Eight-Fifth Approximation for the Path TSP

András Sebő*

CNRS, UJF, Grenoble-INP, Laboratoire G-SCOP, France
Andras.Sebo@g-scop.inpg.fr
http://www.g-scop.grenoble-inp.fr/recherche/

Abstract. We prove the approximation ratio 8/5 for the metric $\{s,t\}$-path-TSP, and more generally for shortest connected T-joins.

The algorithm that achieves this ratio is the simple "Best of Many" version of Christofides' algorithm (1976), suggested by An, Kleinberg and Shmoys (2012), which consists in determining the best Christofides $\{s,t\}$-tour out of those constructed from a family $\mathcal{F}_+$ of trees having a convex combination dominated by an optimal solution x^* of the Held-Karp relaxation. They give the approximation guarantee $\frac{\sqrt{5}+1}{2}$ for such an $\{s,t\}$-tour, which is the first improvement after the 5/3 guarantee of Hoogeveen's Christofides type algorithm (1991). Cheriyan, Friggstad and Gao (2012) extended this result to a 13/8-approximation of shortest connected T-joins, for $|T| \geq 4$.

The ratio 8/5 is proved by simplifying and improving the approach of An, Kleinberg and Shmoys that consists in completing $x^*/2$ in order to dominate the cost of "parity correction" for spanning trees. We partition the edge-set of each spanning tree in $\mathcal{F}_+$ into an $\{s,t\}$-path (or more generally, into a T-join) and its complement, which induces a decomposition of x^*. This decomposition can be refined and then efficiently used to complete $x^*/2$ without using linear programming or particular properties of T, but by adding to each cut deficient for $x^*/2$ an individually tailored explicitly given vector, inherent in x^*.

A simple example shows that the Best of Many Christofides algorithm may not find a shorter $\{s,t\}$-tour than 3/2 times the incidentally common optima of the problem and of its fractional relaxation.

Keywords: traveling salesman problem, path TSP, approximation algorithm, matching, T-join, polyhedron, tree (basis) polytope.

1 Introduction

A Traveling Salesman wants to visit all vertices of a graph $G = (V, E)$, starting from his home $s \in V$, and – since it is Friday – ending his tour at his week-end residence, $t \in V$. Given the nonnegative valued length function $c : E \longrightarrow \mathbb{Q}_+$, he is looking for a shortest $\{s,t\}$-tour, that is, one of smallest possible (total) length.

* Supported by the TEOMATRO grant ANR-10-BLAN 0207 "New Trends in Matroids: Base Polytopes, Structure, Algorithms and Interactions".

M. Goemans and J. Correa (Eds.): IPCO 2013, LNCS 7801, pp. 362–374, 2013.

The Traveling Salesman Problem (TSP) is usually understood as the $s = t$ particular case of the defined problem, where in addition every vertex is visited exactly once. This "minimum length Hamiltonian cycle" problem is one of the main exhibited problems of combinatorial optimization. Besides being *NP*-hard even for very special graphs or lengths [11], even the best up to date methods of operations research, the most powerful computers coded by the brightest programmers fail solving reasonable size problems exactly.

On the other hand, some implementations provide solutions only a few percent away from the optimum on some large "real-life" instances. A condition on the length function that certainly helps both in theory and practice is the triangle inequality. A nonnegative function on the edges that satisfies this inequality is called a *metric* function. The special case of the TSP where G is a complete graph and c is a metric is called the *metric TSP*. For a thoughtful and distracting account of the difficulties and successes of the TSP, see Bill Cook's book [5].

If c is not necessarily a metric function, the TSP is hopeless in general: it is not only *NP*-hard to solve but also to approximate, and even for quite particular lengths, since the Hamiltonian cycle problem in 3-regular graphs is *NP*-hard [11]. The practical context makes it also natural to suppose that c is a metric.

A *ρ-approximation algorithm* for a minimization problem, where $\rho \in \mathbb{R}_+$, $\rho \geq 1$, is a polynomial-time algorithm that computes a solution of value at most ρ times the optimum. The *guarantee* or *ratio* of the approximation is ρ.

The first trace of allowing s and t be different is Hoogeveen's article [16], providing a Christofides type 5/3-approximation algorithm, again in the metric case. There had been no improvement until An, Kleinberg and Shmoys [1] improved this ratio to $\frac{1+\sqrt{5}}{2} < 1.618034$ with a simple algorithm, an ingenious new framework for the analysis, but a technically involved realization.

The algorithm first determines an optimum x^* of the Held-Karp relaxation; writing x^* as a convex combination of spanning trees and applying Christofides' heuristic for each, it outputs the best of the arising tours. For the TSP problem $x^*/2$ dominates *any* possible parity correction, as Wolsey [23] observed, but this is not true if $s \neq t$. However, [1] manages to perturb $x^*/2$, differently for each spanning tree of the constructed convex combination, with small average increase of the length.

