

# A 5.875-approximation for the Traveling Tournament Problem

Stephan Westphal · Karl Noparlik

Published online: 26 January 2012

© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** In this paper we propose an approximation for the Traveling Tournament Problem which is the problem of designing a schedule for a sports league consisting of a set of teams  $T$  such that the total traveling costs of the teams are minimized. It is not allowed for any team to have more than  $k$  home-games or  $k$  away-games in a row. We propose an algorithm which approximates the optimal solution by a factor of  $2 + 2k/n + k/(n - 1) + 3/n + 3/(2 \cdot k)$  which is not more than 5.875 for any choice of  $k \geq 4$  and  $n \geq 6$ . This is the first constant factor approximation for  $k > 3$ .

We furthermore show that this algorithm is also applicable to real-world problems as it produces solutions of high quality in a very short amount of time. It was able to find solutions for a number of well known benchmark instances which are even better than the previously known ones.

**Keywords** Sports scheduling · Traveling Tournament Problem · Approximation algorithms

## 1 Introduction

During the last decades professional sports leagues worldwide have turned into million or sometimes even billion dollar businesses. Soccer in Europe as well as American Football, basketball, baseball or ice hockey in North America attracts thousands of fans inside the stadiums and millions of spectators around the world. A crucial contribution to the success of a season lies in the timetable or schedule of the league which determines which games are arranged when and at which arenas. In doing so, the planers of those leagues have to

---

S. Westphal (✉)

Institute for Numerical and Applied Mathematics, Georg-August University, Lotzestr. 16–18,  
37083 Göttingen, Germany  
e-mail: [s.westphal@math.uni-goettingen.de](mailto:s.westphal@math.uni-goettingen.de)

K. Noparlik

Department of Mathematics, University of Kaiserslautern, Paul-Ehrlich-Str. 14, 67663 Kaiserslautern,  
Germany  
e-mail: [noparlik@mathematik.uni-kl.de](mailto:noparlik@mathematik.uni-kl.de)

balance not only the expectations of the fans but also many requests made by clubs and TV stations.

In this paper we will focus on the Traveling Tournament Problem (TTP) introduced by Easton et al. (2001). It is a quite well-known and practically difficult optimization problem inspired by Major League Baseball. Especially North American sports leagues have an incentive to minimize the travel distance of the participants of a tournament due to the vast expanse of the continent.

### 1.1 Sports scheduling and the traveling tournament problem

*Sports scheduling* in general deals with the design of tournaments. A *single round robin tournament* on  $n$  teams where  $n$  is an even number consists of  $(n - 1)$  rounds (also called *slots*). In each round  $n/2$  games which are themselves ordered pairs of teams take place. Every team has to participate in one game per round and must meet every other team exactly once. It is standard to assume  $n$  to be even since in sports leagues with  $n$  being odd, usually a dummy team is introduced, and whoever plays it has a day off, which is called a *bye*. For scheduling single round robin tournaments a rather general and useful scheme called the *canonical schedule* has been known in sports scheduling literature for at least 30 years (de Werra 1981). It is based on the polygon/circle method, which was first suggested by Kirkman (1847). One can think of Kirkman's method as a long table at which  $n$  players sit such that  $n/2$  players on one side face the other players seated on the other side of the table. Every player plays a match against the person seated directly across the table. The next round of the schedule is obtained when everyone moves one chair to the right with the crucial exception that there exists one person at the end of the table who never moves and always maintains the seat from his or her first round. Note that this method only specifies who plays whom when and not where. The canonical schedule introduced by de Werra defines for each of the encounters specified by the method described above, at whose site they take place such that the number of successive home or away games is minimized (de Werra 1981).

A *double round robin tournament* on  $n$  teams consists of  $2(n - 1)$  rounds and every team must meet every other team twice: once at its own home venue (*home game*) and once at the other team's venue (*away game*). A popular policy in practice is to obtain a double round robin tournament from a single round robin tournament by *mirroring*, that is repeating the matches of round  $k$  for  $k = 1, \dots, n - 1$  in round  $k + n - 1$  with changed home field advantage. Consecutive home games are called a *home stand* and consecutive away games form a *road trip*. The *length* of a home stand or road trip is the number of opponents played (and not the distance traveled).

The *Traveling Tournament Problem (TTP)* as introduced in Easton et al. (2003) is then defined as follows:

*Input:*

- A set  $V = \{1, 2, \dots, n\}$  of  $n$  teams with  $n$  even.
- An  $n \times n$  integer distance matrix  $D$  containing the metric travel distances between the home venues of all teams.
- Integers  $L, k$ , representing lower and upper bounds for the lengths of the teams' home stands and road trips.

*Output:* A double round robin tournament on  $V$  satisfying:

- The length of every home stand and road trip is between  $L$  and  $k$  inclusive.

- No pair of teams plays both of their matches against each other in two successive rounds.
- The total distance traveled by all teams is minimized.

In this paper, we assume that  $L = 1$  which is common in literature and means that we forget about  $L$ . This assumption is reasonable since it is hard to imagine a sports league planner who will insist on forbidding home stands or road trips of length 1 when facing his many conflictive objectives.

