

# Better $s$ - $t$ -Tours by Gao Trees

Corinna Gottschalk<sup>1</sup>(✉) and Jens Vygen<sup>2</sup>

<sup>1</sup> RWTH Aachen University, Aachen, Germany

corinna.gottschalk@oms.rwth-aachen.de

<sup>2</sup> University of Bonn, Bonn, Germany

vygen@or.uni-bonn.de

**Abstract.** We consider the  $s$ - $t$ -path TSP: given a finite metric space with two elements  $s$  and  $t$ , we look for a path from  $s$  to  $t$  that contains all the elements and has minimum total distance. We improve the approximation ratio for this problem from 1.599 to 1.566. Like previous algorithms, we solve the natural LP relaxation and represent an optimum solution  $x^*$  as a convex combination of spanning trees. Gao showed that there exists a spanning tree in the support of  $x^*$  that has only one edge in each narrow cut (i.e., each cut  $C$  with  $x^*(C) < 2$ ). Our main theorem says that the spanning trees in the convex combination can be chosen such that many of them are such “Gao trees” simultaneously at all sufficiently narrow cuts.

## 1 Introduction

The traveling salesman problem (TSP) is one of the best-known NP-hard problems in combinatorial optimization. In this paper, we consider the  $s$ - $t$ -path variant: given a finite metric space  $(V, c)$  and two elements  $s, t \in V$ , the goal is to find a sequence  $v_1, \dots, v_n$  containing every element exactly once and with  $v_1 = s$  and  $v_n = t$ , minimizing  $\sum_{i=1}^{n-1} c(v_i, v_{i+1})$ . For  $s = t$ , this is the well-known metric TSP; but in this paper we assume  $s \neq t$ .

The classical algorithm by Christofides (1976) computes a minimum-cost spanning tree  $(V, S)$  and then does *parity correction* by adding a minimum-cost matching on the vertices whose degree in  $S$  has the wrong parity.

While Christofides’ algorithm is still the best known approximation algorithm for metric TSP (with ratio  $\frac{3}{2}$ ), there have recently been improvements for special cases and variants (see e.g. Vygen’s 2012 survey), including the  $s$ - $t$ -path TSP.

### 1.1 Previous Work

For the  $s$ - $t$ -path TSP, Christofides’ algorithm has only an approximation ratio of  $\frac{5}{3}$  as shown by Hoogeveen (1991). An, Kleinberg and Shmoys (2015) were the first to improve on this and obtained an approximation ratio of  $\frac{1+\sqrt{5}}{2} \approx 1.618$ .

---

This work was done during the trimester program on combinatorial optimization at the Hausdorff Institute for Mathematics in Bonn.

They first solve the natural LP relaxation and represent an optimum solution  $x^*$  as a convex combination of spanning trees. This idea, first proposed by Held and Karp (1970), was exploited earlier for different TSP variants by Asadpour et al. (2010) and Oveis Gharan et al. (2011). Given this convex combination, An, Kleinberg and Shmoys (2015) do parity correction for each of the contributing trees and output the best of these solutions. Sebő (2013) improved the analysis of this *best-of-many Christofides algorithm* and obtained the approximation ratio  $\frac{8}{5}$ . Gao (2015) gave a unified analysis. Vygen (2015) suggested to “reassemble” the trees: starting with an arbitrary convex combination of spanning trees, he computed a different one, still representing  $x^*$ , that avoids certain bad local configurations. This led to the slightly better approximation ratio of 1.599.

In this paper, we will reassemble the trees more systematically to obtain a convex combination with strong global properties. This will enable us to control the cost of parity correction much better, leading to an approximation ratio of 1.566.

This also proves an upper bound of 1.566 on the integrality ratio of the natural LP. The only known lower bound is 1.5. Sebő and Vygen (2014) proved that the integrality ratio is indeed 1.5 for the graph  $s$ - $t$ -path TSP, i.e., the special case of *graph metrics* (where  $c(v, w)$  is the distance from  $v$  to  $w$  in a given unweighted graph on vertex set  $V$ ). Gao (2013) gave a simpler proof of this result, which inspired our work: see Sect. 1.4.

## 1.2 Notation and Preliminaries

Throughout this paper,  $(V, c)$  is the given metric space,  $n := |V|$ , and  $E$  denotes the set of edges of the complete graph on  $V$ . For any  $U \subseteq V$  we write  $E[U]$  for the set of edges with both endpoints in  $U$  and  $\delta(U)$  for the set of edges with exactly one endpoint in  $U$ ; moreover,  $\delta(v) := \delta(\{v\})$  for  $v \in V$ . If  $F \subseteq E$  and  $U \subseteq V$ , we denote by  $(V, F)[U]$  the subgraph  $(U, F \cap E[U])$  of  $(V, F)$  induced by  $U$ . For  $x \in \mathbb{R}_{\geq 0}^E$  we write  $c(x) := \sum_{e=\{v,w\} \in E} c(v, w)x_e$  and  $x(F) := \sum_{e \in F} x_e$  for  $F \subseteq E$ . Furthermore,  $\chi^F \in \{0, 1\}^E$  denotes the characteristic vector of a subset  $F \subseteq E$ , and  $c(F) := c(\chi^F)$  the cost of  $F$ . For  $F \subseteq E$  and  $f \in E$ , we write  $F + f$  and  $F - f$  for  $F \cup \{f\}$  and  $F \setminus \{f\}$ , respectively.

For  $T \subseteq V$  with  $|T|$  even, a  $T$ -join is a set  $J \subseteq E$  for which  $|\delta(v) \cap J|$  is odd if and only if  $v \in T$ . Edmonds (1965) proved that a minimum cost  $T$ -join can be computed in polynomial time. Moreover, the minimum cost of a  $T$ -join is the minimum over  $c(y)$  for  $y$  in the  $T$ -join polyhedron  $\{y \in \mathbb{R}_{\geq 0}^E : y(\delta(U)) \geq 1 \forall U \subseteq V \text{ with } |U \cap T| \text{ odd}\}$  as proved by Edmonds and Johnson (1973).

To obtain a solution for the  $s$ - $t$ -path TSP, it is sufficient to compute a connected multigraph with vertex set  $V$  in which exactly  $s$  and  $t$  have odd degree. We call such a graph an  $\{s, t\}$ -tour. As an  $\{s, t\}$ -tour contains an Eulerian walk from  $s$  to  $t$ , we can obtain a *Hamiltonian  $s$ - $t$ -path* (i.e. an  $s$ - $t$ -path with vertex set  $V$ ) by traversing the Eulerian walk and shortcutting when the walk encounters a vertex that has been visited already. Since  $c$  is a metric, it obeys the triangle inequality. Thus, the resulting path is not more expensive than the  $\{s, t\}$ -tour.

