = z(t_i)$  for brevity. By minimality of  $T$  we have  $z(t_k) > 0$ . The following is our main technical lemma.

**Lemma 2.2.** *Let  $\ell$  be the number of  $t \in T$  with  $z(t) \geq \frac{1}{4}$ , i.e.,  $z_1 \geq \dots \geq z_\ell \geq \frac{1}{4} > z_{\ell+1} \geq \dots \geq z_k$ . Then we have*

$$\frac{1}{4}(k - \ell) \leq \sum_{i=1}^{\ell} z_i.$$

*In other words, the number of terminals with small imbalance is not much more than the sum of large imbalances.*

Assuming this lemma, the main theorem follows immediately, since we have  $\frac{1}{4}k = \frac{1}{4}\ell + \frac{1}{4}(k - \ell) \leq \sum_{i=1}^{\ell} z_i + \sum_{i=1}^{\ell} z_i \leq 2c(\delta^+(s))$ , i.e.,  $k \leq 8c(\delta^+(s))$ .  $\square$

The remainder of this section is devoted to the proof of the technical lemma. Let us first give an outline. We analyze the residual capacity (with respect to  $f$ ) of certain cuts that must be present due to the lexicographic property. First of all, there must be a saturated cut (that is, one of 0 residual in-capacity)  $A \subseteq V$  containing all large terminals (i.e., those with imbalance at least  $1/4$ ) but no small ones (Claim 3). Next, consider an arbitrary *small* terminal  $i$ . Also by the maximality property, it is not possible to increase the value of  $z_i$  to  $1/4$  by rerouting flow from other small terminals to  $i$ . Hence there must be a cut  $B_i$ , disjoint from  $A$ , which contains  $t_i$  as the only terminal and has residual in-capacity less than  $1/4 - z_i$  in  $D \setminus A$  (Claim 4). As an illustration of the argument, let us assume that these sets  $B_i$  are pairwise disjoint. It follows from (1) that the residual in-capacity of  $B_i$  is at least  $1 - z_i$  (Claim 2). Hence every set  $B_i$  must receive  $3/4$  units of its residual in-capacity from  $A$ . On the other hand, (1) upper-bounds the residual out-capacity of  $A$  by  $2 \sum_{i=1}^{\ell} z_i$  (Claim 1). These together give a bound  $\frac{3}{4}(k - \ell) \leq 2 \sum_{i=1}^{\ell} z_i$ . Recall however that we assumed that all sets  $B_i$  are disjoint. Since these sets may in fact overlap, the proof needs to be more careful: instead of sets  $B_i$ , we argue with the sets  $B_i \setminus (\cup_{j \neq i} B_j)$  (nonempty as containing  $t_i$ ), and the union of pairwise intersections  $B^*$ ; thus instead of  $3/8$ , we get a slightly worse constant  $1/4$ .

*Proof of Lemma 2.2.* First note that the claim is trivial if  $\ell = k$ , so assume  $\ell < k$ . For an arc  $e = (u, v)$ , we let  $\overleftarrow{e} = (v, u)$  denote the reverse arc. We define the residual graph  $D_f = (V \cup \{s\}, E_f)$  with  $E_f = \{e \in E : f(e) < c(e)\} \cup \{e : \overleftarrow{e} \in E, f(\overleftarrow{e}) > 0\}$ . The residual capacity for the first set of arcs is defined as  $c_f(e) = c(e) - f(e)$ , and for the second set as  $c_f(e) = f(\overleftarrow{e})$ . For any set  $X \subseteq V$ , and a disjoint  $Y \subseteq V$ , let  $\delta_f^-(X)$ ,  $\delta_f^+(X)$  and  $\delta_f(X, Y)$  denote the capacities

---

<sup>3</sup>That is, the maximum flow value from  $s$  to any proper subset  $T' \subsetneq T$  is smaller than  $c(\delta^+(s))$ .



of the respective cuts in the residual graph  $D_f$  of  $f$ , i.e.,

$$\begin{aligned}\delta_f^-(X) &= \sum_{e \in \delta^-(X)} c_f(e) = c(\delta^-(X)) - f(\delta^-(X)) + f(\delta^+(X)), \\ \delta_f^+(X) &= \sum_{e \in \delta^+(X)} c_f(e) = c(\delta^+(X)) - f(\delta^+(X)) + f(\delta^-(X)), \\ \delta_f(X, Y) &= \sum_{e \in \delta(X, Y)} c_f(e) = c(\delta(X, Y)) - f(\delta(X, Y)) + f(\delta(Y, X)).\end{aligned}$$

The next two claims derive simple bounds from (1) on the residual in-and out-capacities of cuts.

**Claim 1.** *If  $X \subseteq V$ , then  $\delta_f^-(X) + 2 \sum_{t \in T \cap X} z(t) \geq \delta_f^+(X)$ .*

*Proof.*

$$\begin{aligned}\delta_f^-(X) - \delta_f^+(X) &= c(\delta^-(X)) - c(\delta^+(X)) - 2(f(\delta^-(X)) - f(\delta^+(X))) \\ &= c(\delta^-(X)) - c(\delta^+(X)) - 2 \sum_{t \in T \cap X} z(t) \geq -2 \sum_{t \in T \cap X} z(t).\end{aligned}$$

The equality is by flow conservation. The inequality is by (1). The claim follows.  $\square$

**Claim 2.** *Consider  $X \subseteq V$  such that  $X \cap T = \{t_i\}$  for some  $1 \leq i \leq k$ . Then  $\delta_f^-(X) \geq 1 - z_i$ .*

*Proof.*

$$\delta_f^-(X) = c(\delta^-(X)) - (f(\delta^-(X)) - f(\delta^+(X))) \geq 1 - z_i.$$

Here we used (1) and the flow conservation  $f(\delta^-(X)) - f(\delta^+(X)) = z_i$  (as the single sink contained in  $X$  is  $t_i$ ).  $\square$

The next claim shows that the large terminals can be separated from the small ones by a cut of residual in-degree 0. This follows from the lexicographically maximal choice, and is not a particular property of the threshold  $1/4$  (it remains true if we replace  $\ell$  by any  $1 \leq j \leq k$ ).

**Claim 3.** *There exists a set  $A \subseteq V$  with  $A \cap T = \{t_1, \dots, t_\ell\}$  (i.e.,  $A$  contains exactly the large terminals) such that  $\delta_f^-(A) = 0$ , and  $\delta_f^+(A) \leq 2 \sum_{i=1}^\ell z_i$ .*

*Proof.* If  $\ell = 0$ , then we can choose  $A = \emptyset$ . So assume  $0 < \ell < k$ . Consider the maximum flow in the residual graph  $D_f$  from the source set  $\{t_{\ell+1}, \dots, t_k\}$  to the sink set  $\{t_1, \dots, t_\ell\}$ . If its value is positive, then there exists a path  $P$  in  $D_f$  from  $t_i$  to  $t_j$  for some  $i > \ell$  and  $j \leq \ell$  (without loss of generality it contains no other terminals). Set  $\varepsilon = \min\{z(t_i), \min_{(u,v) \in P} c_f(u,v)\} > 0$ . Then the  $s$ - $T$  flow  $f' = f + \varepsilon \cdot \mathbf{1}_P$  has a lexicographically larger imbalance sequence than  $f$  because  $z_i$  is increased without decreasing any other of the large imbalances, a contradiction. So there must be a cut  $A \subseteq V$  with  $A \cap T = \{t_1, \dots, t_\ell\}$  and  $\delta_f^-(A) = 0$ .<sup>4</sup> The second part follows by the first via Claim 1.  $\square$

**Claim 4.** *For any  $\ell + 1 \leq i \leq k$  (i.e.,  $t_i$  is a small terminal) there exists a set  $B_i \subseteq V \setminus A$  such that  $B_i \cap T = \{t_i\}$  and  $\delta_f^-(B_i) - \delta_f(A, B_i) < \frac{1}{4} - z_i$ .*

<sup>4</sup>Note that  $s \notin A$  since  $D_f$  contains a path from  $t_k$  to  $s$ .

*Proof.* If  $k = \ell + 1$ , then we can choose  $B_k = V \setminus A$ ; we have  $\delta_f^-(B_k) - \delta_f(A, B_k) = 0$  since all arcs leaving the source  $s$  are saturated. So assume  $k - \ell \geq 2$ . Consider the maximum flow from the source set  $\{t_{\ell+1}, \dots, t_k\} \setminus \{t_i\}$  to the sink  $t_i$  in the graph  $D_f \setminus (A \cup \{s\})$ .