## 1.2 Previous work

So far, most efforts concerning the TTP have led to a variety of algorithms aiming to minimize the total distance driven by the teams. Kendall et al. (2010) and Rasmussen and Trick (2008) provide a good overview of the work done on the TTP and sports scheduling in general. Just to mention a very few examples, hybrid algorithms with constraint programming (CP) exist by Benoist et al. (2001) who additionally use Lagrange relaxation. Easton et al. (2003) merge CP with integer programming while Henz (2004) combines CP with large neighborhood search. Anagnostopoulos et al. (2003), van Hentenryck and Vergados (2006), Di Gaspero and Schaerf (2007) and Lim et al. (2006) propose neighborhood search-based algorithms, whereas Ribeiro and Urrutia (2007) focus on the special class of constant distance TTP where break maximization is equivalent to travel distance minimization.

Miyashiro et al. (2008) provide a  $2 + (9/4)/(n - 1)$  approximation for the intensively studied special case  $k = 3$  by means of the *Modified Circle Method*, a variation of the canonical schedule. Yamaguchi et al. (2009) obtain an algorithm with approximation ratio  $(2k - 1)/k + O(k/n)$  for  $k \leq 5$  and  $(5k - 7)/(2k) + O(k/n)$  for  $k > 5$ . Again they make use of the canonical schedule, now refined such that the teams are ordered around the ‘table’ such that most of the distances driven are part of a near optimal traveling salesman tour which clearly has positive effects on the length of many distances traveled. As  $k \leq n - 1$ , they showed this way that a constant factor approximation for any choice of  $k$  and  $n$  exists. However, they did not show how this factor looks exactly.

The complexity has been settled quite recently by Thielen and Westphal (2011) who showed that the TTP is strongly *NP*-hard. They use a reduction from 3-satisfiability (3-SAT) by showing that for any instance  $\phi$  of 3-SAT there is an instance of TTP(3) and some  $\xi$  such that there is a feasible schedule of cost at most  $\xi$  if and only if  $\phi$  is solvable.

## 1.3 Our results

Our aim in this work is to approximate the TTP by a constant ratio for arbitrary choices of  $k$  and  $n$ .

Applying the canonical schedule mentioned above, we choose a specific orientation of the underlying graph which ensures that home stands and road trips do not contain more than  $k$  matches and for which the total distance traveled is not too long. Whereas it is common practice to derive the second half of the season by repeating the first half’s games in the same order but with changed home field advantage, it is not suitable here, as road trips or home stands might become too long. Thus, we derive the second half in a slightly different way. Finally, we show that the plan we construct approximates the optimal solution by a factor of  $2 + 2k/n + k/(n - 1) + 3/n + 3/(2 \cdot k)$ . For the case of  $k = 3$  this guarantees an approximation ratio of  $5/2 + 12/(n - 1)$  which is actually not an improvement on the ratio of Miyashiro et al. cited above. But for any choice of  $k \geq 4$  (and thus  $n \geq 6$ ) this yields an approximation ratio of less than 5.875, which is the first constant factor approximation for  $k > 3$ .

Furthermore, we show that this algorithm is also applicable to real-world applications as it produces solutions of high quality in a very short amount of time. We were also able to find solutions for the benchmark instances Galaxy22 to Galaxy40 which are better than the previously known ones.

## 2 Lower bounds

The objective of the TTP, minimizing the total travel distance of all teams during a double round robin tournament, can be estimated by various bounds. One of them is called Independent Lower Bound (ILB) (Easton et al. 2001) and consists of finding the shortest tour for each team individually, independent of the other team constraints (primarily that  $B$  has to be at home when  $A$  visits  $B$  during one of  $A$ 's road trips). Finding an ILB is equivalent to solving a capacitated vehicle routing problem. In this paper we will use an even coarser version of ILB where we focus only on a traveling salesman tour traversing all venues.

**Theorem 1** *Let  $\rho$  be the length of a TSP in  $G$ . Every solution of the TTP has a total length of at least  $n \cdot \rho$ .*

*Proof* Every team has to visit all the other teams. Thus, each team has to travel at least a distance of  $\rho$  which gives a total distance of  $n \cdot \rho$ . □

As in Miyashiro et al. (2008), we denote the sum of the distances of all ordered pairs of teams as  $\Delta = \sum_{i,j \in T} d(i, j)$ . Miyashiro et al. (2008) showed a lower bound of  $2/3 \cdot \Delta$  for the objective function of TTP with  $k = 3$ . We generalize this result for arbitrary  $k$ :

**Theorem 2** *Every solution of the TTP has a total length of at least  $2/k \cdot \Delta$ .*

*Proof* Consider an arbitrary solution and suppose team  $i$  plays  $l \leq k$  consecutive away games at teams  $t_1, t_2, \dots, t_l$ . The distance  $\tilde{d}_i$  driven thereby is

$$\tilde{d}_i = d(i, t_1) + \sum_{j=1}^{l-1} d(t_j, t_{j+1}) + d(t_l, i)$$

Because of the triangle inequality we have  $\tilde{d}_i \geq 2 \cdot d(i, t_j)$  for all  $j$  and thus we have

$$l \cdot \tilde{d}_i \geq 2 \cdot \sum_{j=1}^l d(i, t_j) \implies \tilde{d}_i \geq \frac{2}{k} \cdot \sum_{j=1}^l d(i, t_j)$$