By  $\mathcal{S}$  we denote the set of edge sets of spanning trees in  $(V, E)$ . For  $S \in \mathcal{S}$ ,  $T_S$  denotes the set of vertices whose degree has the wrong parity in  $(V, S)$ , i.e.,

even for  $s$  or  $t$  and odd for  $v \in V \setminus \{s, t\}$ . Christofides' algorithm computes an  $S \in \mathcal{S}$  with minimum  $c(S)$  and adds a  $T_S$ -join  $J$  with minimum  $c(J)$ ; this yields an  $\{s, t\}$ -tour.

### 1.3 Best-of-Many Christofides

Like An, Kleinberg and Shmoys (2015), we begin by solving the natural LP relaxation:

$$\begin{aligned}
 & \min c(x) \\
 & \text{subject to } x(\delta(U)) \geq 2 \quad (\emptyset \neq U \subset V, |U \cap \{s, t\}| \text{ even}) \\
 & \quad \quad \quad x(\delta(U)) \geq 1 \quad (\emptyset \neq U \subset V, |U \cap \{s, t\}| \text{ odd}) \\
 & \quad \quad \quad x(\delta(v)) = 2 \quad (v \in V \setminus \{s, t\}) \\
 & \quad \quad \quad x(\delta(v)) = 1 \quad (v \in \{s, t\}) \\
 & \quad \quad \quad x_e \geq 0 \quad (e \in E)
 \end{aligned} \tag{1}$$

whose integral solutions are precisely the incidence vectors of the edge sets of the Hamiltonian  $s$ - $t$ -paths in  $(V, E)$ . This LP can be solved in polynomial time (either by the ellipsoid method Grötschel et al. (1981) or an extended formulation). Let  $x^*$  be an optimum basic solution; then  $x^*$  has at most  $2n - 3$  positive variables (Goemans 2006). Moreover,  $x^*$  (in fact, every feasible solution) satisfies  $x^*(E) = n - 1$  and  $x^*(E[U]) \leq |U| - 1$  for all  $\emptyset \neq U \subset V$ . Therefore  $x^*$  can be written as convex combination of spanning trees, i.e. as  $x^* = \sum_{S \in \mathcal{S}} p_S \chi^S$ , where  $p$  is a *distribution* on  $\mathcal{S}$ , i.e.,  $p_S \geq 0$  for all  $S \in \mathcal{S}$  and  $\sum_{S \in \mathcal{S}} p_S = 1$ .

As  $x^*$  has at most  $2n - 3$  positive variables, we can assume that  $p_S > 0$  for at most  $2n - 2$  spanning trees  $(V, S)$  by Carathéodory's theorem. Such spanning trees and numbers  $p_S$  can be computed in polynomial time, using either the ellipsoid method or the splitting-off technique (cf. Genova and Williamson 2015).

The best-of-many Christofides algorithm does the following. Compute an optimum solution  $x^*$  for (1) and obtain a distribution  $p$  with  $x^* = \sum_{S \in \mathcal{S}} p_S \chi^S$  as above. For each tree  $(V, S)$  with  $p_S > 0$ , compute a minimum weight  $T_S$ -join  $J_S$ . Then, the multigraph  $(V, S \dot{\cup} J_S)$  is an  $\{s, t\}$ -tour. Output the best of these.

We will fix  $x^*$  henceforth. An important concept in An, Kleinberg and Shmoys (2015) and the subsequent works are the so-called *narrow cuts*, i.e., the cuts  $C = \delta(U)$  with  $\emptyset \neq U \subset V$  and  $x^*(C) < 2$ . We denote by  $\mathcal{C}$  the set of all narrow cuts. We are going to exploit their structure as well.

**Lemma 1 (An, Kleinberg and Shmoys 2015).** *The narrow cuts form a chain: there are sets  $\{s\} = U_0 \subset U_1 \subset \dots \subset U_{\ell-1} \subset U_\ell = V \setminus \{t\}$  so that  $\mathcal{C} = \{\delta(U_i) : i = 0, \dots, \ell\}$ . These sets can be computed in polynomial time.*

We number the narrow cuts  $\mathcal{C} = \{C_0, C_1, \dots, C_\ell\}$  with  $C_i = \delta(U_i)$  ( $i = 0, \dots, \ell$ ).

### 1.4 Gao Trees

Our work was inspired by the following idea of Gao (2013):

**Theorem 1 (Gao 2013).** *There exists a spanning tree  $S \in \mathcal{S}$  with  $x_e^* > 0$  for all  $e \in S$  and  $|C \cap S| = 1$  for all  $C \in \mathcal{C}$ .*

In fact, Gao (2013) showed this for any vector  $x \in \mathbb{R}_{\geq 0}^E$  with  $x(\delta(U)) \geq 1$  for all  $\emptyset \neq U \subset V$  and  $x(\delta(U)) \geq 2$  for all  $\emptyset \neq U \subset V$  with  $|U \cap \{s, t\}|$  even. For graph  $s$ - $t$ -path TSP, one uses only variables corresponding to edges of the given graph. Then every spanning tree has cost  $n - 1$ . The approximation guarantee of  $\frac{3}{2}$  then follows from the fact that for a tree  $(V, S)$  with  $|S \cap C| = 1$  for all  $C \in \mathcal{C}$ , the vector  $\frac{1}{2}x^*$  is in the  $T_S$ -join polyhedron. But, as shown by Gao (2015), for the general  $s$ - $t$ -path TSP there may be no tree as in Theorem 1 whose cost is bounded by the LP value.

Let us call a tree  $S \in \mathcal{S}$  a *local Gao tree at  $C$*  if  $|C \cap S| = 1$ . We call  $S$  a *global Gao tree* if it is a local Gao tree at every narrow cut.

An, Kleinberg and Shmoys (2015) and Sebő (2013) observed that for every distribution  $p$  with  $x^* = \sum_{S \in \mathcal{S}} p_S \chi^S$  and every narrow cut  $C \in \mathcal{C}$ , at least a  $2 - x^*(C)$  fraction of the trees will be local Gao trees at  $C$ . However, in general none of these trees will be a global Gao tree.