We adopt this algorithm and this global framework for the analysis, and develop new tools that essentially change its realization and shortcut the most involved parts. This results in a simpler analysis guaranteeing a solution within 8/5 times the optimum and within less than three pages, at the same time improving Cheriyan, Friggstad and Gao's $13/8 = 1.625$, valid for arbitrary T [4].[1]

We did not fix that the Traveling Salesman visits each vertex exactly once, our problem statement requires only that *every vertex is visited at least once.* This version has been introduced by Cornuéjols, Fonlupt and Naddef [6] and was called the "graphical" TSP. In other words, this version asks for the "shortest spanning

[1] This was the first ratio better than 5/3 for arbitrary T; the proof extended the proof of [1].

Eulerian subgraph" ("tour"), and puts forward an associated polyhedron and its integrality properties, characterized in terms of excluded minors.

This version has many advantages: while the metric TSP is defined on the complete graph, the graphical problem can be sparse, since an edge which is not a shortest path between its endpoints can be deleted; however, it is equivalent to the metric TSP (see Subsection "Tours" below); the length function c does not have to satisfy the triangle inequality; this version has an unweighted special case (all 1 weights), asking for the minimum size (cardinality) of a spanning Eulerian subgraph.

The term "graphic" or "graph-TSP" has eventually been taken by this all 1 special case. We avoid these three terms too close (in Hamming distance) but used in a too diversified way in the literature, different from habits for other problems which also have weighted and unweighted variants called differently.[2]

2 Notation, Terminology and Preliminaries

The set of real numbers is denoted by $\mathbb{R}$; $\mathbb{R}_+$, $\mathbb{Q}_+$ denote the set of non-negative real or rational numbers respectively, and $\mathbb{1}$ denotes the all 1 vector of appropriate dimension. We fix the notation $G = (V, E)$ for the input graph. For $X \subseteq V$ we write $\delta(X)$ for the set of edges with exactly one endpoint in X. If $w : E \longrightarrow \mathbb{R}$ and $A \subseteq E$, then we use the standard notation $w(A) := \sum_{e \in A} w(e)$.

T**-joins**: For a graph $G = (V, E)$ and $T \subseteq V$ a T*-join* in G is a set $F \subseteq E$ such that $T = \{v \in V : |\delta(v) \cap F| \text{ is odd}\}$. For (G, T), where G is connected, it is well-known and easy to see that a T-join exists if and only if $|T|$ is even [18], [17]. When (G, T) or (G, T, c) are given, we assume that G is a connected graph, $|T|$ is even, and $c : E \longrightarrow \mathbb{Q}_+$, where c is called the *length* function, $c(A)$ $(A \subseteq E)$ is the length of A.

Given (G, T, c), the minimum length of a T-join in G is denoted by $\tau(G, T, c)$. A T*-cut* is a cut $\delta(X)$ such that $|X \cap T|$ is odd. It is easy to see that a T-join and a T-cut meet in an odd number of edges. If in addition c is integer, the maximum number of T-cuts so that every edge e is contained in at most $c(e)$ of them is denoted by $\nu(G, T, c)$. By a theorem of Edmonds and Johnson [8], [18] $\tau(G, T, c) = \nu(G, T, 2c)/2$, and a minimum length T-join can be determined in polynomial time. These are useful for an intuition, even if we only use the weaker Theorem 2 below. For an introduction and more about different aspects of T-joins, see [18], [21], [9], [17].

T**-Tours**: A T*-tour* $(T \subseteq V)$ of $G = (V, E)$ is a set $F \subseteq 2E$ such that

(i) F is a T-join of $2G$,
(ii) (V, F) is a connected multigraph,

[2] We do not investigate here these unweighted problems. For comparison, however, let us note the guaranteed ratios for the cardinality versions of the problems: the ratio 3/2 has been reached for the minimum size of a T-tour (see the definition a few lines below), and 7/5 for $T = \emptyset$ [22].

where $2E$ is the multiset consisting of the edge-set E, and the multiplicity of each edge is 2; we then denote $2G := (V, 2E)$. It is not false to think about $2G$ as G with a parallel copy added to each edge, but we find the multiset terminology better, since it allows for instance to keep the length function and its notation $c : E \longrightarrow \mathbb{Q}_+$, or in the polyhedral descriptions to allow variables to take the value 2 without increasing the number of variables; the length of a multi-subset will be the sum of the lengths of the edges multiplied by their multiplicities, with obvious, unchanged terms or notations: for instance the size of a multiset is the sum of its multiplicities; χ_A is the multiplicity vector of A; $x(A)$ is the scalar product of x with the multiplicity vector of A; a *subset of a multiset* A is a multiset with multiplicities smaller than or equal to the corresponding multiplicities of A, etc.

A *tour* is a T-tour with $T = \emptyset$.

The *T-tour problem (TTP)* is to minimize the length of a T-tour for (G, T, c) as input. Denote $\textsc{opt}(G, T, c)$ this minimum. The subject of this work is the TTP in general.

If $F \subseteq E$, we denote by T_F the set of vertices incident to an odd number of edges in F; if F is a spanning tree, $F(T)$ denotes the *unique T-join of F;* accordingly, $F(s, t) := F(\{s, t\})$ is the (s, t)-path of F.