If its value is at least  $\frac{1}{4} - z_i$ , then we will get a contradiction by increasing the imbalance of  $t_i$  to at least  $\frac{1}{4}$  without changing any of the large imbalances. Namely, let  $g$  be a flow from  $\{t_{\ell+1}, \dots, t_k\} \setminus \{t_i\}$  to  $t_i$  of value  $\frac{1}{4} - z_i$ . Consider the vector  $f + g$ . There are two possible cases:

- If  $f + g$  is still an  $s$ - $T$  flow, i.e., if for all  $j$  we have  $g(\delta^+(t_j)) - g(\delta^-(t_j)) \leq z(t_j)$ , then it has a lexicographically larger imbalance sequence than  $f$ , a contradiction.
- Otherwise pick the maximum  $\alpha > 0$  such that  $f + \alpha g$  is still an  $s$ - $T$  flow, i.e., for all  $j$  we have  $\alpha(g(\delta^+(t_j)) - g(\delta^-(t_j))) \leq z(t_j)$ , with equality for some  $j$ . This means that  $f + \alpha g$  is an  $s$ - $T$  flow where at least one terminal  $t_j$  has zero imbalance, i.e., it can be removed from the set  $T$ , contradicting its minimality.

So there must be a cut  $B_i \subseteq V \setminus (A \cup \{s\})$  such that  $B_i \cap T = \{t_i\}$  and

$$\frac{1}{4} - z_i > \delta_f(V \setminus (A \cup \{s\}) \cup B_i, B_i) = \delta_f^-(B_i) - \delta_f(A, B_i) - \delta_f(s, B_i).$$

The claim follows by  $\delta_f(s, B_i) = 0$ , which holds since all edges in  $\delta^+(s)$  are saturated in  $f$ .  $\square$

The argument uses the bound  $\delta_f^+(A) \leq 2 \sum_{i=1}^{\ell} z_{\ell}$  and the fact that all the  $B_i$ 's must receive a large part of their residual in-degrees from  $A$ . Since the sets  $B_i$  overlap, we have to take their intersections into account. Let us therefore define

$$B^* := \bigcup_{i,j > \ell, i \neq j} (B_i \cap B_j) \subseteq V \setminus (A \cup \{s\})$$

as the set of vertices contained in at least two sets  $B_i$ . Let  $\alpha^* := \delta_f(A, B^*)$ ,  $\alpha_i := \delta_f(A, B_i \setminus B^*)$ , and  $\beta_i := \delta_f(B^*, B_i \setminus B^*)$  for each  $\ell + 1 \leq i \leq k$ .

**Claim 5.** For each  $\ell + 1 \leq i \leq k$  we have  $\frac{3}{4} < \alpha_i + \beta_i$ .

*Proof.* Note that  $(B_i \setminus B^*) \cap T = \{t_i\}$ , and thus  $\delta_f^-(B_i \setminus B^*) \geq 1 - z_i$  by Claim 2. From this we can see

$$\begin{aligned} 1 - z_i &\leq \delta_f^-(B_i \setminus B^*) \\ &\leq \left( \delta_f^-(B_i) - \delta_f(A, B_i) \right) + \delta_f(A, B_i \setminus B^*) + \delta_f(B^*, B_i \setminus B^*) \\ &< \frac{1}{4} - z_i + \alpha_i + \beta_i, \end{aligned}$$

where the second inequality follows because an edge entering  $B_i \setminus B^*$  either enters  $B_i$  from outside of  $A$ , or enters  $B_i \setminus B^*$  from  $B^*$ , or enters  $B_i \setminus B^*$  from  $A$ .  $\square$

For the residual in-degree of the set  $B^*$ , we apply the trivial bound

$$\delta_f^-(B^*) \leq \delta_f(A, B^*) + \sum_{i=\ell+1}^k \left( \delta_f^-(B_i) - \delta_f(A, B_i) \right) \leq \alpha^* + \frac{1}{4}(k - \ell).$$

The last estimate is by the choice of the sets  $B_i$  in Claim 4. For the residual out-degree, we get

$$\delta_f^+(B^*) \geq \sum_{i=\ell+1}^k \beta_i > \frac{3}{4}(k - \ell) - \sum_{i=\ell+1}^k \alpha_i,$$

using Claim 5. Applying Claim 1 to  $B^*$  and noting that  $B^* \cap T = \emptyset$  gives  $\delta_f^-(B^*) \geq \delta_f^+(B^*)$ . Putting Claim 3 and the above two bounds together, we conclude that

$$2 \sum_{i=1}^{\ell} z_i \geq \delta_f^+(A) \geq \alpha^* + \sum_{i=\ell+1}^k \alpha_i \geq \frac{1}{2}(k - \ell). \quad (2)$$

Lemma 2.2 now follows.  $\square$

### 3 Algorithm for Local-Connectivity ATSP

We prove our main result in this section. Our claim for ATSP follows from solving Local-Connectivity ATSP:

**Theorem 3.1.** *There is a polynomial-time 100-light algorithm for Local-Connectivity ATSP on graphs with two edge weights.*

Together with Theorem 1.2, this implies our main result:

**Theorem 3.2.** *For any graph with two edge weights, the integrality gap of its Held-Karp relaxation is at most 500. Moreover, we can find an 901-approximate tour in polynomial time.*

The factor 500 comes from  $5 \cdot 100$ . In Theorem 1.2, we select  $\varepsilon$  such that  $(9 + \varepsilon) \cdot 100 \leq 901$ . Our proof of Theorem 3.1 proceeds as outlined in Section 1.2. In Section 3.1, we give an algorithm for calculating lb and define the *split graph* which will be central for finding light solutions. In Section 3.2, we then show how to use these concepts to solve Local-Connectivity ATSP for any given partitioning of the vertices.

Recall that the edges are partitioned into the set  $E_0$  of cheap edges and the set  $E_1$  of expensive edges. Set  $x^*$  to be an optimal solution to the Held-Karp relaxation. We start by noting that the problem is easy if  $x^*$  assigns very small total fractional value to expensive edges. In that case, we can easily reduce the problem to the unweighted case which was solved in [Sve15].

**Lemma 3.3.** *There is a polynomial-time 6-light algorithm for Local-Connectivity ATSP for graphs where  $x^*(E_1) < 1$ .*

*Proof.* If  $x^*(E_1) = 0$ , then just apply the standard 3-light polynomial-time algorithm for unweighted graphs [Sve15]. So suppose that  $0 < x^*(E_1) < 1$ . Then clearly the graph  $(V, E_0)$  is strongly connected, i.e. every pair of vertices is connected by a directed path of cheap edges (of length at most  $n - 1$ ). Thus each expensive edge  $(u, v)$  can be replaced by such a  $u$ - $v$ -path  $P(u, v)$ . Let us obtain a new circulation  $x'$  from  $x^*$  by replacing all expensive edges in this way, i.e.,

$$x' = x^*|_{E_0} + \sum_{(u,v) \in E_1} x^*_{(u,v)} \cdot \mathbb{1}_{P(u,v)}.$$

To bound the cost of  $x'$ , note that  $x^*(E_0) = x^*(E) - x^*(E_1) > n - 1 > (n - 1)x^*(E_1)$  and thus

$$w(x') \leq w_0 \cdot x^*(E_0) + (n - 1)w_0 \cdot x^*(E_1) \leq 2w_0 \cdot x^*(E_0) \leq 2w(x^*).$$

By construction,  $x'$  is a feasible solution for the Held-Karp relaxation and  $\text{supp}(x') \subseteq E_0$ . Therefore we can use it in the standard 3-light polynomial-time algorithm for the unweighted graph  $(V, E_0)$ . Together with the bound  $w(x') \leq 2w(x^*)$  this gives a 6-light algorithm.  $\square$

For the rest of this section, we thus assume  $x^*(E_1) \geq 1$ . Our objective is to define a function  $\text{lb} : V \rightarrow \mathbb{R}_+$  such that  $\text{lb}(V) \leq \text{OPT} = w(x^*)$  and then show how to, given a partition  $V = V_1 \cup \dots \cup V_k$ , find an Eulerian set of edges  $F$  which crosses all  $V_i$ -cuts and is  $\mathcal{O}(1)$ -light with respect to the defined lb function.

### 3.1 Calculating lb and constructing the split graph

First, we use our flow decomposition technique to find a small set of *terminals*  $T$  such that it is possible to route a certain flow  $f$  from endpoints of all expensive edges to  $T$ . Next, we use  $f$  and  $T$  to calculate the function  $\text{lb}$  and to construct a *split graph*  $G_{\text{sp}}$ , where each vertex of  $G$  is split into two.

**Finding terminals  $T$  and flow  $f$ .** We use Theorem 2.1 to obtain a small-enough set of terminals  $T$  and a flow  $f$  which takes all flow on expensive edges to this set  $T$ . More precisely, we have the following corollary of Theorem 2.1.

**Lemma 3.4.** *There exist a vertex set  $T \subseteq V$  and a flow  $f : E \rightarrow \mathbb{R}_+$  from source set  $\{\text{tail}(e) : e \in E_1\}$  to sink set  $T$  of value  $x^*(E_1)$  such that:*

- $|T| \leq 8x^*(E_1)$ ,
- $f \leq x^*$ ,
- $f$  saturates all expensive edges, i.e.,  $f(e) = x_e^*$  for all  $e \in E_1$ ,
- for each  $t \in T$ ,  $f(E_0 \cap \delta^+(t)) = 0$  and  $f(\delta^-(t)) > 0$ .

Moreover,  $T$  and  $f$  can be computed in polynomial time.

