An Improved Analysis of the Mömke–Svensson Algorithm for Graph-TSP on Subquartic Graphs (Journal Version)

Alantha Newman

HAL Id: hal-03049522
https://hal.science/hal-03049522

Submitted on 7 Jan 2021

An improved analysis of the Mömke-Svensson algorithm for graph-TSP on subquartic graphs[*]

Alantha Newman[†]

March 4, 2020

Abstract

Mömke and Svensson presented a beautiful new approach for the traveling salesman problem on a graph metric (graph-TSP), which yields a 4/3-approximation guarantee on subcubic graphs as well as a substantial improvement over the 3/2-approximation guarantee of Christofides' algorithm on general graphs. The crux of their approach is to compute an upper bound on the minimum cost of a circulation in a particular network, $C(G, T)$, where G is the input graph and T is a carefully chosen spanning tree. The cost of this circulation is directly related to the number of edges in a tour output by their algorithm. Mucha subsequently improved the analysis of the circulation cost, proving that Mömke and Svensson's algorithm for graph-TSP has an approximation ratio of at most 13/9 on general graphs.

This analysis of the circulation is local, and vertices with degree four or five can contribute the most to its cost. Thus, hypothetically, there could exist a subquartic graph (a graph with degree at most four at each vertex) for which Mucha's analysis of the Mömke-Svensson algorithm is tight. We show that this is not the case and that Mömke and Svensson's algorithm for graph-TSP has an approximation guarantee of at most 25/18 on subquartic graphs. To prove this, we present different methods to upper bound the minimum cost of a circulation on the network $C(G, T)$. Our approximation guarantee holds for all graphs that have an optimal solution for a standard linear programming relaxation of graph-TSP with subquartic support.

1 Introduction

The metric traveling salesman problem (TSP) is one of the most well-known problems in the field of combinatorial optimization and approximation algorithms. Given a complete graph, $G = (V, E)$, with nonnegative edge weights that satisfy the triangle inequality, the goal is to compute a minimum cost tour of G that visits each vertex exactly once. Christofides' algorithm yields a tour with cost no more than 3/2 times that of an optimal tour [Chr76]. It remains a major open problem to improve upon this approximation factor.

In the past few years, there have been many exciting developments relating to *graph-TSP*. In this setting, we are given an unweighted graph $G = (V, E)$, and the goal is to find the shortest tour that visits each vertex *at least* once. This problem is equivalent to the special case of metric TSP where the shortest path distances in G define the metric. It is also equivalent to the problem of finding a connected, spanning, Eulerian multigraph in G with the minimum number of edges.

A promising approach to improving upon the factor of $3/2$ for metric TSP is to round a linear programming relaxation known as the Held-Karp relaxation [HK70]. A lower bound of $4/3$ on its integrality gap can be demonstrated using a family of graph-TSP instances. Since it is widely conjectured that its integrality gap is also upper bounded by $4/3$, proving this for graph-TSP would be a step towards a more comprehensive understanding of the relaxation and would hopefully provide insights applicable to metric TSP. However, even in this special case of metric TSP, graph-TSP had also long resisted significant progress before the recent spate of results.

1.1 Recent progress on graph-TSP

In 2005, Gamarnik, Lewenstein and Sviridenko presented an algorithm for graph-TSP on cubic, 3-edge-connected graphs with an approximation factor of $3/2 - 5/389$ [GLS05], thus proving that Christofides' approximation guarantee of $3/2$ is not the best possible factor for this class of graphs. Their approach is based on finding a cycle cover for which they can upper bound the number of components. This general approach was also taken by Boyd, Sitters, van der Ster and Stougie who combined it with polyhedral ideas to obtain approximation guarantees of $4/3$ for cubic graphs and $7/5$ for subcubic graphs (i.e., graphs with degree at most three at each vertex) [BSvdSS14]. Shortly afterwards, Gharan, Saberi and Singh proved that a subtle modification of Christofides' algorithm has an approximation guarantee of $3/2 - \epsilon_0$ for graph-TSP on general graphs, where ϵ_0 is a fixed constant with value approximately 10^{-12} [GSS11].

Mömke and Svensson then presented a beautiful new approach for graph-TSP, which resulted in a substantial improvement over the $3/2$-approximation guarantee of Christofides [MS16]. Their approach also leads to a simple $\frac{4}{3}$-approximation algorithm for subcubic graphs. We will discuss their algorithm in more detail in Section 1.2, since our paper is directly based on their approach. Ultimately, they were able to prove an approximation guarantee of 1.461 for graph-TSP. Mucha subsequently gave an improved analysis, thereby proving that Mömke and Svensson's algorithm for graph-TSP actually has an approximation ratio of at most $13/9$ [Muc14]. Sebő and Vygen introduced an approach for graph-TSP based on ear decompositions and matroid intersection, which incorporated the techniques of Mömke and Svensson, and improved the approximation ratio to $7/5$, where it currently stands [SV14]. For the special case of k-regular graphs, Vishnoi gave an algorithm for graph-TSP with an approximation guarantee that approaches 1 as k increases [Vis12].

1.2 Mömke-Svensson's approach to graph-TSP

Christofides' algorithm for graph-TSP finds a spanning tree of the graph and adds to it a J-join, where J is the set of vertices that have odd degree in the spanning tree.[1] Since the spanning

---

[1] A J-join of a graph $G = (V, E)$ is a subgraph $F \subset E$ such that the degree of each vertex in $J \subseteq V$ is odd in F and the degree of each vertex not in J is even (and can be zero) in F. An *odd-join* of G is a J-join where $J = V$.

tree is connected, the resulting subgraph is clearly connected, and since the J-join corrects the parity of the spanning tree, the resulting subgraph is Eulerian. In contrast, the approach of Mömke and Svensson is based on removing an odd-join of the graph, which yields a possibly disconnected Eulerian subgraph. Thus, to maintain connectivity, one must double, rather than remove, some of the edges in the odd-join. The key step in proving the approximation guarantee of the algorithm is to show that many edges will actually be removed and relatively few edges will be doubled, resulting in a connected, Eulerian subgraph with few edges. First, Mömke and Svensson design a circulation network, $C(G, T)$, which is constructed based on the input graph G, an optimal solution for a linear programming relaxation for graph-TSP, and a carefully chosen spanning tree T (see Appendix A.1 in [MS16]). Using techniques of Naddef and Pulleyblank [NP81], Mömke and Svensson show how to sample an odd-join of size $|E'|/3$, where E' is the edge set of a 2-vertex-connected subgraph of G, chosen via the circulation network $C(G, T)$. After removing this odd-join from G, the number of doubled edges that are added back to guarantee connectivity (while maintaining the Eulerian property) is directly related to the minimum cost of a circulation of $C(G, T)$. Lemma 4.2 from [MS16] relates this cost to the size of a solution output by their algorithm.

Lemma 1. [**MS16**] *Given a 2-vertex-connected graph G and a depth-first-search (DFS) tree T of G, let z^* be a circulation for $C(G, T)$ with cost $c(z^*)$. Then there is a spanning Eulerian multigraph in G with at most $\frac{4}{3}n + \frac{2}{3}c(z^*)$ edges.*

We discuss the circulation network $C(G, T)$ further in Section 2. For the moment, we emphasize that if one can prove a better upper bound on the value of $c(z^*)$, then this directly implies an improved upper bound on the number of edges in a tour output by Mömke and Svensson's algorithm.

1.3 Our contribution

We consider the graph-TSP problem for *subquartic* graphs (i.e., graphs in which each vertex has degree at most four). The best-known approximation guarantee for these graphs is inherited from the general case, even when the graph is 4-regular, and is therefore 7/5 due to Sebő and Vygen. For subquartic graphs, we give an improved upper bound on the minimum cost of a circulation for $C(G, T)$. Using Lemma 1, this leads to an improved approximation guarantee of 25/18 for graph-TSP on these graphs. Before we give an overview of our approach, we first explain our motivation for studying graph-TSP on this restricted class of graphs.

As mentioned in Section 1.1, graph-TSP is now known to be approximable to within 4/3 for subcubic graphs. So, on the one hand, trying to prove the same guarantee for subquartic graphs is arguably a natural next step. Additionally, it is a well-motivated problem to study the graph-TSP on sparse graphs, because the support of an optimal extreme point solution for the standard linear programming relaxation (reviewed in Section 2.1) has at most $2n - 1$ nonzero edges (see Theorem 4.9 in [CFN85]). Thus, any graph that corresponds to the support of such an optimal solution for the standard linear program has average degree less than four.