The *sum* of two (or more) multisets is a multiset whose multiplicities are the sums of the two corresponding multiplicities. If $X, Y \subseteq E$, $X + Y \subseteq 2E$ and $(V, X + Y)$ is a multigraph. Given (G, T), $F \subseteq E$ such that (V, F) is connected,

and a $T_F \triangle T$-join J_F, the multiset $F + J_F$ is a T-tour.

the notation "$\triangle$" stays for the *symmetric difference* (mod 2 sum of sets). This simple operation is the tool for "parity correction".

In [22] T-tours were introduced under the term *connected T-joins*. (This was a confusing term, since T-joins have only 0 or 1 multiplicities.) Even if the main target remains $|T| \leq 2$, the arguments concerning this case often lead out to problems with larger T.

By "Euler's theorem" a subgraph of $2G$ is a tour or $\{s, t\}$-tour if and only if its edges can be ordered to form a closed "walk" or a walk from s to t, that visits every vertex of G at least once, and uses every edge as many times as its multiplicity.

Linear Relaxation: We adopt the polyhedral background and notations of [22], which itself is the adaptation of the so-called "Held-Karp" [15] relaxation to our slightly different context, for slightly improved comfort.

Let $G = (V, E)$ be a graph. For a partition $\mathcal{W}$ of V we introduce the notation $\delta(\mathcal{W}) := \bigcup_{W \in \mathcal{W}} \delta(W)$, that is, $\delta(\mathcal{W})$ is the set of edges that have their two endpoints in different classes of $\mathcal{W}$.

Let G be a connected graph, $T \subseteq V$ with $|T|$ even. Denote

$$P(G,T) := \Big\{ x \in \mathbb{R}^E : \ x(\delta(W)) \geq 2 \text{ for all } \emptyset \neq W \subset V \text{ with } |W \cap T| \text{ even}, \\ x(\delta(\mathcal{W})) \geq |\mathcal{W}| - 1 \text{ for all partitions } \mathcal{W} \text{ of } V, \\ 0 \leq x(e) \leq 2 \text{ for all } e \in E \Big\}.$$

Let $x^* \in P(G,T)$ minimize $c^\top x$ on $P(G,T)$.

Fact 1: Given (G,T,c), $\text{OPT}(G,T,c) \geq \min_{x\in P(G,T)} c^\top x = c^\top x^*$.

Indeed, if F is a T-tour, χ_F satisfies the defining inequalities of $P(G,T)$.

The following theorem is essentially the same as Schrijver [21, page 863, Corollary 50.8].

Theorem 1. *Let $x \in \mathbb{R}^E$ satisfy the inequalities*

$$x(\delta(\mathcal{W})) \geq |\mathcal{W}| - 1 \text{ for all partitions } \mathcal{W} \text{ of } V,$$
$$0 \leq x(e) \leq 2 \text{ for all } e \in E.$$

Then there exists a set $\mathcal{F}_+$, $|\mathcal{F}_+| \leq |E|$ of spanning trees and coefficients $\lambda_F \in \mathbb{R}, \lambda_F > 0$, $(F \in \mathcal{F}_+)$ so that

$$\sum_{F\in\mathcal{F}_+} \lambda_F = 1, \qquad x \geq \sum_{F\in\mathcal{F}_+} \lambda_F \chi_F,$$

and for given x as input, $\mathcal{F}_+$, λ_F $(F \in \mathcal{F}_+)$ can be computed in polynomial time.

Proof. Let x satisfy the given inequalities. If $(2 \geq)x(e) > 1$ $(e \in E)$, introduce an edge e' parallel to e, and define $x'(e') := x(e) - 1$, $x'(e) := 1$, and $x'(e) := x(e)$ if $x(e) \leq 1$. Note that the constraints are satisfied for x', and $x' \leq \mathbb{1}$. Apply Fulkerson's theorem [10] (see [21, page 863, Corollary 50.8]) on the blocking polyhedron of spanning trees: x' is then a convex combination of spanning trees, and by replacing e' by e in each spanning tree containing e'; applying then Carathéodory's theorem, we get the assertion. The statement on polynomial solvability follows from Edmonds' matroid partition theorem [7], or the ellipsoid method [13]. □

Note that the inequalities in Theorem 1 form a subset of those that define $P(G,T)$. In particular, any optimal solution $x^* \in P(G,T)$ for input (G,T,c) satisfies the conditions of the theorem. Fix $\mathcal{F}_+$, λ_F provided by the theorem for x^*, that is,

$$\sum_{F\in\mathcal{F}_+} \lambda_F \chi_F \leq x^*.$$

We fix the input (G,T,c) and keep the definitions x^*, $\mathcal{F}_+$, λ_F until the end of the paper.

It would be possible to keep the Held-Karp context of [1] for $s \neq t$ where metrics in complete graphs are kept and only Hamiltonian paths are considered (so the condition $x(\delta(v)) = 2$ if $v \neq s$, $v \neq t$ is added), or the corresponding generalization in [4] for $T \neq \emptyset$. However, we find it more comfortable to have in mind only (G,T,c), where c is the given function which is not necessarily a metric, and G is the original (connected) graph that is not necessarily the complete graph, and T is only required to have even size, with $T = \emptyset$ allowed. The only price to pay for this is to have $\sum_{F\in\mathcal{F}_+} \lambda_F \chi_F \leq x^*$ without the irrelevant

"=". The paper can be also read with the classical Held-Karp definition in mind at the price of minor technical adjustments.

