Solving the Social Golfer Problem with a GRASP

Markus Triska · Nysret Musliu

Received: date / Accepted: date

Abstract The Social Golfer Problem (SGP) is a combinatorial optimization problem that exhibits a lot of symmetry and has recently attracted significant attention. This paper presents a greedy randomized adaptive search procedure (GRASP) for the SGP. We introduce a novel greedy heuristic based on the intuitive concept of freedom among players, which significantly improves results obtained by local search alone. In particular, our method is the first metaheuristic technique that can solve the original problem instance optimally. We show that our approach is also highly competitive with other metaheuristic and constraint-based methods on many other benchmark instances from the literature.

Keywords sports scheduling · combinatorial design · finite geometry

1 Introduction

The Social Golfer Problem (SGP) is a combinatorial optimization problem derived from a question that was posted to sci.op-research in May 1998:

32 golfers play golf once a week, and always in groups of 4. For how many weeks can they play such that no two players play together more than once in the same group?

The research herein is partially conducted within the competence network Softnet Austria (www.soft-net.at) and funded by the Austrian Federal Ministry of Economics (bm:wa), the province of Styria, the Steirische Wirtschaftsförderungsgesellschaft mbH (SFG), and the city of Vienna in terms of the center for innovation and technology (ZIT).

Markus Triska (es)
Database and Artificial Intelligence Group
Vienna University of Technology
triska@dbai.tuwien.ac.at

Nysret Musliu
Database and Artificial Intelligence Group
Vienna University of Technology
musliu@dbai.tuwien.ac.at
The problem is readily generalized to the following decision problem: Is it possible to schedule \( n = g \times p \) golfers in \( g \) groups of \( p \) players for \( w \) weeks such that no two golfers play in the same group more than once? An instance of the SGP is then denoted by a triple \( g-p-w \) of natural numbers. In practice, one is interested in the corresponding optimization problem that asks for the maximum number \( w^* \) of weeks that admits a solution for given \( g \) and \( p \), since such a solution also implies solutions for all instances \( g-p-w \) with \( w \leq w^* \).

Clearly, not all combinations of parameters admit a solution, and upper bounds are easy to determine. For example, in the original problem, \( 8-4-w \), \( w \) can be no more than 10: Suppose \( w \geq 11 \), and observe the schedule of an arbitrary but fixed player \( \alpha \): Each week, \( \alpha \) plays in a group with 3 (distinct) other players. To play 11 weeks, \( \alpha \) would have to partner \( 3 \cdot 11 > 32 \) players.

The SGP exhibits many symmetries: Weeks, groups within weeks, and players within groups, can all be ordered arbitrarily. In addition, players can be assigned arbitrary names. Due to its highly constrained and symmetric nature, the SGP has attracted much attention from the constraint programming community and has lead to the development of powerful but complex dynamic symmetry breaking schemes (Barnier and Brisset 2005; Petrie and Smith 2004). The SGP is included as problem number 10 in CSPLib, a benchmark library for constraints (Gent and Walsh 1999). No constraint solver was so far able to solve the \( 8-4-10 \) instance, although a solution is known to exist. In addition to being a hard and interesting benchmark for constraint solvers, the SGP and closely related problems arise in many practical applications such as encoding, encryption and covering problems (Douglas R. Stinson 1994; Gordon et al. 1995; Hsiao et al. 1970).

2 Related Work

Research on a problem that is very closely related to the SGP actually dates back to Euler, who considered an instance of the SGP in a different context: Euler asked whether two orthogonal Latin squares of order 6 exist, which has become known as “Euler’s Officer Problem” (Colbourn and Dinitz 1996). In terms of the SGP, this corresponds to solving the \( 6-6-4 \) instance, which is now known to be impossible. As another special case of the SGP, the \( 5-3-7 \) instance also has a long history and is known as Kirkman’s schoolgirl problem (Barnier and Brisset 2005).

In general, the task of finding \( w \) mutually orthogonal Latin squares (MOLS) of order \( q \) is equivalent to solving the SGP instance \( q-q-(w+2) \). MOLS play an important role in the design of statistical experiments, coding theory and cryptography, and several (in)existence results and construction methods for specific instances are known from design theory, a branch of discrete mathematics. Harvey and Winterer have compared and successfully applied several of these interesting techniques to the SGP (Harvey and Winterer 2005).