However, our actual motivation for studying graphs with degree at most four has more to do with understanding the Mömke-Svensson algorithm than with an abstract interest in subquartic

---

For example, a perfect matching in a cubic graph is an odd-join.

graphs. The basic approach to computing an upper bound on the minimum value of $C(G,T)$ used in both [MS16] and [Muc14] is to specify flow values on the edges of a particular circulation network that are functions of an optimal solution for the linear programming relaxation for graph-TSP on the graph G. The cost of the circulation obtained using these values can be analyzed in a local, vertex by vertex manner. Mucha showed that vertices with degree four or five potentially increase the cost of the circulation the most [Muc14]. In fact, one could hypothetically construct a tight example for Mucha's analysis of the Mömke-Svensson algorithm on a graph where each vertex has degree at most four (or where each vertex has degree at most five). It therefore seems worthwhile to determine if the cost of the circulation can be improved on subquartic graphs. Our results actually hold for a slightly more general class of graphs than subquartic graphs: they hold for any graph that has an optimal solution for the standard linear programming relaxation of graph-TSP with subquartic support.

1.4 Organization

In Section 2.1, we discuss the standard linear programming relaxation for graph-TSP, and in Section 2.2, we discuss the circulation network $C(G,T)$ and how it can be used to find a short tour. In Section 3, we show that if, for a subquartic graph, the optimal solution for the linear program has value equal to the number of vertices in G, then the network $C(G,T)$ has a circulation with cost zero, implying that the Mömke-Svensson algorithm has an approximation ratio of 4/3. This observation provides us with some intuition as to how one might attempt to design a better circulation for general subquartic graphs.

In Section 4, we describe three different methods to obtain feasible circulations. In Section 4.1, we detail the method used by Mömke and Svensson and Mucha, which becomes somewhat simpler in the special case of subquartic graphs. This method directly uses values from the optimal solution for the linear program to obtain a feasible solution for the circulation network. In Section 4.2, we present a new method that "rounds" the values from the optimal solution for the linear program. The latter circulation alone leads to an improved analysis over 13/9 for subquartic graphs, but it does not improve on the best-known guarantee of 7/5. However, as we show in Section 4.3, if we take the best of these two circulations, we can show that at least one of the circulations will lead to an approximation guarantee of at most 46/33. Next, in Section 5 we consider a third method based on extreme point structure to upper bound the optimal cost of a circulation. Combining all three analyses, we obtain an approximation guarantee of 25/18.

We remark that our notation differs from that in [MS16] and [Muc14], even though we are using exactly the same circulation network and we use their approach for obtaining the feasible circulation described in Section 4.1. This different notation allows us to more easily analyze the tradeoff between the different circulations.

2 Preliminaries: Notation and definitions

For $S \subset V$, let $\delta(S) \subset E$ denote the set of edges with exactly one endpoint in S. For $S_1, S_2 \subset V$ such that S_1 and S_2 are disjoint, let (S_1, S_2) denote the edges with exactly one endpoint in S_1 and

the other endpoint in S_2. Throughout this paper, we make use of the following well-studied linear programming relaxation for graph-TSP.

2.1 Linear program for graph-TSP

For a graph $G = (V, E)$, the following linear program is a relaxation of graph-TSP. We refer the reader to Section 2 of [MS16] for a discussion of its derivation and history.

$$\min \sum_{e \in E} y_e$$
$$y(\delta(S)) \geq 2 \quad \text{for } \emptyset \neq S \subset V, \qquad\qquad (LP(G))$$
$$y \geq 0.$$

We denote the feasible region of this linear program by $LP(G)$, and we denote the value of an optimal solution by $OPT_{LP}(G)$. If $x \in LP(G)$ and the sum of coordinates of x equals $OPT_{LP}(G)$, then we say that x is an *optimal solution for* $(LP(G))$. Let n denote the number of vertices in V.

We want to exploit certain properties of an extreme point of $(LP(G))$ such as the fact that it has sparse support. An extreme point $x^* \in LP(G)$ has at most $2n - 1$ edges (see Theorem 4.9 in [CFN85]). In fact, we can obtain a more accurate bound in terms of the number of *heavy* vertices.

Definition 1. *A vertex $v \in V$ is called* heavy *with respect to $y \in LP(G)$ if $y(\delta(v)) > 2$. Let $H(y)$ denote the set of heavy vertices in V with respect to y.*

It is known that the number of nonzero edges in the support of an extreme point $x^* \in LP(G)$ is at most the number of tight constraints in a maximal laminar family, which can have size at most $2n - |H(x^*)| - 1$, since out of the tight constraints in a maximal laminar family, at most $n - 1$ of them can be attributed to tight sets containing of two or more vertices [CFN85].

For the sake of simplicity, we also want to use an optimal solution $x \in LP(G)$ such that $x \leq 1$. The following lemma shows that we can find such an optimal solution (but it might not be an extreme point).

Lemma 2. *Let $G = (V, E)$ be a 2-edge-connected graph. Then there exists $x \in LP(G)$, $x \leq 1$ such that $\sum_{e \in E} x_e = OPT_{LP}(G)$.*

Proof. Let $x \in LP(G)$ be an extreme point such that $\sum_{e \in E} x_e = OPT_{LP}(G)$. Suppose that there is some $x_e > 1$ for $e \in E$. First, we observe that edge e must belong to some tight cut (i.e., there exists $S \subset V$ such that $e \in \delta(S)$ and $x(\delta(S)) = 2$). Since G is 2-edge-connected, each cut crossing edge e must include at least one other edge. If x_e did not belong to any tight cut, then we could decrease the value of x_e and obtain a smaller solution, which is a contradiction to the optimality of x.

Next, we show that edge e can be in at most one tight cut. Towards a contradiction, suppose that e belongs to at least two tight cuts. Consider the cuts $(S \cup A, V \setminus (S \cup A))$ and $(S \cup B, V \setminus (S \cup B))$, where S, A and B are disjoint and $x(\delta(S \cup A)) = x(\delta(S \cup B)) = 2$. Suppose that $e = ij$ and $i \in S$

and $j \in V \setminus (S \cup A \cup B)$. Then edge e crosses both these cuts (i.e., $e \in \delta(S \cup A)$ and $e \in \delta(S \cup B)$).

$$
\begin{aligned}
\delta(S \cup A) &= (S, V \setminus (S \cup A \cup B)) + (S, B) + (A, B) + (A, V \setminus (S \cup A \cup B)), \\
\delta(S \cup B) &= (S, V \setminus (S \cup A \cup B)) + (S, A) + (A, B) + (B, V \setminus (S \cup A \cup B)).
\end{aligned}
$$

Then we have

$$
\begin{aligned}
x(\delta(S \cup A)) + x(\delta(S \cup B)) &= 2 \cdot x(S, V \setminus (S \cup A \cup B)) + 2 \cdot x(A, B) + x(B, S) + x(A, S) \quad (1) \\
&\quad + x(A, V \setminus (S \cup A \cup B)) + x(B, V \setminus (S \cup A \cup B)).
\end{aligned}
$$

Since both of these cuts are tight and since $x_e > 1$ and $e \in (S, V \setminus (S \cup A \cup B))$, it follows that

$$
x(\delta(S \cup A)) + x(\delta(S \cup B)) - 2 \cdot x(S, V \setminus (S \cup A \cup B)) \;<\; 2.
$$

By (1), this implies that

$$
2 \cdot x(A, B) + x(B, S) + x(A, S) + x(A, V \setminus (S \cup A \cup B)) + x(B, V \setminus (S \cup A \cup B)) \;<\; 2. \qquad (2)
$$

However, we know that $x(A \cup B, \; V \setminus (A \cup B)) \geq 2$.

$$
\begin{aligned}
x(A \cup B, V \setminus (A \cup B)) &= x(A, S) + x(A, V \setminus (S \cup A \cup B)) \qquad (3) \\
&\quad + x(B, S) + x(B, V \setminus (S \cup A \cup B)).
\end{aligned}
$$

Since the quantity in (3) is at most the quantity on the left-hand side of (2), it must be strictly less than 2, which is a contradiction. We can conclude that the edge e occurs in at most one tight cut.