Last, we state a well-known theorem of Edmonds and Johnson for the blocking polyhedron of T'-joins in the form we will use it. (The notation T is now fixed for our input (G, T, c), and the theorem will be applied for several different T' in the same graph.)

Theorem 2. [8], (cf. [18], [21]) *Given (G, T', c), let*

$$Q_+(G, T') := \{x \in \mathbb{R}^E : x(C) \geq 1 \textit{ for each } T'\textit{-cut } C,\ x(e) \geq 0 \ \textit{ for all } e \in E\}.$$

A shortest T'-join can be found in polynomial time, and if $x \in Q_+(G, T')$,

$$\tau(G, T', c) \leq c^\top x.$$

Christofides for T-tours: A 2-approximation algorithm for the TTP is trivial by taking a minimum length spanning tree F and doubling the edges of a $T_F \triangle T$-join of F, that is, of $F(T_F \triangle T)$. It is possible to do better by adapting Christofides' algorithm [3], which is usually stated in terms of matchings and in the context of the metric TSP. It can quite easily be generalized to T-tours once the relation of the latter to the metric TSP is clear:

Minimizing the length of a tour or $\{s,t\}$-tour is equivalent to the metric TSP problem or its path version (with all degrees 2 except s and t of degree 1, that is, a shortest Hamiltonian cycle or path). Indeed, any length function of a connected graph can be replaced by a function on the complete graph with lengths equal to the lengths of shortest paths (metric completion): then a tour or an $\{s,t\}$-tour can be "shortcut" to a sequence of edges with all inner degrees equal to 2. Conversely, if in the metric completion we have a shortest Hamiltonian cycle or path we can replace the edges by paths and get a tour or $\{s,t\}$-tour.

For $T = \emptyset$, Christofides [3] is equivalent to first determining a minimum length spanning tree F to assure connectivity, and then adding to it a shortest T_F-join. The straightforwardf approximation guarantee $3/2$ of this algorithm has not been improved ever since. A *Christofides type algorithm* for general T *adds a shortest $T_F \triangle T$-join instead.*

We finish the discussion of TTP with a proof of the $5/3$-approximation ratio for Christofides's algorithm. Watch the partition of the edges of a spanning tree into a T-join – if $T = \{s,t\}$, an $\{s,t\}$ path – and the rest of the tree in this proof! For $\{s,t\}$-paths this ratio was first proved by Hoogeveen [16] slightly differently (see for T-tours in the Introduction of [22]), and in [14] in a similar way, as pointed out to me by David Shmoys.

Proposition: *Let (G, T, c) be given, and let F be an arbitrary shortest spanning tree. Then $\tau(G, T_F \triangle T, c) \leq \frac{2}{3}\textsc{opt}(G, T, c)$.*

Proof. $\{F(T), F \setminus F(T)\}$ is a partition of F into a T-join and a $T \triangle T_F$-join (see Fig. 1). The shortest T-tour K has a T_F-join F' by connectivity, so $\{F', K \setminus F'\}$ is a partition of K to a T_F-join and a $T_F \triangle T$-join.

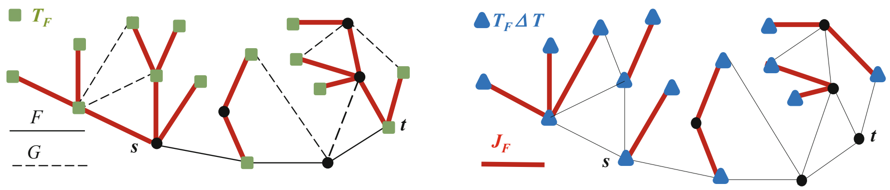

Fig. 1. One of many: $T_F \triangle T$**-joins**, in F (left), minimum in G (right), J_F; $T := \{s, t\}$

If either $c(F \setminus F(T)) \leq \frac{2}{3}c(F)$ or $c(K \setminus F') \leq \frac{2}{3}c(K)$, then we are done, since both are $T \triangle T_F$-joins. If neither hold, then we use the $T \triangle T_F$-join $F(T) \triangle F'$. Since $c(F(T)) \leq \frac{1}{3}c(F) \leq \frac{1}{3}\textsc{opt}(G, T, c)$ and $c(F') \leq \frac{1}{3}c(K) = \frac{1}{3}\textsc{opt}(G, T, c)$, we have $c(F(T) \triangle F') \leq c(F(T)) + c(F') \leq \frac{2}{3}\textsc{opt}(G, T, c)$. $\square$

When $T = \emptyset$ ($s = t$) Wolsey [23] observed that $x^*/2 \in Q_+(G, T)$ and then by the last inequality of Theorem 2 parity correction costs at most $c^\top x^*/2$, so Christofides's tour is at most $3/2$ times $c^\top x^*$; in [1], [4] $\textsc{opt}(G, T, c)$ is replaced by $c^\top x^*$ in the Proposition see also the remark after Fact 2 below.