*Proof.* We construct  $G'$  to be  $G$  with a new vertex  $s$ , where the tail of every expensive edge is redirected to be  $s$ . Formally,  $V(G') = V \cup \{s\}$  and  $E(G') = E_0 \cup \{(s, \text{head}(e)) : e \in E_1\}$ . The capacity vector  $c$  is obtained from  $x^*$  by just following this redirection, i.e., for any edge  $e' \in E(G')$  we define  $c(e') = x_e^*$ , where  $e \in E(G)$  is taken to be the preimage of  $e'$  in  $G$ .

Clearly  $c(\delta^+(s)) = x^*(E_1)$ , and  $s \in V(G')$  has no incoming edges. To see that condition (1) of Theorem 2.1 is satisfied, recall that for every  $\emptyset \neq S \subsetneq V(G)$  we have  $x^*(\delta^+(S)) = x^*(\delta^-(S)) \geq 1$ ; redirecting the tail of some edges to  $s$  can only reduce the outdegree or increase the indegree of  $S$ , i.e.,  $c(\delta^-(S)) \geq x^*(\delta^-(S)) = x^*(\delta^+(S)) \geq c(\delta^+(S))$ . This gives condition (1) for all sets  $S \subsetneq V(G') - s$ ; for  $S = V(G') - s$ , note that  $c(\delta^-(S)) = x^*(E_1) \geq 1 = \max\{1, 0\} = \max\{1, c(\delta^+(S))\}$  since we assumed that  $x^*(E_1) \geq 1$ .

From Theorem 2.1 we obtain a vertex set  $T$  with  $|T| \leq 8x^*(E_1)$  and a flow  $f' : E(G') \rightarrow \mathbb{R}_+$  from  $s$  to  $T$  of value  $x^*(E_1)$  with  $f' \leq c$ . We can assume  $f'(\delta^+(t)) = 0$  for all  $t \in T$ : in a path-cycle decomposition of  $f'$  we can remove all cycles and terminate every path at the first terminal it reaches. The flow  $f$  is obtained by mapping  $f'$  back to  $G$ , i.e., taking each  $f'(e)$  to be  $f'(e')$ , where  $e'$  is the image of  $e$ . Note that  $f'$  must saturate all outgoing edges of  $s$ , so  $f$  saturates all expensive edges. For the last condition, the part  $f(E_0 \cap \delta^+(t)) = 0$  is implied by  $f'(\delta^+(t)) = 0$ , and for the part  $f(\delta^-(t)) > 0$ , note that if  $f(\delta^-(t)) = 0$ , then we could have removed  $t$  from  $T$ .

Note that such a set  $T$  can be found in polynomial time: starting from  $T = V$  (for which the required flow exists: consider  $f' = c|_{\delta^+(s)}$ , the restriction of  $c$  to  $\delta^+(s)$ ), we remove vertices from  $T$  one by one until we obtain a minimal set  $T$  such that there exists a flow of value  $x^*(E_1)$  from  $s$  to  $T$ .  $\square$

**Definition of  $\text{lb}$ .** We set  $\text{lb} : V \rightarrow \mathbb{R}_+$  to be a scaled-down variant of  $\overline{\text{lb}} : V \rightarrow \mathbb{R}_+$  which is defined as follows:

$$\overline{\text{lb}}(v) := \begin{cases} w_0 \cdot x^*(\delta^-(v)) & \text{if } v \notin T, \\ w_0 \cdot x^*(\delta^-(v)) + w_1 \cdot \lceil f(\delta^-(t)) \rceil & \text{if } v \in T. \end{cases}$$

The definition of  $\text{lb}$  is now simply  $\text{lb}(v) = \overline{\text{lb}}(v)/10$ . The scaling-down is done so as to satisfy  $\text{lb}(V) \leq \text{OPT}$  (see Lemma 3.5). Clearly we have  $\overline{\text{lb}}(v) \geq w_0$  for all  $v \in V$  and  $\overline{\text{lb}}(t) \geq w_1 + w_0 \geq w_1$  for terminals  $t \in T$ .

The intuition behind this setting of  $\overline{\text{lb}}$  is that we want to pay for each expensive edge  $e \in E_1$  in the terminal  $t \in T$  which the flow  $f$  “assigns” to  $e$ . Indeed, in the split graph we will reroute flow (using  $f$ ) so as to ensure that any path which traverses  $e$  must then visit such a terminal  $t$  to offset the cost of the expensive edge. As for the total cost of  $\overline{\text{lb}}$ , note that if we removed the rounding from its definition, then we would get  $\overline{\text{lb}}(V) \leq 2 \text{OPT}$ , since

$$w_0 \cdot \sum_{v \in V} x^*(\delta^-(v)) + w_1 \cdot \sum_{v \in T} f(\delta^-(t)) = w_0 \cdot x^*(E_0) + w_0 \cdot x^*(E_1) + w_1 \cdot x^*(E_1) \leq 2w(x^*)$$

(here we used that  $f$  is of value  $x^*(E_1) = \sum_{t \in T} f(\delta^-(t))$ ). So, similarly to the 3-light algorithm for unweighted metrics in [Sve15], the key is to argue that rounding does not increase this cost too much. For this, we will take advantage of the small size of  $T$ . Details are found in the proof of the following lemma.

**Lemma 3.5.**  $\overline{\text{lb}}(V) \leq 10 \cdot \text{OPT}$ .

*Proof.* The bound follows from elementary calculations:

$$\begin{aligned} \overline{\text{lb}}(V) &= w_0 \cdot \sum_{v \in V} x^*(\delta^-(v)) + w_1 \cdot \sum_{t \in T} [f(\delta^-(t))] \\ &\leq w_0 \cdot \sum_{v \in V} x^*(\delta^-(v)) + w_1 \cdot \sum_{t \in T} (f(\delta^-(t)) + 1) \\ &\leq w_0 \cdot x^*(E) + w_1 \cdot (x^*(E_1) + |T|) \\ &\leq w(x^*) + w_1 \cdot 9x^*(E_1) \\ &\leq 10w(x^*) \end{aligned}$$

(recall that  $|T| \leq 8x^*(E_1)$  by Lemma 3.4). □

**Construction of the split graph.** The next step is to reroute flow so as to ensure that all expensive edges are “paid for” by the  $\text{lb}$  at terminals. To this end, we define a new *split graph* and a *split circulation* on it (see also Fig. 1 for an example).

**Definition 3.6.** *The split graph  $G_{\text{sp}}$  is defined as follows. For every  $v \in V$  we create two copies  $v^0$  and  $v^1$  in  $V(G_{\text{sp}})$ . For every cheap edge  $(u, v) \in E_0$ :*

- *if  $x^*(u, v) - f(u, v) > 0$ , create an edge  $(u^0, v^0)$  in  $E(G_{\text{sp}})$  with  $x_{\text{sp}}^*(u, v) = x^*(u, v) - f(u, v)$ ,*
- *if  $f(u, v) > 0$ , create an edge  $(u^1, v^1)$  in  $E(G_{\text{sp}})$  with  $x_{\text{sp}}^*(u, v) = f(u, v)$ .*

*For every expensive edge  $(u, v) \in E_1$  we create one edge  $(u^0, v^1)$  in  $E(G_{\text{sp}})$  with  $x_{\text{sp}}^*(u, v) = f(u, v)$ . Finally, for each  $t \in T$  we create an edge  $(t^1, t^0)$  in  $E(G_{\text{sp}})$  with  $x_{\text{sp}}^*(t^1, t^0) = f(\delta^-(t))$ .*

*The new edges are weighted as follows: images of edges in  $E_0$  have weight  $w_0$ , the images of edges in  $E_1$  have weight  $w_1$ , and the new edges  $(t^1, t^0)$  have weight 0. Let us denote the new weight function by  $w_{\text{sp}}$ .*

*Vertices  $v^0$  will be called free vertices and vertices  $v^1$  will be called debt vertices. Edges entering a free vertex will be called free edges, and those entering a debt vertex will be called debt edges.*

A fundamental consequence of our construction is the following.

**Fact 3.7.** *Consider any path  $P$  in  $G_{\text{sp}}$  such that the first edge of  $P$  is a debt edge and the last one is a free edge or an expensive edge. Then  $P$  must go through a terminal, i.e., it must contain an edge  $(t^1, t^0)$  for some  $t \in T$ .  $\square$*

We also have the following two properties by design.

**Fact 3.8.** *The vector  $x_{\text{sp}}^* : E(G_{\text{sp}}) \rightarrow \mathbb{R}_+$  is a circulation in  $G_{\text{sp}}$ .*

*Proof.* Intuitively, this is because our rerouting corresponds to taking every path  $v_0, v_1, \dots, v_k$  in a path-cycle decomposition of  $f$  (where  $(v_0, v_1)$  is an expensive edge and  $v_k \in T$  is a terminal) and, rather than placing it on the “free level” in  $G_{\text{sp}}$  (i.e., mapping it to  $v_0^0, v_1^0, \dots, v_k^0$ ), instead routing it as follows:  $v_0^0, v_1^1, v_2^1, \dots, v_{k-1}^1, v_k^1, v_k^0$ . We can think of the edge  $(v_0^0, v_1^1)$  as incurring a debt, and of the edge  $(v_k^1, v_k^0)$  as discharging this debt.