Therefore, the only case to consider is when e belongs to exactly one tight cut. Since there is at least one other edge (call it f) besides e crossing this cut (since G is 2-edge-connected) and x_f must have value strictly less than 1, we can increase x_f and decrease x_e. Since the cut is still tight, the solution is still feasible and has the same value as the original solution. Observe that if $x_e = 1 + \eta$, we can simply decrease x_e to 1 and increase x_f by η. Indeed, if this increase of x_e and decrease of x_f results in another cut, which is crossed by edge e, becoming infeasible (i.e., having value strictly less than 2), then there is some value η' (for $0 < \eta' < \eta$) such that after decreasing x_e by η' and increasing x_f by η', we have two tight cuts crossed by edge e. Since this cannot happen by our previous arguments, we conclude that we can simply decrease x_e to 1 and increase x_f by η.

Thus, for each edge e with $x_e > 1$, we can simply find a tight cut (which is a minimum cut and we can enumerate the minimum cuts in polynomial time), decrease x_e to 1 and increase another edge so that this cut remains tight. Since we can do this at most $|E|$ times, we claim that this procedure can be performed efficiently. □

We can actually assume that G is 2-vertex-connected (see Lemma 2.1 from [MS16]). We can also assume that there is an optimal extreme point $x^* \in LP(G)$ whose support is E. (If not, we can restrict G to the support of x^*, which does not increase the value $OPT_{LP}(G)$. Should the resulting graph not be 2-vertex-connected, we can break it up into 2-vertex-connected components, find an optimal extreme point solution for each component and repeat.) Thus, we have the following corollary of Lemma 2.

Corollary 1. *Let $G = (V, E)$ be a 2-vertex-connected graph. Suppose there exists an extreme point $x^* \in LP(G)$ with support E such that $\sum_{e \in E} x_e^* = OPT_{LP}(G)$. Then there is an $x \in LP(G)$ such that the following properties hold:*

(i) $\sum_{e \in E} x_e = \sum_{e \in E} x_e^* = OPT_{LP}(G)$.

(ii) The support of x and the support of x^* are the same.

(iii) The support of x and x^* contains $|E| \leq 2n - |H(x^*)| - 1$ edges.

(iv) $x \leq 1$.

For the rest of this paper, we assume that $G = (V, E)$ is 2-vertex-connected and that x^* is an optimal extreme point of $(LP(G))$ with support E. Moreover, we fix $x \in LP(G)$ to have the properties stated in Corollary 1. We will refer to the set of values $\{x_e\}$ for $e \in E$ as x-values. Let $\sum_{e \in E} x_e = OPT_{LP}(G) = (1 + \epsilon)n$ for some ϵ, where $0 \leq \epsilon \leq 1$.

Definition 2. *The excess x-value $\epsilon(v)$ at a vertex v is the amount by which the total value on the incident edges exceeds 2 (i.e., $\epsilon(v) = x(\delta(v)) - 2$).*

The following fact will be useful in our analysis. If $OPT_{LP}(G) = (1 + \epsilon)n$, then

$$\sum_{v \in V} x(\delta(v)) \quad = \quad \sum_{v \in V} (2 + \epsilon(v)) \quad = \quad 2(1 + \epsilon)n.$$

This implies,

$$\sum_{v \in V} \epsilon(v) \quad = \quad 2\epsilon n. \tag{4}$$

2.2 Spanning trees and circulations

Let us recall some useful definitions from the approach of Mömke and Svensson [MS16] that we use throughout this paper.

Definition 3. *Let $y \in LP(G)$. A greedy DFS tree chosen with respect to y is a spanning tree formed via a depth-first search of G. If there is a choice as to which edge to traverse next, the edge with the highest y-value is chosen.*

For a given graph G and a solution $y \in LP(G)$, let T denote a greedy DFS tree with respect to y. Suppose T has root r, and let $E(T)$ denote the edges in T (i.e., tree edges). We orient $E(T)$ to be an arborescence with root r, and we orient $B(T) := E \setminus E(T)$ "backwards," that is, so that each edge in $B(T)$ forms a directed cycle with a path of the tree. This is possible since T is a DFS tree. We use the notation (i, j) to denote an edge directed from i to j. Note that once we have fixed a tree T, all edges in E can be viewed as directed edges. When we wish to refer to an undirected edge in E, we use the notation $ij \in E$. With respect to the greedy DFS tree T, we have the following definitions.

Definition 4. *An* internal *node in T is a vertex that is neither the root of T nor a leaf in T. We use T_{int} to denote this subset of vertices.*

Definition 5. *A* branch *vertex in T is a vertex with at least two outgoing tree edges.*

Note that the root of T cannot be a branch vertex since G is 2-vertex-connected.

Definition 6. *An* expensive *vertex is a vertex in T_{int} with two incoming edges that belong to $B(T)$. We use T_{exp} to denote this subset of vertices.*

As we will see in Lemma 4, in a subquartic graph, expensive vertices are the only vertices (besides the root) that can contribute to the cost of $C(G, T)$ (which is of interest due to Lemma 1). For the sake of simplicity, we sometimes ignore the contribution of the root in our calculations, since the contribution of the root to the cost of $C(G, T)$ is negligible (at most 2).

Fact 1. *The number of expensive vertices is bounded as follows: $|T_{exp}| \leq |B(T)|/2 \leq n/2$.*

Lemma 3. *If G is subquartic, then a branch vertex in T_{int} is not expensive.*

Proof. In a graph with vertex degree at most four, a branch vertex can have at most one incoming back edge and therefore cannot be expensive. $\square$

Definition 7. *A* tree cut *is the partition of the vertices of the tree T induced when we remove an edge $(u, v) \in T$.*

For each edge $(i, j) \in B(T)$, let $b(i, j) \leq 1$ be a nonnegative value.

Definition 8. *Consider a tree cut corresponding to edge $(u, v) \in T$, and remove all back edges $(w, u) \in B(T)$, where w belongs to the subtree of v in T. We say that the remaining back edges that cross this tree cut* cover *the cut. If the total b-value of the edges that cover the cut is at least 1, then we say that this tree cut is* satisfied *by b.*

We extend this definition to the vertices of T.

Definition 9. *A vertex v in T is* satisfied *by b if for each incident outgoing edge in T, the corresponding tree cut is satisfied by b. On the other hand, if there is at least one incident outgoing edge whose corresponding tree cut is not satisfied by b, then the vertex v is* unsatisfied *by b.*

Mömke and Svensson define a circulation network, $C(G, T)$ (see Appendix A.1 of [MS16]), and use the cost of a feasible circulation to upper bound the length of a TSP tour in G. (See Lemma 1.)[2]

Lemma 4. *Let G be a subquartic graph and T be a DFS tree. Let $b : B(T) \to [0,1]$. If each internal vertex in T is satisfied by b, then there is a feasible integer circulation of $C(G,T)$ whose cost (not including the contribution of the root) is upper bounded by the following function:*

$$\sum_{j \in T_{exp}} \max \left\{ 0, \ \left(\sum_{i:(i,j) \in B(T)} b(i,j) \right) - 1 \right\}. \tag{5}$$

Proof. The purpose of the circulation network $C(G,T)$ in [MS16] is to find a subset of back edges, $B' \subseteq B(T)$, so that $T \cup B'$ is 2-vertex-connected and so that the following cost function (stated here for a subquartic graph) is minimized:

$$\sum_{j \in T_{exp}} \max \left\{ 0, \ \left(\sum_{i:(i,j) \in B'} 1 \right) - 1 \right\}. \tag{6}$$

If $b' : B(T) \to \{0,1\}$ is an integral function and every internal vertex in T is satisfied by b', then the edges with b'-value 1 form a set B' such that $T \cup B'$ is 2-vertex-connected. To obtain an upper bound on the cost of such a 2-vertex-connected subgraph, we can use any fractional values $b : B(T) \to [0,1]$ such that each internal vertex in T is satisfied by b. This is made explicit in the formulation of $LP(G,T)$ and the subsequent discussion on page 11 of [MS16]; observe that these fractional b-values are a feasible solution for $LP(G,T)$. In the case of subquartic graphs, the only vertices that can contribute to the objective function in $LP(G,T)$ (and therefore to the cost function (5)) are the expensive vertices, since the maximum value allowed on an edge is 1. Thus, in this special case, the simplified cost function (5) is equivalent to the objective function for $LP(G,T)$ and hence for the cost of the circulation for the network $C(G,T)$ used in [MS16]. For a formal description of $C(G,T)$ and for the equivalence between the objective function of $LP(G,T)$ and the cost of a circulation of $C(G,T)$, we refer the reader to the proof of Lemma 4.1 in [MS16]. $\square$

From Lemma 1, we see that if we find a circulation with cost zero for $C(G,T)$, then G has a TSP tour of length at most $\frac{4}{3}n$. We note that in [Muc14], it is shown that for a graph G for which $OPT_{LP}(G) = n$, there is a circulation with cost at most $\frac{n}{6}$ for $C(G,T)$, which results in a TSP tour of length at most $\frac{13}{9}n$.