Best of Many Christofides Algorithm (BOM) [1]: Input (G, T, c).

Determine x^* [13] using [2], see [22].
(Recall: x^* is an optimal solution of $\min_{x \in P(G,T)} c^\top x$.)

Determine $\mathcal{F}_+$. (see Theorem 1 and its proof.)

Determine the best *parity correction* for each $F \in \mathcal{F}_+$,
that is, a shortest $T_F \triangle T$-join J_F [8], [17].

Output that $F + J_F$ ($F \in \mathcal{F}_+$) for which $c(F + J_F)$ is minimum.

The objective value of the T-tour that the BOM algorithm outputs will be upper bounded by the average of the spanning trees in $\mathcal{F}_+$ weighted by the coefficients λ_F ($F \in \mathcal{F}$). The following is a usual, but elegant and useful way of thinking about and working with this average. Our use of it merely notational:

Random Sampling: The coefficient λ_F of each spanning tree $F \in \mathcal{F}_+$ in the convex combination dominated by x^* (see Theorem 1) will be *interpreted as a probability distribution of a random variable* $\mathcal{F}$,

$$\Pr(\mathcal{F} = F) := \lambda_F$$

whose values are spanning trees of G, *and*

$$\mathcal{F}_+ = \{F \subseteq E : F \text{ spanning tree of } G,\ \Pr(\mathcal{F} = F) > 0\}.$$

The notations for spanning trees will also be used for random variables whose values are spanning trees. For instance $\mathcal{F}(s, t)$ denotes the random variable whose value is $F(s, t)$ precisely when $\mathcal{F} = F$. Another example is $\chi_\mathcal{F}$, a random variable whose value is χ_F when $\mathcal{F} = F$. Similarly, $T_\mathcal{F} = T_F$ when $\mathcal{F} = F$.

$$\text{Define } R := \min_{F\in\mathcal{F}_+} \frac{c(F)+\tau(G,T_F\triangle T,c)}{c^\top x^*} \le \frac{E[c(\mathcal{F})+\tau(G,T_{\mathcal{F}}\triangle T,c)]}{c^\top x^*} \le$$

$$\le 1+\frac{E[\tau(G,T_{\mathcal{F}}\triangle T,c)]}{c^\top x^*}.$$

Clearly, R is an upper bound for the guarantee of BOM. The main result of the paper is $R \le 8/5$, and we have just observed that this is implied by $E[\tau(G,T_{\mathcal{F}}\triangle T,c)] \le 3/5c^\top x^*$ (Theorem 3).

3 Proving the New Ratio

In this section we prove the result of the paper, the approximation ratio 8/5 for path TSP, achieved by the BOM algorithm. For the simplicity of reading we substitute $\{s,t\}$ for T, and $F(s,t)$ for $F(T)$ without any other change in the proof. The experienced reader can simply change back each occurrence of $\{s,t\}$ to T, and $F(s,t)$ to $F(T)$.

We use now the probability notation for defining two vectors that will be extensively used:

$$p^*(e) := \Pr(e\in\mathcal{F}(s,t)); \quad q^*(e) := \Pr(e\in\mathcal{F}\setminus\mathcal{F}(s,t)) \ (e\in E).$$

Fact 2: $E[\chi_{\mathcal{F}(s,t)}]=p^*$, $E[\chi_{\mathcal{F}\setminus\mathcal{F}(s,t)}]=q^*$, $x^*\ge E[\chi_{\mathcal{F}}]=p^*+q^*$. □

Introducing p^* and q^* and observing this fact lead us to a version of the Proposition (end of the previous section) about the expectation of parity correction versus the linear optimum: $E[\tau(G,T_{\mathcal{F}}\triangle\{s,t\},c)] \le \frac{2}{3}c^\top q^*$, implying that BOM *outputs a tour of length at most* $\frac{5}{3}c^T x^*$. (Let us sketch the proof (even though we prove our sharper bound in full details below): this inequality follows from $E[\tau(G,T_{\mathcal{F}}\triangle T,c)] \le \min\{c^\top q^*, c^\top x^* - \frac{c^\top q^*}{2}\}$, which in turn holds because q^* is the mean value of the parity correcting $\mathcal{F}\setminus\mathcal{F}(s,t)$, whereas $c^\top x^* - \frac{c^\top q^*}{2} \ge \frac{c^\top x^*}{2} + \frac{c^\top p^*}{2}$ sums to at least 1 on *every* cut, so the last inequality of Theorem 2 can be applied to both.)

Key definitions, key lemma, key theorem: Define

$$\mathcal{Q} := \{Q \text{ is a cut: } x^*(Q) < 2\}.$$

Every $Q\in\mathcal{Q}$ *is an* $\{s,t\}$*-cut*, since non-$\{s,t\}$-cuts C are required to have $x(C)\ge 2$ in the definition of $P(G,\{s,t\})$.[3] Define $x^Q \in \mathbb{Q}_+^E$ with

$$x^Q(e) := \Pr(\{e\} = Q\cap\mathcal{F}).$$

We have from this definition directly that *the support (set of nonzero edges) of* x^Q *is* Q, *and* $x^Q(Q)=\sum_{e\in Q}x^Q(e)=\sum_{e\in Q}\Pr(\{e\}=Q\cap\mathcal{F}) = \Pr(|Q\cap\mathcal{F}|=1)$.