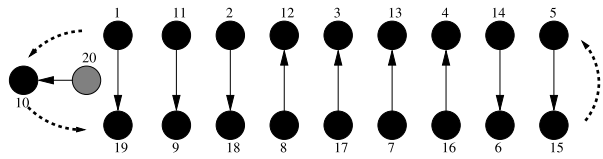
Summing up over all tours driven yields the desired lower bound of  $2/k \cdot \Delta$  for the total distance driven by the teams in any solution. □

## 3 Construction of the tournament

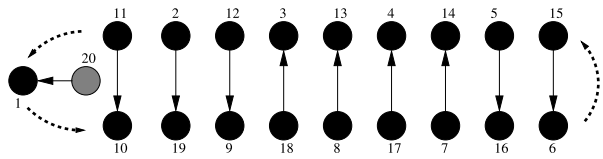
For  $i \in V$  let  $s(i) = \sum_{j \in V} d(i, j)$  be the *star-weight* of  $i$ . Since

$$\sum_{i \in V} s(i) = \sum_{i \in V, j \in V} d(i, j) = \Delta$$

**Fig. 1** Example for slot 1 with  $n = 20, k = 4$  and  $l = 2$



**Fig. 2** Example for slot 2 with  $n = 20, k = 4$  and  $l = 2$



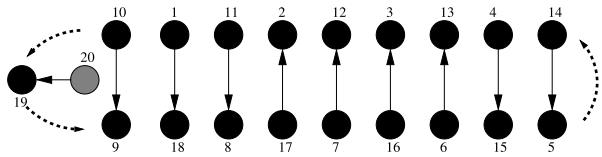
there has to be one  $j \in V$  for which  $s(i) \leq \Delta/n$ . Let  $T_{heu}$  be a tour through all of the teams' venues which has been found by applying the well known heuristic by Christofides (1976). Therefore, we know that this tour is not more than 1.5 times longer than the shortest possible tour. We furthermore assume that the teams are named in a way such that  $T_{heu}$  traverses them in the order  $1, 2, \dots, n$  and that  $n$  is the team with minimum star weight. Given this tour we construct a solution of the TTP in the following way. For  $n = 20$  the games of the first two rounds of the season are displayed in Figs. 1 and 2. The figures corresponding to other choices of  $n$  can be derived analogously. A solid arc  $(u, v)$  in this digraph means that team  $u$  is playing against team  $v$  in the arena of team  $v$ . The games of the other rounds can be derived analogously by changing the positions of the teams counterclockwise. The only arc which changes its orientation during one half of the season is the arc incident to node  $n$  which changes its orientation every  $k$ th match. This way, the season starts for team 4 with a tour visiting the teams 16, 17, 18 and 19 before coming home and then playing against the teams 1, 2 and 3. Then, it starts off again to play against 20, 5, 6, 7, and then a home stand again consisting of matches against 8, 9, 10, 11. Finally, there is a last road trip including 12 and 13 and a last home stand with 14 and 15. It is clear that no team has home stands or road trips which are longer than  $k$  matches. And it is also clear that every two teams have met each other during this first  $n - 1$  games. The full schedule for the first half of the season is shown in Table 1.

In order to construct a full tournament, it remains to construct the second half of the season. If we just repeated the first  $n - 1$  matches with changed locations (changed the orientation of the arcs), we would obtain a solution, in which every pair of teams met twice and these two games took place at different sites. Furthermore, no half of the season would contain a road trip or a home stand longer than  $k$  matches. However, this solution could contain road trips and home stands being longer than  $k$ . For example, the team 4 we considered above would start into the second half of the tournament with a home stand of length 4 after having ended the first half with two home stands. In order to get rid of this problem, we start the second half with the match of round  $n - 2$ , succeeded by the matches of the rounds  $n - 1, 1, 2, \dots, n - 3$  in this order (see Figs. 3 and 4). The double round robin tournament obtained this way contains neither road trips nor home stands longer than  $k$ . To see this, assume for the sake of a contradiction that there is a team  $t$  which has a road trip longer than  $k$ . It is clear from construction that no half of the season completely contains such a tour. Thus, the tour has to include the rounds  $n - 1$  and  $n$ . In case  $t$  has away-games at both of these rounds, the other matches involving these opponents will be home-games for  $t$ . By construction, these games will take place on the rounds  $n - 2$  and  $n + 1$  which means that the road trip had only a length of 2, contradicting the assumption. The case for home stands that are too long follows along the same lines.

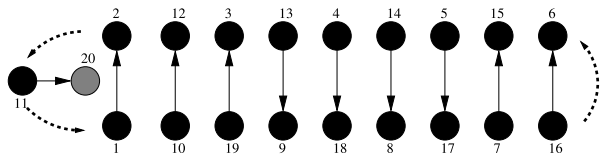
**Table 1** Whole schedule for the first half of the season according to the example with  $n = 20$ ,  $k = 4$  and  $l = 2$