2.3 Circulations from feasible LP solutions

How do we choose a function $b : B(T) \to [0,1]$ so that every internal vertex is satisfied by b? One natural approach is to begin with a feasible solution $y \in LP(G)$. The following useful lemma shows that certain vertices will be satisfied by y.

Lemma 5. *For $y \in LP(G)$, $y \leq 1$, let T be a greedy DFS tree chosen with respect to y, and let $y(i,j) = y_{ij}$ for all back edges in $B(T)$. Then for any $v \in T_{int}$, v is satisfied by y if $y(\delta(v)) = 2$ or if v is a nonbranch vertex with no outgoing back edge (e.g., v is an expensive vertex if G is subquartic).*

9

Proof. First, we prove that v is satisfied by y if $y(\delta(v)) = 2$. Let (v, t) be an outgoing tree edge from vertex v. Partition the edges of $\delta(v)$ into the following two sets: the first set E_1 consists of edge (v, t) and the incoming back edges from the subtree rooted at t. Let E_2 denote the remaining edges in $\delta(v)$. Observe that at least one of the sets E_1 and E_2 has total y-value at most 1, since these two sets form a partition of $\delta(v)$. Let $B' \subset B(T)$ denote the back edges that cover the tree cut corresponding to (v, t). Observe that both $B' \cup E_1$ and $B' \cup E_2$ are edge sets that cross cuts in G (which we assume to be the support of y). In other words, let V_t denote the vertices in the subtree of T rooted at t (including vertex t). Then $B' \cup E_1 = \delta(V_t)$ and $B' \cup E_2 = \delta(V_t \cup v)$. Since either E_1 or E_2 has total y-value at most 1, we can conclude that B' has y-value at least 1, and therefore the tree cut corresponding to (v, t) is satisfied by y. By repeating this argument for each outgoing tree edge (in the case that v is a branch vertex), we can conclude that v is satisfied by y.

Now let us consider the case in which v is a nonbranch vertex with no outgoing back edge (i.e., all back edges are incoming). Then the edge set E_2 consists of a single edge and we can conclude that the y-value of the edges in B' is at least 1. $\qquad\square$

The following lemma pertains to vertices not satisfied by a feasible solution $y \in LP(G)$.

Lemma 6. *For a subquartic graph G, and $y \in LP(G)$, $y \leq 1$, let T be a greedy DFS tree chosen with respect to y and let $y(i, j) = y_{ij}$ for all back edges in $B(T)$. Then for any $u \in T_{int}$, u has at most one outgoing tree edge whose corresponding tree cut is not satisfied by y.*

Proof. Let (u, v) be a tree edge such that there is no back edge coming into vertex u from the subtree of T rooted at v. Then observe that since the y-value of edge (u, v) is at most 1, the total y-value of the edges that cover this tree cut must be at least 1, since the covering edges plus the edge (u, v) have y-value at least 2 (by the constraints in $(LP(G))$). Thus, the tree cut corresponding to edge (u, v) must be satisfied by y.

If u has only one outgoing tree edge, then the lemma holds. If u is a branch vertex with two outgoing tree edges, there is at most one incoming back edge to vertex u. Thus, the tree cut corresponding to at least one outgoing tree edge, say (u, v), is satisfied by y, since u has no incoming back edge from the subtree rooted at v. Moreover, if u has three outgoing tree edges, then there is no incoming back edge to vertex u and therefore the tree cuts corresponding to each of the three outgoing tree edges are satisfied by y. $\qquad\square$

3 Subquartic graphs: $OPT_{LP}(G) = n$

We now show that in the special case when $OPT_{LP}(G) = n$ and G is subquartic, there is a circulation with cost zero. Recall the $\{x_{ij}\}$ values from Corollary 1, and assume that $\sum_{ij \in E} x_{ij} = n$. Note that if $|E| = n$, then each edge in E must have x-value 1. Thus, G is a Hamilton cycle. If $|E| > n$, then we can choose a greedy DFS tree T with respect to x such that each edge $ij \in E$ with x-value $x_{ij} = 1$ (a "1-edge") belongs to T.

Lemma 7. *If $OPT_{LP}(G) = n$ and $|E| > n$, then there is a greedy DFS tree chosen with respect to x such that all 1-edges belong to $E(T)$.*

Proof. Observe that a vertex with degree at least three can have at most one incident 1-edge, since the x-value at each vertex is exactly 2 when $OPT_{LP}(G) = n$ (i.e., $x(\delta(v)) = 2$ for all $v \in V$). By the assumptions in the lemma, there is some vertex, say i, with degree at least three, which therefore has at most one incident 1-edge. Thus, we choose i to be the root of the greedy DFS tree. If i is incident to a 1-edge, then this 1-edge belongs to the resulting tree by the rules defining the construction of a greedy DFS tree.

Suppose that after we are done constructing the greedy DFS tree, there is a back edge (u, v) that has x-value 1. Then when vertex v was visited in the depth-first search, it should have traversed this edge as the next tree edge. Otherwise, the edge it did traverse/add to the tree also had an x-value of 1, which is a contradiction because as an internal vertex with an incoming back edge, vertex v has degree at least three. $\square$

For the rest of Section 3, let T denote a greedy DFS tree in which all 1-edges are tree edges.

Lemma 8. *When $OPT_{LP}(G) = n$, and each back edge $(i, j) \in B(T)$ is assigned value $f(i, j) = 1/2$, then each vertex in T_{int} is satisfied by f.*

Proof. Set $x(i, j) = x_{ij}$ for each back edge. Then each tree cut is satisfied by x, because each vertex $v \in V$ has $x(\delta(v)) = 2$, and we can therefore apply Lemma 5. Since there are no 1-edges in the set of back edges, this implies that each tree cut must in fact be covered by at least two edges. Thus, setting $f(i, j) = 1/2$ results in each tree cut being satisfied by f. $\square$

We remark that Lemmas 7 and 8 hold for general graphs.

Lemma 9. *If G is subquartic and $OPT_{LP}(G) = n$, setting $f(i, j) = 1/2$ for each edge $(i, j) \in B(T)$ yields a circulation with cost zero.*

Proof. This follows from the fact that each vertex in T_{int} has in-degree at most two and therefore the total f-value coming into a vertex is at most one. Thus, the circulation value is zero. (Note that the root can contribute $1/2$ to the circulation, but there exists a minimum cost circulation that is integral, and its cost will therefore still be zero.) $\square$

Theorem 1. *If G is subquartic and $OPT_{LP}(G) = n$, then G has a TSP tour of length at most $4n/3$.*

4   Subquartic graphs: General case

In this section, we consider the general case of subquartic graphs. For a subquartic graph $G = (V, E)$, suppose $OPT_{LP}(G) = (1 + \epsilon)n$ for some $\epsilon > 0$. Recall the $\{x_{ij}\}$ values from Corollary 1. Let T be a greedy DFS tree chosen with respect to x (Definition 3), and let $x(i, j) = x_{ij}$ for all back edges in $B(T)$. As shown in Lemma 4, the only vertices that can add to the cost function are the expensive vertices (Definition 6). We find the following terminology convenient.

Definition 10. *A vertex $v \in T_{int}$ that is satisfied by x is called* LP-satisfied.

Definition 11. *A vertex $v \in T_{int}$ that is not satisfied by x is called* LP-unsatisfied.

The following corollaries follow from Lemma 5.

Corollary 2. *An expensive vertex is LP-satisfied.*

Corollary 3. *An LP-unsatisfied vertex is heavy (i.e., it belongs to $H(x)$).*

The reason we emphasize that an LP-unsatisfied vertex is heavy is that we can use the excess x-value of this vertex to pay for the increased value on an edge that covers the unsatisfied tree cut corresponding to one of its incident outgoing edges so that this tree cut becomes satisfied. We also wish to use the excess x-value of an expensive vertex to pay for some of its contribution to the cost function incurred by the back edges coming into the vertex. For each vertex v, we want to use the quantity $\epsilon(v)$ at most once in this payment scheme. This will be guaranteed by the fact that LP-unsatisfied vertices and expensive vertices are disjoint sets (Corollary 2).