Now we proceed to give a formal proof. For all  $v \in V$  we will prove that  $x_{\text{sp}}^*(\delta^+(v^0)) = x_{\text{sp}}^*(\delta^-(v^0))$  and  $x_{\text{sp}}^*(\delta^+(v^1)) = x_{\text{sp}}^*(\delta^-(v^1))$ .

Suppose  $v = t \in T$ . Then  $x_{\text{sp}}^*(\delta^-(t^1)) = f(\delta^-(t))$ . Also  $x_{\text{sp}}^*(\delta^+(t^1)) = x_{\text{sp}}^*(t^1, t^0) = f(\delta^-(t))$  because of the property that  $f(E_0 \cap \delta^+(t)) = 0$ . For  $t^0$  we have

$$\begin{aligned} x_{\text{sp}}^*(\delta^-(t^0)) &= f(\delta^-(t)) + (x^* - f)(E_0 \cap \delta^-(t)) \\ &= f(E_1 \cap \delta^-(t)) + x^*(E_0 \cap \delta^-(t)) \\ &= x^*(\delta^-(t)) \end{aligned}$$

and

$$\begin{aligned} x_{\text{sp}}^*(\delta^+(t^0)) &= (x^* - f)(E_0 \cap \delta^+(t)) + x^*(E_1 \cap \delta^+(t)) \\ &= x^*(E_0 \cap \delta^+(t)) + x^*(E_1 \cap \delta^+(t)) \\ &= x^*(\delta^+(t)) \end{aligned}$$

where the second inequality again follows by  $f(E_0 \cap \delta^+(t)) = 0$ . This implies that  $x_{\text{sp}}^*(\delta^-(t^0)) = x^*(\delta^-(t)) = x^*(\delta^+(t)) = x_{\text{sp}}^*(\delta^+(t^0))$ .

We turn to the case  $v \notin T$ . Note that since all expensive edges are saturated by  $f$ , and their tails are exactly the sources of  $f$ , we have

$$f(E_0 \cap \delta^+(v)) = f(E_0 \cap \delta^-(v)) + x^*(E_1 \cap \delta^-(v)). \quad (3)$$

Since the left hand side is equal to  $x_{\text{sp}}^*(\delta^+(v^1))$  and the right hand side is  $x_{\text{sp}}^*(\delta^-(v^1))$ , we have proved our claim for  $v^1$ . For  $v^0$ , note that the incoming edges of  $v_0$  are all cheap, and the incoming flow is

$$x_{\text{sp}}^*(\delta^-(v^0)) = (x^* - f)(E_0 \cap \delta^-(v)).$$

As for outgoing flow,  $v^0$  is also the tail of all expensive edges whose tail in  $G$  was  $v$ , i.e.,

$$\begin{aligned} x_{\text{sp}}^*(\delta^+(v^0)) &= (x^* - f)(E_0 \cap \delta^+(v)) + x^*(E_1 \cap \delta^+(v)) \\ &= x^*(\delta^+(v)) - f(E_0 \cap \delta^+(v)) \\ &= x^*(\delta^-(v)) - f(E_0 \cap \delta^-(v)) - x^*(E_1 \cap \delta^-(v)) \\ &= x^*(E_0 \cap \delta^-(v)) - f(E_0 \cap \delta^-(v)) \\ &= (x^* - f)(E_0 \cap \delta^-(v)) \\ &= x_{\text{sp}}^*(\delta^-(v^0)), \end{aligned}$$

where the third equality is by (3).  $\square$

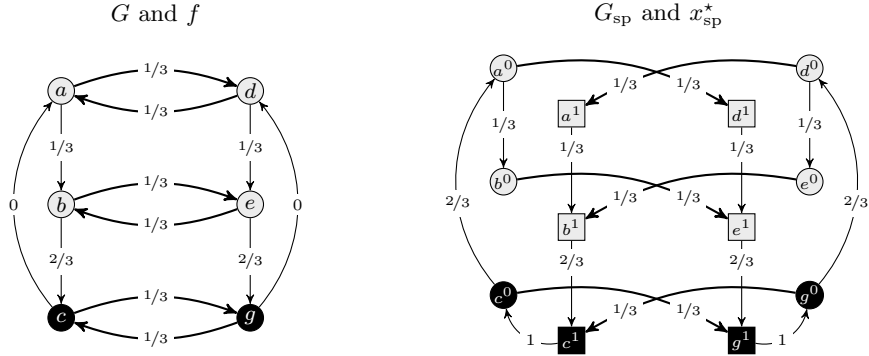


Figure 1: An example of the construction of  $G_{\text{sp}}$  and  $x_{\text{sp}}^*$  from  $G, x^*$  and  $f$ . Here  $x^*$  is  $1/3$  for the expensive edges (depicted as thick) and  $2/3$  for the remaining (cheap) edges; the set  $T$  of terminals of the flow  $f$  is depicted in black.

**Fact 3.9.** For each proper subset  $U \subset V$  we have  $x_{\text{sp}}^*(\delta^-(\{v^0, v^1 : v \in U\})) = x_{\text{sp}}^*(\delta^+(\{v^0, v^1 : v \in U\})) \geq 1$ . In other words, the image of every cut in  $G$  is still crossed by at least one unit of  $x_{\text{sp}}^*$ .

*Proof.* This follows since  $x^*(\delta^+(U)) = x^*(\delta^-(U)) \geq 1$ , and contracting all pairs  $v^0, v^1$  would yield back  $G$  and  $x^*$ .  $\square$

### 3.2 Solving Local-Connectivity ATSP

Now our algorithm is given a partition  $V = V_1 \cup \dots \cup V_k$  of the original vertex set. The objective is to output a set of edges  $F$  which crosses all  $V_i$ -cuts and is  $\mathcal{O}(1)$ -light with respect to our lb function.

We are aiming for a similar construction as in the unit-weight case: based on the split graph  $G_{\text{sp}}$ , we construct an integer circulation problem with an auxiliary vertex  $A_i$  representing a certain subset  $U_i \subseteq V_i$  for every  $1 \leq i \leq k$ . We then map its solution back to the original graph and patch up the loose endpoints inside every  $U_i$  by a path. However, we have to account for the following difficulties: (i) an edge leaving  $U_i$  should also leave  $V_i$ ; (ii) debt should not disappear inside  $U_i$ : if the edge entering it carries debt but the edge leaving does not, we must make sure this difference can be charged to a terminal in  $U_i$ ; (iii) the path used inside  $U_i$  must pay for all expensive edges it uses.

All three issues can be appropriately tackled by defining an *auxiliary graph* inside  $V_i$ . Edges of the auxiliary graph represent paths containing one expensive edge and one terminal (which can pay for themselves); however, these paths may not map to paths in the split graph. We select the subset  $U_i \subseteq V_i$  as a sink component in the auxiliary graph. For convenience, Figs. 2 and 3 give an overview of the different steps, graphs and flows used by our algorithm.

**Construction of auxiliary graphs and modification of split graph.** Our first step is to construct an *auxiliary graph* for each component  $V_i$ . The strong-connectivity structure of this graph will guide our algorithm.

**Definition 3.10.** The auxiliary graph  $G_i^{\text{aux}}$  is a graph with vertex set  $V_i$  and the following edge set: for  $u, v \in V_i$ ,  $(u, v) \in E(G_i^{\text{aux}})$  if any of the following three conditions is satisfied:

- there is a cheap edge  $(u, v) \in E_0 \cap G[V_i]$  inside  $V_i$ , or

| name              | graph            | circulation | integral | obtained from the previous by        |
|-------------------|------------------|-------------|----------|--------------------------------------|
| $x^*$             | $G$              | yes         | no       | solving LP (Section 1.1)             |
| $x_{\text{sp}}^*$ | $G_{\text{sp}}$  | yes         | no       | splitting (Definition 3.6)           |
| $x'_{\text{sp}}$  | $G'_{\text{sp}}$ | yes         | no       | redirecting edges to $A_i$           |
| $y'_{\text{sp}}$  | $G'_{\text{sp}}$ | yes         | yes      | rounding to integrality (Lemma 3.12) |
| $y_{\text{sp}}$   | $G_{\text{sp}}$  | no          | yes      | redirecting edges back from $A_i$    |
| $y$               | $G$              | no          | yes      | mapping back to $G$                  |
| $F$               | $G$              | yes         | yes      | adding walks $P_i$                   |

Figure 2: This table summarizes the various circulations and pseudo-flows that appear in our algorithm, in order.

- *there is a  $u$ - $v$ -path in  $G[V_i]$  whose first edge is expensive and all other edges are cheap, and  $v \in T$  is a terminal – we then call the edge  $(u, v) \in E(G_i^{\text{aux}})$  a postpaid edge – or*
- *there is a  $u$ - $v$ -path in  $G[V_i]$  whose last edge is expensive and all other edges are cheap, and  $u \in T$  is a terminal – we then call the edge  $(u, v) \in E(G_i^{\text{aux}})$  a prepaid edge.*

Define the preimage of such an edge  $(u, v) \in E(G_i^{\text{aux}})$  to be a shortest path inside  $V_i$  as above (in the first case, a single edge).