### 1.5 Our Contribution

Our main contribution is a new structural result: Starting from an arbitrary distribution of trees representing  $x^*$ , we can compute a new distribution in which a sufficient number of trees are local Gao trees simultaneously for all sufficiently narrow cuts. For example, if  $x^*(C) = \frac{3}{2}$  for all  $C \in \mathcal{C}$ , at least half of our new distribution will be made of global Gao trees. Here is our main structure theorem:

**Theorem 2.** *For every feasible solution  $x^*$  of (1), there are  $S_1, \dots, S_r \in \mathcal{S}$  and  $p_1, \dots, p_r > 0$  with  $\sum_{j=1}^r p_j = 1$  such that  $x^* = \sum_{j=1}^r p_j \chi^{S_j}$  and for every  $C \in \mathcal{C}$  there exists a  $k \in \{1, \dots, r\}$  with  $\sum_{j=1}^k p_j \geq 2 - x^*(C)$  and  $|C \cap S_j| = 1$  for all  $j = 1, \dots, k$ .*

Note that this result immediately implies Theorem 1, simply by taking  $S_1$ .

In the next section we will prove Theorem 2. In Sect. 3 we show how to obtain such a distribution in polynomial time. Finally, in Sect. 4, we explain how this leads to an improved approximation guarantee of the best-of-many Christofides algorithm.

The intuition is as follows. For each tree  $S$  in our list, we find a vector  $y^S$  in the  $T_S$ -join polyhedron and use it to bound the cost of parity correction. We aim to bound the average cost of these vectors by an as small as possible multiple of  $c(x^*)$ . Following Sebő (2013), we design  $y^S$  as follows. If  $S = I_S \cup J_S$ , where  $I_S$  is the  $s$ - $t$ -path in  $S \in \mathcal{S}$ , then  $\chi^{J_S}$  is in the  $T_S$ -join polyhedron. Moreover,  $\frac{1}{2}x^*$  satisfies all constraints of the  $T_S$ -join polyhedron except those of narrow cuts  $C$  with  $|S \cap C|$  even. We choose  $y^S$  as a convex combination of  $\chi^{J_S}$  and  $\frac{1}{2}x^*$  but need to add a correction term for even narrow cuts. For this we can use some edges in  $I_{S'}$ , possibly of a different tree  $S'$ . The  $s$ - $t$ -paths of the early trees in our list, which are global Gao trees and thus do not need correction at narrow cuts, can thus help pay for the late trees.

## 2 Proof of the Structure Theorem

We will start with an arbitrary convex combination  $x^* = \sum_{S \in \mathcal{S}} p_S \chi^S$  where the  $p_S$  ( $S \in \mathcal{S}$ ) are rational. If  $r$  is a common denominator, then we can take  $rp_S$  copies of  $S$  and write  $x^* = \frac{1}{r} \sum_{j=1}^r \chi^{S_j}$ .

Starting from this list  $S_1, \dots, S_r \in \mathcal{S}$ , we will successively exchange a pair of edges in two of the trees. We will first satisfy the properties for the first tree  $S_1$ , then for  $S_2$ , and so on. For each  $S_j$ , we will work on the narrow cuts  $C_1, \dots, C_{\ell-1}$  in this order; note that  $|C_0 \cap S_j| = |C_\ell \cap S_j| = 1$  always holds for all  $j = 1, \dots, r$  due to  $x^*(\delta(s)) = x^*(\delta(t)) = 1$ . In the following we write

$$\theta_i := \lceil r(2 - x^*(C_i)) \rceil$$

for  $i = 0, \dots, \ell$ . Note that  $\theta_0 = r \geq \theta_i$  for all  $i = 1, \dots, \ell$  because  $x^*(C_i) \geq 1 = x^*(C_0) = x^*(\delta(s))$ . Our goal is to obtain  $|S_j \cap C_i| = 1$  whenever  $j \leq \theta_i$ .

Note that  $|S_j \cap C_i| = 1$  implies that  $(V, S_j)[U_i]$  is connected, and we will first obtain this weaker property by Lemma 6, before obtaining  $|S_j \cap C_i| = 1$  by Lemma 7.

We need a few preparations. Throughout this section, we use indices  $g, h, i, i'$  for cuts and  $j, j', k$  for trees. We first observe that a sufficient number of trees is connected in every section between two narrow cuts.

**Lemma 2.** *Let  $S_1, \dots, S_r \in \mathcal{S}$  such that  $x^* = \frac{1}{r} \sum_{j'=1}^r \chi^{S_{j'}}$ . Let  $0 \leq h < i \leq \ell$  and  $1 \leq j \leq r$  with  $j \leq \theta_h$  and  $j \leq \theta_i$ . Let  $M = U_i \setminus U_h$  or  $M = U_i$ . Then there exists an index  $k \geq j$  such that  $(V, S_k)[M]$  is connected.*

**Proof.** Assume the above is not true. Then  $|E[M] \cap S_{j'}| \leq |M| - 1$  for all  $j' < j$ , and  $|E[M] \cap S_{j'}| \leq |M| - 2$  for  $j' \geq j$ . Therefore,

$$\begin{aligned} x^*(E[M]) &= \frac{1}{r} \sum_{j'=1}^r |E[M] \cap S_{j'}| \\ &\leq \frac{1}{r} ((j-1)(|M|-1) + (r-j+1)(|M|-2)) = |M| - 2 + \frac{j-1}{r}. \end{aligned}$$

On the other hand,  $x^*(E[M]) = \frac{1}{2} (\sum_{v \in M} x^*(\delta(v)) - x^*(\delta(M)))$ .

If  $M = U_i \setminus U_h$ , we therefore have  $x^*(E[M]) = |M| - \frac{1}{2}(x^*(C_h) + x^*(C_i)) + x^*(C_h \cap C_i) \geq |M| - \frac{1}{2}(x^*(C_h) + x^*(C_i))$ .

If  $M = U_i$ , we have  $x^*(E[M]) = |M| - \frac{1}{2} - \frac{1}{2}x^*(C_i) = |M| - \frac{1}{2}(1 + x^*(C_i))$ .