4.1   The x-circulation

In this section, we use the x-values to obtain an upper bound on the cost of a circulation, essentially following the arguments of Mömke and Svensson [MS16] and Mucha [Muc14]. We present the analysis here, since we refer to it in Section 4.3 when we analyze the cost of taking the best of two circulations. Also, the arguments can be somewhat simplified due to the subquartic structure of the graph, which is useful for our analysis.

Recall that for each back edge (i, j) in $B(T)$, we have $x(i, j) = x_{ij}$. (For a vertex $j \in T_{int} \setminus T_{exp}$, we can actually set $x(i, j) = 1$, since there is at most one incoming back edge to vertex j and this does not change the worst-case analysis.)

Definition 12. *For each vertex $j \in T_{exp}$, let $x_{min}(j) \leq x_{max}(j)$ denote the x-values of the two incoming back edges to vertex j. Let $c_x(j) = x_{min}(j) + x_{max}(j) - 1 - \epsilon(j)$.*

We will show that there is a function $x' : B(T) \to [0, 1]$ such that each vertex in T_{int} is satisfied by x' and the cost of the circulation can be bounded by

$$\sum_{j \in T_{exp}} \max \left\{ 0, \left(\sum_{i:(i,j)\in B(T)} x'(i,j) \right) - 1 \right\} \ \leq \ \sum_{j \in T_{exp}} \max\{0, c_x(j)\} + \sum_{j \in T_{int}} \epsilon(j). \tag{7}$$

Claim 1. *For an expensive vertex $j \in T_{exp}$, the following holds:*

$$2 \cdot x_{max}(j) + x_{min}(j) \leq 2 + \epsilon(j).$$

Proof. By the construction of T, we note that the x-value of the tree edge leaving vertex j must be at least $x_{max}(j)$. Thus, the above inequality holds. ◇

Claim 2. *The value $c_x(j)$ can be upper bounded as follows:*

$$c_x(j) \ \leq \ \frac{x_{min}(j)}{2} - \frac{\epsilon(j)}{2} \ \leq \ 1 - x_{min}(j).$$

Proof. For a vertex $j \in T_{exp}$, we can use Claim 1 to show

$$
\begin{aligned}
c_x(j) \; = \; x_{max}(j) + x_{min}(j) - 1 - \epsilon(j) \; &\leq \; (2 + \epsilon(j) - x_{min}(j))/2 + x_{min}(j) - 1 - \epsilon(j) \\
&= \; \frac{x_{min}(j)}{2} - \frac{\epsilon(j)}{2}.
\end{aligned}
$$

Claim 1 also implies

$$
\frac{x_{min}(j)}{2} - \frac{\epsilon(j)}{2} \; \leq \; 1 - x_{max}(j) \; \leq \; 1 - x_{min}(j).
$$

$\diamond$

Lemma 10. *For a vertex $j \in T_{exp}$, $c_x(j) \leq 1/3$.*

Proof. By Claim 2, we have

$$
c_x(j) \; \leq \; \min\left\{ \frac{x_{min}(j)}{2}, \; 1 - x_{min}(j) \right\}.
$$

This implies that $c_x(j) \leq 1/3$, which occurs when $x_{min}(j) = 2/3$, as shown by Mucha [Muc14]. $\square$

To make the circulation feasible, we need to increase the x-values of some of the back edges in $B(T)$ so that all of the LP-unsatisfied vertices become satisfied. By Corollary 3, these vertices are heavy. Thus, we will use the "extra" $\epsilon(v)$ for an LP-unsatisfied vertex v to "pay" for increasing the x-value on an appropriate back edge. For ease of notation, we now set $x'(i,j) := x(i,j)$ for all $(i,j) \in B(T)$. We will update these x'-values so that each LP-unsatisfied vertex is satisfied by x'.

Consider an LP-unsatisfied vertex $v \in T$ and let (v,t) denote the outgoing tree edge corresponding to the unsatisfied tree cut. Let $B' \subseteq B(T)$ be the set of back edges that cover the tree cut corresponding to edge (v,t). Partition the edges in $\delta(v)$ into the following two sets: the first set E_1 consists of edge (v,t) and the incoming back edges from the subtree rooted at t. Let E_2 denote the remaining edges in $\delta(v)$. Note that $x(E_1) + x(E_2) = 2 + \epsilon(v)$ and since v is LP-unsatisfied, $\epsilon(v) > 0$. From this and the following facts

$$
\begin{aligned}
x(E_1) + x(B') \; &\geq \; 2, \\
x(E_2) + x(B') \; &\geq \; 2,
\end{aligned}
$$

we can conclude that

$$
x(B') \; \geq \; 1 - \frac{\epsilon(v)}{2}.
$$

Let $(i,j) \in B'$ be an arbitrary edge in B', which exists because G is 2-vertex-connected. We will update the value of $x'(i,j)$ as follows:

$$
x'(i,j) := \min\{1, \; x'(i,j) + \epsilon(v)/2\}.
$$

We use this recursive notation because a back edge's value can be increased multiple times in the process of satisfying all LP-unsatisfied vertices. The following lemma follows by the construction of the x'-values and by Corollary 2.

13

Lemma 11. *The cost of satisfying all of the LP-unsatisfied vertices is at most $\sum_{j \in T_{int} \backslash T_{exp}} \epsilon(j)/2$. In other words,*

$$\sum_{(u,v) \in B(T)} (x'(u,v) - x(u,v)) \leq \sum_{j \in T_{int} \backslash T_{exp}} \frac{\epsilon(j)}{2}.$$

Since all vertices in T are now satisfied by x', the x'-values can be used to compute an upper bound on the cost of a feasible circulation of $C(G,T)$.

Lemma 12. *The function $x' : B(T) \to [0,1]$ corresponds to a feasible circulation of $C(G,T)$ with cost (not including the contribution of the root) at most*

$$\sum_{j \in T_{exp}} \max\{0, \ c_x(j)\} + \sum_{j \in T_{int}} \epsilon(j).$$

Proof. By construction, every vertex in T_{int} is satisfied by x'. Thus, the x'-values correspond to a feasible circulation of $C(G,T)$. The cost of the circulation based on the x'-values is

$$\sum_{j \in T_{exp}} \max\left\{0, \left(\sum_{i:(i,j) \in B(T)} x'(i,j)\right) - 1\right\} \leq \sum_{j \in T_{exp}} \max\left\{0, \left(\sum_{i:(i,j) \in B(T)} x(i,j)\right) - 1\right\}$$
$$+ \sum_{(u,v) \in B(T)} (x'(u,v) - x(u,v)).$$

We have

$$\sum_{j \in T_{exp}} \max\left\{0, \left(\sum_{i:(i,j) \in B(T)} x(i,j)\right) - 1\right\} = \sum_{j \in T_{exp}} \max\{0, x_{max}(j) + x_{min}(j) - 1\}$$
$$\leq \sum_{j \in T_{exp}} (\max\{0, x_{max}(j) + x_{min}(j) - 1 - \epsilon(j)\} + \epsilon(j))$$
$$\leq \sum_{j \in T_{exp}} \max\{0, \ c_x(j)\} + \sum_{j \in T_{exp}} \epsilon(j).$$

Combining the above inequality with Lemma 11 proves the lemma. $\square$

Theorem 2. *When G is subquartic and $OPT_{LP}(G) = (1 + \epsilon)n$, there is a feasible circulation for $C(G,T)$ with cost at most $n/6 + 2\epsilon n + 2$.*

Proof. The number of expensive vertices is at most $n/2$ (Fact 1) and each expensive vertex $j \in T_{exp}$ can add at most $1/3 + \epsilon(j)$ to the cost function (Lemma 10). Each vertex $j \in T_{int} \backslash T_{exp}$ can add at most $\epsilon(j)$ to the cost function. Using the fact that $\sum_{j \in T_{int}} \epsilon(j) \leq 2\epsilon n$ and the fact that the contribution of the root is at most 2 yields the theorem. $\square$

4.2 The f-circulation

Now we describe a new method to obtain a feasible circulation: We show how to obtain values $f'(i,j)$ for each edge $(i,j) \in B(T)$ such that each vertex in T_{int} is satisfied by f'. The values will be used to demonstrate an improved upper bound on the cost of a circulation of $C(G,T)$ when G is a subquartic graph. In this section, we will prove the following theorem, which implies that the Mömke-Svensson algorithm has an approximation guarantee of $17/12$ for graph-TSP on subquartic graphs.