[3] $\mathcal{Q}$ is defined in [1], where its defining vertex-sets are proved to form a chain if $T=\{s,t\}$; in [4] $\mathcal{Q}$ is proved to form a laminar family for general T. These properties are not needed any more.

Lemma: *Let $Q \in \mathcal{Q}$. Then*

(lower bound) $$\mathbb{1}^\top x^Q = x^Q(Q) \geq 2 - x^*(Q),$$

(upper bound) $$\sum_{Q \in \mathcal{Q}} x^Q \leq p^*.$$

Proof. If Q is an arbitrary cut of G (not necessarily in $\mathcal{Q}$), $x^*(Q) = E[|\mathcal{F} \cap Q|] \geq \Pr(|Q \cap \mathcal{F}| = 1) + 2\Pr(|Q \cap \mathcal{F}| \geq 2) = 2 - \Pr(|Q \cap \mathcal{F}| = 1)$, and by the preliminary identity this is equal to $2 - x^Q(Q)$ proving the lower bound for $x^Q(Q)$.

To see the upper bound let us check

$$\sum_{Q \in \mathcal{Q}} \Pr(Q \cap \mathcal{F} = \{e\}) \leq \Pr(e \in \mathcal{F}(s,t)).$$

Indeed, since Q is an $\{s,t\}$-cut, it has a common edge with every $\{s,t\}$-path, so the event $Q \cap \mathcal{F} = \{e\}$ implies $e \in \mathcal{F}(s,t)$; moreover, if $Q_1, Q_2 \in \mathcal{Q}$ are distinct, then the events $Q_1 \cap \mathcal{F} = \{e\}$ and $Q_2 \cap \mathcal{F} = \{e\}$ mutually exclude one another, since for $\mathcal{F} = F$ the set of edges joining the two components of $F \setminus \{e\}$ cannot be equal both to Q_1 and to Q_2. So the left hand side is the probability of the union of disjoint events all of which imply the event on the right hand side, proving the inequality.

Finally, recall $\Pr(e \in \mathcal{F}(s,t)) = p^*(e)$, finishing the proof. □

Theorem 3. $E[\tau(G, T_{\mathcal{F}} \triangle \{s,t\}, c)] \leq \frac{3}{5} c^\top x^*$.

Proof. We have two upper bounds for $\tau(G, T_F \triangle \{s,t\}, c)$ $(F \in \mathcal{F})$ (Fig. 1). The first is $c(F \setminus F(s,t))$, which is an upper bound because $F \setminus F(s,t)$ is a $T_F \triangle \{s,t\}$-join. The second upper bound will follow as an application of the last inequality of Theorem 2 to a vector z_F, whose feasibility for $P(G, \{s,t\})$ follows from the lower bound of the Lemma, while the length expectation $E[z_{\mathcal{F}}]$ can be bounded from above by the upper bound of the Lemma. In Case 1 the first bound is small in average, and when it is too large, the second bound is turning out to be small in average (Case 2).

Case 1: $c^\top q^* \leq 3/5\, c^\top x^*$. Then we are done, since

$$E[\tau(G, T_{\mathcal{F}} \triangle \{s,t\}, c)] \leq E[c(\mathcal{F} \setminus \mathcal{F}(s,t))] = c^\top q^* \leq 3/5\, c^\top x^*.$$

Case 2: $c^\top q^* \geq 3/5\, c^\top x^*$. Then by Fact 2, $c^\top p^* \leq 2/5\, c^\top x^*$.

In this case we construct for every $F \in \mathcal{F}$ a vector $z_F \in Q_+(G, T_F \triangle \{s,t\})$ (Claim). Then the last inequality of Theorem 2 establishes $\tau(G, T_{\mathcal{F}} \triangle \{s,t\}, c)] \leq c^\top z_F$, and therefore $E[\tau(G, T_{\mathcal{F}} \triangle \{s,t\}, c)] \leq E[c^\top z_{\mathcal{F}}]$.

Since $c^T z_F$ will be bounded in terms of $c^\top p^*$, itself bounded from above by $\frac{2}{5} c^\top x^*$ in our Case 2, $E[c^\top z_{\mathcal{F}}] \leq \frac{3}{5} c^\top x^*$ will follow, establishing the theorem.

Claim: $z_F := \frac{4}{9}x^* + \frac{1}{9}\Big(\chi_F + \sum_{Q\in\mathcal{Q}, |Q\cap F|\geq 2} \frac{x^Q}{\Pr(|Q\cap\mathcal{F}|\geq 2)}\Big) \in Q_+(G, T_F\triangle\{s,t\})$.