Now,  $j \leq \theta_h$  and  $j \leq \theta_i$  implies  $1 \leq x^*(C_h) < 2 - \frac{j-1}{r}$  and  $x^*(C_i) < 2 - \frac{j-1}{r}$ . Thus, in both cases,

$$x^*(E[M]) > |M| - \frac{1}{2}(2 - \frac{j-1}{r} + 2 - \frac{j-1}{r}) = |M| - 2 + \frac{j-1}{r},$$

a contradiction. □

Next, a similar argument shows that, for any pair of narrow cuts, sufficiently many trees have no edge in their intersection:

**Lemma 3.** *Let  $S_1, \dots, S_r \in \mathcal{S}$  such that  $x^* = \frac{1}{r} \sum_{j'=1}^r \chi^{S_{j'}}$ . Let  $0 \leq h < i \leq \ell$  and  $1 \leq j \leq r$  with  $j \leq \theta_h$  and  $j \leq \theta_i$ . Then there exists an index  $k \geq j$  such that  $S_k \cap C_h \cap C_i = \emptyset$ .*

**Proof.** Using  $x^*(C_{i'}) < 2 - \frac{j-1}{r}$  for  $i' \in \{h, i\}$  and  $x^*(\delta(U)) \geq 2$  for  $|U \cap \{s, t\}|$  even, we obtain

$$x^*(C_h \cap C_i) = \frac{1}{2}(x^*(C_h) + x^*(C_i) - x^*(\delta(U_i \setminus U_h))) < 2 - \frac{j-1}{r} - 1 = \frac{r-j+1}{r}.$$

Therefore,  $\frac{1}{r} \sum_{j'=1}^r |S_{j'} \cap C_h \cap C_i| = x^*(C_h \cap C_i) < \frac{r-j+1}{r}$ , i.e. at most  $r-j$  trees can contain an edge in  $C_h \cap C_i$ .  $\square$

Finally, as mentioned already in Sect. 1.4, many trees are local Gao trees at a narrow cut:

**Lemma 4.** *Let  $S_1, \dots, S_r \in \mathcal{S}$  such that  $x^* = \frac{1}{r} \sum_{j'=1}^r \chi^{S_{j'}}$ . Let  $1 \leq i \leq \ell - 1$  and  $j \leq \theta_i$  with  $|C_i \cap S_j| \geq 2$ . Then there exists an index  $k > \theta_i$  with  $|C_i \cap S_k| = 1$ .*

**Proof.** Suppose there exists no such  $k$ . Then we get  $rx^*(C_i) = \sum_{j'=1}^r |C_i \cap S_{j'}| \geq (r - \theta_i + 1)2 + \theta_i - 1 = 2r - \lceil r(2 - x^*(C_i)) \rceil + 1 > 2r - r(2 - x^*(C_i)) = rx^*(C_i)$ , which is a contradiction.  $\square$

Now we proceed to the main components of the proof of Theorem 2. Recall that, in order to obtain  $|S_j \cap C_i| = 1$ , we plan to first get the weaker condition that  $(V, S_j)[U_i]$  is connected. While it is not, we exchange a pair of edges with a later tree. We split the proof into two parts, beginning with the following lemma:

**Lemma 5.** *Let  $1 \leq i \leq \ell - 1$  and  $M \subseteq U_i$  and  $S_j, S_k \in \mathcal{S}$  such that  $(V, S_j)[M]$  is disconnected and  $(V, S_k)[M]$  is connected and  $|S_j \cap \delta(U_i \setminus M)| \leq 1$ . Then there exist edges  $e \in S_j$  and  $f \in S_k$  such that  $S_j - e + f \in \mathcal{S}$  and  $S_k + e - f \in \mathcal{S}$  and  $e \notin E[U_i]$  and  $f \in E[M]$ .*

**Proof.** Let  $A_1, \dots, A_q$  be the vertex sets of the connected components of  $(V, S_j)[M]$ ; note that  $q \geq 2$  (For illustrations, see Fig. 1).

Let  $F := S_k \cap \bigcup_{p=1}^q (\delta(A_p) \setminus \delta(M))$  be the set of edges of  $S_k$  between the sets  $A_1, \dots, A_q$ . Note that  $F \subseteq E[M]$ . For  $p = 1, \dots, q$  let  $B_p$  be the set of vertices reachable from  $A_p$  in  $(V, S_k \setminus F)$ . Trivially,  $A_p \subseteq B_p$  for all  $p$ , and  $\{B_1, \dots, B_q\}$  is a partition of  $V$  because  $(V, S_k)[M]$  is connected and  $(V, S_k)$  is a tree.

Let  $Y$  be the union of the edge sets of the unique  $v$ - $w$ -paths in  $S_j$  for all  $v, w \in M$ . Note that  $Y \subseteq E[V \setminus (U_i \setminus M)]$  because  $|S_j \cap \delta(U_i \setminus M)| \leq 1$ .

**Claim:** There exists an index  $p \in \{1, \dots, q\}$  and an edge  $e \in Y \cap \delta(B_p)$  such that for every  $p' \in \{1, \dots, q\} \setminus \{p\}$  and  $v' \in A_{p'}$ ,  $v'' \in A_p$ , the  $v'$ - $v''$ -path in  $S_j$  contains  $e$ .

To prove the Claim, observe that  $(V, Y)$  consists of a tree and possibly isolated vertices. Choose an arbitrary root  $z$  in this tree and take an edge  $e \in Y \cap \bigcup_{p'=1}^q \delta(B_{p'})$  with maximum distance from  $z$ .

Let  $D := \{v \in V : v \text{ is reachable from } z \text{ in } (V, Y) \text{ but not in } (V, Y - e)\}$ . We will show that  $D \cap M = A_p$  for some  $p \in \{1, \dots, q\}$ . This will immediately imply that  $p$  and  $e$  satisfy the properties of the Claim.

As  $e \in Y$ , there are  $v, w \in M$  such that the  $v$ - $w$ -path in  $S_j$  contains  $e$ . Then exactly one of these two vertices belongs to  $D$ , so w.l.o.g.  $v \in D \cap M$ . Let  $p$  be the index with  $v \in A_p$ .

Since  $(V, S_j)[A_p]$  is connected, this implies  $A_p \subseteq D$ . Next,  $(D, Y \cap E[D])$  is a tree and by the choice of  $e$ , it contains no edge from  $\bigcup_{p'=1}^q \delta(B_{p'})$ . Therefore,  $D \subseteq B_p$  and hence,  $D \cap M = A_p$ . The Claim is proved.