Theorem 3. *When G is subquartic and $OPT_{LP}(G)= (1+\epsilon)n$, there is a feasible circulation for $C(G,T)$ with cost at most $n/8 + 2\epsilon n + 2$.*

Consider a vertex $j \in T_{exp}$. If both incoming back edges have f-value $1/2$, then this vertex will not contribute anything to the cost of the circulation. Thus, on a high level, our goal is to find f-values that are as close to $1/2$ as possible, while at the same time not creating any additional unsatisfied vertices. For example, in Lemma 10, we saw that a vertex can contribute as much as $1/3$ to the cost of the circulation, when $x_{min} = 2/3$. If we could decrease x_{max} and x_{min}, we could decrease the cost of the x-circulation. This might, in turn, cause some LP-satisfied vertices to become unsatisfied. We can avoid this situation by strategically increasing some other x-values. The f-value therefore corresponds to a decreased x-value if the x-value is high, and an increased x-value if the x-value is low.

The remaining issue that needs to be addressed is that the f-values corresponding to decreased x-values may cause some of the LP-unsatisfied vertices to become more unsatisfied than they were by the x-values. However, note that in Section 4.1, we only used $\epsilon(j)/2$ to satisfy an LP-unsatisfied vertex j. We can actually use up to $\epsilon(j)$. This observation allows us to decrease the x-values while still ensuring that all vertices are satisfied. We use the rules depicted in Figure 1 to determine the values $f : B(T) \to [0, 1]$.

$$
\begin{array}{rcl}
x_{ij} > 3/4 & \Rightarrow & f(i,j) = 2x_{ij} - 1, \\
x_{ij} < 1/4 & \Rightarrow & f(i,j) = 2x_{ij}, \\
1/4 \leq x_{ij} \leq 3/4 & \Rightarrow & f(i,j) = 1/2.
\end{array}
$$

Figure 1: Rules for constructing the f-values from the x-values.

Lemma 13. *If a vertex v is LP-satisfied, then it is satisfied by f.*

Proof. Let $B' \subseteq B(T)$ denote the set of back edges that covers a particular tree cut. If B' consists of a single edge with x-value 1, then the f-value of this edge will also be 1. Let us now suppose the set B' contains multiple edges, whose total x-value is at least 1. Consider the following three cases: First, suppose B' contains at least two edges with x-value at least $1/2$. In this case, the f-value on each of these edges remains at least $1/2$. Second, if the set B' contains only edges that have x-value at most $1/2$, then the total f-value is at least the total x-value, since the f-value does not decrease in this case.

The third case is when B' contains only one edge with x-value at least $1/2$. Suppose that this edge e has value $x_e = 1 - \gamma \geq 1/2$. The remaining edges in B' must have total x-value at least γ. If at least one of these edges' x-value is at least $1/4$, then we are done (because this edge will have f-value $1/2$). Thus, all the edges in the set $B' \setminus e$ must have x-value less than $1/4$. In this case, the total f-value for these edges is at least 2γ. Note that the f-value of edge e is at least $1 - 2\gamma$. $\square$

Definition 13. *For each vertex $j \in T_{exp}$, let $c_f(j) = \sum_{i:(i,j)\in B(T)} f(i,j) - 1 - \epsilon(j)$.*

For ease of notation, set $f'(i,j) := f(i,j)$ for all $(i,j) \in B(T)$.

Lemma 14. *For an LP-unsatisfied vertex $v \in T_{int}$, if we increase by the amount $\epsilon(v)$ the f'-value of an edge that covers its unsatisfied tree cut, then vertex v will be satisfied by f'.*

Proof. We will argue, as we did in Section 4.1, that each LP-unsatisfied vertex $v \in T_{int} \setminus T_{exp}$ can be satisfied by increasing the f'-value of a single back edge that covers the unsatisfied tree cut corresponding to one of its outgoing tree edges, (v,t).

Let $B' \subseteq B(T)$ be the set of back edges that cover the tree cut corresponding to edge (v,t). Partition the edges in $\delta(v)$ into the following two sets: the first set E_1 consists of edge (v,t) and the incoming back edges from the subtree rooted at t. Let E_2 denote the remaining edges in $\delta(v)$. Then we have

$$
\begin{aligned}
x(E_1) + x(B') &\geq 2, \\
x(E_2) + x(B') &\geq 2.
\end{aligned}
$$

Since v is LP-unsatisfied, $x(B') = 1 - \gamma < 1$. Thus,

$$
x(E_1) + x(E_2) \geq 2 + 2\gamma.
$$

So $\epsilon(v) \geq 2\gamma$. Therefore, since the f-value of the edges in B' is at least $1 - 2\gamma$, the amount 2γ is sufficient to "correct" the f-values so that v is satisfied by f'. $\square$

The following lemma follows by the construction of the f'-values and by Corollary 2.

Lemma 15. *The cost of satisfying all of the LP-unsatisfied vertices is at most $\sum_{j\in T_{int}\setminus T_{exp}} \epsilon(j)$. In other words,*

$$
\sum_{(u,v)\in B(T)} (f'(u,v) - f(u,v)) \leq \sum_{j\in T_{int}\setminus T_{exp}} \epsilon(j).
$$

Lemma 16. *For a vertex $j \in T_{exp}$, $c_f(j) \leq \frac{1}{4}$.*

Proof. We break the proof into the following claims.

Claim 3. *For $j \in T_{exp}$, if $x_{min}(j) \geq 1/2$ or if $x_{max}(j) \leq 3/4$, then $c_f(j) \leq 0$.*

Proof. We consider the following three cases.

Case (i) First, we consider the case in which $x_{min}(j) \geq 3/4$. Then, the total f-value of the back edges coming into vertex j is

$$
\begin{aligned}
c_f(j) + 1 + \epsilon(j) &= 2 \cdot x_{max}(j) - 1 + 2 \cdot x_{min}(j) - 1 \\
&\leq (2 + \epsilon(j) - x_{min}(j)) + 2 \cdot x_{min}(j) - 2 \\
&= x_{min}(j) + \epsilon(j).
\end{aligned}
$$

The inequality follows from Claim 1. This implies that

$$ c_f(j) \leq x_{min}(j) - 1 \leq 0. $$

Case (ii) Now let us consider the case when $x_{max}(j) \geq 3/4$ and $1/2 \leq x_{min}(j) \leq 3/4$. The total f-value of the incoming back edges is

$$
\begin{aligned}
c_f(j) + 1 + \epsilon(j) = 2 \cdot x_{max}(j) - 1 + \frac{1}{2} &= 2 \cdot x_{max}(j) - \frac{1}{2} \\
&\leq 3/2 - x_{min}(j) + \epsilon(j).
\end{aligned}
$$

The inequality follows from Claim 1. This implies that

$$ c_f(j) \leq 1/2 - x_{min}(j). $$

Since $x_{min}(j) \geq 1/2$, this implies that $c_f(j) \leq 0$.

Case (iii) Now let us consider the case when $x_{max}(j) \leq 3/4$. Note that in this case, the f-value for each incoming back edge is at most $1/2$. Thus, $c_f(j) \leq 0$. $\diamond$

It remains to examine the case when $x_{max}(j) > 3/4$ and $0 < x_{min}(j) \leq 1/2$. This is the only situation when $c_f(j)$ can be positive.

Claim 4. *If $x_{max}(j) \geq 3/4$ and $0 < x_{min}(j) \leq 1/2$, then $c_f(j) \leq \min\{x_{min}(j), \ 1/2 - x_{min}(j)\}$.*

Proof. **Case (iv)** Now let us consider the case when $x_{max}(j) \geq 3/4$ and $1/4 \leq x_{min}(j) < 1/2$. Applying Claim 1, we see that the total f-value of the incoming back edges is

$$ c_f(j) + 1 + \epsilon(j) = 2 \cdot x_{max}(j) - 1 + \frac{1}{2} \leq \frac{3}{2} - x_{min}(j) + \epsilon(j). $$

Therefore,

$$ c_f(j) \leq \frac{1}{2} - x_{min}(j) \leq x_{min}(j). $$

Case (v) Now let us consider the case when $x_{max}(j) \geq 3/4$ and $0 < x_{min}(j) < 1/4$. Applying Claim 1, the total f-value of the incoming back edges is

$$ c_f(j) + 1 + \epsilon(j) = 2 \cdot x_{max}(j) - 1 + 2 \cdot x_{min}(j) \leq x_{min}(j) + 1 + \epsilon(j). $$

Therefore,

$$c_f(j) \;\leq\; x_{min}(j) \;\leq\; \frac{1}{2} - x_{min}(j).$$

$\diamond$

Claims 3 and 4 show that $c_f(j) \leq \min\{x_{min}(j), 1/2 - x_{min}(j)\}$ for $0 \leq x_{min}(j) \leq 1/2$ and $c_f(j) \leq 0$ otherwise. Thus, $c_f(j) \leq 1/4$. $\quad\square$

We can now prove Theorem 3.