Indeed, we check that the inequalities defining $Q_+(G, T_F\triangle\{s,t\})$ (see Theorem 2) are all satisfied by z_F. Let C be a $T_F\triangle\{s,t\}$-cut.
First, if $C\notin\mathcal{Q}$, then regardless of whether it is a $T_F\triangle\{s,t\}$-cut or not,

$$z_F(C) \geq 2\,\frac{4}{9} + \frac{1}{9} = 1,$$

because then $x^*(C)\geq 2$, and by the connectivity of F, $\chi_F(C)\geq 1$.
Second, if $C\in\mathcal{Q}$ and it is a $T_F\triangle\{s,t\}$-cut, denoting $z := x^C(C)$:

$$z_F(C) \geq \frac{4}{9}(2-z) + \frac{1}{9}(2 + \frac{z}{1-z}) \geq 1,$$

because after evaluating z_F at C, the first term is $\frac{4}{9}x^*(C)$, and then we apply the lower bound of the Lemma. For the second term (first term of the second parenthesis), by the connectivity of F, this time $\chi_F(C)\geq 2$, since $|F\cap C|=1$ would imply that C is a T_F-cut, which is impossible, since it is an $\{s,t\}$-cut. (A T_F-cut which is a $\{s,t\}$-cut is not a $T_F\triangle\{s,t\}$-cut.) Last (for the first inequality), since $C\in\mathcal{Q}$, the expression

$$\frac{x^C}{\Pr(|C\cap\mathcal{F}|\geq 2)}$$

is among the terms of the definition of z_F, and *leaving only this term of the last* $\sum$ of this definition and recalling $z := x^C(C) = \Pr(|C\cap\mathcal{F}| = 1)$, the checking of the first inequality is finished.

The second inequality ≥ 1 follows now from $0<z<1$ and that in this interval the unique minimum of the function in variable z to bound is at $z = 1/2$, when its value is 1, finishing the proof of the claim.

Now by the Claim the last inequality of Theorem 2 can be applied to z_F:

$$E[\tau(G, T_{\mathcal{F}}\triangle\{s,t\}, c)] \leq E[c^\top z_{\mathcal{F}}] \leq \frac{4}{9}c^\top x^* + \frac{1}{9}\,c^\top x^* +$$

$$+\frac{1}{9}\sum_{F\in\mathcal{F}_+}\lambda_F \sum_{Q\in\mathcal{Q},|Q\cap F|\geq 2}\frac{c^\top x^Q}{\Pr(|Q\cap\mathcal{F}|\geq 2)}.$$

Exchanging the summation signs in this double-sum:

$$\sum_{F\in\mathcal{F}_+}\lambda_F \sum_{Q\in\mathcal{Q},|Q\cap F|\geq 2}\frac{c^\top x^Q}{\Pr(|Q\cap\mathcal{F}|\geq 2)} = \sum_{Q\in\mathcal{Q}}\Pr(|Q\cap\mathcal{F}|\geq 2)\frac{c^\top x^Q}{\Pr(|Q\cap\mathcal{F}|\geq 2)},$$

and now applying the upper bound of the Lemma and then the bound of our Case 2, we get that this expression is equal to:

$$c^\top\sum_{Q\in\mathcal{Q}} x^Q \leq c^\top p^* \leq \frac{2}{5}c^\top x^*.$$

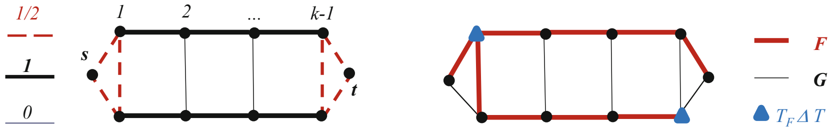

Fig. 2. The approximation guarantee cannot be improved below 3/2 with BOM. This example is essentially the same as the more complicated one in [22, Fig. 3] providing the same lower bound for a more powerful algorithm in the cardinality case. $|V| = 2k$, $\text{OPT}(G, T, \mathbb{1}) = c^\top x^* = 2k - 1$ (left). BOM output (right): $3k - 2$ if $\mathcal{F}_+$ consists of the thick (red) tree and its central symmetric image. There are more potential spanning trees for $\mathcal{F}_+$, but $\tau(G, T_F \triangle T, \mathbb{1}) \geq k - 2$ for each, so $c(F + J_F) \geq 3k - 3$ for each, and with any $T_F \triangle T$-join J_F.

So we finally got that

$$E[\tau(G, T_{\mathcal{F}} \triangle \{s,t\}, c)] \leq \frac{4}{9} c^\top x^* + \frac{1}{9} c^\top x^* + \frac{1}{9}\frac{2}{5} c^\top x^* = \frac{3}{5} c^\top x^*. \qquad \square$$

In Fig. 2 the optimum and the LP optimum are the same, but the BOM algorithm cannot decrease the approximation guarantee below 3/2.

Some of the questions that arise may be more hopefully tractable than the famous questions of the field:

Can the guarantee of BOM be improved for this problem or for other variants of the TSP ?

Namely are the results of [22] 3/2-*approximating minimum size T-tours or* 7/5-*approximating tours* be obtained *by BOM ?*

Could the new methods of analysis that have appeared in the last two years make the so far rigid bound of 3/2 move down at least for *shortest* 2-*edge-connected multigraphs* ?