The computational complexity of the SGP is currently unknown. Some instances are easily solved using construction methods from design theory, but such methods are typically restricted to certain families of instances. However, from a result derived by Colbourn in (Colbourn 1984), one can show that the completion problem of the SGP, i.e., deciding whether a partially filled schedule can be completed to conflict-free one, is NP-complete.
Existing metaheuristic approaches towards the SGP include local search with tabu-lists (Dotú and Hentenryck 2005) and an evolutionary approach (Cotta et al. 2006). Dotú and Hentenryck also use a constructive heuristic, but instead of a greedy heuristic, they use a construction method from design theory that is known to trivially solve instances of the form $p-p-(p+1)$ when $p$ is prime. However, the heuristic is not easy to generalize or randomize meaningfully.

Other approaches include a SAT encoding (Gent and Lynce 2005), which so far is not competitive with other methods, and various constraint-based formulations including sophisticated dynamic symmetry breaking techniques (Barnier and Brisset 2005; Fahle et al. 2001).

For the original problem, state-of-the-art constraint solvers can currently generate solutions for no more than 9 weeks, using either set-based formulations or by posting more constraints than strictly necessary to break symmetries. Whether there is a solution for 10 weeks was an open question that was answered – six years after the problem was posed – in the affirmative by Aguado, who used a combination of design-theoretic techniques, backtracking search and instance-specific considerations to construct an explicit solution (Aguado 2004). Unfortunately, his approach does not generalize to other instances.

3 Modeling the SGP

Assuming the correct group sizes, there are essentially only two constraints in the SGP:

1. Each player plays exactly once each week.
2. Each pair of golfers can play in the same group at most once.

These constraints can be enforced using many different formulations of the problem, which is one of the reasons it is so interesting. On the most abstract level, all constraints can be expressed using a set-based formulation, in which the sets $G_{ij}$ denote groups of golfers that play in week $i$. The constraints are then as follows, written as first-order formulae in the language of set theory, with the players being represented by the numbers $1, \ldots, n$:

\[ G_{ij} \subseteq \{1, \ldots, n\} \land |G_{ij}| = p \quad \text{for all } 1 \leq i \leq w, 1 \leq j \leq g \quad (1) \]

\[ \bigcup_{1 \leq i \leq g} G_{ij} = \{1, \ldots, n\} \quad \text{for all } 1 \leq i \leq w \quad (2) \]

\[ |G_{ij} \cap G_{i'j'}| \leq 1 \quad \text{for all } 1 \leq i < i' \leq w, 1 \leq j, j' \leq g \quad (3) \]

In most actual implementations, the abstract concept of sets is simulated by the use of variables $G_{ijk}, 1 \leq G_{ijk} \leq n$, that denote which player plays in week $i$, group $j$ and position $k$. The constraints are then either expressed using built-in primitives provided by solvers or encoded manually. This is also the model we chose. We call the set of variables $G_{ijk}$ and their respective values $v(G_{ijk})$ a configuration. There are several interesting observations to be made about this model:

- The model admits many identical solutions due to its inherent symmetries. For example, the order of groups or of players within groups does not distinguish actual solutions, but leads to distinct solutions in this model.
Most of the symmetries can in fact be removed by imposing additional requirements on solutions. For example, players within groups could be required to occur in ascending order (Barnier and Brisset 2005).

There is one kind of symmetry that cannot be removed statically in this formulation: Players can be renumbered arbitrarily.

While removing symmetries can dramatically reduce computation time in constraint-based approaches, it can make metaheuristic approaches significantly slower, since it actually removes solutions. Therefore, we do not break any symmetries in our approach, which also has a local search component.

We remark that this is not the only sensible way to encode the problem, and in fact there is a model that allows to statically break the symmetry arising from player permutations: Let \( M_{(w \cdot g) \times n} \) denote a Boolean \((w \cdot g) \times n\) matrix. Each column corresponds to a player, and each row corresponds to one group. The Boolean value at position \((i, j)\) denotes whether player \(j\) plays in group \(i\). The necessary constraints on \(M\) are quite obvious. The symmetry arising from player permutations can now be broken by imposing a lexicographic “less than” constraint on columns. Similarly, the symmetry among groups can be broken by imposing a lexicographic “less than” constraint on the rows corresponding to the groups of each week. However, a drawback of this model is that it uses \(g\) times more variables than the model we use, which we also consider more natural.