Proof of Theorem 3. We show that the function $f' : B(T) \to [0,1]$ corresponds to a feasible circulation of $C(G,T)$ with cost (not including the root, which adds 2) at most

$$\sum_{j \in T_{exp}} \max\{0, \; c_f(j)\} + \sum_{j \in T_{int}} \epsilon(j). \tag{8}$$

By construction, every vertex in T_{int} is satisfied by f'. Thus, the f'-values correspond to a feasible circulation of $C(G,T)$. The cost of the circulation based on the f'-values is

$$\sum_{j \in T_{exp}} \max\left\{0, \left(\sum_{i:(i,j)\in B(T)} f'(i,j)\right) - 1\right\} \;\leq\; \sum_{j \in T_{exp}} \max\left\{0, \left(\sum_{i:(i,j)\in B(T)} f(i,j)\right) - 1\right\}$$
$$+ \sum_{(u,v)\in B(T)} (f'(u,v) - f(u,v)).$$

We have

$$\sum_{j \in T_{exp}} \max\left\{0, \left(\sum_{i:(i,j)\in B(T)} f(i,j)\right) - 1\right\} \;\leq\; \sum_{j \in T_{exp}} \max\{0, \; c_f(j)\} + \sum_{j \in T_{exp}} \epsilon(j).$$

Combining the above inequality with Lemma 15 proves that the cost of the circulation corresponding to f' is at most (8). By Lemma 16, we have

$$\sum_{j \in T_{exp}} \max\{0, c_f(j)\} + \sum_{j \in T_{int}} \epsilon(j) \;\leq\; |T_{exp}| \cdot \frac{1}{4} + \sum_{j \in T_{int}} \epsilon(j)$$
$$\leq\; \frac{n}{8} + 2\epsilon n,$$

where the last inequality follows from Fact 1. $\quad\square$

4.3 Combining the x- and the f-circulations

We can classify each vertex j in T_{exp} according to the value of $x_{min}(j)$. Intuitively, if many vertices contribute a lot, say $1/3$ to the x-circulation, then they will not contribute a lot to the f-circulation, and vice versa. The entries in the table below follow from Claims 3 and 4 and Lemma 10.

18

$x_{min}(j)$	$c_x(j)$	$c_f(j)$
$[0, 1/4]$	$x_{min}(j)/2$	$x_{min}(j)$
$[1/4, 1/2]$	$x_{min}(j)/2$	$1/2 - x_{min}(j)$
$[1/2, 1]$	$\leq 1/3$	0

Theorem 4. *When G is subquartic and $OPT_{LP}(G) = (1 + \epsilon)n$, there is a feasible circulation for $C(G, T)$ with cost at most $|B(T)|/11 + 2\epsilon n \leq n/11 + 2\epsilon n + 2$.*

Proof. We can compute the cost of the x-circulation and the cost of the f-circulation for $C(G, T)$. We will show that the minimum of the two costs is upper bounded by the guarantee in the theorem. In other words, our goal is to show that for some $\alpha \in [0, 1]$, the following inequality holds:

$$\alpha \sum_{j \in T_{int}} c_x(j) + (1 - \alpha) \sum_{j \in T_{int}} c_f(j) \leq \frac{n}{11}. \tag{9}$$

Let $\beta \in [0, 1]$ represent the fraction of vertices in T_{exp} for which $x_{min}(j) \in [0, 1/2]$. We may assume without loss of generality that $x_{min}(j) \geq \frac{1}{4}$ for all j; if $x_{min}(j) < \frac{1}{4}$, we can substitute $x_{min}(j)$ with $1/2 - x_{min}(j)$, which will preserve the value of $c_f(j)$ and increase the value of $c_x(j)$ (i.e., we only make the left-hand side of (9) higher). Let $(1 - \beta)$ be the remaining fraction of the vertices, for which $x_{min}(j) \in (1/2, 1]$.

Let $\bar{x}_{min}$ denote the average value of $x_{min}(j)$ for the β-fraction of the vertices in T_{exp} with $x_{min}(j) \in [1/4, 1/2]$. Note that $\beta \cdot \bar{x}_{min}/2$ is the average contribution of these vertices to the x-circulation and that $\beta(1/2 - \bar{x}_{min})$ is the average contribution of these vertices to the f-circulation. We can take the following convex combination of the x- and f-circulations to obtain the following inequality:

$$\frac{6}{11}\left(\beta \cdot \frac{\bar{x}_{min}}{2} + (1 - \beta) \cdot \frac{1}{3}\right) + \frac{5}{11}\left(\beta(\frac{1}{2} - \bar{x}_{min})\right) \leq \frac{2}{11}. \tag{10}$$

Since (10) holds when $\bar{x}_{min} \geq 1/4$, and $x_{min}(j) \in [1/4, 1]$ by assumption, we can conclude that the average contribution of a vertex in T_{exp} to the circulation is at most $2/11$. By Fact 1, since there are at most $|B(T)|/2$ vertices in T_{exp}, the worst-case cost of the circulation (not including the contribution of the root) is $|B(T)|/11 + 2\epsilon n \leq n/11 + 2\epsilon n$. □

5 The h-circulation

In this section, we define a third circulation using an optimal extreme point x^* of $(LP(G))$ as defined in Corollary 1. First, we fix a greedy DFS tree T^* chosen with respect to x^*. Then we assign each back edge in $B(T^*)$ a value of $\frac{1}{2}$, which we will refer to as the h-values. If each vertex is satisfied by h, then we have found a circulation with zero cost. Otherwise, for each tree cut that is not satisfied by h, we increase the h-value on the back edge covering this tree cut to 1. Note that these are exactly the tree cuts that are covered by a single back edge. More formally, we have the following definitions.

Definition 14. *With respect to a DFS tree T^*, we call a tree cut* poor *if it is covered by only one back edge. Moreover, we call the respective back edge* costly.

The costly back edges are the ones with h-value 1. All other back edges have h-value $\frac{1}{2}$. Then the cost of the circulation is at most

$$\frac{1}{2} \cdot (\text{number of costly back edges}). \tag{11}$$

Recall that $H(x^*)$ is the set of vertices that are heavy with respect to x^*. Let $k = |H(x^*)|$. We define $\epsilon^*(v) = x^*(\delta(v)) - 2$.

Lemma 17. *When G is subquartic and $OPT_{LP}(G) = (1 + \epsilon)n$, there is a feasible circulation for $C(G, T^*)$ with cost (not including the contribution of the root) at most $k + 2\epsilon n$.*

Since $|E| \le 2n - k - 1$, observe that $|B(T^*)| \le n - k$. To prove Lemma 17, we will relate the quantity k to the number of costly back edges. We remark that while x and x^* have the same support, the heavy sets $H(x)$ and $H(x^*)$ might not be the same (and might not have the same cardinality). We find it easier to relate the number of costly back edges to the number of heavy vertices in $H(x^*)$. Before we prove Lemma 17, we show how it can be used to obtain an improved bound on graph-TSP in subquartic graphs.

Theorem 5. *When G is subquartic and $OPT_{LP}(G)(1 + \epsilon)n$, there is a feasible circulation either for $C(G, T)$ or for $C(G, T^*)$ with cost at most $n/12 + 2\epsilon n + 2$.*

Proof. By Theorem 4, we see that there is a circulation for $C(G, T)$ of cost at most $|B(T)|/11 + 2\epsilon n + 2$. Note that $|B(T)| = |B(T^*)| = |E| - (n - 1) = n - k$ (see Corollary 1). By Lemma 17, we also have a circulation for $C(G, T^*)$ of cost at most $k + 2\epsilon n + 2$. These two quantities are equal when $k = n/12$. This proves the theorem: When $k \le n/12$, the bound on the circulation for $C(G, T^*)$ established via Lemma 17 is at most $n/12 + 2\epsilon n + 2$. Alternatively, when $k > n/12$, we have

$$\frac{|B(T)|}{11} \le \frac{(n - k)}{11} < \frac{(n - n/12)}{11} = n/12,$$

and the bound on the circulation for $C(G, T)$ established via Theorem 4 is at most $n/12 + 2\epsilon n + 2$. □

Theorem 6. *The approximation guarantee of the Mömke-Svensson algorithm on subquartic graphs is at most $25/18$.*

Proof. Applying Lemma 1 and Theorem 5, we can compute an upper bound on the cost of a TSP tour:

$$\frac{\frac{4n}{3} + \frac{2}{3}\left(\frac{n}{12} + 2\epsilon n\right)}{(1 + \epsilon)n} \le \frac{\frac{25}{18} + \frac{4\epsilon}{3}}{(1 + \epsilon)} \le \frac{25}{18}.$$

□

It remains to prove Lemma 17. Let k_0 denote the number of vertices for which $0 < \epsilon^*(v) < 1$, and let k_1 denote the number of vertices for which $\epsilon^*(v) \geq 1$. Then $k = k_0 + k_1$. Additionally, we have

$$2\epsilon n \;=\; \sum_{v \in V} \epsilon^*(v) \;\geq\; k_1. \tag{12}$$

The following lemma implies Lemma 17.