Teams/Rounds	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	-19	20	2	3	4	-5	-6	-7	-8	9	10	-11	-12	13	14	15	16	-17	-18
2	-18	-19	-1	20	3	4	5	-6	-7	-8	-9	10	11	-12	-13	14	15	16	17
3	17	18	-19	-1	-2	-20	4	5	6	-7	-8	-9	-10	11	12	-13	-14	15	16
4	16	17	18	19	1	2	3	-20	5	6	7	-8	-9	-10	-11	12	13	-14	-15
5	-15	-16	17	18	19	1	-2	-3	-4	20	6	7	8	-9	-10	-11	-12	13	14
6	14	15	-16	-17	18	19	1	2	-3	-4	-5	20	7	8	9	-10	-11	-12	-13
7	-13	-14	15	16	-17	-18	19	1	2	3	-4	-5	-6	-20	8	9	10	-11	-12
8	-12	-13	-14	-15	16	17	-18	-19	1	2	3	4	-5	-6	-7	-20	9	10	11
9	11	12	-13	-14	-15	-16	17	18	-19	-1	2	3	4	5	-6	-7	-8	20	10
10	20	11	12	13	-14	-15	-16	-17	18	19	-1	-2	3	4	5	6	-7	-8	-9
11	-9	-10	20	12	13	14	-15	-16	-17	-18	19	1	-2	-3	4	5	6	7	-8
12	8	-9	-10	-11	-20	13	14	15	-16	-17	-18	-19	1	2	-3	-4	5	6	7
13	7	8	9	10	11	12	-20	14	15	16	-17	-18	-19	1	2	-3	-4	5	6
14	-6	7	8	9	10	-11	-12	-13	20	15	16	-17	-18	-19	-1	2	3	-4	-5
15	5	-6	-7	8	9	10	11	-12	-13	-14	20	16	17	18	-19	-1	-2	3	4
16	-4	5	6	-7	-8	9	10	11	12	-13	-14	-15	-20	17	18	19	-1	-2	-3
17	-3	-4	-5	6	7	-8	-9	10	11	12	13	-14	-15	-16	-20	18	19	1	-2
18	2	-3	-4	-5	-6	7	8	-9	-10	11	12	13	14	-15	-16	-17	20	19	1
19	1	2	3	-4	-5	-6	-7	8	9	-10	-11	12	13	14	15	-16	-17	-18	20
20	-10	-1	-11	-2	12	3	13	4	-14	-5	-15	-6	16	7	17	8	-18	-9	-19

**Fig. 3** Example for slot  $n - 1$  with  $n = 20, k = 4$  and  $l = 2$



**Fig. 4** Example for slot  $n$  with  $n = 20, k = 4$  and  $l = 2$



By looking at the figures presented above, one can see that every home stand or road trip is defined by a set of consecutive arcs pointing in the same direction. We call such a set of arcs a *block*. Furthermore, any orientation of the arcs defining the schedule gives rise to a feasible schedule, as long as the blocks do not contain more than  $k$  arcs. The leftmost block is not even allowed to contain more than  $k - 1$  arcs because of the games team  $n$  is involved in. As long as we obey these rules for the maximum sizes of blocks stated above, we will always obtain a feasible plan for any choice of orientations of the arcs defining the tournament.

In the following, we consider  $k$  different orientations. The main difference between them is the width of the rightmost block. For  $l \in \{1, \dots, k\}$  let  $O_l$  be the orientation in which the rightmost block has width  $l$ , the blocks in the middle all have width  $k$  and the leftmost block contains the rest (see Fig. 5). In case this leads to the leftmost block containing exactly  $k$  arcs, we change the orientation of the edge  $(u_1, v_1)$ , such that the arc incident to team  $n$  cannot prolong the road trips induced by this block to have a length of  $k + 1$  matches. The left- and rightmost arcs in a block always define the first and the last match of a trip.

### 4 Costs of the tournament

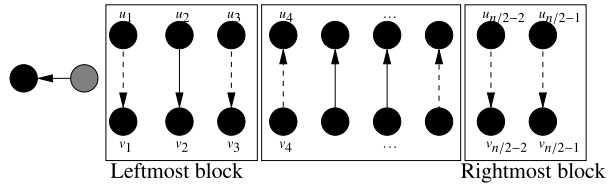
In this section we will prove an upper bound for the total length of the tours defined by the tournament constructed in the previous section.

We assume that every team  $t$  having an away game against team  $n$  will drive home first before driving to team  $n$ 's site and drives home after having played that match. By construction,  $t$  has a home game before or after that game anyway. We just obtain one more visit home this way. By the triangle inequality, the costs incurred this way are only higher than before. Furthermore, we will apply the triangle inequality a second time by assuming that every team drives home after the last game of the first half if it is not already at home. Let the nodes of the underlying graph be denoted as  $u_1, u_2, \dots, u_{n/2-2}$  and  $v_1, v_2, \dots, v_{n/2-2}$  (see Fig. 5).

In the following we will estimate the distances related to the constructed tournament separately:

1.  $C_h$ —the costs related to home-games of team  $n$
2.  $C_a$ —the costs related to away-games of team  $n$
3.  $C_s$ —the costs related to the first rounds of the season-halves and the costs of returning home after the last rounds of the season-halves

**Fig. 5** The blocks defined by the orientation  $O_2$



4.  $C_l$ —the other costs incurred by the edge  $(u_1, v_1)$
5.  $C_r$ —the other costs incurred by the edge  $(u_{n/2-1}, v_{n/2-1})$
6.  $C_o$ —the other costs

$C_h$ —The costs related to home-games of team  $n$ : Every other team plays against team  $n$  once. As we can assume by application of the triangle inequality that all teams come from their home venues to play against team  $n$  and return to their home venues after the game, we know that the cost incurred thereby is at most

$$C_h \leq \sum_{i=1}^{n-1} d(i, n) + d(n, i) = 2 \cdot s(n) \leq 2 \cdot \Delta/n$$

where the last follows from the assumption of  $n$  being the node with the smallest star-weight.