### 4 A GRASP for the SGP

The basic structure of a greedy randomized adaptive search procedure (GRASP) for a minimization problem is as follows (Feo and Resende 1995):

1. \( f^* = \infty \)
2. Repeat until stopping criterion:
   (a) Generate a greedy randomized candidate solution \(x\)
   (b) Find local optimum \(x_l\) with local search starting from \(x\)
   (c) If \(f(x_l) < f^*\) then
      (i) \(f^* = f(x_l)\)
      (ii) \(x^* = x_l\)

We adhere to this general scheme as follows:

- our stopping criterion is the discovery of a conflict-free schedule (or a time-out, whichever comes first)
- our randomised greedy heuristic generates initial configurations with maximal “freedom”, which we introduce in the next section
- our local search component tries to find a conflict-free schedule.

We explain the details of our greedy heuristic and local search component in the next sections.

### 5 A Novel Greedy Heuristic for the SGP

We now describe a new greedy heuristic for the SGP, which is based on the concept of freedom of sets of players. Let \(C\) be a partial configuration. For an arbitrary player \(x\),
we denote with \( P_C(x) \) the potential partner-set of \( x \) with respect to \( C \), i.e., the set of players that \( x \) can still partner in any group, assuming \( C \) as given. In other words, \( P_C(x) \) is the set of all players, minus the players that \( x \) has already partnered in any group of \( C \). For any set \( S \) of players, we denote with \( \varphi_C(S) \) the freedom of \( S \) with respect to \( C \), and define it as the cardinality of the intersection of the potential partner-sets of all players in \( S \), i.e.: 

\[
\varphi_C(S) = \left| \bigcap_{x \in S} P_C(x) \right|
\]

(4)

Informally, the freedom of a set of players denotes how many players they can “partner together”.

The greedy heuristic now proceeds as follows: Let us first suppose that \( p \) is even. Then the task of scheduling the players into groups in each week can also be seen as scheduling pairs of players into groups. To produce an initial configuration, the heuristic visits the weeks one after another. A single week is produced as follows: The week’s groups are traversed one after another. For each pair of adjacent positions in a group, the heuristic needs to select a pair of players still remaining to be scheduled in the current week. It selects the pair having maximal freedom with respect to the current partial configuration. In addition, there is a parameter \( \gamma \), with \( 0 \leq \gamma \leq 1 \), that can be used to randomize the heuristic: In the case of ties, a random choice is made among the pairs of players having maximal freedom with probability \( \gamma \). With probability \( 1 - \gamma \), pairs are regarded as ordered, with the numerically smaller player first, and the lexicographically smallest pair is selected. After a pair of players was selected and is placed into a group, a large number is subtracted as a penalty from that pair’s freedom in further weeks, to discourage that pair from being selected again in a different group. Other than that, the heuristic pays no attention to potential conflicts in a group, and never undoes a choice of pairs.

The remaining case is when \( p \) is odd. Here, the heuristic can still work with pairs of players, except for the last player in each group. With probability \( \gamma \), that player is randomly selected from all players that are still remaining to be scheduled in that week. With probability \( 1 - \gamma \), the numerically smallest remaining player is selected.

The heuristic is readily generalized from pairs to larger sets of players, although there is a clear trade-off between maximizing freedom of groups and efficiency. The intuition behind maximizing the freedom among players of a group is to “make room” for good local moves, which we discuss in the next section.

6 The Local Search Component

We aimed to keep the local search component as simple as possible. We chose the approach underlying the memetic algorithm (Cotta et al. 2006) and tabu-search (Dotu and Hentenryck 2005) as our basis, and then simplified it further. In particular, we eliminated the restart component, since the search is now restarted within the more general GRASP scheme. Also, based on experiments with different lengths of tabu lists, we fixed the length of tabu lists to 10 instead of imposing random limits.

Let \( C \) denote a configuration. The triple \((i, j, k)\) is a conflict position iff there is a \( k' \neq k \) such that there is a week \( i' \neq i \) with \( v(G_{ijk}) = v(G_{i'lm}) \) and \( v(G_{ij'k}) = v(G_{i'l'm}) \) for any \( l, m \) and \( n \), i.e., wherever a player is in the same group with another
player more than once. Let \( f(C) \) denote the number of conflict positions in \( C \). A swap affecting position \((i,j,k)\) means the exchange of the values of \( G_{ijk} \) with that of any other variable \( G_{i'j'k'} \), \( j \neq j' \). Note that swaps preserve constraints (1) and (2).