Now, for each  $i$  consider a decomposition of  $G_i^{\text{aux}}$  into strongly connected components.<sup>5</sup> Let  $U_i \subseteq V_i$  be the vertex set of a sink component in this decomposition. That is, there is no edge from  $U_i$  to  $V_i \setminus U_i$  in the auxiliary graph  $G_i^{\text{aux}}$ . Note that  $G_i^{\text{aux}}$  is constructed based only on the original graph  $G$  and not the split graph  $G_{\text{sp}}$ . However, we will solve Local-Connectivity ATSP by solving an integral circulation problem on  $G'_{\text{sp}}$ : a modification of the split graph  $G_{\text{sp}}$ , described as follows.

For each  $i$ , define  $U_i^{\text{sp}} = \{v^0, v^1 : v \in U_i\} \subseteq V(G_{\text{sp}})$  to be the set of vertices in the split graph corresponding to  $U_i$ . (Note that  $U_i^{\text{sp}}$  may not be strongly connected in  $G_{\text{sp}}$ .) We are going to reroute part of the  $x_{\text{sp}}^*$  flow going in and out of  $U_i^{\text{sp}}$  to a new auxiliary vertex  $A_i$ . While the 3-light algorithm for unit-weight graphs rerouted flow from all boundary edges of a component  $U_i$  (see Section 1.2), here we will be more careful and choose only a subset of boundary edges of  $U_i^{\text{sp}}$  to be rerouted.

To this end, select a subset of edges  $X_i^- \subseteq \delta^-(U_i^{\text{sp}})$  with  $x_{\text{sp}}^*(X_i^-) = 1/2$  such that either all edges in  $X_i^-$  are debt edges, or all are free edges. This is possible since  $x_{\text{sp}}^*(U_i^{\text{sp}}) \geq 1$  by Fact 3.9.<sup>6</sup>

We define the set of outgoing edges  $X_i^+ \subseteq \delta^+(U_i^{\text{sp}})$  to be, intuitively, the edges over which the flow that entered  $U_i^{\text{sp}}$  by  $X_i^-$  exits  $U_i^{\text{sp}}$ . That is, consider an arbitrary cycle decomposition of the circulation  $x_{\text{sp}}^*$ , and look at the set of cycles containing the edges in  $X_i^-$ . We define  $X_i^+$  as the set of edges on these cycles that first leave  $U_i^{\text{sp}}$  after entering  $U_i^{\text{sp}}$  on an edge in  $X_i^-$ ; clearly,  $x_{\text{sp}}^*(X_i^+) = 1/2$ .<sup>7</sup>

Let  $g_i$  denote the flow on these cycles connecting the heads of edges in  $X_i^-$  and the tails of edges in  $X_i^+$ . We will use the following claim later in the construction.

**Fact 3.11.** *Assume all edges in  $X_i^-$  are debt edges but  $e \in X_i^+$  is a free edge or an expensive edge. Then there exists a path in  $G_{\text{sp}}[U_i]$  between a vertex  $t^0$  (for some terminal  $t \in T$ ) and the tail of  $e$ , made up of only cheap edges.*

<sup>5</sup>Note that we decompose the vertex set  $V_i$ , but with respect to the edge set  $E(G_i^{\text{aux}})$ , not  $E(G[V_i])$ .

<sup>6</sup>To obtain exactly  $1/2$ , we might need to break an edge up into two copies, dividing its  $x_{\text{sp}}^*$ -value between them appropriately, and include one copy in  $X_i^-$  but not the other; we omit this for simplicity of notation, and assume there is such an edge set with exactly  $x_{\text{sp}}^*(X_i^-) = 1/2$ .

<sup>7</sup>Again, we might need to split some edges into two copies.



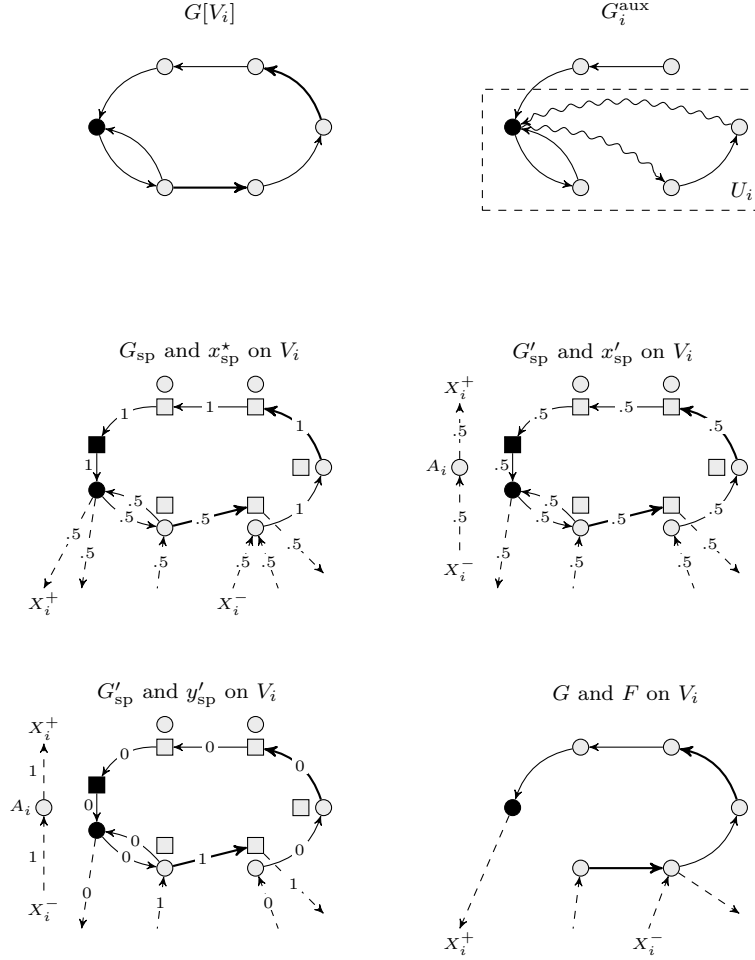


Figure 3: An illustration of the different graphs and flows (restricted to a single component  $V_i$ ) used in our algorithm for Local-Connectivity ATSP. The black vertices depict terminals. In the split graphs, we depict debt vertices by squares and free vertices by circles.

*Proof.* Consider the cycle in the cycle decomposition that contains  $e$ ; it enters  $U_i$  on a debt edge. Using Fact 3.7, this cycle fragment must contain an edge of the form  $(t^1, t^0)$ ; pick the last such edge. All edges that follow are free and cheap.  $\square$

We now transform  $G_{\text{sp}}$  into a new graph  $G'_{\text{sp}}$  and  $x_{\text{sp}}^*$  into new circulation  $x'_{\text{sp}}$  as follows. For every set  $V_i$  in the partition we introduce a new auxiliary vertex  $A_i$  and redirect all edges in  $X_i^-$  to point to  $A_i$  and those in  $X_i^+$  to point from  $A_i$ . We further subtract the flow  $g_i$  inside  $U_i^{\text{sp}}$ ; hence the resulting vector  $x'_{\text{sp}}$  will be a circulation, with  $x'_{\text{sp}}(\delta^-(A_i)) = 1/2$ . If  $X_i^-$  is a set of free edges, then we will say that  $A_i$  is a free vertex, otherwise we say that it is a debt vertex.

**Transforming  $x'_{\text{sp}}$  into an integral flow and obtaining our solution  $F$ .** In the next step we round  $x'_{\text{sp}}$  to integrality while respecting degrees of vertices:

**Lemma 3.12.** *There exists an integral circulation  $y'_{\text{sp}}$  on  $G'_{\text{sp}}$  satisfying the following conditions:*

- $y'_{\text{sp}}(\delta^-(v)) \leq \lceil 2x_{\text{sp}}^*(\delta^-(v)) \rceil$  for each  $v \in V(G_{\text{sp}})$ ,

- $y'_{\text{sp}}(\delta^-(A_i)) = 1$  for each  $i$ .

Such a circulation  $y'_{\text{sp}}$  can be found in polynomial time.

*Proof.* The bounds are integral, and there exists a fractional circulation which satisfies them, namely  $2x'_{\text{sp}}$ .  $\square$

We will now transform  $y'_{\text{sp}}$  into an Eulerian set of edges  $F$  in the original graph  $G$ . We can think of this as a three-stage process.

First, we map all edges adjacent to the auxiliary vertices  $A_i$  back to their preimages in  $G_{\text{sp}}$ , obtaining from  $y'_{\text{sp}}$  an integral *pseudo-flow*  $y_{\text{sp}}$  in  $G_{\text{sp}}$ . (We use the term *pseudo-flow* as now, some vertices may not satisfy flow conservation.)

Second, we contract the two copies  $v^0$  and  $v^1$  of every vertex  $v \in V$ , thus mapping all edges back to their preimages in  $G$ . (We remove all edges  $(t^1, t^0)$  for  $t \in T$ .) This creates an integral pseudo-flow  $y$  in  $G$ .