Now, take an index  $p$  and an edge  $e$  as in the Claim. Consider the path  $P$  in  $S_k$  that connects the endpoints of  $e$ . Since  $e \in \delta(B_p)$ ,  $P$  has an edge  $f \in \delta(B_p) \cap S_k = \delta(A_p) \cap F$ . Thus,  $(V, S_k + e - f)$  is a tree. The path in  $S_j$  that connects the endpoints of  $f$  contains  $e$  by the Claim. Thus,  $(V, S_j - e + f)$  is a tree. We have  $f \in E[M]$  since  $f \in F$ , and  $e \notin E[U_i]$  as  $e \in Y \cap \delta(B_p)$ .  $\square$

Now we make  $(V, S_j)[U_i]$  connected without destroying previously obtained properties:

**Lemma 6.** *Let  $S_1, \dots, S_r \in \mathcal{S}$  such that  $x^* = \frac{1}{r} \sum_{j'=1}^r \chi^{S_{j'}}$ . Let  $1 \leq j \leq r$  and  $1 \leq i \leq \ell - 1$  such that  $j \leq \theta_i$  and  $|S_j \cap C_h| = 1$  for all  $h < i$  with  $j \leq \theta_h$ . Then we can find  $\hat{S}_1, \dots, \hat{S}_r \in \mathcal{S}$  in polynomial time such that  $x^* = \frac{1}{r} \sum_{j'=1}^r \chi^{\hat{S}_{j'}}$ , and  $\hat{S}_{j'} = S_{j'}$  for all  $j' < j$ , and  $|\hat{S}_j \cap C_h| = 1$  for all  $h < i$  with  $j \leq \theta_h$ , and  $(V, \hat{S}_j)[U_i]$  is connected.*

**Proof.** Assume that  $(V, S_j)[U_i]$  is disconnected, i.e.,  $|S_j \cap E[U_i]| < |U_i| - 1$ . Let  $h$  be the largest index smaller than  $i$  with  $j \leq \theta_h$ . Such an index must exist because  $\theta_0 \geq \theta_i \geq j$ .

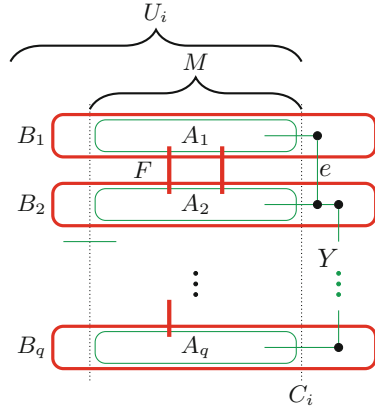
**Case 1:**  $|S_j \cap E[U_i \setminus U_h]| < |U_i \setminus U_h| - 1$ .

Let  $M := U_i \setminus U_h$ . Note that  $|S_j \cap \delta(U_i \setminus M)| = |S_j \cap C_h| = 1$ . Since  $(V, S_j)[M]$  is not connected, by Lemma 2 there exists an index  $k > j$  such that  $(V, S_k)[M]$  is connected.

Now we apply Lemma 5 and obtain two trees  $\hat{S}_j := S_j - e + f$  and  $\hat{S}_k := S_k + e - f$  with  $e \notin E[U_i]$  and  $f \in E[M]$ .

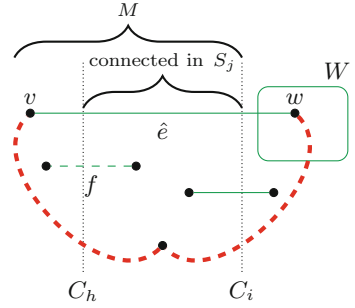
We have  $|\hat{S}_j \cap E[U_i]| = |S_j \cap E[U_i]| + 1$  and  $|\hat{S}_j \cap C_{h'}| \leq |S_j \cap C_{h'}|$  for all  $h' \leq h$  and hence  $|\hat{S}_j \cap C_{h'}| = 1$  for  $h' \leq h$  with  $j \leq \theta_{h'}$ . Note that  $j > \theta_{i'}$  for all  $h < i' < i$ , so a new edge  $f$  in such cuts  $C_{i'}$  does no harm.

We replace  $S_j$  and  $S_k$  by  $\hat{S}_j$  and  $\hat{S}_k$  and leave the other trees unchanged. If  $(V, \hat{S}_j)[U_i]$  is still not connected, we iterate.



**Fig. 1.** Edges of tree  $S_j$  are green/thin, edges of  $S_k$  are red/bold. Note that  $e$  could belong to  $C_i$  (but not to  $\delta(U_i \setminus M)$ ). (Color figure online)

**Case 2:**  $|S_j \cap E[U_i \setminus U_h]| = |U_i \setminus U_h| - 1$ .  $(V, S_j)[U_h]$  is connected since  $|S_j \cap C_h| = 1$ . Moreover,  $(V, S_j)[U_i \setminus U_h]$  is connected, but  $(V, S_j)[U_i]$  is disconnected. Therefore,  $S_j$  must contain an edge in  $C_h \cap C_i$  and  $S_j \cap C_h \subset S_j \cap C_i$  and  $(V, S_j)[U_i]$  has exactly two connected components:  $U_h$  and  $U_i \setminus U_h$  (For illustrations, see Fig. 2).



**Fig. 2.** Tree  $S_j$  green/thin,  $\hat{S}_{k'}$  red/bold. The dashed edge  $f$  is added to  $S_j$  by applying Lemma 5. (Color figure online)

**Case 2a:**  $h > 0$ . Let  $g$  be the largest index smaller than  $h$  with  $j \leq \theta_g$ . Set  $M := U_i \setminus U_g$ . Note that  $|S_j \cap \delta(U_i \setminus M)| = |S_j \cap C_g| = 1$ . By Lemma 2 there exists an index  $k \geq j$  with  $(V, S_k)[M]$  connected. As  $(V, S_j)[M]$  is not connected, we have  $k > j$ .

**Case 2b:**  $h = 0$ . Set  $M := U_i$ . Note that  $|S_j \cap \delta(U_i \setminus M)| = |S_j \cap \delta(\emptyset)| = 0$ . By Lemma 2, there exists an index  $k \geq j$  such that  $(V, S_k)[M]$  is connected. Since  $(V, S_j)[M]$  is not connected,  $k > j$ .