$C_a$ —The costs related to away-games of team  $n$ : Analogously, to the estimation of the home-games of team  $n$ , we can upper bound the costs incurred by the away games by first assuming that team  $n$  always returns home after each away-game. This way, we derive the same upper bound of  $2 \cdot \Delta/n$  for the costs  $C_a$  incurred by the away-games of team  $n$ .

$C_s$ —The costs related to the first rounds of the season-halves and the costs of returning home after the last rounds of the season-halves: In the first round of the season,  $n/2$  teams have to travel to their opponents. We do not consider the game that team  $n$  is involved in, as we have already taken care of these costs above. So, there are  $n/2 - 1$  distances traveled left which correspond directly to the vertical arcs of Fig. 1. After the games of round  $n - 1$  the first half of the season is over, and we assume that all teams drive home. The second half of the season starts with the matches which have already taken place at round  $n - 2$  and it ends with the second leg of the game of round  $n - 3$ . Observe, that the orientation of the arcs does not have an effect on the total distance driven. It only affects the question who is driving which is not of interest here. In the example mentioned above, for team 4 these are the teams 16, 15, 14 and 13. If team 4 did not start the season this way but with a match against team 15, then we would need to consider the distances to the teams 15, 14, 13 and 12. This way we obtain  $n - 1$  different choices for the first and last trips of the two halves of the season. Furthermore, it is easy to see that each edge of  $(\{1, \dots, n - 1\} \times \{1, \dots, n - 1\})$  is part of at most four of these choices. So, summing up the distances of the  $n - 1$  different possible choices for round 1, we obtain a total of at most

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n-1} 4d(i, j) = 2\Delta - 4 \cdot s(n)$$

So, there has to be a choice for which we can estimate

$$C_s \leq 2 \cdot (\Delta - 2 \cdot s(n))/(n - 1)$$



$C_l$ —The costs incurred by the edge  $(u_1, v_1)$ : As we assumed that every team’s trip to team  $n$  starts at the home-site and leads back there after the match, there is always a trip ending or starting with a trip along the edge  $(u_1, v_1)$ . Apparently, these are always trips between teams being neighbors on the heuristically obtained tour  $T_{Heu}$ . As these teams will meet in both halves of the games, the edges have to be counted twice and the cost incurred on that arc can thus be estimated as

$$C_l \leq 2d(n - 1, 1) + 2 \sum_{i=1}^{n-2} d(i, i + 1) \leq 2 \cdot d(T_{Heu})$$

$C_r$ —The costs incurred by the edge  $(u_{n/2-1}, v_{n/2-1})$ : In the first half of the season, the edge  $(u_{n/2-1}, v_{n/2-1})$  always marks the end of a trip, whereas it stands for the beginning of a trip in the second half of the season. The costs incurred in both halves together can be estimated as follows.

$$\begin{aligned} C_r &= 2 \cdot \left( \sum_{i=1}^{n/2} d(i, i + n/2 - 1) + \sum_{i=n/2+1}^n d(i, i - n/2) \right) \\ &= \sum_{i=1}^{n/2} (d(i, i + n/2 - 1) + d(i + n/2 - 1, i)) + \sum_{i=n/2+1}^n (d(i, i - n/2) + d(i - n/2, i)) \\ &\leq \sum_{i=1}^{n/2} \text{opt}_i + \sum_{i=n/2+1}^n \text{opt}_i = \text{opt} \end{aligned} \tag{1}$$

with  $\text{opt}_i$  denoting the length of team  $i$  driven in an optimal solution of total length  $\text{opt}$ . Every possible solution has to contain a trip for any team  $i \in \{1, \dots, n/2\}$  which covers team  $i + n/2 - 1$ . For the length of this trip is not longer than  $d(i, i + n/2 - 1) + d(i + n/2 - 1, i)$  and we can make similar observations for the other teams as well, inequality 1 follows.

$C_o$ —The other costs: As already mentioned earlier in this paper, we do not only consider the orientation of the arcs as displayed in Figs. 1–3. Instead, we will consider  $k$  different orientations. The difference between them is the width of the rightmost block, the block including the arc  $(u_{n/2-1}, v_{n/2-1})$  or resp.  $(v_{n/2-1}, u_{n/2-1})$ . For  $l \in \{1, \dots, k\}$  let  $O_l$  be the orientation in which the rightmost block has width  $l$ , the blocks in the middle have width  $k$  and the leftmost block contains the rest. In case, this leads to the leftmost block containing exactly  $k$  arcs, we change the orientation of the edge  $(u_1, v_1)$ , such that the arc incident to team  $n$  cannot prolong the road trips induced by this block to have a length of  $k + 1$  matches.