Lemma 18. *When G is subquartic and $OPT_{LP}(G) = (1 + \epsilon)n$, there is a feasible circulation for $C(G, T^*)$ with cost at most $k_0 + \frac{3}{2}k_1 \leq k + k_1$.*

Proof. We set up the following scheme to account for the costly back edges.

Definition 15. *We say a vertex j is directly charged for a costly back edge $f = (u, v)$ if $(j, t) \in E(T^*)$ corresponds to a poor tree cut covered by f and vertex j is charged for f.*

Definition 16. *We say a vertex v is indirectly charged for a costly back edge $f = (u, v)$ if vertex v is charged for f.*

Let $f = (u, v)$ be a costly back edge in $B(T^*)$. We use the following rules to charge f to a heavy vertex.

1. If $x_f^* < 1$, then directly charge edge f to some heavy vertex j such that $(j, t) \in E(T^*)$ and (j, t) corresponds to a poor tree cut covered by f.

2. If $x_f^* \geq 1$, then indirectly charge edge f to vertex v.

Now we will show that for each vertex j, exactly one of the following statements holds, which shows that $C(G, T^*)$ has a circulation cost of at most $k + k_1$.

(i) Vertex j is not charged.

(ii) Vertex j is charged once or twice and $\epsilon^*(j) > 0$.

(iii) Vertex j is charged three times and $\epsilon^*(j) \geq 1$.

Claim 5. *If vertex j is charged at least once, then $\epsilon^*(j) > 0$.*

Proof. If j is directly charged for back edge $f = (u, v)$, then f is the single back edge covering the poor tree cut corresponding to an outgoing tree edge. Since $x_f^* < 1$, this tree cut is not satisfied by x^*. Thus, by Lemma 5, we can conclude that $j \in H(x^*)$, which implies $x^*(j) > 2$ and $\epsilon^*(j) > 0$. If j is indirectly charged, then j is incident to at least two edges with x^*-value at least 1, and so $\epsilon^*(j) > 0$. $\diamond$

Claim 6. *A vertex j is charged at most three times.*

Proof. A vertex j can be directly charged at most once for each outgoing tree edge and can be indirectly charged for each incoming back edge. In a subquartic graph, the number of outgoing tree edges plus the number of incoming back edges is at most three. $\diamond$

21

Claim 7. *If a vertex j is charged three times, then $\epsilon^*(j) \geq 1$.*

Proof. If j is not a branch vertex, then the only way that j can be charged three times is if j is indirectly charged twice and directly charged once. If a vertex j is indirectly charged twice, it must have two incoming back edges with x^*-value at least 1. Thus, $\epsilon^*(j) > 1$.

Now consider the case in which vertex j is a branch vertex with three outgoing tree edges. Let e_1, e_2 and e_3 denote three edges in $E(T^*)$ outgoing from vertex j. Since j cannot be indirectly charged in this case, it must be directly charged three times. Let $f_i \in B(T^*)$ denote the lone back edge covering the tree cut corresponding to edge e_i. Then we have $x^*_{f_i} + x^*_{e_i} \geq 2$ (since e_i and f_i cross a cut in G) and $x^*_{f_i} < 1$ implying $x^*_{e_i} > 1$ for $i \in \{1, 2, 3\}$. It follows that $\epsilon^*(j) > 1$.

The last case to consider is when vertex j is a branch vertex with two outgoing tree edges e_1 and e_2. In this case, if j is charged three times, then it is directly charged twice and indirectly charged once. Assume that the back edge b coming into vertex j comes from the subtree rooted at t_1, where $e_1 = (j, t_1)$. Then vertex j is indirectly charged for b, which implies that $x^*_b \geq 1$ and $x^*_{e_1} \geq 1$ (since T^* is a greedy DFS tree). Suppose f_2 is the back edge covering the poor tree cut corresponding to e_2. Then $x^*_{f_2} + x^*_{e_2} \geq 2$ (since e_2 and f_2 cross a cut in G) and $x^*_{f_2} < 1$, implying that $x^*_{e_2} > 1$. Therefore, $x^*_{e_2} + x^*_{e_1} + x^*_b > 3$, which implies $\epsilon^*(j) > 1$. $\diamond$

This concludes the proof of Lemma 18. $\square$

6   Conclusions

It is possible that the Mömke-Svensson algorithm or some close variant is a $\frac{4}{3}$-approximation algorithm for graph-TSP. It is also possible that some graph with maximum degree five could demonstrate that the approximation ratio of $\frac{13}{9}$ is tight, although we believe this to be unlikely. A key step in the algorithm is to find a minimum cost circulation of the network $C(G, T)$, which is an easy problem (i.e., it can be solved efficiently), but the key difficulty is to relate the minimum cost of this circulation to the minimum cost of a tour.

Previous work used solutions for $(LP(G))$ to bound the minimum cost of a circulation. In this paper, we explored two new approaches to analyze this cost. In the first approach, we showed how to round the LP values resulting in an improved circulation cost in subquartic graphs. Our approach does not immediately extend to general graphs or even to graphs with maximum degree five, because we try to find a circulation with values on the back edges close to $1/2$, whereas this is the worst case for graphs with maximum degree five as shown by Mucha [Muc14]. However, perhaps a different scheme for modifying the LP values could be designed for such instances.

The second approach is based on extreme point structure: an extreme point solution for $(LP(G))$ with k heavy vertices has at most $n - k$ back edges with respect to any DFS tree. This is true for general graphs, and having fewer back edges leads directly to an improved bound on the circulation cost via the analysis of Mucha. In the case of subquartic graphs, we are able to balance this with another analysis that bounds the cost of a circulation in terms of k. In general graphs, this latter step may also be possible, but it seems that new ideas are needed.

Finally, we remark that besides the flexibility in analyzing the circulation cost, which we have explored in this paper, there is also the possibility to choose the DFS tree more strategically in order to obtain an improved analysis of the algorithm.

Acknowledgements

References

[BSvdSS14] Sylvia Boyd, René Sitters, Suzanne van der Ster, and Leen Stougie. The traveling salesman problem on cubic and subcubic graphs. *Mathematical Programming*, 144(1-2):227–245, 2014.

[CFN85] Gérard Cornuéjols, Jean Fonlupt, and Denis Naddef. The traveling salesman problem on a graph and some related integer polyhedra. *Mathematical Programming*, 33(1):1–27, 1985.

[Chr76] Nicos Christofides. Worst-case analysis of a new heuristic for the travelling salesman problem. Technical Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, 1976.

[GLS05] David Gamarnik, Moshe Lewenstein, and Maxim Sviridenko. An improved upper bound for the TSP in cubic 3-edge-connected graphs. *Operations Research Letters*, 33(5):467–474, 2005.

[GSS11] Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. A randomized rounding approach to the traveling salesman problem. In *Proceedings of 52nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 550–559, 2011.

[HK70] Michael Held and Richard M. Karp. The traveling-salesman problem and minimum spanning trees. *Operations Research*, 18(6):1138–1162, 1970.

[MS16] Tobias Mömke and Ola Svensson. Removing and adding edges for the traveling salesman problem. *Journal of the ACM*, 63(1):2, 2016.

[Muc14] Marcin Mucha. 13/9-approximation for graphic TSP. *Theory of Computing Systems*, 55(4):640–657, 2014.

[NP81] Denis Naddef and William R. Pulleyblank. Matchings in regular graphs. *Discrete Mathematics*, 34(3):283–291, 1981.

[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.

[Vis12]   Nisheeth K. Vishnoi. A permanent approach to the traveling salesman problem. In *Proceedings of 53rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 76–80, 2012.