Since the in- and out-degree of  $A_i$  were exactly 1 in  $y'_{\text{sp}}$ , now (in  $y$ ) in each component  $U_i$  there is a pair of vertices  $u_i, v_i$  which are the head and tail, respectively, of the mapped-back edges adjacent to  $A_i$ . These are the only vertices where flow conservation in  $y$  can be violated.<sup>8</sup> As the third step, to repair this, we route a walk  $P_i$  from  $u_i$  to  $v_i$ , as described below. Our Eulerian set of edges  $F \subseteq E$  which we finally return is the integral pseudo-flow  $y$  plus the union (over  $i$ ) of all such walks  $P_i$ , i.e.,  $\mathbb{1}_F = y + \sum_i \mathbb{1}_{P_i}$ .

It remains to describe how we route these paths. Fix  $i$ . Recall that  $U_i$  is strongly connected in  $G_i^{\text{aux}}$ . We distinguish two cases:

- If  $A_i$  is a free vertex or the edge exiting  $A_i$  in  $y'_{\text{sp}}$  (in  $G'_{\text{sp}}$ ) is a debt edge, then select a shortest  $u_i$ - $v_i$ -path in  $G_i^{\text{aux}}$ , map each edge of this path to its preimage path (see Definition 3.10) and concatenate them to obtain a  $u_i$ - $v_i$ -walk  $P_i$  in  $V_i$ .<sup>9</sup>
- If  $A_i$  is a debt vertex but the edge exiting  $A_i$  in  $y'_{\text{sp}}$  (in  $G'_{\text{sp}}$ ) is a free edge, then by Fact 3.11 there is a terminal  $t$  inside  $U_i$ , with a path from  $t$  to  $v_i$  using only cheap edges.<sup>10</sup> Proceed as above to obtain a  $u_i$ - $t$ -walk and then append this cheap  $t$ - $v_i$ -path to it, obtaining a  $u_i$ - $v_i$ -walk  $P_i$  in  $V_i$ .

This concludes the description of the algorithm. In Sections 3.2.1 and 3.2.2 we prove that the returned Eulerian set of edges  $F$  has the properties we desire, i.e.,

**Lemma 3.13.** *For every connected component  $\tilde{G}$  of  $(V, F)$  we have  $w(\tilde{G}) \leq 10 \cdot \overline{\text{lb}}(\tilde{G})$ .*

**Lemma 3.14.** *For every component  $V_i$  we have  $|\delta_F^+(V_i)| \geq 1$ .*

Lemmas 3.5 and 3.13 together prove that our algorithm is 100-light with respect to lb.

### 3.2.1 Bounding the cost – proof of Lemma 3.13

For this section, let us fix  $\tilde{G}$  to be a connected component of  $(V, F)$ . We want to prove the following lightness claim:  $w(\tilde{G}) \leq 10 \cdot \overline{\text{lb}}(\tilde{G})$ .

Intuitively, our solution is cheap because we built the split graph  $G_{\text{sp}}$  so as to ensure that any circulation in  $G_{\text{sp}}$  which roughly respects the degree bounds given by  $x_{\text{sp}}^*$  has low cost, and

<sup>8</sup>It is violated unless  $u_i = v_i$ .

<sup>9</sup>Note that this walk may exit  $U_i$ , but it will stay inside  $V_i$ .

<sup>10</sup>Map the path given by Fact 3.11 from  $G_{\text{sp}}$  to  $G$ .

because our pseudoflow  $y$  is not too far from being a circulation (in particular, the cost of walks  $P_i$  can be accounted for). We make this argument precise below.

First, recall that the edges in  $F$  are edges in the integral pseudo-flow  $y$  and the edges of the walks  $P_1, P_2, \dots, P_k$ . For a walk  $P_i$ , we write  $P_i \subseteq \tilde{G}$  if  $P_i$  is contained in  $\tilde{G}$  (note that a walk is either contained in or disjoint from  $\tilde{G}$  because  $\tilde{G}$  is a connected component of  $(V, F)$ ). Hence

$$w(\tilde{G}) = \sum_{e \in E(\tilde{G})} y_e w(e) + \sum_{i: P_i \subseteq \tilde{G}} w(P_i). \quad (4)$$

We start by analyzing the first term:  $\sum_{e \in E(\tilde{G})} y_e w(e)$ . Recall that  $y$  is obtained by mapping  $y_{\text{sp}}$  from  $G_{\text{sp}}$  to  $G$ . Let  $\tilde{G}_{\text{sp}}$  denote the pre-image of  $\tilde{G}$  in  $G_{\text{sp}}$ . It will be convenient to define the *debt* of  $y_{\text{sp}}$  as the number of “unpaid” expensive edges in this component:

$$\text{debt} = \sum_{e \in E(\tilde{G}_{\text{sp}}): w_{\text{sp}}(e) = w_1} y_{\text{sp}}(e) - \sum_{t \in V(\tilde{G}) \cap T} y_{\text{sp}}(t^1, t^0).$$

That is, the debt of  $y_{\text{sp}}$  on  $\tilde{G}_{\text{sp}}$  is the difference between the  $y_{\text{sp}}$ -flow on all expensive edges inside  $\tilde{G}_{\text{sp}}$  (those edges “incur debt”) and the  $y_{\text{sp}}$ -flow on all edges  $(t^1, t^0)$  inside  $\tilde{G}_{\text{sp}}$  (those edges “discharge debt”). Note also that, by the definition of  $y$  from  $y_{\text{sp}}$ ,  $\sum_{e \in E(\tilde{G}_{\text{sp}}): w_{\text{sp}}(e) = w_1} y_{\text{sp}}(e) = \sum_{e \in E(\tilde{G}): w(e) = w_1} y(e)$ .

By the construction of  $G_{\text{sp}}$  we have the following upper bound on debt.

**Lemma 3.15.** *Let  $e_i^-$  and  $e_i^+$  be the incoming and outgoing edges of  $A_i$  in  $y'_{\text{sp}}$ . For any  $i$ , define*

$$\text{bad}(i) = \begin{cases} 1 & \text{if } e_i^- \text{ is a debt edge and } e_i^+ \text{ is a free edge,} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\text{debt} \leq \sum_{i: P_i \subseteq \tilde{G}} \text{bad}(i)$$

Moreover, if  $y_{\text{sp}}$  is a circulation, then  $\text{debt} = 0$ .

*Proof.* In  $G_{\text{sp}}$ , the only edges of the form  $(u^0, v^1)$  are the expensive edges, and the only edges of the form  $(u^1, v^0)$  are edges  $(t^1, t^0)$  with  $t \in T$ . Hence, if  $y_{\text{sp}}$  is a circulation, then

$$\begin{aligned} \sum_{e \in E(\tilde{G}_{\text{sp}}): w_{\text{sp}}(e) = w_1} y_{\text{sp}}(e) &= y_{\text{sp}}(\delta^-(\{v^1 : v \in V(\tilde{G})\})) \\ &= y_{\text{sp}}(\delta^+(\{v^1 : v \in V(\tilde{G})\})) = \sum_{t \in V(\tilde{G}) \cap T} y_{\text{sp}}(t^1, t^0), \end{aligned}$$

which shows that  $\text{debt} = 0$  in this case.

Now, if  $y_{\text{sp}}$  is not a circulation, then one can imagine transforming it into one by connecting the “dangling” endpoints of  $e_i^-$  and  $e_i^+$  in  $G_{\text{sp}}$  using a new “virtual” edge for each  $i$ ; this edge would be from a debt vertex to a free vertex if and only if  $e_i^-$  is a debt edge and  $e_i^+$  is a free edge. Hence, by the calculations above, debt is upper-bounded by the number of such new edges in our component. (We do not have equality as the new edges may also be introduced from free vertices to debt vertices.)  $\square$

Using the above lemma, degree bounds and basic calculations, we upper-bound the term  $\sum_{e \in E(\tilde{G})} y_e w(e)$ :

**Lemma 3.16.** *We have  $\sum_{e \in E(\tilde{G})} y_e w(e) \leq 6 \bar{\text{lb}}(\tilde{G})$ .*

*Proof.* We begin by writing:

$$\begin{aligned} \sum_{e \in E(\tilde{G})} y_e w(e) &\leq w_0 \sum_{e \in E(\tilde{G})} y_e + w_1 \sum_{e \in E(\tilde{G}_{\text{sp}}): w_{\text{sp}}(e) = w_1} y_{\text{sp}}(e) \\ &= w_0 \sum_{e \in E(\tilde{G})} y_e + w_1 \left( \text{debt} + \sum_{t \in V(\tilde{G}) \cap T} y_{\text{sp}}(t^1, t^0) \right) \\ &= w_0 \sum_{v \in V(\tilde{G})} y(\delta^-(v)) + w_1 \sum_{t \in V(\tilde{G}) \cap T} y'_{\text{sp}}(t^1, t^0) + w_1 \cdot \text{debt}. \end{aligned}$$