Note that in both cases 2a and 2b,  $(V, S_j)[M]$  is disconnected. Now we apply Lemma 5 to  $S_j$  and  $S_k$  and obtain two trees  $\hat{S}_j := S_j - e + f$  and  $\hat{S}_{k'} := S_k + e - f$  with  $e \notin E[U_i]$  and  $f \in E[M]$ . We replace  $S_j$  and  $S_k$  by  $\hat{S}_j$  and  $\hat{S}_{k'}$  and leave the other trees unchanged. Then  $(V, \hat{S}_j)[U_i]$  is connected. We have  $|\hat{S}_j \cap C_{h'}| = 1$  for all  $h' < h$  with  $j \leq \theta_{h'}$ , but we may have  $|\hat{S}_j \cap C_h| = 2$  since  $E[M] \cap C_h \neq \emptyset$ .

Assume  $|\hat{S}_j \cap C_h| = 2$  (otherwise we are done). Then  $\hat{S}_j \cap C_h \cap C_i = S_j \cap C_h \cap C_i = S_j \cap C_h$ , and this set contains precisely one edge  $\hat{e} = \{v, w\}$  (where  $v \in U_h$  and  $w \in V \setminus U_i$ ). By Lemma 3 there exists an index  $k' > j$  with  $\hat{S}_{k'} \cap C_h \cap C_i = \emptyset$ .

Let  $W$  be the set of vertices reachable from  $w$  in  $(V, \hat{S}_j \setminus C_i)$ . Since  $(V, \hat{S}_j)[U_i]$  is connected,  $\hat{S}_j \cap \delta(W) = \{\hat{e}\}$ . The unique path in  $(V, \hat{S}_{k'})$  from  $v$  to  $w$  contains at least one edge  $\hat{f} \in \delta(W)$ . Note that  $\hat{f} \notin C_h$  by the choice of  $\hat{S}_{k'}$ . We replace  $\hat{S}_j$  and  $\hat{S}_{k'}$  by  $\hat{\hat{S}}_j := \hat{S}_j - \hat{e} + \hat{f}$  and  $\hat{\hat{S}}_{k'} := \hat{S}_{k'} + \hat{e} - \hat{f}$ . Then  $(V, \hat{\hat{S}}_j)[U_i]$  is still connected and  $|\hat{\hat{S}}_j \cap C_{h'}| = 1$  for all  $h' < i$  with  $j \leq \theta_{h'}$ . □

**Lemma 7.** Let  $S_1, \dots, S_r \in \mathcal{S}$  such that  $x^* = \frac{1}{r} \sum_{j'=1}^r \chi^{S_{j'}}$ . Let  $1 \leq i \leq \ell - 1$  and  $j \leq \theta_i$  such that  $(V, S_j)[U_i]$  is connected and  $|S_j \cap C_h| = 1$  for all  $h < i$  with  $j \leq \theta_h$ . Then we can find  $\hat{S}_1, \dots, \hat{S}_r \in \mathcal{S}$  in polynomial time such that  $x^* = \frac{1}{r} \sum_{j'=1}^r \chi^{\hat{S}_{j'}}$  and  $\hat{S}_{j'} = S_{j'}$  for all  $j' < j$  and  $|\hat{S}_j \cap C_h| = 1$  for all  $h \leq i$  with  $j \leq \theta_h$ .

**Proof.** Suppose  $|S_j \cap C_i| \geq 2$ . Then by Lemma 4 there exists an index  $k > i$  with  $|S_k \cap C_i| = 1$ . We will swap a pair of edges, reducing  $|S_j \cap C_i|$  and increasing  $|S_k \cap C_i|$  while maintaining the other properties (For illustrations, see Fig. 3). Let  $S_k \cap C_i = \{\{x, y\}\}$  with  $x \in U_i$  and  $y \in V \setminus U_i$ . Let  $A$  be the set of vertices reachable from  $y$  in  $(V, S_j \setminus C_i)$ . Note that  $A \cap U_i = \emptyset$ . We have  $|\delta(A) \cap S_j \cap C_i| = 1$  because  $(V, S_j)[U_i]$  is connected. So let  $e = \{v, w\} \in (S_j \cap C_i) \setminus \delta(A)$ , with  $v \in U_i$  and  $w \in V \setminus U_i$ . Let  $B$  be the set of vertices reachable from  $w$  in  $(V, S_j \setminus C_i)$ . We

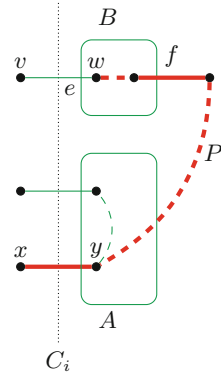


have  $w \in B$ ,  $y \in A$ , and  $A \cap B = \emptyset$  by the choice of  $e$ . Consider the path  $P$  in  $S_k$  from  $w$  to  $y$ . Note that  $P$  does not contain any vertex in  $U_i$  because  $|S_k \cap C_i| = 1$ . But  $P$  contains at least one edge  $f \in \delta(B)$ .

We will swap  $e$  and  $f$ . Since  $S_k \cap C_i = \{\{x, y\}\}$ , the  $w$ - $v$ -path in  $S_k$  contains  $P$ . Therefore,  $\hat{S}_k := S_k + e - f$  is a tree. On the other hand, the path in  $S_j$  connecting the endpoints of  $f$  must use an edge in  $\delta(B)$ . Since  $S_j \cap E[U_i]$  is connected and  $S_j \cap (\delta(B) \setminus C_i) = \emptyset$ ,  $e$  is the only edge in  $\delta(B) \cap S_j$  and thus,  $\hat{S}_j := S_j + f - e$  is a tree.

Since  $f \in E[V \setminus U_i]$  and  $e \in C_i$ , we have  $|\hat{S}_j \cap C_i| = |S_j \cap C_i| - 1$  and  $|\hat{S}_j \cap C_h| = |S_j \cap C_h|$  for all  $h < i$  with  $j \leq \theta_h$ . Moreover,  $(V, \hat{S}_j)[U_i]$  is still connected. As before, we replace  $S_j$  and  $S_k$  by  $\hat{S}_j$  and  $\hat{S}_k$  and leave the other trees unchanged. If  $|\hat{S}_j \cap C_i| > 1$ , we iterate.  $\square$

Now the proof of Theorem 2 is a simple induction. We scan the indices of the trees  $j = 1, \dots, r$  in this order. For each  $j$ , we consider all narrow cuts  $C_i$  with  $j \leq \theta_i$ . Since  $x^*(\delta(s)) = x^*(\delta(t)) = 1$ , we always have  $|S_j \cap C_0| = 1$  and  $|S_j \cap C_\ell| = 1$  for all  $j = 1, \dots, r$ . Now let  $i \in \{1, \dots, \ell - 1\}$  with  $j \leq \theta_i$ . Assuming  $|S_j \cap C_h| = 1$  for all  $h < i$  with  $j \leq \theta_h$ , we first apply Lemma 6 and then Lemma 7. The new tree then satisfies  $|S_j \cap C_h| = 1$  for all  $h \leq i$  with  $j \leq \theta_h$ , and  $S_1, \dots, S_{j-1}$  remain unchanged.