Acknowledgment. Many thanks to the three IPCO reviewers' criticism concerning the presentation, and to IPCO's page limitation. The resulting reorganization process has eventually led to essential simplification, promptly and efficiently proof-read by Guyslain Naves and Zoli Szigeti, in the last hours. I also used many suggestions of the participants of the Cargèse meeting of Combinatorial Optimization (September 2012) and in particular of Attila Bernáth, R. Ravi, David Shmoys, as well as of Joseph Cheriyan and Zoli Király, concerning a preliminary version[4].

[4] The full manuscript with more motivations and more on the history and the reasons of the particular choices, as well as an analysis of the (im)possibility of mixing several methods, is planned to be submitted to a journal, and to be first presented in "Cahiers Leibniz":

`https://cahiersleibniz.g-scop.grenoble-inp.fr/apps/WebObjects/CahiersLeibnizApplication.woa/`

The previous version of the proof and some of the information that had to be deleted because of space limitation of IPCO can be accessed in the draft at `http://arxiv.org/abs/1209.3523v3`. Nevertheless, the present version is self-contained.

References

1. An, H.-C., Kleinberg, R., Shmoys, D.B.: Improving Christofides' algorithm for the s-t path TSP. In: Proceedings of the 44th Annual ACM Symposium on Theory of Computing (2012) (to appear)
2. Barahona, F., Conforti, M.: A construction for binary matroids. Discrete Mathematics 66, 213–218 (1987)
3. Christofides, N.: Worst-case analysis of a new heuristic for the traveling salesman problem. Technical Report 388, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh (1976)
4. Cheriyan, J., Friggstad, Z., Gao, Z.: Approximating Minimum-Cost Connected T-Joins, arXiv:1207.5722v1 (cs.DS) (2012)
5. Cook, W.J.: In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation. Princeton University Press (2012)
6. Cornuéjols, G., Fonlupt, J., Naddef, D.: The traveling salesman problem on a graph and some related integer polyhedra. Mathematical Programming 33, 1–27 (1985)
7. Edmonds, J.: Submodular functions, matroids and certain polyhedra. In: Guy, R., Hanani, H., Sauer, N., Schönheim, J. (eds.) Proceedings of the Calgary International Conference on Combinatorial Structures and Their Applications 1969, pp. 69–87. Gordon and Breach, New York (1970)
8. Edmonds, J., Johnson, E.L.: Matching, Euler tours and the Chinese postman. Mathematical Programming 5, 88–124 (1973)
9. Frank, A.: Connections in Combinatorial Optimization. Oxford University Press (2011)
10. Fulkerson, D.R.: Blocking Polyhedra. In: Graph Theory and Its Applications (Proceedings Advanced Seminar Madison, Wisconsin, 1969; Harris, B. (ed.)), pp. 93–112. Academic Press, New York (1970)
11. Garey, M.R., Johnson, D.S., Tarjan, R.E.: The planar Hamiltonian circuit problem is NP-complete. SIAM Journal on Computing 5, 704–714 (1976)
12. Gharan, S.O., Saberi, A., Singh, M.: A randomized rounding approach to the traveling salesman problem. In: Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science, pp. 550–559 (2011)
13. Grötschel, M., Lovász, L., Schrijver, A.: The ellipsoid method and its consequences in combinatorial optimization. Combinatorica 1(2), 169–197 (1981)
14. Guttmann-Beck, N., Hassin, R., Khuller, S., Raghavachari, B.: Approximation Algorithms with Bounded Performance Guarantees for the Clustered Traveling Salesman Problem. Algorithmica 28, 422–437 (2000)
15. Held, M., Karp, R.M.: The traveling-salesman problem and minimum spanning trees. Operations Research 18, 1138–1162 (1970)
16. Hoogeveen, J.A.: Analysis of Christofides' heuristic, some paths are more difficult than cycles. Operations Research Letters 10(5), 291–295 (1991)
17. Korte, B., Vygen, J.: Combinatorial Optimization, 5th edn. Springer (2012)
18. Lovász, L., Plummer, M.D.: Matching Theory. Akadémiai Kiadó, North-Holland, Budapest, Amsterdam (1986)
19. Mömke, T., Svensson, O.: Approximating graphic TSP by matchings. In: Proceedings of the 52nd Annual Symposium on Foundations of Computer Science, pp. 560–569 (2011)

20. Mucha, M.: $\frac{13}{9}$-approximation for graphic TSP. In: Proceedings of the 29th International Symposium on Theoretical Aspects of Computer Science, pp. 30–41 (2012)
21. Schrijver, A.: Combinatorial Optimization. Springer (2003)
22. Sebő, A., Vygen, J.: Shorter Tours by Nicer Ears: 7/5-approximation for graphic TSP, 3/2 for the path version, and 4/3 for two-edge-connected subgraphs, arXiv:1201.1870v3 (cs.DM) (2012)
23. Wolsey, L.A.: Heuristic analysis, linear programming and branch and bound. Mathematical Programming Study 13, 121–134 (1980)