We bound the first term. Note that for each  $v \in V$  we have  $y(\delta^-(v)) = y_{\text{sp}}(\delta^-(\{v^0, v^1\}))$  and  $x^*(\delta^-(v)) = x_{\text{sp}}^*(\delta^-(\{v^0, v^1\}))$ . Lemma 3.12 guarantees that  $y'_{\text{sp}}(\delta^-(v^j)) \leq \lceil 2x_{\text{sp}}^*(\delta^-(v^j)) \rceil$  for  $j \in \{0, 1\}$ . This implies that  $y'_{\text{sp}}(\delta^-(\{v^0, v^1\})) \leq 2x^*(\delta^-(v)) + 2$ , because:

- If  $v \in T$ , then  $v^1$  has only one outgoing edge in  $G_{\text{sp}}$ , which goes to  $v^0$ , and thus for any circulation  $c$  in  $G_{\text{sp}}$  we have  $c(\delta^-(\{v^0, v^1\})) = c(\delta^-(v^0))$ . It follows that  $y'_{\text{sp}}(\delta^-(\{v^0, v^1\})) = y'_{\text{sp}}(\delta^-(v^0)) \leq \lceil 2x_{\text{sp}}^*(\delta^-(v^0)) \rceil \leq 2x_{\text{sp}}^*(\delta^-(v^0)) + 1 = 2x_{\text{sp}}^*(\delta^-(\{v^0, v^1\})) + 1 = 2x^*(\delta^-(v)) + 1$ .
- If  $v \notin T$ , then there are no edges between  $v^0$  and  $v^1$  in  $G_{\text{sp}}$ , and thus for any circulation  $c$  in  $G_{\text{sp}}$  we have  $c(\delta^-(\{v^0, v^1\})) = c(\delta^-(v^0)) + c(\delta^-(v^1))$ . It similarly follows that  $y'_{\text{sp}}(\delta^-(\{v^0, v^1\})) = y'_{\text{sp}}(\delta^-(v^0)) + y'_{\text{sp}}(\delta^-(v^1)) \leq \lceil 2x_{\text{sp}}^*(\delta^-(v^0)) \rceil + \lceil 2x_{\text{sp}}^*(\delta^-(v^1)) \rceil \leq 2x_{\text{sp}}^*(\delta^-(v^0)) + 2x_{\text{sp}}^*(\delta^-(v^1)) + 2 = 2x_{\text{sp}}^*(\delta^-(\{v^0, v^1\})) + 2 = 2x^*(\delta^-(v)) + 2$ .

As  $y_{\text{sp}}$  is the same as  $y'_{\text{sp}}$  except for the edges redirected from the auxiliary vertices in  $G_{\text{sp}}$  (which may increase the in-degree of a vertex by at most 1),

$$y(\delta^-(v)) = y_{\text{sp}}(\delta^-(\{v^0, v^1\})) \leq y'_{\text{sp}}(\delta^-(\{v^0, v^1\})) + 1 \leq 2x^*(\delta^-(v)) + 3 \leq 5x^*(\delta^-(v)),$$

where we used  $x^*(\delta^-(v)) \geq 1$  for the last inequality. Therefore

$$w_0 \sum_{e \in E(\tilde{G})} y_e = w_0 \sum_{v \in V(\tilde{G})} y(\delta^-(v)) \leq 5 \cdot w_0 \sum_{v \in V(\tilde{G})} x^*(\delta^-(v)).$$

We bound the second term similarly:

$$y'_{\text{sp}}(t^1, t^0) = y'_{\text{sp}}(\delta^-(t^1)) \leq \lceil 2x_{\text{sp}}^*(\delta^-(t^1)) \rceil \leq 2 \lceil x_{\text{sp}}^*(\delta^-(t^1)) \rceil = 2 \lceil f(\delta^-(t)) \rceil.$$

Plugging both in we get:

$$\begin{aligned} \sum_{e \in E(\tilde{G})} y_e w(e) &\leq 5 \cdot w_0 \sum_{v \in V(\tilde{G})} x^*(\delta^-(v)) + 2 \cdot w_1 \sum_{t \in (\tilde{G}) \cap T} \lceil f(\delta^-(t)) \rceil + w_1 \cdot \text{debt} \\ &\leq 5 \cdot \left( w_0 \sum_{v \in V(\tilde{G})} x^*(\delta^-(v)) + w_1 \sum_{t \in V(\tilde{G}) \cap T} \lceil f(\delta^-(t)) \rceil \right) + w_1 \cdot \text{debt} \\ &= 5 \bar{\text{lb}}(\tilde{G}) + w_1 \cdot \text{debt}. \end{aligned}$$

(For the final equality, recall the definition of the function  $\overline{\text{lb}}$ .) We will be done if we show that  $w_1 \cdot \text{debt} \leq \overline{\text{lb}}(\tilde{G})$ . By Lemma 3.15 we have  $\text{debt} \leq \sum_{i:P_i \subseteq \tilde{G}} \text{bad}(i)$ . Whenever  $\text{bad}(i) = 1$ , Fact 3.7 implies that  $P_i$  contains a terminal, and so  $\overline{\text{lb}}(P_i) \geq w_1$ . Consequently,

$$w_1 \cdot \text{debt} \leq \sum_{i:P_i \subseteq \tilde{G}, \text{bad}(i)=1} w_1 \leq \sum_{i:P_i \subseteq \tilde{G}, \text{bad}(i)=1} \overline{\text{lb}}(P_i) \leq \overline{\text{lb}}(\tilde{G}).$$

In the last inequality we used that the different  $P_i$ 's are disjoint subsets of  $\tilde{G}$  (each  $P_i$  is contained in  $V_i$ ). The statement follows.  $\square$

It remains to argue that the cost of the walks  $P_i$  can be accommodated.

**Lemma 3.17.** *For every  $i$  we have  $w(P_i) \leq 4 \cdot \overline{\text{lb}}(P_i)$ .*

*Proof.* This holds because in  $P_i$ , each vertex has low indegree and expensive edges can be offset against terminals. Concretely, we have the following two claims:

**Claim 6.** *For each  $v \in V(P_i)$  we have  $|\delta_{P_i}^-(v)| \leq 4$ . (In particular,  $|E(P_i)| \leq 4|V(P_i)|$ .)*

*Proof.* This follows from the fact that we select a *shortest*  $u_i$ - $v_i$ -path (or  $u_i$ - $t$ -path) in  $G_i^{\text{aux}}$  (see Definition 3.10). This implies that each vertex  $v$  appears as an internal vertex on at most one preimage path of a prepaid edge (otherwise we could shortcut the path in  $G_i^{\text{aux}}$ ). Same applies for postpaid edges. And similarly, each  $v$  appears as head of at most one preimage path of any kind (single cheap edge, prepaid or postpaid). This means that  $v$  has indegree at most 3 on the walk created from preimages of edges in  $G_i^{\text{aux}}$ . The path from  $t$  to  $u_i$  can contribute a fourth incoming edge.  $\square$

**Claim 7.** *The number of expensive edges on  $P_i$  is at most twice the number of terminals on  $P_i$ , i.e.,  $|E_1 \cap E(P_i)| \leq 2 \cdot |T \cap V(P_i)|$ .*

*Proof.* Expensive edges appear only in preimage paths of prepaid or postpaid edges, one per such preimage path, and such a path also contains a terminal (as its head or tail). A terminal can only appear as head of one prepaid and as tail of one postpaid edge (otherwise, again, we could shortcut the path in  $G_i^{\text{aux}}$ ).  $\square$

Having these two claims, we can bound the cost:

$$\begin{aligned} w(P_i) &\leq w_0 \cdot |E(P_i)| + w_1 \cdot |E_1 \cap E(P_i)| \\ &\leq w_0 \cdot 4|V(P_i)| + w_1 \cdot 2|T \cap V(P_i)| \\ &\leq 4(w_0 \cdot |V(P_i)| + w_1 \cdot |T \cap V(P_i)|) \\ &\leq 4 \cdot \overline{\text{lb}}(P_i). \end{aligned} \quad \square$$

Now we have all the tools needed to prove our lightness claim, i.e., to upper-bound (4) by  $10 \cdot \overline{\text{lb}}(\tilde{G})$ :

$$\begin{aligned} w(\tilde{G}) &= \sum_{e \in E(\tilde{G})} y_e w(e) + \sum_{i:P_i \subseteq \tilde{G}} w(P_i) \\ &\leq 6 \cdot \overline{\text{lb}}(\tilde{G}) + \sum_{i:P_i \subseteq \tilde{G}} 4 \cdot \overline{\text{lb}}(P_i) \\ &\leq 6 \cdot \overline{\text{lb}}(\tilde{G}) + 4 \cdot \overline{\text{lb}}(\tilde{G}) \\ &= 10 \cdot \overline{\text{lb}}(\tilde{G}), \end{aligned}$$

where line 2 is by Lemmas 3.16 and 3.17, and in line 3 we use that all walks  $P_i$  are disjoint and in  $\tilde{G}$ . This completes the proof of Lemma 3.13.  $\square$

### 3.2.2 Crossing the cuts – proof of Lemma 3.14

In this section we prove that our solution  $F \subseteq E$  crosses each component  $V_i$ . Let us recall that  $U_i$  is a sink strongly-connected component of the graph  $G_i^{\text{aux}}$ , so there is no edge in  $G_i^{\text{aux}}$  from  $U_i$  to  $V_i \setminus U_i$ . However, there can be such edges in  $G$  and therefore it might not be the case that the edge in  $F$  which leaves  $U_i$  also leaves  $V_i$ . We will however argue that  $F$  does contain some edge leaving  $V_i$ . Assume towards a contradiction that this is not true, i.e., that  $V_i$  is a union of connected components of the graph  $(V, F)$ .