A local search iteration starting from a configuration \( C \) that satisfies constraints (1) and (2) proceeds as follows: First, determine \( f = f(C) \). If \( f = 0 \), we have found a solution and are done. Otherwise, of all swaps affecting conflict positions, make the swap that leads to a configuration \( C' \) with minimal \( f(C') \) among all considered configurations (breaking ties lexicographically).

In addition, we associate a tabu list with each week that stores all pairs of players that were exchanged within the last 10 iterations in that week. Swaps that affect a pair of players in the corresponding week’s tabu list are only considered if they result in an improvement over the best solution found so far. Also, if there was no improvement for 4 iterations, two random swaps are made.

7 Experimental Results

Our implementation consists of two programs: The first one, written in Prolog, generates initial configurations for given \( g, p, w \) and \( \gamma \) according to our greedy heuristic. We benefit from arbitrary precision arithmetic to represent sets of players as bit vectors. This lets us efficiently intersect sets by using fast bitwise operations. Determining a set’s cardinality is thus also very efficient.

The second program, written in C++, is the local search component. It is started with parameters \( g, p, w \), and can also read an initial configuration.

We executed 10 runs for each instance \( g-p-w \) presented here, using an Apple MacBook with a 2.16 GHz Intel Core 2 Duo CPU and 1GB RAM. Conceptually, a run of a single instance proceeds as follows: First, the greedy heuristic is used to generate five initial configurations with varying \( \gamma \), including 0, 0.1, 0.2, and two values drawn at random between 0.3 and 1. The time it takes to generate an initial configuration for the benchmark instances is negligible, i.e., at most half a second. Then the local search component is run with each of those starting configurations for at most one minute. If no solution is found in these tries, the search is restarted with that initial configuration that yielded the minimum number of conflicts (and smallest \( \gamma \), in the case of ties) while it was run, and it is then run to completion.

We chose 20 minutes as the maximum running time of the algorithm, since this is also the limit used in the benchmarks of the memetic algorithm (Cotta et al. 2006) on a similar machine configuration, with which we wanted to allow a fair comparison. We considered an instance solved if a solution was found in at least one of the 10 runs we performed. With a single exception, all tested instances were consistently solved well within the 20 minutes time limit. The only exception was the instance 10–6–7, which was solved only in two runs out of ten within the time limit.

Our results are shown in Fig. 1. For various values of \( g \) and \( p \), we show groups of three bars, which denote the maximum \( w \) such that the instance could be solved with (from left to right): Our GRASP scheme, the best memetic algorithm (Cotta et al. 2006), and local search alone (Dotú and Hentenryck 2005). The latter values are similar to those we obtain if we run the local search component in isolation. In particular, we cannot even solve the 8–4–9 instance, let alone 8–4–10, without our greedy heuristic.

For each instance, the thin horizontal lines show the (optimistic) upper bound and the
Fig. 1 Solved number of weeks for $g$ equal to (a) 6, (b) 7, (c) 8, (d) 9 and (e) 10, with various values of $p$. Each group of three bars represents the maximum $w$ obtained by (from left to right): Our GRASP, the best memetic approach proposed in (Cotta et al. 2006), and local search alone (Dottu and Hentenryck 2005). The thin horizontal lines show the best $w$ found with constraint solvers and design theory (Harvey 2002) and optimistic upper bounds, respectively.
best solution obtained with a mix of constraint-based formulations and basic design-theoretic techniques (Harvey 2002), respectively. Table 1 shows average running times for selected instances with $\gamma = 0$.

It is clear from these figures that our approach is highly competitive on other instances besides the original problem as well: On all tested instances, it finds solutions for as many weeks as the best variant of the memetic algorithm (surpassing it on $8-4-10$ and $8-8-5$), and surpasses plain local search and constraint-based solutions in many cases.

### Table 1 Running times for selected instances ($\gamma = 0$)

<table>
<thead>
<tr>
<th>instance</th>
<th>time</th>
<th>instance</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>5−3−7</td>
<td>0.2s</td>
<td>8−8−5</td>
<td>37s</td>
</tr>
<tr>
<td>8−3−10</td>
<td>0.4s</td>
<td>9−3−11</td>
<td>0.2s</td>
</tr>
<tr>
<td>8−4−9</td>
<td>1.2s</td>
<td>9−4−9</td>
<td>5.6s</td>
</tr>
</tbody>
</table>

#### 8 New Solutions for the 8−4−10 Instance

The 8−4−10 instance of the SGP is of particular interest due to two reasons: First, it is the optimal solution for “the” Social Golfer Problem in the original sense, which is problem number 10 in CSPLib, a benchmark library for constraints (Gent and Walsh 1999). Second, being on the verge of solvability, the instance was previously thought to be amenable only to constraint solvers due to its highly constrained nature (previous metaheuristics could only solve for eight weeks). However, even the most sophisticated constraint solvers are currently unable to solve the instance. In contrast, by using the GRASP scheme outlined in this paper, solutions for this instance are readily generated. Two such solutions are depicted in Fig. 2. They were found by varying the randomization factor of the greedy heuristic. Computation time was 11 and 4 minutes, respectively. We used McKay’s *dreadnaut* program (McKay 1990) on the Levi graphs (Colbourn and Dinitz 1996) of the two solutions to verify that they are not isomorphic. The fact that we could obtain structurally different solutions is an indication that the greedy heuristic is meaningful and does not work only by accident.

It is left to explain why the greedy heuristic works so well on the 8−4−10 instance. While we are currently unable to give an analytical justification, we believe that observing the solution process can give an important indication.

To see an effect of the greedy heuristic, consider first Fig. 3, which shows what happens when the search is started from the trivial initial configuration of simply lining up the players in order for each week. For comparison, Fig. 4 shows different states of the local search component starting from a greedy initial configuration with $\gamma = 0$, with conflict positions highlighted. Initially, every position is a conflict position in both cases.

When contrasting the distribution of conflicts in the two figures, one effect of the greedy initial heuristic is apparent: Conflicts become more concentrated, and some weeks become conflict-free very early. In contrast, with a bad initial heuristic (Fig. 3), remaining conflicts are dispersed throughout all weeks. We believe that Fig. 4 gives a valuable suggestion on how the SGP could be successfully approached with a completely different local search method, which explicitly encodes a behaviour that is similar to the
Fig. 2 Two new non-isomorphic solutions for the 8–4–10 instance

one found in this case. For example, one could build conflict-free groups incrementally, while exchanging players in existing weeks or dropping already built groups on occasion.
**Fig. 3** Instance 8–4–10, local search using a trivial initial configuration. Conflicts after (a) 0, (b) 10, (c) 100, (d) 500 iterations.

**Fig. 4** Instance 8–4–10, local search with our greedy initial configuration ($\gamma = 0$). Conflicts after (a) 0, (b) 10, (c) 100, (d) 500 iterations.
9 Conclusions and Future Work

We have presented a GRASP scheme for the Social Golfer Problem. We introduced a novel greedy heuristic based on the intuitive concept of freedom among golfers. The heuristic is readily randomized and generalized, and was shown to improve results obtained by local search alone. In particular, we have obtained new solutions for the 8−4−10 instance. This makes our approach the first metaheuristic method that solves the original problem optimally, and also surpasses current constraint solvers on this instance. In addition, our approach is among the simplest and was shown to be highly competitive with other metaheuristic and constraint-based techniques on other instances as well.

Several interesting questions and opportunities for future research arise from our results: First, what other good greedy heuristics are there for the SGP and related problems? Second, it seems interesting to use the schedules produced by our greedy heuristic as initial configurations in combination with a proposed SAT formulation from the literature (Gent and Lynce 2005), and to use existing SAT solvers as the local search component. This could further reduce the effort for solving the problem. Third, we considered only one among several options for neighbourhoods and evaluation functions in the local search component, and it would be interesting to compare this with other variants. Observing the solution process of heuristics that are known to work well can give valuable hints on what else to try.

An interesting property of our greedy heuristic is that its dual can be successfully used in complete backtracking algorithms, or in combination with a constraint solver. To show this, we implemented a complete backtracking search with early pruning that always schedules that pair of players next that has the least freedom. For the 8−4−9 instance, this almost instantly yields a solution (although some backtracking is necessary), without any constraint solving or more complicated techniques at work. Thus, the heuristic solves the problem as well as any constraint solver currently can. Extending the heuristic from pairs to triples yields solutions for Kirkman’s schoolgirl problem (Barnier and Brisset 2005).

In fact, the method of always choosing the variable with smallest degree of freedom next is well-known in the constraint programming community, and it is typically provided by solvers under the name “first-fail”. As we have shown in this paper, metaheuristic approaches can similarly benefit from taking the freedom of sets of variables into account. In the case of the SGP, it sufficed to restrict this reasoning to the initial greedy heuristic to obtain very competitive results. However, it could be very advantageous to include such considerations also when evaluating configurations in the local search component.

10 Acknowledgements

We thank the anonymous reviewers for their helpful comments.

References