In every half, every team  $i$  is associated to one of the nodes  $v_1, v_2, \dots, v_{n/2-1}$  exactly once. When it is associated to node  $v_j$  it plays against the team  $(i + j - 1) \bmod (n - 1) + 1$  which is associated to node  $u_j$  at that time. In case the edge  $(u_j, v_j)$  marks the first or the last game of a road trip in the first or the second half of the tournament, we call this edge a *home-edge* (the dashed arcs in Fig. 5). If the home-edge corresponds to the beginning of a trip in the first half of the season, it marks the end of a tour in the second half of the season. Therefore, the distance associated with this edge is driven exactly twice in the corresponding tournament. Let us have a closer look at the costs which are being incurred by teams traveling along the home-edges. Since every direct travel from or to  $i$ ’s home site can only happen via exactly one home-edge, and as there are at most two orientations in which some edge  $(u_j, v_j)$  is a home-edge, the overall costs incurred by the home edges

is at most  $2\Delta$ . It still remains to estimate the distances traveled which are not from or to the traveling teams' home sites. A trip which visits  $l$  teams consists of two drives along home-edges and  $l - 1$  drives inbetween. By construction, these  $l - 1$  rides are driven along edges which are part of the heuristically obtained tour  $T_{Heu}$ . Let  $u_j$  be a node which does not represent the beginning of a trip. Whenever a team  $i$  is assigned to this node, there is another team  $l$  visiting  $i$  after having played an away match at the team  $i - 1$ , the predecessor of  $i$  in  $T_{Heu}$ . Thus, for any node  $u_j$  or  $v_j$  which does not represent the beginning of a trip, we can estimate the sum of the distances driven to get to the teams assigned to this node as no more than  $d(T_{Heu})$ . Since there are no more than  $n/2 - 2$  such nodes, the distances driven here are not more than  $(n - 2)d(T_{Heu})$ .

For there are  $k$  different orientations, there has to be one with total distance incurred by the home-edges not more than

$$C_o \leq \frac{2\Delta + (n - 2)d(T_{Heu})}{k}$$

### 5 The approximation ratio

If we choose the parameters in the above mentioned ways, we obtain an approximation ratio of

$$\begin{aligned} & \frac{C_h + c_a + C_s + C_l + C_r + C_o}{\text{opt}} \\ & \leq \frac{2\Delta/n + 2\Delta/n + 2 \cdot (\Delta - 2s(n))/(n - 1) + 2 \cdot d(T_{heu}) + \text{opt} + \frac{2\Delta + (n-2)d(T_{Heu})}{k}}{\text{opt}} \\ & = \frac{2\Delta/n + 2\Delta/n + 2 \cdot (\Delta - 2s(n))/(n - 1)}{2/k \cdot \Delta} + \frac{2 \cdot d(T_{heu})}{n \cdot d(T_{opt})} + 1 \\ & \quad + \frac{2/k \cdot \Delta}{2/k \cdot \Delta} + \frac{(n - 2)/k \cdot d(T_{Heu})}{n \cdot d(T_{opt})} \\ & \leq \frac{4\Delta/n + 2 \cdot \Delta/(n - 1)}{2/k \cdot \Delta} + \frac{3}{n} + 1 + 1 + \frac{(n - 2)/k \cdot 3/2 \cdot d(T_{opt})}{n \cdot d(T_{opt})} \\ & \leq \frac{2/n + 1/(n - 1)}{1/k} + \frac{3}{n} + 2 + 3/(2 \cdot k) \\ & = 2k/n + k/(n - 1) + \frac{3}{n} + 2 + 3/(2 \cdot k) \end{aligned}$$

As  $k \leq n - 1$ , this bound cannot be larger than  $5 + \frac{3}{n} + 3/(2 \cdot k)$  which is not more than 5.875 for  $k \geq 4$  and  $n \geq 6$ .

### 6 Computational results

In order to demonstrate this algorithm's applicability to real-world instances, we applied it to the well-known benchmark-instances provided on Trick's site (Trick 2010). At the core of every timetable generated this way lies the solution of a Traveling Salesman Problem. For that task, we applied the software LKH in the version 2.0.3, a very effective implementation of the Lin-Kernighan heuristic developed by Helsgaun (2012).

After the generation of the timetables, we finally tried to improve the obtained solutions by changing the venues of single matches. More precisely, we check for every pair of teams whether it would be less expensive and feasible if they swapped their home/away roles for the matches between them. Whenever this is the case, we apply this change and look for more moves of that kind until no further improvement is possible this way. This neighborhood has been applied very successfully in several other approaches (Anagnostopoulos et al. 2003; Ribeiro and Urrutia 2007; Di Gaspero and Schaerf 2007) where it is called *SwapHomes*. We found it could help to improve our schedules, too. *ling Salesman Problem*. For that task, we applied the software LKH in the version 2.0.3, a very effective implementation of the Lin-Kernighan heuristic developed by Helsgaun (2012). After the generation of the timetables, we finally tried to improve the obtained solutions by changing the venues of single matches. More precisely, we check for every pair of teams whether it would be less expensive and feasible if they swapped their home/away roles for the matches between them. Whenever this is the case, we apply this change and look for more moves of that kind until no further improvement is possible this way. This neighborhood has been applied very successfully in several other approaches (Anagnostopoulos et al. 2003; Ribeiro and Urrutia 2007; Di Gaspero and Schaerf 2007) where it is called *SwapHomes*. It could help to improve our schedules, too.

The results of our computational experiments are displayed in Table 2. The first three columns contain the instances' names, the value of their best known feasible solutions (UB) and the corresponding lower bounds (LB) as they could be found at Trick (2010) on October 27, 2010. Next, there are two columns which show the total distance driven according to the approximation and the relative gap compared to the best solutions known so far. Finally, the last two columns show the objective function after the execution of the simple greedy 2-Opt-heuristic mentioned above.