Fix  $i$ . Consider the (only) edge in  $y'_{\text{sp}}$  (in  $G'_{\text{sp}}$ ) with tail  $A_i$ ; let  $e_{\text{sp}}$  be its image in  $y_{\text{sp}}$  (in  $G_{\text{sp}}$ ) and  $e$  its image in  $y$  (in  $G$ ). We have  $e_{\text{sp}} \in X_i^+$ ; the tail of  $e$  is  $v_i \in U_i$ , and the head of  $e$  is in  $V_i \setminus U_i$  (since we assumed that no edge of  $F$  leaves  $V_i$ ).

The following claim is the first, simplest example of how we use the structure of  $G_i^{\text{aux}}$  to reason about our solution  $F$ .

**Claim 8.** *Any edge from  $U_i$  to  $V_i \setminus U_i$  in  $G$  is expensive. (Therefore  $e$  is expensive, and  $e_{\text{sp}}$  is a debt edge.)*

*Proof.* If there was such a cheap edge, it would also appear in  $G_i^{\text{aux}}$ . However, there is no edge in  $G_i^{\text{aux}}$  from  $U_i$  to  $V_i \setminus U_i$ .  $\square$

Define  $(V_i \setminus U_i)^{\text{sp}} = \{v^0, v^1 : v \in V_i \setminus U_i\} \subseteq V(G_{\text{sp}})$ . We can now start traversing  $y_{\text{sp}}$  inside  $(V_i \setminus U_i)^{\text{sp}}$  like an Eulerian graph (since  $y_{\text{sp}}$  satisfies flow conservation in  $(V_i \setminus U_i)^{\text{sp}}$ ), starting from  $e_{\text{sp}}$ , until we return to  $U_i^{\text{sp}}$ .

**Claim 9.** *We will not reach a terminal before returning to  $U_i^{\text{sp}}$ . (Therefore we will return on a debt edge  $e'_{\text{sp}}$ .)*

*Proof.* If there was a terminal  $t \in T \cap (V_i \setminus U_i)$  such that there is a path from  $\text{head}(e_{\text{sp}})$  (a debt vertex) to  $t^1$  in  $G_{\text{sp}}[V_i]$ , then (assuming without loss of generality that there is no other terminal on this path) all edges of this path are cheap. Therefore  $e$  together with the image of this path in  $G$  would give rise to a postpaid edge  $(\text{tail}(e), t)$  from  $U_i$  to  $V_i \setminus U_i$  in  $G_i^{\text{aux}}$ , a contradiction. Since  $\text{head}(e_{\text{sp}})$  is a debt vertex and we do not visit a terminal, we will return on a debt edge.  $\square$

Now we distinguish two cases. Suppose that the debt edge  $e'_{\text{sp}}$  on which we return to  $U_i^{\text{sp}}$  is *not* the image in  $G_{\text{sp}}$  of the (only) edge in  $y'_{\text{sp}}$  (in  $G'_{\text{sp}}$ ) with head  $A_i$ . This means that we can keep following  $y$  inside  $U_i^{\text{sp}}$  until we exit  $U_i^{\text{sp}}$  again via some edge  $e''_{\text{sp}}$  (maybe  $e''_{\text{sp}} = e_{\text{sp}}$ ). By Claim 8,  $e''_{\text{sp}}$  is expensive. So we have followed a path (a segment of  $y$ ) which contains a debt edge and later an expensive edge; this means (see Fact 3.7) that it must also contain a terminal in between. Pick the last such terminal  $t$ ; then the segment of  $y$  between  $t^0$  and  $\text{tail}(e''_{\text{sp}})$  must consist of cheap (free) edges. This gives rise to a prepaid edge  $(t, \text{head}(e''_{\text{sp}}))$  from  $U_i$  to  $V_i \setminus U_i$  in  $G_i^{\text{aux}}$ , a contradiction.

So suppose instead that the debt edge  $e'_{\text{sp}}$  on which we return to  $U_i^{\text{sp}}$  is, in fact, the image in  $G_{\text{sp}}$  of the (only) edge in  $y'_{\text{sp}}$  (in  $G'_{\text{sp}}$ ) with head  $A_i$ . This means that  $e'_{\text{sp}} \in X_i^-$ . Recall that by definition,  $X_i^-$  consists only of debt or only of free edges, and therefore every edge in  $X_i^-$  must be a debt edge. By Fact 3.11,  $G_{\text{sp}}[U_i]$  contains a path from a terminal  $t^0$  via cheap edges to the expensive edge  $e_{\text{sp}}$ . Again, the image of this path in  $G$  gives rise to a prepaid edge  $(t, \text{head}(e_{\text{sp}}))$  from  $U_i$  to  $V_i \setminus U_i$  in  $G_i^{\text{aux}}$ , a contradiction. This concludes the proof of Lemma 3.14.  $\square$

## Acknowledgment

An earlier version of this paper has appeared in the proceedings of IPCO 2016 (the 18th Conference on Integer Programming and Combinatorial Optimization). This work has been published in Mathematical Programming Series B. The final publication is available at Springer via <http://dx.doi.org/10.1007/s10107-017-1195-7>.

## References

- [AG15] Nima Anari and Shayan Oveis Gharan. Effective-resistance-reducing flows and asymmetric TSP. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2015.
- [AGK<sup>+</sup>98] Sanjeev Arora, Michelangelo Grigni, David R. Karger, Philip N. Klein, and Andrzej Woloszyn. A polynomial-time approximation scheme for weighted planar graph TSP. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 1998*, pages 33–41, 1998.
- [AGM<sup>+</sup>10] Arash Asadpour, Michel X. Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. An  $O(\log n / \log \log n)$ -approximation algorithm for the asymmetric traveling salesman problem. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010*, pages 379–389, 2010.
- [BK06] Piotr Berman and Marek Karpinski.  $8/7$ -approximation algorithm for  $(1, 2)$ -TSP. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 641–648, 2006.
- [Blä04] Markus Bläser. A  $3/4$ -approximation algorithm for maximum ATSP with weights zero and one. In *Approximation, Randomization, and Combinatorial Optimization, Algorithms and Techniques, 7th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX*, pages 61–71, 2004.
- [Chr76] Nicos Christofides. Worst-case analysis of a new heuristic for the travelling salesman problem. Technical report, DTIC Document, 1976.
- [ES14] Jeff Erickson and Anastasios Sidiropoulos. A near-optimal approximation algorithm for asymmetric TSP on embedded graphs. In *30th Annual Symposium on Computational Geometry (SOCG)*, page 130, 2014.
- [FGM82] Alan M. Frieze, Giulia Galbiati, and Francesco Maffioli. On the worst-case performance of some algorithms for the asymmetric traveling salesman problem. *Networks*, 12(1):23–39, 1982.
- [GKP95] Michelangelo Grigni, Elias Koutsoupias, and Christos H. Papadimitriou. An approximation scheme for planar graph TSP. In *36th Annual Symposium on Foundations of Computer Science, FOCS 1995*, pages 640–645, 1995.
- [GS11] Shayan Oveis Gharan and Amin Saberi. The asymmetric traveling salesman problem on graphs with bounded genus. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 967–975. SIAM, 2011.

- [GSS11] Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. A randomized rounding approach to the traveling salesman problem. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011*, pages 550–559, 2011.
- [KLS15] Marek Karpinski, Michael Lampis, and Richard Schmied. New inapproximability bounds for TSP. *J. Comput. Syst. Sci.*, 81(8):1665–1677, 2015.
- [MS11] Tobias Mömke and Ola Svensson. Approximating graphic TSP by matchings. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011*, pages 560–569, 2011.
- [Muc12] Marcin Mucha. 13/9-approximation for graphic TSP. In *29th International Symposium on Theoretical Aspects of Computer Science, STACS 2012*, pages 30–41, 2012.
- [PY93] Christos H. Papadimitriou and Mihalis Yannakakis. The traveling salesman problem with distances one and two. *Math. Oper. Res.*, 18(1):1–11, 1993.
- [SV14] András Sebő and Jens Vygen. Shorter tours by nicer ears: 7/5-approximation for the graph-TSP, 3/2 for the path version, and 4/3 for two-edge-connected subgraphs. *Combinatorica*, 34(5):597–629, 2014.
- [Sve15] Ola Svensson. Approximating ATSP by relaxing connectivity. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2015. URL: <http://arxiv.org/abs/1502.02051>.
- [WS11] David P. Williamson and David B. Shmoys. *The Design of Approximation Algorithms*. Cambridge University Press, 2011.