**Fig. 3.** Tree  $S_j$  green/thin,  $S_k$  red/bold (Color figure online)

### 3 Obtaining the Distribution in Polynomial Time

So far it was not clear whether the number  $r$  of trees in our distribution can be polynomially bounded. In this section we show two solutions to this question.

First, one can start with an arbitrary distribution  $p$  with at most  $2n - 2$  trees  $S$  with  $p_S > 0$  and round the coefficients down to integral multiples of  $\frac{\epsilon}{2n^2}$  for a sufficiently small constant  $\epsilon > 0$ , and then scale up all coefficients so that their sum is 1 again. This way we will get a vector  $x$  close to  $x^*$  that we can write as  $x = \sum_{j=1}^r \frac{1}{r} \chi^{S_j}$ , where  $r \leq \frac{2n^2}{\epsilon}$  and  $S_j \in \mathcal{S}$  for  $j = 1, \dots, r$ . It is not difficult to show that  $x \in \mathbb{R}_{\geq 0}^E$  satisfies the properties

$$x(\delta(s)) = x(\delta(t)) = 1 \quad \text{and} \quad x^*(F) - \epsilon \leq x(F) \leq x^*(F) + \epsilon \quad \text{for all } F \subseteq E. \quad (2)$$

In the full version of this paper [arXiv:1511.05514] we show that the proof in Sect. 2 also works in this case. More precisely:

**Theorem 3.** *Given  $S_1, \dots, S_r \in \mathcal{S}$ , a feasible solution  $x^*$  of (1) and  $\epsilon \geq 0$  such that  $x = \frac{1}{r} \sum_{j=1}^r \chi^{S_j}$  satisfies (2), we can find  $\hat{S}_1, \dots, \hat{S}_r \in \mathcal{S}$  in polynomial time such that  $x = \frac{1}{r} \sum_{j=1}^r \chi^{\hat{S}_j}$ , and for every  $C \in \mathcal{C}$  there exists a  $k \in \{1, \dots, r\}$  with  $\frac{k}{r} \geq 2 - x^*(C) - \epsilon$  and  $|C \cap \hat{S}_j| = 1$  for all  $j = 1, \dots, k$ .*

This is sufficient to obtain the claimed approximation ratio if  $\epsilon$  is chosen small enough. However, Kanstantsin Pashkovich [private communication, 2015] suggested a more elegant solution: Theorem 2 implies a stronger version in which  $r$  can be chosen to be less than  $2n^2$  and the trees and  $p$  can be found in polynomial time. We now explain how.

As before, fix an optimum basic solution  $x^*$  of (1). Let  $\xi^1 > \dots > \xi^k = 1$  be the distinct values among  $\{x^*(C) : C \in \mathcal{C}\}$  and  $\xi^0 := 2$ . Note that  $k \leq \ell \leq n - 2$  by Lemma 1. Then Theorem 2 implies that the polytope defined by

$$\begin{aligned} \sum_{h=1}^k (\xi^{h-1} - \xi^h) x^h &= x^* \\ x^h(C_i) &= 1 && (1 \leq h \leq k, 0 \leq i \leq \ell, x^*(C_i) \leq \xi^h) \\ x^h(E) &= n - 1 && (1 \leq h \leq k) \\ x^h(E[U]) &\leq |U| - 1 && (1 \leq h \leq k, \emptyset \neq U \subset V) \\ x_e^h &\geq 0 && (1 \leq h \leq k, e \in E) \end{aligned}$$

is nonempty: if  $S_1, \dots, S_r$  and  $p_1, \dots, p_r$  are as in Theorem 2, then

$$x^h = \sum_{j=1}^r \frac{\max\left\{0, \min\left\{2 - \xi^h, \sum_{j' \leq j} p_{j'}\right\} - \max\left\{2 - \xi^{h-1}, \sum_{j' < j} p_{j'}\right\}\right\}}{\xi^{h-1} - \xi^h} \chi^{S_j}$$

defines a feasible solution. We can find a vector  $x$  in this polytope in polynomial time by the ellipsoid method because the separation problem can be solved in polynomial time. Then, writing each  $x^h$  as convex combination  $x^h = \sum_{j=1}^{2n-2} p_j^h \chi^{S_j^h}$  of  $2n - 2$  spanning trees, and setting  $S_{(h-1)(2n-2)+j} := S_j^h$  and  $p_{(h-1)(2n-2)+j} := (\xi^{h-1} - \xi^h) p_j^h$  for  $h = 1, \dots, k$  and  $j = 1, \dots, 2n - 2$ , yields a decomposition  $x^* = \sum_{j=1}^{k(2n-2)} p_j \chi^{S_j}$  with at most  $(2n - 2)(n - 2)$  spanning trees and the properties of Theorem 2.

### 4 Analysis of the Approximation Ratio

In this section, we will analyze the best-of-many Christofides algorithm on a distribution as in Theorem 2. We follow the framework from Vygen (2015) (based on An, Kleinberg and Shmoys 2015 and Sebő 2013); see the end of Sect. 1 for an intuition. In particular, we use the following definition and lemma from Vygen (2015):

**Definition 1.** *Given numbers  $0 \leq \gamma_S \leq 1$  for  $S \in \mathcal{S}$  and  $\beta < \frac{1}{2}$ , we define the benefit of  $(S, C) \in \mathcal{S} \times \mathcal{C}$  to be  $b_{S,C} := \min\left\{\frac{\beta(2-x^*(C))}{1-2\beta}, \gamma_S\right\}$  if  $|S \cap C|$  is even,  $b_{S,C} := 1 - \gamma_S$  if  $|S \cap C| = 1$ , and  $b_{S,C} = 0$  otherwise.*

**Lemma 8.** *Let  $0 \leq \beta < \frac{1}{2}$  and  $0 \leq \gamma_S \leq 1$  for  $S \in \mathcal{S}$ . Let  $p$  be a distribution on  $\mathcal{S}$  with  $x^* = \sum_{S \in \mathcal{S}} p_S \chi^S$ . If*

$$\sum_{S \in \mathcal{S}} p_S b_{S,C} \geq \frac{\beta}{1-2\beta} (2 - x^*(C)) \sum_{S \in \mathcal{S}: |S \cap C| \text{ even}} p_S \tag{3}$$

for all  $C \in \mathcal{C}$ , then the best-of-many Christofides algorithm run on the trees  $S \in \mathcal{S}$  with  $p_S > 0$  returns a solution of cost at most  $(2 - \beta)c(x^*)$

We now show how to set the  $\gamma$ -constants in order to maximize the benefits, with the ultimate goal to choose  $\beta$  as large as possible.