The running time is very reasonable. Only Galaxy40, Galaxy38, Galaxy36, and Brazil24 took between one and two seconds. The others were solved in shorter time. The computation of all the schedules together took only 13.61 seconds on a regular desktop computer with 2.4 GHz and 4 GB RAM using only a single thread. The running time of the Lin-Kernighan heuristic is already included in these figures.

Comparing the values of our solution with the known lower bounds, we observe that the optimality gap we experience here seems to be somewhere between 15% and 20% and that this gap does not seem to increase with the number of teams increasing. (Actually, in case of the NFL-instances, it does decrease from 21% for NFL16 to 15% for NFL32.)

This behavior is very different to that of the approaches considered so far. The instances with only a few number of teams are by now all solved optimally, but with the number of teams increasing, the gaps between the known upper and lower bounds grow as well. For this reason, our algorithm does not provide better solutions than the ones already known for leagues with only a few teams. But with the number of teams increasing this situation changes slightly. For the well studied problem NFL32 our generated solution is only 5% more expensive than the best solution known so far. We could even find better solutions for the instances Galaxy22, Galaxy24, Galaxy26, Galaxy28, Galaxy30, Galaxy32, Galaxy34, Galaxy36, Galaxy38, and Galaxy40.

We also generated schedules for  $TTP(k)$  with  $k = 4, 5, 6$  (see Table 3). In these cases, we could lower the total distances driven on average by 13% in the case of  $k = 4$ , and by 18% and 23% for  $k = 5$  and  $k = 6$ , respectively.

**Table 2** Computational results for TTP(3)

Instance	LB	UB	Gap	Alg.	Gap(UB)	Alg + 2-Opt	Gap(UB)
Galaxy40		258899		249400	−4%	249230	−4%
Galaxy38		214546		212431	−1%	210787	−2%
Galaxy36		177090		174229	−2%	173827	−2%
Galaxy34		153530		147817	−4%	147817	−4%
Galaxy32		123510		122202	−1%	120683	−2%
Galaxy30		101774		97039	−5%	96979	−5%
Galaxy28		81911		77611	−5%	77577	−5%
Galaxy26		62613		61443	−2%	60962	−3%
Galaxy24		48434		45996	−5%	45910	−5%
Galaxy22		36152		35870	−1%	35767	−1%
Galaxy20		26289		28322	8%	27780	6%
Galaxy18		20907		21613	3%	21528	3%
Galaxy16		14900		16244	9%	16175	9%
Galaxy14	10255	10918	6%	13000	19%	12613	16%
Galaxy12	7034	7197	2%	8227	14%	8131	13%
Galaxy10	4535	4535	0%	5270	16%	5270	16%
NFL32	836031	914620	9%	970427	6%	962631	5%
NFL30	688875	739697	7%	788641	7%	786965	6%
NFL28	560697	609788	9%	653523	7%	652106	7%
NFL26	495982	536792	8%	595778	11%	588890	10%
NFL24	431226	463657	8%	501308	8%	498651	8%
NFL22	378813	402534	6%	446977	11%	445282	11%
NFL20	316721	332041	5%	393500	19%	385565	16%
NFL18	272834	282258	3%	327990	16%	325215	15%
NFL16	223800	231483	3%	272457	18%	270253	17%
NL16	249477	261687	5%	301728	15%	300744	15%
NL14	183354	188728	3%	236692	25%	230874	22%
NL12	108629	110729	2%	125887	14%	125086	13%
NL10	59436	59436	0%	70149	18%	69958	18%
NL8	39721	39721	0%	50521	27%	47128	19%
Super 14	557354	571632	3%	776848	36%	758382	33%
Super 12	453860	463876	2%	557219	20%	547023	18%
Super 10	316329	316329	0%	408950	29%	408950	29%
Brazil24		500756		540127	8%	536218	7%

## 7 Conclusions and outlook

In this paper, we present the first constant factor approximation for  $TTP(k)$  with  $k > 3$ . Furthermore, we show that this algorithm is also applicable to real-world applications as it produces solutions of high quality in a very short amount of time. It was able to find solutions for the benchmark instances Galaxy22, Galaxy24, Galaxy26, Galaxy28, Galaxy30, Galaxy32, Galaxy34, Galaxy36, Galaxy38, and Galaxy40 which are better than the previously known ones.

**Table 3** Comparison of computational results for TTP( $k$ ) with  $k \in \{3, 4, 5, 6\}$ 

Instance	$k = 3$	$k = 4$	$k = 5$	$k = 6$
Galaxy40	249230	216863	197972	187762
Galaxy38	210787	183485	167055	157613
Galaxy36	173827	154286	141957	130661
Galaxy34	147817	132513	121076	113055
Galaxy32	120683	104433	98196	92327
Galaxy30	96979	86506	79198	76432
Galaxy28	77577	69249	63956	60569
Galaxy26	60962	54367	49798	47519
Galaxy24	45910	40832	38252	35341
Galaxy22	35767	31989	30084	27844
Galaxy20	27780	24350	22060	21436
Galaxy18	21528	19890	18294	17485
Galaxy16	16175	14365	13816	12863
Galaxy14	12613	11015	10638	10074
Galaxy12	8131	7722	7480	7371
NFL32	962631	809725	737306	674746
NFL30	786965	685848	612329	579228
NFL28	652106	565601	513224	479321
NFL26	588890	507154	461236	429653
NFL24	498651	436531	401375	365126
NFL22	445282	390626	360905	332744
NFL20	385565	332649	298316	290757
NFL18	325215	292881	264251	253576
NFL16	270253	235936	224367	207604
NL16	300744	263745	247209	228885
NL14	230874	203782	187499	179239
NL12	125086	116634	112713	109144
Super 14	758382	648891	587686	533056
Super 12	547023	486546	440843	423627
Brazil24	536218	464921	421079	387046