**Lemma 9.** *Let  $S_1, \dots, S_r \in \mathcal{S}$  such that  $x^* = \frac{1}{r} \sum_{j=1}^r \chi^{S_j}$ ,  $r$  is even, and for every  $C \in \mathcal{C}$  there exists a  $k \in \{1, \dots, r\}$  with  $\frac{k}{r} \geq 2 - x^*(C)$  and  $|C \cap S_j| = 1$  for all  $j = 1, \dots, k$ .*

We set  $\delta := 0.126$ , and  $\gamma_{S_j} = \delta$  if  $j \leq \frac{r}{2}$  and  $\gamma_{S_j} = 1 - \delta$  otherwise. Choose  $\beta$  such that  $\frac{\beta}{1-2\beta} = 3.327$ . Then

$$\frac{1}{r} \sum_{j=1}^r b_{S_j, C_i} \geq 3.327 (2 - x^*(C_i)) \frac{1}{r} |\{j : |S_j \cap C_i| \text{ even}\}| \tag{4}$$

for all  $i = 0, \dots, \ell$ .

For the proof of this lemma, see the full version of this paper [arXiv:1511.05514]. Now we obtain our approximation guarantee:

**Theorem 4.** *There is a 1.566-approximation algorithm for the  $s$ - $t$ -path TSP.*

**Proof.** Let  $x^*$  be an optimal solution to (1). Let  $S_1, \dots, S_{r'}$   $\in \mathcal{S}$  and rational  $p_1, \dots, p_{r'} > 0$  as in Theorem 2. We showed in Sect. 3 how to obtain this in polynomial time. Therefore, we can apply Lemma 9 to the trees  $S_1, \dots, S_{r'}$  with appropriate multiplicities and obtain inequality (3) for  $\frac{\beta}{1-2\beta} = 3.327$ . Equivalently,  $\beta = \frac{3 \cdot 327}{7.654} > 0.434$ .

Thus, the conditions in Lemma 8 are met and best-of-many Christofides yields an approximation ratio of at most  $2 - \beta < 1.566$ . □

## 5 Conclusion

The approximation ratio can probably be improved slightly by choosing the  $\gamma_{S_j}$  differently, but still depending only on  $\frac{j}{r}$ . However, using an analysis based on Lemma 8, one cannot obtain a better approximation ratio than  $\frac{14}{9}$  because the benefit can never be more than one and there can be cuts  $C$  with  $x^*(C) = \frac{3}{2}$  and  $\sum_{S \in \mathcal{S}: |S \cap C| \text{ even}} p_S = \frac{1}{2}$ ; therefore  $\frac{\beta}{1-2\beta} \leq 4$ . Our ratio is already close to this threshold.

On the other hand, it is not impossible that the best-of-many Christofides algorithm on a distribution like the one obtained in Theorem 2, or even on an arbitrary distribution, has a better approximation ratio, maybe even  $\frac{3}{2}$ .

**Acknowledgement.** We thank Kanstantsin Pashkovich for allowing us to include his idea described in the second half of Sect. 3.

## References

- An, H.-C., Kleinberg, R., Shmoys, D.B.: Improving Christofides' algorithm for the  $s$ - $t$  path TSP. *J. ACM* **62**, Article 34, 34:1–34:28 (2015)
- Asadpour, A., Goemans, M.X., Mądry, A., Oveis Gharan, S., Saberi, A.: An  $O(\log n / \log \log n)$ -approximation algorithm for the asymmetric traveling salesman problem. In: Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2010), pp. 379–389 (2010)
- 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)
- Edmonds, J.: The Chinese postman's problem. *Bull. Oper. Res. Soc. Am.* **13**, B-73 (1965)
- Edmonds, J., Johnson, E.L.: Matching, Euler tours and the Chinese postman. *Math. Program.* **5**, 88–124 (1973)
- Gao, Z.: An LP-based  $\frac{3}{2}$ -approximation algorithm for the  $s$ - $t$  path graph traveling salesman problem. *Oper. Res. Lett.* **41**, 615–617 (2013)
- Gao, Z.: On the metric  $s$ - $t$  path traveling salesman problem. *SIAM J. Discrete Math.* **29**, 1133–1149 (2015)
- Genova, K., Williamson, D.P.: An experimental evaluation of the best-of-many Christofides' algorithm for the traveling salesman problem. In: Bansal, N., Finocchi, I. (eds.) Algorithms – ESA 2015. LNCS, pp. 570–581. Springer, Heidelberg (2015)
- Goemans, M.X.: Minimum bounded-degree spanning trees. In: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), pp. 273–282 (2006)
- Grötschel, M., Lovász, L., Schrijver, A.: The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica* **1**, 169–197 (1981)
- Held, M., Karp, R.M.: The traveling-salesman problem and minimum spanning trees. *Oper. Res.* **18**, 1138–1162 (1970)
- Hoogeveen, J.A.: Analysis of Christofides' heuristic: some paths are more difficult than cycles. *Oper. Res. Lett.* **10**, 291–295 (1991)
- Oveis Gharan, S., 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 (FOCS 2011), pp. 550–559 (2011)
- Sebő, A.: Eight-fifth approximation for the path TSP. In: Goemans, M., Correa, J. (eds.) IPCO 2013. LNCS, vol. 7801, pp. 362–374. Springer, Heidelberg (2013)
- Sebő, A., Vygen, J.: Shorter tours by nicer ears:  $7/5$ -approximation for graph-TSP,  $3/2$  for the path version, and  $4/3$  for two-edge-connected subgraphs. *Combinatorica* **34**, 597–629 (2014)
- Vygen, J.: New approximation algorithms for the TSP. *OPTIMA* **90**, 1–12 (2012)
- Vygen, J.: Reassembling trees for the traveling salesman. *SIAM J. Discrete Math.* To appear. [arXiv:1502.03715](https://arxiv.org/abs/1502.03715)