The reason for this behavior might be that for leagues with only a small number of teams, it is possible to scan a big portion of the search space in a sophisticated way and this is basically how many algorithmic approaches for the TTP work. As our algorithm only does one smart guess, we cannot hope to find solutions as good as those which can be found by enumerative approaches for a small number of teams, but we can expect to find reasonable results even for big leagues in a very short amount of time.

For the future, it would be promising to implement some more local-search techniques. We have only implemented a greedy heuristic making use of *SwapHomes* which is just one out of several neighborhoods.

Furthermore, one could think about how to apply these concepts to cases in which other constraints are placed on the schedule. For example, how does the approximation ratio change, if certain matches have to be put in certain rounds or some teams cannot play at home at the same time?

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Anagnostopoulos, A., Michel, L., Van Hentenryck, P., & Vergados, Y. (2003). A simulated annealing approach to the traveling tournament problem. In *Proceedings CPAIOR'03*, Montreal.
- Ball, B. C., & Webster, D. B. (1977). Optimal scheduling for even-numbered team athletic conferences. *AIIE Transactions*, 9, 161–169.
- Benoist, T., Laburthe, F., & Rottembourg, B. (2001). Lagrange relaxation and constraint programming collaborative schemes for traveling tournament problems. In *Proceedings CPAIOR'01*, Wye College (Imperial College), Ashford, Kent, UK.
- Christofides, N. (1976). *Worst-case analysis of a new heuristic for the traveling salesman problem*. Report 388, Graduate School of Industrial Administration, CMU.
- de Werra, D. (1981). Scheduling in sports. In P. Hansen (Ed.), *Studies on graphs and discrete programming* (pp. 381–395). Amsterdam: North-Holland.
- Di Gaspero, L., & Schaerf, A. (2007). A composite-neighborhood tabu search approach to the traveling tournament problem. *Journal of Heuristics*, 13, 189–207.
- Easton, K., Nemhauser, G., & Trick, M. (2001). The traveling tournament problem: description and benchmarks. *Lecture Notes in Computer Science*, 2239, 580–585.
- Easton, K., Nemhauser, G., & Trick, M. (2003). Solving the traveling tournament problem: a combined integer programming and constraint programming approach. In E. Burke, P. De Causmaecker (Eds.), *Lecture notes in computer science: Vol. 2740. Practice and theory of automated timetabling IV* (pp. 100–109). Berlin: Springer.
- Helsgaun, K. LKH. <http://www.akira.ruc.dk/~keld/research/LKH/>.
- Henz, M. (2004). Playing with constraint programming and large neighborhood search for traveling tournaments. In E. Burke, M. Trick (Eds.), *Proceedings PATAT 2004* (pp. 23–32).
- Kendall, G., Knust, S., Ribeiro, C. C., & Urrutia, S. (2010). Scheduling in sports: an annotated bibliography. *Computers & Operations Research*, 37, 1–19.
- Kirkman, T. P. (1847). On a problem in combinations. *The Cambridge and Dublin Mathematical Journal*, 2, 191–204.
- Lim, A., Rodrigues, B., & Zhang, X. (2006). A simulated annealing and hill-climbing algorithm for the traveling tournament problem. *European Journal of Operational Research*, 174, 1459–1478.
- Miyashiro, R., Matsui, T., & Imahori, S. (2008). An approximation algorithm for the traveling tournament problem. In *The 7th international conference on the practice and theory of automated timetabling (PATAT 2008)*.
- Rasmussen, R. V., & Trick, M. A. (2008). Round robin scheduling—a survey. *European Journal of Operational Research*, 188, 617–636.
- Ribeiro, C. C., & Urrutia, S. (2007). Heuristics for the mirrored traveling tournament problem. *European Journal of Operational Research*, 179(3), 775–787.
- Thielen, C., & Westphal, S. (2011). Complexity of the traveling tournament problem. *Theoretical Computer Science*, 412, 345–351.
- Trick, M. (October, 2010). Challenge traveling tournament instances. <http://mat.gsia.cmu.edu/TOURN/>.
- van Hentenryck, P., & Vergados, Y. (2006). Traveling tournament scheduling: a systematic evaluation of simulated annealing. In J. C. Beck, B. M. Smith (Eds.), *Lecture notes in computer science: Vol. 3990. Integration of AI and OR techniques in constraint programming for combinatorial optimization problems* (pp. 228–243). Berlin: Springer.
- Yamaguchi, D., Imahori, S., Miyashiro, R., & Matsui, T. (2009). *An improved approximation algorithm for the traveling tournament problem*. Mathematical engineering technical report, METR09-42, Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo.