

# Quadratic Assignment Problems in Cost Function Networks

Guidio Sewa<sup>1,2</sup>[0009–0003–7869–0793], Coralie Marchau<sup>4</sup>[0000–0000–0000–0000],  
David Allouche<sup>1</sup>[0000–0003–3549–4592], Simon de Givry<sup>1,2</sup>[0000–0002–2242–0458],  
George Katsirelos<sup>2,3</sup>[0000–0002–3727–6698], and Thomas  
Schiex<sup>1,2</sup>[0000–0001–6049–3415]

<sup>1</sup> Université de Toulouse, MIAT, INRAE, France

<sup>2</sup> Université de Toulouse, ANITI, France

{guidio.sewa,david.allouche,simon.de-givry,thomas.schiex}@inrae.fr

<sup>3</sup> MIA Paris, INRAE, France gkatsi@gmail.com

<sup>4</sup> IMT Atlantique, Nantes, France coralie.marchau@imt-atlantique.fr

**Abstract.** The Quadratic Assignment Problem (QAP) consists of finding a permutation that minimizes a quadratic objective function. Exact methods generally rely on a branch-and-bound procedure, the efficiency of which depends heavily on the quality of its lower bound. In integer linear programming, several bounds have been investigated, exhibiting different trade-offs between speed and quality. The Gilmore-Lawler bound appears to be the most commonly used in practice. It requires solving a linear assignment problem (LAP) for each variable-value pair. We show how to obtain this bound using Singleton Node Consistency (SNC) and LAP. In Cost Function Networks (CFNs), we propose a reformulation that transforms the result of applying LAP to a given variable-value pair into cost functions of arity 1 and 2, which can be added to the original problem. Combined with existing lower bounds for CFNs, including EDAC and a recent CFN propagator for ALLDIFFERENT, this method (SNC-LAP-GLB), used as a preprocessing, significantly increases the initial lower bound and accelerates the search, resulting in competitive results on the QAPLIB benchmark. We then propose an extension of the ALLDIFFERENT propagator for the Global Cardinality Constraint. It allows us to exploit variable symmetries on some challenging QAPLIB instances, thus improving the results. Last, we report some preliminary results obtained by an automatic configuration tool for tuning the solver parameters.

**Keywords:** Linear Assignment Problem · ALLDIFFERENT · Global Cardinality Constraint.

## 1 Introduction

The quadratic assignment problem (QAP) was introduced in [26] as a mathematical model for the location of indivisible economical activities. *“We want to assign  $n$  facilities to  $n$  locations with the cost being proportional to the flow*

between the facilities multiplied by the distances between the locations plus, eventually, costs for placing the facilities at their respective locations. The objective is to allocate each facility to a location such that the total cost is minimized” [8]. It is a very challenging problem and even MIP solvers struggle to scale beyond the smallest instances with less than thirty variables. The problem has many real-world applications, and has attracted the attention of the community [30].

The QAP is naturally modeled as a cost function network (CFN) [10, 11] with a single permutation constraint, and unary and binary cost functions to capture the costs of linear (on facility-location pairs) and quadratic terms (on pairs of facility-location pairs), respectively. Branch-and-bound methods for solving CFNs rely on soft local consistency (SLC) techniques [10]. The idea of SLC is to reformulate a problem into an equivalent one by rewriting the objective function. The rewritten objective function has a constant term which provides an explicit lower bound and may have infinite cost for some values, indicating they are pruned. The reformulation is computed by moving costs from the quadratic/binary level to the linear/unary level, and then to the constant term. This reformulation allows for better communication between the quadratic terms and the permutation (ALLDIFFERENT) constraint, for which a dedicated SLC based on solving the polynomial-time linear assignment problem (LAP) has recently been proposed [40].

This observation was already known in QAP and so-called reduction methods have been developed with exactly the same goal [8]. In particular, the Lagrangian relaxation proposed in [20] can be related to optimal soft arc consistency in CFNs [10]. It dominates the Gilmore-Lawler lower bound (GLB), which is a decomposition approach used in branch-and-bound methods for QAP because of its computation speed.

In crisp constraint networks, several extensions of arc consistency methods have been proposed to get better value pruning [37]. One technique is singleton arc consistency (SAC) [14, 3, 13, 29, 5]. In this paper, we propose an adaptation to CFNs of singleton node consistency (SNC), a weaker form of SAC. We show how it can be used to simulate GLB, by providing valid reformulations and even get improved lower bounds when used in conjunction with existing SLC techniques such as *existential directional arc consistency* (EDAC) [12, 10] during preprocessing.

In the following, we introduce necessary background on QAP and CFNs. Then, we present our singleton node consistency approach with SLC reformulations. Two iterative algorithms are proposed: SNC-LAP-GLB and a variant SNC-LAP-GREEDY. Next, we exploit symmetries present in some specific QAP instances using an inverted formulation of the problem and a global cardinality constraint (GCC). Following our work on ALLDIFFERENT, and the relationship with the semi-assignment problem (SAP), we develop SLC techniques for the case of GCC in closed form. Replacing LAP by SAP in the previous algorithms, we obtain SNC-SAP-GLB and SNC-SAP-GREEDY. Last, we give detailed experimental results for these four algorithms compared to CP and OR approaches on the QAPLIB benchmark and conclude.

The closest related work in CFNs is *virtual singleton arc consistency via super-reparametrizations* [16, 17]. However, a super-reparametrization does not reformulate the problem into an equivalent one. Also, SNC with the greedy strategy applied on binary CFNs (without the ALLDIFFERENT constraint) implies the *existential arc consistent* (EAC) property [12].

## 2 Background

### 2.1 Linear Assignment Problem

The Linear Assignment Problem (LAP) of size  $n$  is defined by

$$l = \min \sum_{i=1}^n \sum_{u=1}^n c_{iu} x_{iu} \quad (1)$$

$$\sum_{u=1}^n x_{iu} = 1 \quad \forall i \in \{1, \dots, n\} \quad (2)$$

$$\sum_{i=1}^n x_{iu} = 1 \quad \forall u \in \{1, \dots, n\} \quad (3)$$

$$x_{iu} \in \{0, 1\} \quad \forall i, u \in \{1, \dots, n\}^2 \quad (4)$$

The constraint matrix Eq.(2)–(3) being totally unimodular, it can be solved by linear programming algorithms in polynomial time, or *e.g.*, using the Hungarian [27] or LAPJV [24] algorithms in  $O(n^3)$  time.

### 2.2 Quadratic Assignment Problem

A Quadratic Assignment Problem (QAP) is defined by a set of  $n$  facilities and a set of  $n$  locations. For each pair  $(i, j)$  of facilities, a flow  $a_{ij}$  is specified, and for each pair  $(u, v)$  of locations, a distance  $b_{uv}$  is specified. Finally, there is a cost  $c_{iu}$  of locating facility  $i$  at location  $u$ . The problem is to assign all facilities to different locations while minimizing the total cost, assuming a unit cost per supply and unit of distance.

Formally, a QAP of size  $n$  is defined by

$$q = \min \sum_{i=1}^n \sum_{j=1}^n \sum_{u=1}^n \sum_{v=1}^n a_{ij} b_{uv} x_{iu} x_{jv} + \sum_{i=1}^n \sum_{u=1}^n c_{iu} x_{iu} \quad (5)$$

$$\sum_{u=1}^n x_{iu} = 1 \quad \forall i \in \{1, \dots, n\} \quad (6)$$

$$\sum_{i=1}^n x_{iu} = 1 \quad \forall u \in \{1, \dots, n\} \quad (7)$$

$$x_{iu} \in \{0, 1\} \quad \forall i, u \in \{1, \dots, n\}^2 \quad (8)$$

where coefficients  $a$ ,  $b$ , and  $c$  are assumed to be non-negative integers. W.l.o.g., we assume  $a_{ii} = b_{uu} = 0$  for all pairs  $(i, u)$ . If not,  $c_{iu}$  can be updated by  $c_{iu} \leftarrow c_{iu} + a_{ii}b_{uu}$ . We have  $x_{iu} = 1$  if facility  $i$  is assigned to location  $u$ , otherwise,  $x_{iu} = 0$ . The problem is NP-hard, as it contains the Traveling Salesman Problem as a special case [8].

If we ignore the quadratic terms in the objective, we get a LAP. Solving this gives a basic lower bound for QAP.

### 2.3 Gilmore-Lawler Bound

The Gilmore-Lawler bound (GLB) [21, 43] is a strengthened variant of the simple LAP bound. It considers solving first the following  $n^2$  specific LAPs  $LAP_{GLB}^{iu}$ , for each pair  $(i, u) \in \{1, \dots, n\}^2$ :<sup>5</sup>

$$LAP_{GLB}^{iu} : l_{iu} = \min \sum_{j=1, j \neq i}^n \sum_{v=1, v \neq u}^n a_{ij}b_{uv}x_{jv} \quad (9)$$

$$\sum_{v=1, v \neq u}^n x_{jv} = 1 \quad \forall j \in \{1, \dots, n\} \setminus \{i\} \quad (10)$$

$$\sum_{j=1, j \neq i}^n x_{jv} = 1 \quad \forall v \in \{1, \dots, n\} \setminus \{u\} \quad (11)$$

$$x_{jv} \in \{0, 1\} \quad \forall j \in \{1, \dots, n\} \setminus \{i\}, v \in \{1, \dots, n\} \setminus \{u\} \quad (12)$$

Observe that each  $LAP_{GLB}^{iu}$  objective function Eq.(9) involves a different part of the  $a$  and  $b$  coefficients, which means that the individual bounds we compute can be added to the coefficient of the corresponding linear term  $x_{iu}$ .

The GLB is obtained by solving this final LAP:

$$glb = \min \sum_{i=1}^n \sum_{u=1}^n (l_{iu} + c_{iu})x_{iu} \quad (13)$$

$$\sum_{u=1}^n x_{iu} = 1 \quad \forall i \in \{1, \dots, n\} \quad (14)$$

$$\sum_{i=1}^n x_{iu} = 1 \quad \forall u \in \{1, \dots, n\} \quad (15)$$

$$x_{iu} \in \{0, 1\} \quad \forall i, u \in \{1, \dots, n\}^2 \quad (16)$$

### 2.4 QAP as a Weighted Constraint Satisfaction Problem

A Cost Function Network is a triplet  $P = (X, D, F)$  where  $X = \{x_1, \dots, x_n\}$  is a set of  $n$  decision variables with finite domain  $D_i$  for variable  $x_i$ . An assignment

<sup>5</sup> These LAPs can all be solved together in  $O(n^2 \log(n))$  time (see *e.g.*, [43, 8]). However, it assumes a decomposition of their objective function as a product of two vectors, which is not preserved by the problem reformulations done in Section 3.

$I$  to a set of variables  $S$  is a mapping from each variable  $x_i \in S$  to a value in  $D_i$ . Let  $\ell(S) = \prod_{i \in S} D_i$  denote the set of all possible assignments for  $S$ . An assignment is complete if  $S = X$ , otherwise it is partial.  $F$  is a set of cost functions, each cost function  $f_S \in F$  applies to a subset  $S$  of  $X$  and returns a non-negative integer cost depending on its variable assignment. Returning an infinite cost means a forbidden assignment. A cost function returning only zero or infinite costs defines a hard constraint.

The binary Weighted Constraint Satisfaction Problem (WCSP) is defined by:<sup>6</sup>

$$\min \sum_{i=1}^n \sum_{j>i}^n f_{ij}(x_i, x_j) + \sum_{i=1}^n f_i(x_i) + f_\emptyset \quad (17)$$

$$x_i \in D_i \quad \forall i \in \{1, \dots, n\} \quad (18)$$

where  $f_\emptyset, f_i, f_{ij}$  are respectively called empty, unary, and binary cost functions.  $f_\emptyset$  is a constant that serves as a trivial problem lower bound. The problem is NP-hard [11]. Complete solving approaches are usually based on branch-and-bound, whose efficiency greatly depends on the quality of its lower bound [10]. In the past, several so-called *Soft Local Consistency* (SLC) methods, including NC, EDAC [12, 10], and *FØIC* [33, 40], have been proposed in order to reformulate  $P$  into an equivalent problem  $P'$  having the same cost for any complete assignment, but a better  $f_\emptyset$ .

The simplest SLC method is Node Consistency (NC): if all domain values of a variable  $x_i$  have a nonzero unary cost, then increase  $f_\emptyset \leftarrow f_\emptyset + \min_{x_i \in D_i} f_i(x_i)$  and decrease the unary cost function by this minimum:  $\forall x_i \in D_i, f_i(x_i) \leftarrow f_i(x_i) - \min_{x_i \in D_i} f_i(x_i)$ . Moreover, if a problem upper bound  $ub$  is available, then NC will also prune all domain values  $x_i \in D_i$  for which it holds that  $f_i(x_i) + f_\emptyset \geq ub$ . EDAC is a stronger SLC method which enforces NC and more, but is limited to binary (and ternary) cost functions. *FØIC* has been introduced for the propagation of global constraints in CFNs (KNAPSACK [33], ALLDIFFERENT [40]). The aim of *FØIC* is to enforce that for each nonunary cost function  $f_S, |S| > 1$ , it admits a partial assignment  $I \in \ell(S)$  such that  $f_S(I) + \sum_{x_i \in S} f_i(I[\{x_i\}]) = 0$ . EDAC is stronger than *FØIC* on binary cost functions.

A QAP of size  $n$  can be formulated as a WCSP  $P = (X, D, F)$  with:

$$D_i = \{1, \dots, n\} \quad \forall i \in \{1, \dots, n\} \quad (19)$$

$$f_{ij}(x_i = u, x_j = v) = a_{ij}b_{uv} + a_{ji}b_{vu} \quad \forall i, j, u, v \in \{1, \dots, n\}^4, i < j, u \neq v \quad (20)$$

$$f_{ij}(x_i = u, x_j = u) = \infty \quad \forall i, j, u \in \{1, \dots, n\}^3 \quad (21)$$

$$f_i(x_i = u) = c_{iu} \quad \forall i, u \in \{1, \dots, n\}^2 \quad (22)$$

$$f_\emptyset = 0 \quad (23)$$

<sup>6</sup> Later, we use the term WCSP indifferently to refer to the CFN or its associated minimization problem.

Here, decision variables are associated to facilities and values to locations. So  $\{x_i = u\}$  in the WCSP represents  $x_{iu} = 1$  in the QAP. We will see in Section 4 that associating variables to locations instead of facilities allows to exploit symmetries in some cases.

Using a ALLDIFFERENT( $X$ ) hard global constraint instead of Eq.(21) gives better propagation and stronger lower bounds as shown in [40]. We call this QAP model with the ALLDIFFERENT,  $P_{\text{alldiff}}$ .<sup>7</sup>

**Reformulation with the ALLDIFFERENT Constraint** The combination of the ALLDIFFERENT constraint and the unary cost functions  $f_i \forall i \in \{1, \dots, n\}$  is exactly the LAP of Eq.(1-4) where  $c_{iu} \equiv f_i(x_i = u)$ . We have shown in [40] a reformulation of  $P_{\text{alldiff}}$  by solving the corresponding LAP, then adding its optimum  $l$  to the problem lower bound,  $f_{\emptyset} \leftarrow f_{\emptyset} + l$ , and the reduced costs  $rc(i, u)$  in the LAP become the new unary costs:  $f_i(x_i = u) \leftarrow rc(i, u)$  for all  $i, u \in \{1, \dots, n\}$ <sup>2</sup>. This is a valid reformulation because the LAP contains only equality constraints [40, Theorem 1] [32]. Applying this reformulation enforces exactly *FØIC* on the ALLDIFFERENT constraint. After the reformulation,  $P_{\text{alldiff}}$  has a better lower bound  $f_{\emptyset}$ , which can be used by the branch-and-bound method to prune the search when  $f_{\emptyset} \geq ub$  or prune domains by NC when  $f_i(x_i) + f_{\emptyset} \geq ub$ . We have shown [40] that these pruning rules are stronger than the ones developed for the Weighted ALLDIFFERENT constraint [39, 9].

### 3 Singleton Node Consistency and Linear Assignment Problem

Singleton Arc Consistency (SAC) has been proposed in the context of Constraint Satisfaction Problems. The main idea is to test each partial assignment  $\{x_i = u\} \forall x_i \in X, u \in D_i$  and propagate using arc consistency. If a domain wipe-out occurs, then  $u$  can be removed from  $D_i$  [14]. Refinements of this base algorithm improve practical or worst case performance [3, 13, 29, 5].

Here, our goal is to increase the unary costs by using search and *FØIC*, a weaker SLC method than soft arc consistency methods such as EDAC.

Based on the previously-defined QAP problem  $P_{\text{alldiff}}$ , we define the following subproblem  $P_{GLB}^{iu}$  where the value  $u$  is assigned to the variable  $x_i$  and only binary

<sup>7</sup> In practice, we add Eq.(21) as a redundant binary constraint to the ALLDIFFERENT constraint only if the function  $f_{ij}(x_i = u, x_j = v)$  returns a nonzero cost for some  $u \neq v$ , otherwise the function  $f_{ij}$  can be discarded which makes the WCSP having less binary cost functions.

cost functions (or half quantities of them) related to  $\{x_i = u\}$  are kept:<sup>8</sup>

$$P_{GLB}^{iu} : l'_{iu} = \min \sum_{j < i} f_{ji}(x_j, x_i = u) + \frac{1}{2} \sum_{j > i} f_{ij}(x_i = u, x_j) \quad (24)$$

$$\text{ALLDIFFERENT}(X \setminus \{x_i\}) \quad (25)$$

$$x_j \in D_j \setminus \{u\} \quad \forall j \in \{1, \dots, n\} \setminus \{i\} \quad (26)$$

Our main observation is that we can reformulate the original problem  $P_{\text{alldiff}}$  using the solution of the modified  $LAP_{GLB}^{iu}$ . In  $P_{\text{alldiff}}$ , the optimum  $l'_{iu}$  is added to the unary cost  $f_i(x_i = u)$  and the reduced costs  $rc(j, v)$  update binary costs:

$$f_i(x_i = u) \leftarrow f_i(x_i = u) + l'_{iu} \quad (27)$$

$$f_{ji}(x_j = v, x_i = u) \leftarrow rc(j, v) \quad \forall j < i, v \in D_j \quad (28)$$

$$f_{ij}(x_i = u, x_j = v) \leftarrow \frac{f_{ij}(x_i = u, x_j = v)}{2} + rc(j, v) \quad \forall j > i, v \in D_j \quad (29)$$

**Theorem 1.** *The updates in Equations (27)–(29) preserve equivalence.*

*Proof (Sketch).* The proof is the same as for the ALLDIFFERENT constraint [40] but now the modified costs are conditioned on the partial assignment  $\{x_i = u\}$ . After conditioning on  $\{x_i = u\}$ , the lower bound, which is a constant term, becomes a linear term and the reduced costs, which are linear, become quadratic terms.  $\square$

Algorithm 1, SNC-LAP-GLB, iteratively reformulates the problem  $P_{\text{alldiff}}$  (lines 7-13) by solving the modified  $LAP_{GLB}^{iu}$  (line 5) for all  $(i, u)$  and applying the updates (27)–(29). After that, it enforces  $F\emptyset IC$  on ALLDIFFERENT (line 16) to get a lower bound, which it improves with EDAC on  $P_{\text{alldiff}}$  (line 17). EDAC moves costs between unary and binary cost functions, which may invalidate previous optimal LAPs, thus it iterates until a stopping condition is met.

We show next that SNC-LAP-GLB is no weaker than GLB.

**Theorem 2.** *For any QAP with symmetric flow and distance matrices, the SNC-LAP-GLB bound is no smaller than the GLB bound.*

*Proof.* We prove first that for all variables  $i$ ,  $f_i(x_i = u) \geq l_{iu}$ , where  $l_{iu}$  is defined in Equation (9). We show this by induction on the variable  $i$ .

For  $i = 1$ , due to our assumption of symmetric flow and distance matrices, enforcing  $F\emptyset IC$  on  $P_{GLB}^{1u}$  is equivalent to solving  $LAP_{GLB}^{1u}$  for all  $u$ . We have  $\frac{f_{1j}(x_1=u, x_j=v)}{2} = \frac{a_{1j}b_{uv} + a_{j1}b_{vu}}{2} = a_{1j}b_{uv}$ , thus  $l'_{1u} = l_{1u}$ .

Assume the inductive hypothesis holds for all  $j \leq i$ . We show that it holds for  $i + 1$ . If the reduced costs of  $LAP_{GLB}^{iu}$  are all null, then in the  $i$ th reformulation of  $P_{\text{alldiff}}$ , we have  $f_{ij}(x_i = u, x_j = v) \leftarrow \frac{f_{ij}(x_i=u, x_j=v)}{2} \forall j > i, v \in D_j$ .

<sup>8</sup> This is why we call our approach singleton node consistency because we do not want interactions with other binary cost functions that would require introducing new higher-arity (ternary) cost functions to ensure a correct reformulation.

Otherwise, binary costs are higher due to positive reduced costs. Thus, the first term  $f_{ji+1}(x_j, x_{i+1} = u) \forall j < i + 1$  in Eq.24 of  $P_{GLB}^{i+1u}$  is greater than or equal to  $a_{ji+1}b_{vu}$  after the previous  $i$  reformulations. The second term  $f_{i+1j}(x_{i+1} = u, x_j) \forall j > i + 1$  has not been modified by the previous reformulations, and it is also greater than or equal to  $a_{i+1j}b_{uv}$ .

Since the final step of SNC-LAP-GLB (line 16) solves a LAP with linear costs at least as great as those used in the final step of GLB, we get that the optimum of the former LAP is at least as great as the latter, as required.  $\square$

---

**Algorithm 1** SNC-LAP-GLB( $P_{\text{alldiff}}$ ): Enforcing singleton node consistency with GLB-like LAP on a quadratic assignment problem  $P_{\text{alldiff}}$ .

---

```

1: repeat
2:   for all  $x_i \in X$  do
3:     for all  $u \in D_i$  do
4:       /* Temporarily assign  $x_i = u$  and enforce  $F\emptyset IC$  on  $P_{GLB}^{iu}$  */
5:        $(l'_{iu}, rc) = \text{SOLVE-LAP}_{GLB}^{iu}(\{f_{ji} \mid j < i\} \cup \{\frac{f_{ij}}{2} \mid j > i\})$ 
6:       /* Reformulate  $P_{\text{alldiff}}$  */
7:       for all  $j < i, v \in D_j$  do
8:          $f_{ji}(x_j = v, x_i = u) \leftarrow rc(j, v)$ 
9:       end for
10:      for all  $j > i, v \in D_j$  do
11:         $f_{ij}(x_i = u, x_j = v) \leftarrow \frac{f_{ij}(x_i = u, x_j = v)}{2} + rc(j, v)$ 
12:      end for
13:       $f_i(x_i = u) \leftarrow f_i(x_i = u) + l'_{iu}$ 
14:    end for
15:  end for
16:  Enforce  $F\emptyset IC$  on ALLDIFFERENT( $X$ )
17:  Enforce EDAC and  $F\emptyset IC$  on  $P_{\text{alldiff}}$ 
18: until stopping condition

```

---

The time complexity of one pass over all the variables at line 2 is in  $O(n^5)$ , assuming  $O(n^3)$  for solving  $LAP_{GLB}^{iu}$ . The space complexity is the same as for representing  $P_{\text{alldiff}}$ , in  $O(n^4)$ , using tables for binary cost functions. In our implementation, the stopping condition is that the lower bound  $f_\emptyset$  has not increased sufficiently after one pass over all the variables, specifically if  $\frac{f_\emptyset - f_\emptyset^{prev}}{f_\emptyset} \leq 1e - 4$ .

### 3.1 Improving Further the Iterated Gilmore-Lawler Bound

The intuitive idea of GLB is to distribute the binary costs evenly across the unary costs of all variables before solving the final LAP. We propose another strategy that will try to move all the unary and binary costs towards a single variable in order to increase its own unary cost function as much as possible. We call this strategy GREEDY. We define the following subproblem  $P_{Greedy}^{iu}$  where we



assign  $x_i = u$  and keep only unary and binary cost functions related to  $\{x_i = u\}$ :

$$P_{Greedy}^{iu} : s_{iu} = \min \sum_{j < i} f_{ji}(x_j, x_i = u) + \sum_{j > i} f_{ij}(x_i = u, x_j) + \sum_{j \neq i} f_j(x_j) \quad (30)$$

$$\text{ALLDIFFERENT}(X \setminus \{x_i\}) \quad (31)$$

$$x_j \in D_j \setminus \{u\} \quad \forall j \in \{1, \dots, n\} \setminus \{i\} \quad (32)$$

Enforcing  $F\emptyset IC$  on this problem is equivalent to solving  $LAP_{Greedy}^{iu}$ , an LAP with a linear objective function given by Eq.30.

---

**Algorithm 2** SNC-LAP-GREEDY( $P_{\text{alldiff}}$ ): Enforcing singleton node consistency with Greedy LAP on a quadratic assignment problem  $P_{\text{alldiff}}$ .

---

```

1: repeat
2:   for all  $x_i \in X$  do
3:     /* Move unary costs into binary cost functions related to  $x_i$  */
4:     for all  $x_j \in X \setminus \{x_i\}, v \in D_j$  do
5:       for all  $u \in D_i$  do
6:          $f_{ij}(x_i = u, x_j = v) \leftarrow f_{ij}(x_i = u, x_j = v) + f_j(x_j = v)$ 
7:       end for
8:        $f_j(x_j = v) \leftarrow 0$ 
9:     end for
10:    for all  $u \in D_i$  do
11:      /* Temporarily assign  $x_i = u$  and enforce  $F\emptyset IC$  on  $P_{Greedy}^{iu}$  */
12:       $(s_{iu}, rc) = \text{SOLVE-LAP}_{Greedy}^{iu}(\{f_{ij} \mid j \neq i\})$ 
13:      /* Reformulate  $P_{\text{alldiff}}$  */
14:      for all  $x_j \in X \setminus \{x_i\}, v \in D_j$  do
15:         $f_{ij}(x_i = u, x_j = v) \leftarrow rc(j, v)$ 
16:      end for
17:       $f_i(x_i = u) \leftarrow f_i(x_i = u) + s_{iu}$ 
18:    end for
19:    Enforce EDAC and  $F\emptyset IC$  on  $P_{\text{alldiff}}$ 
20:  end for
21: until stopping condition

```

---

We define Algorithm 2, SNC-LAP-GREEDY, which reformulates  $P_{\text{alldiff}}$  by repeatedly solving  $LAP_{Greedy}^{iu}$  (line 12) for all  $(i, u)$  until a stopping condition is met (the same as for SNC-LAP-GLB). After singleton tests and reformulations for all domain values of the target variable (for-loop line 10), we enforce EDAC on the binary cost functions of  $P_{\text{alldiff}}$  and  $F\emptyset IC$  on ALLDIFFERENT (line 19).

We give the following theorem without proof, as it is largely the same as the proof of Theorem 1. The main difference is that we cannot directly use the unary costs of  $P_{\text{alldiff}}$  inside  $LAP_{Greedy}^{iu}$ , it has to be done through the binary cost functions related to  $(i, u)$ . Therefore, we move the unary costs to the binary cost

functions related to  $x_i$  before solving  $LAP_{Greedy}^{iu}$  (lines 4-9). Alg. 2 has the same worst-case time and space complexity as Alg. 1. However, the greedy algorithm does not guarantee a better bound than SNC-LAP-GLB or GLB.

**Theorem 3.** SNC-LAP-GREEDY *preserves equivalence.*

Pruned values  $v \in D_j$  found by enforcing  $F\emptyset IC$  on  $P_{GLB}^{iu}$  in Alg. 1 or  $P_{Greedy}^{iu}$  in Alg. 2 become forbidden tuples in  $P_{alldiff}$ . If a value  $v$  is pruned by all singleton tests ( $\forall u \in D_i$ ), it can be removed from  $P_{alldiff}$ , as in SAC and constructive disjunction [42, 6]. Because EDAC is equivalent to classical AC on forbidden tuples [10], Alg. 1 (resp. Alg. 2) prunes such values at line 17 (resp. 19).

*Example 1.* Consider a CFN with 3 variables  $\{x_1, x_2, x_3\}$  with 3 values each, an ALLDIFFERENT constraint over them, and 3 binary cost functions  $f_{ij}(x_i = u, x_j = v) = 2 \forall i, j, u, v \in \{1, 2, 3\}^4, i < j, u \neq v$  (we omit unary and binary tuples that have cost 0). Both SNC-LAP-GLB and SNC-LAP-GREEDY produce an optimal lower bound  $f_\emptyset$  of 6 and all the unary and binary cost functions are set to zero after reformulation.

## 4 Exploiting Symmetries in the Flow Matrix of the QAP

For a subset of the QAPLIB instances, it has been shown [19] that is possible to group facilities into equivalence classes. Two facilities  $i$  and  $j$  are equivalent if the following conditions are met:

$$a_{ij} = a_{ji} \tag{33}$$

$$a_{ik} = a_{jk}, a_{ki} = a_{kj} \quad \forall k \in \{1, \dots, n\} \setminus \{i, j\} \tag{34}$$

$$c_{iu} = c_{ju} \quad \forall u \in \{1, \dots, n\} \tag{35}$$

$$b_{uv} = b_{vu} \quad \forall u, v \in \{1, \dots, n\}^2 \tag{36}$$

In CFNs, such variables are said interchangeable. For any solution of  $P_{alldiff}$ , we can permute the values of  $x_i$  and  $x_j$  to get a new solution with the same cost.

In order to exploit the symmetries, we need to invert the CFN model.<sup>9</sup> In the inverted model, variables are associated to locations and values to facilities. The resulting model is called  $P_{alldiff}^\tau$ .

Let us assume we have  $d$  equivalent classes  $\{E_1, \dots, E_d\}$  of size  $m_k = |E_k|, k \in \{1, \dots, d\}$ . We define  $P_{gcc}^\tau$ , the inverted QAP model of size  $n$ , with reduced domains of size  $d$ , grouping all facilities/values of the same equivalence class  $E_k$  into a single representative value  $\varphi(k)$ , and replacing ALLDIFFERENT by a Global Cardinality Constraint (GCC) [23]. In GCC, we set the upper bound on the

<sup>9</sup> It has been observed [19] that symmetries can only be exploited in the inverted model and not in the original model.

number of occurrences of  $\varphi(k)$  to  $m_k$ . We define  $P_{gcc}^\tau = (X, D, F)$ :

$$D_i = \{1, \dots, d\} \quad \forall i \in \{1, \dots, n\} \quad (37)$$

$$f_{ij}(x_i = k, x_j = k') = \begin{aligned} & a_{\varphi(k)\varphi(k')} b_{ij} + \\ & a_{\varphi(k')\varphi(k)} b_{ji} \end{aligned} \quad \forall i, j \in \{1, \dots, n\}^2, i < j, k, k' \in \{1, \dots, d\}^2 \quad (38)$$

$$f_i(x_i = k) = c_{\varphi(k)i} \quad \forall i \in \{1, \dots, n\}, k \in \{1, \dots, d\} \quad (39)$$

$$f_\emptyset = 0 \quad (40)$$

$$\text{GCC}(X, \{m_1, \dots, m_d\}) \quad (41)$$

By definition,  $d \leq n$  and  $\sum_{k=1}^d m_k = n$ . The GCC is in *closed form*, i.e., all variables must take a value in  $\{1, \dots, d\}$  and each value  $k$  must be used exactly  $m_k$  times.

**Reformulation with the Global Cardinality Constraint** The combination of the GCC constraint in closed form and the unary cost functions  $f_i \forall i \in \{1, \dots, n\}$  corresponds exactly to the Semi-Assignment Problem (SAP) [25, 8]:

$$\text{SAP} : o = \min \sum_{i=1}^n \sum_{k=1}^d c_{\varphi(k)i} x_{ik} \quad (42)$$

$$\sum_{k=1}^d x_{ik} = 1 \quad \forall i \in \{1, \dots, n\} \quad (43)$$

$$\sum_{i=1}^n x_{ik} = m_k \quad \forall k \in \{1, \dots, d\} \quad (44)$$

$$x_{ik} \in \{0, 1\} \quad \forall i \in \{1, \dots, n\}, k \in \{1, \dots, d\} \quad (45)$$

where 0/1 variable  $x_{ik} = 1$  corresponds to  $x_i = k$  in  $P_{gcc}^\tau$ . This problem can be solved efficiently by a modified version of LAPJV algorithm in  $O(dn^2)$  time [25]. As with the ALLDIFFERENT constraint, we can reformulate  $P_{gcc}^\tau$ . The optimum  $o$  of the SAP is added to the problem lower bound  $f_\emptyset \leftarrow f_\emptyset + o$  and the reduced costs  $rc(i, k)$  of each variable  $x_{ik}$  in the SAP become the new unary costs:  $f_i(x_i = k) \leftarrow rc(i, k), i \in \{1, \dots, n\}, k \in \{1, \dots, d\}$ . This reformulation enforces *F0IC* on the GCC. This offers the same advantages compared to GCC with costs [36] as *F0IC* on ALLDIFFERENT does compared to Weighted ALLDIFFERENT.

We can apply SNC on  $P_{gcc}^\tau$  by modifying Algorithm 1 (resp. Alg. 2) to use SAP on GCC instead of LAP on ALLDIFFERENT. The resulting algorithms SNC-SAP-GLB (resp. SNC-SAP-GREEDY) have a time complexity in  $O(d^2n^3)$ .

## 5 Experimental Results

We implemented in TOULBAR2, an open-source C++ exact WCSP solver, a propagator for enforcing  $F\emptyset IC$  on the GCC in closed form, based on the modified LAPJV algorithm [25].<sup>10</sup> TOULBAR2 already includes a propagator for enforcing  $F\emptyset IC$  on ALLDIFFERENT [40]. We implemented four methods combining each of the two propagators with each of the two strategies GLB (Alg. 1) and Greedy (Alg. 2). Due to their relatively high cost, we apply them only in preprocessing.

We compared four versions of TOULBAR2 having different preprocessing:

- no preprocessing (TOULBAR2 with default options), called LAP in the results when applied to QAP models with the ALLDIFFERENT constraint (same results as tb2+LAPJV in [40]) and SAP when applied to a model with GCC,
- SNC-LAP-GLB using the GLB strategy (options  $-S -glb=1$ ),
- SNC-LAP-GREEDY using the Greedy strategy (options  $-S -glb=0$ ),
- SNC-LAP-GLB followed by SNC-LAP-GREEDY (options  $-S -glb=2$ ), called SNC-LAP-GLB-GREEDY in the results.

For each version, after preprocessing, TOULBAR2 continues the search, enforcing EDAC [12] on unary and binary cost functions, and  $F\emptyset IC$  on ALLDIFFERENT and GCC, at every node of a hybrid best/depth-first branch-and-bound search [2], using binary branching and Maximum Cardinality Search for the DAC ordering.<sup>11</sup> TOULBAR2 used the *dom/wdeg* variable ordering heuristic [7] with last conflict [28] and EAC *support value* ordering heuristic [10, 41], combined with solution phase saving [15]. Unless reported otherwise, all tests were run on a 2.5GHz Intel Xeon E5-2680 (2014), using one thread per CPU, and with a CPU-time limit of 1,200 seconds.

We took the 132 smallest of the 136 instances from the QAPLIB.<sup>12</sup> We compared the four versions of TOULBAR2 against Google OR-Tools CP-SAT, an open-source state-of-the-art constraint programming solver, and IBM CPLEX, a state-of-the-art integer programming solver.<sup>13</sup> Unary and binary cost functions are encoded using the support encoding for CPLEX [22], and to binary and ternary constraints in extension with extra objective variables for CP-SAT. The permutation constraint is encoded for CPLEX as binary constraints enforcing that any pair of two variables cannot take the same value. It is encoded as an ALLDIFFERENT constraint in CP-SAT and TOULBAR2. The GCC constraint is only tested using TOULBAR2 in the inverted model  $P_{gcc}^{\tau}$ .<sup>14</sup>

<sup>10</sup> <https://github.com/toulbar2/toulbar2>

<sup>11</sup> TOULBAR2 options  $-d: -o -O=-1$

<sup>12</sup> <http://coral.ise.lehigh.edu/wp-content/uploads/2014/07/qapdata.tar.gz>, sizes less than 100 variables.

<sup>13</sup> <https://github.com/google/or-tools> CP-SAT v9.14 in free-search mode, CPLEX version 22.1.1.0 with non-premature stop parameters  $EPAGAP=EPGAP=EPINT=0$ . All single-threaded.

<sup>14</sup> CP-SAT uses a decomposition of GCC into linear constraints. Our preliminary tests showed worse results on the inverted model with the decomposed GCC.

| Preprocessing      | AvgInitialLbGap | AvgSncIterations | AvgSncTimes       |
|--------------------|-----------------|------------------|-------------------|
| GLB                | 30.46           | N/A              | N/A               |
| SNC-LAP-GLB-Greedy | <b>24.13</b>    | 352.53           | 60.76 sec.        |
| SNC-LAP-Greedy     | 26.38           | 368.67           | 73.08 sec.        |
| SNC-LAP-GLB        | 26.32           | <b>160.38</b>    | <b>25.46 sec.</b> |
| LAP                | 83.63           | N/A              | N/A               |

Table 1: Initial lower bound gap percent., nb. of iter., and av. time for QAPLIB.

| Gap                | mean        | median      | std         | Score        | Nb. instances |
|--------------------|-------------|-------------|-------------|--------------|---------------|
| Cplex              | 16.41       | 9.35        | 20.38       | 25.00        | 109           |
| CP-SAT             | 5.12        | 2.29        | 11.00       | 33.00        | 122           |
| LAP                | 4.01        | 1.88        | 5.27        | 41.50        | <b>132</b>    |
| SNC-LAP-Greedy     | 1.44        | <b>0.00</b> | <b>2.29</b> | 83.00        | 131           |
| SNC-LAP-GLB        | 1.50        | <b>0.00</b> | 2.56        | 89.00        | <b>132</b>    |
| SNC-LAP-GLB-Greedy | <b>1.42</b> | <b>0.00</b> | 2.33        | <b>92.00</b> | 131           |

Table 2: Optimality gap percentages on the original model for QAPLIB.

### 5.1 Comparison of Initial Lower Bounds

Tab. 1 gives average initial lower bound gap  $(1 - \frac{lb}{bestsolknown})$  found by the original GLB method and by the four preprocessing methods on  $P_{alldiff}$ . It also reports the average number of iterations on variables (for-loop line 2) and CPU-time of SNC methods. Without preprocessing, LAP had a poor lower bound. SNC methods were much better, all being superior on average to original GLB. SNC with the GLB or Greedy strategy got the same quality of lb, but SNC-LAP-GREEDY was 2-3x slower to converge than SNC-LAP-GLB (see also Fig.2a). Hopefully, SNC-LAP-GLB-GREEDY is a bit faster than SNC-LAP-GREEDY thanks to the reformulations initially made by SNC-LAP-GLB, and it got the best lb gap. Similar results are observed on the other models  $P_{alldiff}^\tau$  and  $P_{gcc}^\tau$ .

For instance *tai100a*, SNC-LAP-GLB (resp. SNC-LAP-GLB-GREEDY) found an initial lower bound of 15,864,722 in 196 seconds and 300 iterations (resp. 15,931,389 in 405s, 500 iter.), which are greater than the original GLB bound of 15,824,355 and the best bound of 15,844,731 reported on QAPLIB website using a semidefinite relaxation-matrix splitting bound [35].<sup>15</sup>

### 5.2 Comparison of Solving Times

On the original model  $P_{alldiff}$ , LAP solved 33 instances, whereas all SNC methods solved more than 54. SNC-LAP-GLB-GREEDY got the best results (56 solved). On the inverted model  $P_{alldiff}^\tau$ , without exploiting symmetries, performances decreased for LAP (23 solved) and SNC-LAP-GREEDY (50 solved), and remained stable for the two other SNCs. By exploiting symmetries in  $P_{gcc}^\tau$ , we observed a very nice improvement for all the methods, SNC-SAP-GLB being the best (67 solved). The largest solved instance is *lipa90b* with size  $n = 90$  in 87 seconds

<sup>15</sup> <https://coral.ise.lehigh.edu/data-sets/qaplib/qaplib-problem-instances-and-solutions/#Ta>

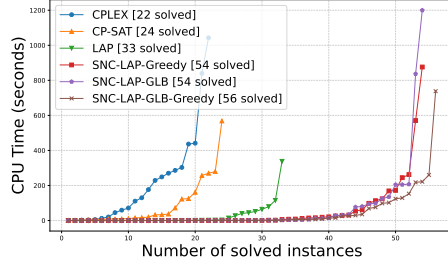
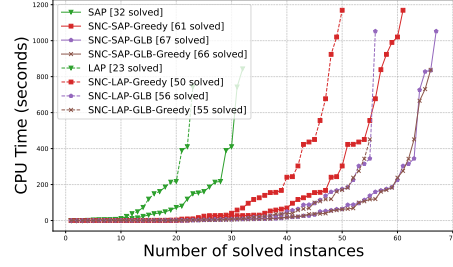
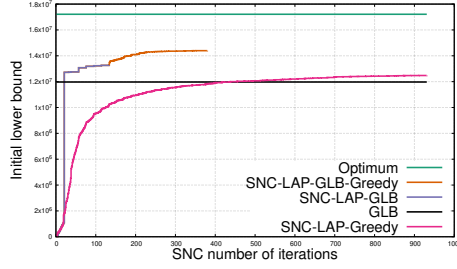
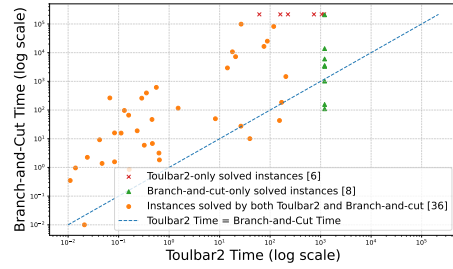
(a) Original model ( $P_{alldiff}$ )(b) Inverted models ( $P_{alldiff}^\tau$  and  $P_{gcc}^\tau$ )

Fig. 1: Cactus plots on the original and inverted models for QAPLIB benchmark.

(a) Initial lower bounds obtained on QAPLIB instance *els19* using the original model. SNC-LAP-GLB-GREEDY took 0.69s and 380 iter. in preprocessing and then tb2 solved in 5.82s.

(b) TOULBAR2 vs Branch-and-Cut [18] runtime comparison. Note the drastically different hardware and timeouts (20 minutes for tb2 / 2.5 days for B&amp;C) makes comparison difficult.

Fig. 2: Detailed results for the QAPLIB benchmark.

and 100 nodes by SNC-SAP-GLB (72s for its preprocessing). As previously reported [40], CFN methods performed much better than CP-SAT (24 solved) and default OR methods (CPLEX, 22 solved) (Fig. 1).

### 5.3 Comparison of Solution Quality

In Table 2, we report optimality gap distribution ( $\frac{ub-best}{ub}$ ), XCSP3 competition score [40], and number of instances with at least one solution found on  $P_{alldiff}$  (similar results are observed on the other models). The best average gap of 1.42% is obtained by SNC-LAP-GLB-GREEDY. However, it could not finish its preprocessing in less than 20 minutes for one instance (*tai100b*). SNC-LAP-GLB found a solution for all 132 instances with an average gap of 1.5%. It is clearly better than LAP (4%), CP-SAT (5%), and CPLEX (16%).

### 5.4 Comparison with a Dedicated Branch-and-Cut Approach

Fig.2b gives a log-log scatter plot between our best approach using TOULBAR2 (whatever the chosen SNC preprocessing or QAP model) and a branch-and-cut

| <i>Instance</i> | <i>n</i> | <i>d</i> | OPT       | Cplex [19]      | SNC-SAP-GLB-GREEDY |
|-----------------|----------|----------|-----------|-----------------|--------------------|
| <i>esc16a</i>   | 16       | 9        | 68        | <b>0.35</b>     | 0.65               |
| <i>esc16b</i>   | 16       | 7        | 292       | 3.07            | <b>2.06</b>        |
| <i>esc16c</i>   | 16       | 12       | 160       | 130.98          | <b>73.14</b>       |
| <i>esc16d</i>   | 16       | 12       | 16        | <b>0.51</b>     | 77.22              |
| <i>esc16e</i>   | 16       | 8        | 28        | <b>0.05</b>     | 1.11               |
| <i>esc16f</i>   | 16       | 1        | 0         | <b>0.00</b>     | <b>0.00</b>        |
| <i>esc16g</i>   | 16       | 9        | 26        | <b>0.04</b>     | 0.18               |
| <i>esc16h</i>   | 16       | 5        | 996       | 0.23            | <b>0.11</b>        |
| <i>esc16i</i>   | 16       | 10       | 14        | <b>0.18</b>     | 7.88               |
| <i>esc16j</i>   | 16       | 7        | 8         | <b>0.03</b>     | 0.13               |
| <i>esc32c</i>   | 32       | 10       | 642       | <b>9,643.82</b> | -                  |
| <i>esc32d</i>   | 32       | 13       | 200       | <b>2,973.26</b> | -                  |
| <i>esc32e</i>   | 32       | 6        | 2         | <b>0.04</b>     | 106.33             |
| <i>esc32g</i>   | 32       | 7        | 6         | <b>0.06</b>     | 5.86               |
| <i>esc64a</i>   | 64       | 15       | 116       | <b>509.87</b>   | -                  |
| <i>tai64c</i>   | 64       | 2        | 1,855,928 | 18,250.40       | <b>995.51</b>      |

Table 3: Comparison on 16 QAPLIB instances with flow matrix symmetry and using the inverted model ( $P_{gcc}^\tau$ ). A “-” means CPU-time limit reached.

(B&C) method dedicated for QAP [18], using the results reported in that paper. Among the 71 instances TOULBAR2 could solve in less than 20 minutes, we kept 42 of size less than 30, removing *esc* and *lipa* families as done in B&C results. B&C could solve 44 instances with  $n \leq 30$ , but often taking much more time than TOULBAR2 (up to 3 orders of magnitude speedups on *bur26* and *chr* families).

### 5.5 Comparison with a Dedicated MILP Approach

In [19], a dedicated MILP formulation is proposed, exploiting symmetry in the flow matrix. In Table 3, we report their results obtained using CPLEX 12.2 with 8 threads on a quad core Intel Xeon CPU 3.2 GHz with 16 GB RAM. We give also the results obtained by our approach (SNC-SAP-GLB-GREEDY) using TOULBAR2 with the parallel version of its hybrid best/depth-first branch-and-bound search method [4] on a 8-core Intel Xeon E5-2687W v4 3.5 GHz with 256 GB. The CPU-time limit is fixed to 10 hours. The MILP approach produced the best results, solving 16 instances and TOULBAR2 only 13. However, on the large *tai64c* instance, our approach was 18 times faster than CPLEX, and developed less search nodes (162,790,320 compared to 1,216,074,081).

## 6 Automatic parameter configuration with IRACE

As shown in Table 4, different modeling approaches will yield the best results depending on the instance. Even in the same subfamily, there can be a lot of performance diversity (like in the *esc\**, *nug\** or *tai\** instances). As such, it would be interesting to know when to use a modelisation or another. Similarly, this problem exists around the choice of the solver’s parameters. Depending on the instance and it’s modelisation, the better configuration will fluctuate. So, using the right configuration to solve a modelisation could allow it to solve more instances, showing its full potential. In this section, we will be searching for a better configuration for the  $P_{alldiff}$  model.

| Instances | $P_{\text{alldiff}}$ | $P_{\text{alldiff}}^\tau$ | $P_{\text{knapsack}}$ | $P_{\text{binary}}$ |
|-----------|----------------------|---------------------------|-----------------------|---------------------|
| bur26*    |                      | <b>8</b>                  |                       |                     |
| chr*      | <b>14</b>            |                           |                       |                     |
| els19     |                      | <b>1</b>                  |                       |                     |
| esc*      | <b>8</b>             | 4                         | 1                     | 5                   |
| had*      | 1                    | <b>3</b>                  |                       |                     |
| kra*      | <b>2</b>             | 1                         |                       |                     |
| lipa*     | 1                    | <b>15</b>                 |                       |                     |
| nug*      | 6                    | <b>9</b>                  |                       |                     |
| rou*      | <b>3</b>             |                           |                       |                     |
| scr*      | <b>3</b>             |                           |                       |                     |
| sko*      | <b>12</b>            | 1                         |                       |                     |
| ste36*    |                      | <b>3</b>                  |                       |                     |
| tai*      | <b>14</b>            | 11                        |                       | 1                   |
| tho*      |                      | <b>2</b>                  |                       |                     |
| wil*      | 1                    | 1                         |                       |                     |

Table 4: Comparison of number of instances best solved (in terms of solution quality and then solving time) on four different problem modelling:  $P_{\text{alldiff}}$ ,  $P_{\text{alldiff}}^\tau$  (both solved by SNC-LAP-GLB in preprocessing and LAP during search), a decomposition of AllDifferent into binary constraints  $P_{\text{binary}}$ , and by adding *at-least-one* constraints to the decomposition  $P_{\text{knapsack}}$  [40].

A solver has a configuration space  $\Theta$  in which all valid instantiations of its  $p$  parameters exist. And a problem space  $\mathcal{X}$  in which all the instances to solve are. Considering a function  $f(\theta \in \Theta, x \in \mathcal{X})$  that gives the performance of the configuration ( $\theta$ ) on the instance ( $x$ ), we search for:  $\theta^* = \min_{\theta \in \Theta} \mathbb{E}[f(\theta, \mathcal{X})]$ . There are many ways to do it. In this paper, we chose to use the automatic configurator IRACE [31].

IRACE is a configurator using iterative racing that asks for a set of instances, the solver to configure, a description of its configuration space, and IRACE hyperparameters (amongst them, the budget allocated to the search). It works by making sets of configurations run on an increasing number of instances. In a regular manner, statistical tests are used to eliminate the configurations that are dominated.

It is possible to give some configurations to the initial candidate set. All other configurations are generated by using uniform sampling of the configuration space during the first run, and then progressively steering them toward elite configurations during subsequent runs. See Algorithm 3.

We decided to explore six TOULBAR2 categorical parameters mostly related to search and use default values for the rest (except option *-o* which is always added):

- Weighted degree variable ordering heuristic [7]: no (option *-q:*), or yes (option *-q=1000000*),
- Weighted degree initial value for each nonunary cost function: 1 (*-m:*), mean (*-m=1*) or median (*-m=2*) cost on the valid tuples returned by the function,
- Solution-based phase saving value ordering heuristic [15]: no (*-solr:*), or yes (*-solr*),
- Dichotomic branching on large domains: no (*-d:*), or yes (*-d=1*, splitting domains in the middle of their range),



---

**Algorithm 3** IRACE  $(\Theta, \mathcal{X}, f, B, p, m, e, T^{first}, T^{each})$ : configurator by iterative racing.

---

```

1:  $n \leftarrow \lfloor 2 + \log_2(p) \rfloor$ 
2:  $B_1 \leftarrow \lfloor \frac{B}{n} \rfloor$ 
3:  $N_1 \leftarrow \lfloor \frac{B_1}{T^{first} + T^{each}} \rfloor$ 
4:  $\Theta_1 \leftarrow N_1$  configurations, those indicated by the user (can be none)  $\cup$  uniform
   sampling on the configuration space  $\Theta$ 
5:  $\mathcal{X}_1 \leftarrow \lfloor \frac{B_1}{N_1} \rfloor$  uniform sampling of the instances
6: for all  $i \in \{1, \dots, n\}$  do
7:   if  $B_i \geq T^{first} + T^{each} \times \min(5, i) \wedge |\Theta_i| \geq m$  then
8:      $r \leftarrow 0$ 
9:     for all  $x \in \mathcal{X}_i$  do
10:      for all  $\theta \in \Theta_i$  do
11:        Run TOULBAR2 to collect  $f(\theta, x)$ 
12:         $r \leftarrow r + 1$ 
13:      end for
14:      After at least  $T^{first}$ , and every  $T^{each}$  afterward, instances, use statistical
        tests to eliminate statistically inefficient configurations in  $\Theta_i$ 
15:    end for
16:    while  $r + |\Theta_i| \leq B_i \wedge |\Theta_i| \geq m$  do
17:       $\mathcal{X}_i \leftarrow \mathcal{X}_i \cup \{x_{new}\}$ , one random instance in the training set
18:      for all  $\theta \in \Theta_i$  do
19:        Run TOULBAR2 on  $x_{new}$  to collect  $f(\theta, x_{new})$ 
20:         $r \leftarrow r + 1$ 
21:      end for
22:      Every  $T^{each}$  instances, use statistical tests to eliminate statistically ineffi-
        cient configurations in  $\Theta_i$ 
23:    end while
24:  end if
25:   $B \leftarrow B - r$ 
26:  if  $i = n \wedge B \geq T^{first} + T^{each} \times \min(5, i + 1)$  then
27:     $n \leftarrow n + 1$  /* continue another iteration */
28:  end if
29:   $B_{i+1} \leftarrow \lfloor \frac{B}{n-i} \rfloor$ 
30:   $N_{i+1} \leftarrow \lfloor \frac{B_{i+1}}{T^{first} + T^{each} \times \min(5, i+1)} \rfloor$ 
31:   $\Theta_{elite} \leftarrow$  the best  $e$  configurations left in  $\Theta_i$ 
32:   $\Theta_{i+1} \leftarrow \Theta_{elite} \cup \{(N_{i+1} - N_i)$  new configurations with sampling biased by using
    the configurations in  $\Theta_{elite}\}$ 
33:   $\mathcal{X}_{i+1} \leftarrow \mathcal{X}_i \cup \{N_{i+1}$  new random instances $\}$ 
34: end for
35: return the best  $e$  configurations of the last  $\Theta_i$  for testing (+ potentially all elite
    configurations if the option is activated)

```

---

- DAC variable ordering heuristics [10]: 8 heuristics from  $-O=-1$  to  $-O=-8$ ,
- Singleton node consistency in preprocessing: no preprocessing ( $-nopre$ ), SNC-LAP-GREEDY ( $-S -glb=0$ ), SNC-LAP-GLB ( $-S glb=1$ ), SNC-LAP-GLB-GREEDY ( $-S -glb=2$ ).

The configuration space size  $|\Theta| = 768$ . Our baseline configuration  $\theta^{def}$  is  $-q=1000000 -m: -solr -d: -O=-1 -S -glb=1$  and it is always added to the initial configuration set of IRACE. Each run of the solver has a CPU time limit of 20 minutes on a single core of a 20-core Intel i7-12700H 4.7GHz and 64GB of RAM (using runsolver [38] to ensure the respect of those constraints).

For the training set  $\mathcal{X}$ , we begun by taking away the 20 % largest instances for each subfamily of the QAPLIB benchmark (rounded superior, so, if we have 2 problems, 1 will be put in the test set). We are left with 99 instances. Since many of them would be long to solve and some subfamilies are large (reducing the chances of the smaller families being picked), we reduce the set further by taking away half of the largest instances of each of the remaining subfamily that have more than 2 instances. Consequently, the training set contains 53 instances. All other instances are in the test set.

The performance of a configuration  $\theta$  on instance  $x$  is given by  $f(\theta, x)$ . We chose a relatively simple function that linearly combines two criteria: solution cost  $c$  and solving time  $t$  (rounded up to seconds). The first criterion is normalized by the SOTA results ( $c^{sota}$ ).<sup>16</sup> If no solution is found then the solving time is twice the CPU time limit. Formally, we have  $f(\theta, x) = 1200 \frac{(c - c^{sota})}{(c^{sota} + 1)} + t$ .

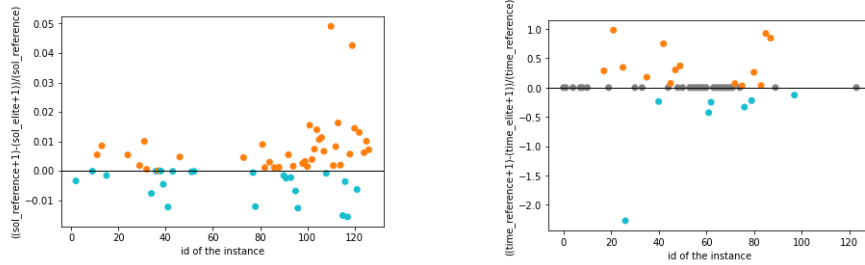
Default hyperparameter values of IRACE were chosen, except for:

- number of executions of IRACE: 5 (using seed values: 42 (2 times), 68 (3 times)),
- maximum number of TOULBAR2 executions  $B = 500$  (IRACE budget),
- minimum number of iterations/races  $n = 4$ ,
- minimum number of configurations to continue racing  $m = 1$ ,
- final number of elite configurations to output  $e = 4$ ,
- statistical test: paired t-test with no correction,
- first test done after  $T^{first} = 4$  instances,
- next tests are performed every  $T^{each} = 2$  instances,
- deterministic mode: TOULBAR2 is using a deterministic complete branch-and-bound method,
- parallelization activated: up to 17 cores can be used in parallel,
- elitist activated: it allows to keep a configuration that should be eliminated for a bit longer if it was an elite previously, reducing the risk of losing a good configuration,
- elitist new instances: at the beginning of each iteration, we begin with 5 new instances to allow for more exploration of the instance space,
- elitist test limit: we allow for 2 successive elimination tests to fail to eliminate any instances, otherwise, we consider the remaining configurations too similar and we want to finish this race to begin with a new one (with new configurations).

<sup>16</sup> <https://qaplib.mgi.polymtl.ca>

With the trace of IRACE (the results that the configurations have had on the instances) and the 17 different elites found during the different executions, we compared how each of those configurations held against the baseline on the instances that were used during the different executions of IRACE.

Figure 3 shows the gap between the results of the baseline configuration  $\theta^{def}$  and a particular elite configuration  $\theta_1^{irace}$  ( $-q: -m: -solr: -d: -O=-2 -S -glb=2$ ) on the instances that were explored during the combined executions of IRACE (training and testing). The orange dots represents the instances where  $\theta_1^{irace}$  was the best configuration. The blue those where  $\theta^{def}$  was the best. The grey represent a full equality on the instance. In the case where an instance reaches the same objective value (for example because  $\theta^{def}$  and  $\theta_1^{irace}$  found the optimum), we do not print them on the left subfigure and will print only their difference on time on the right subfigure. When they have a difference on the objective value, it means they didn't find the optimum and are in time-out, so, we only show their point on the left subfigure showing the relative objective value gap to the baseline. Some instances were not encountered by IRACE on the elite configuration, we ignore those instances since no comparison can be done with this data.



(a) Comparison on relative objective value gap to baseline. (b) Comparison on relative time gap to baseline.

Fig. 3: Comparison of handcrafted ( $\theta^{def}$ ) / elite ( $\theta_1^{irace}$ ) on QAPLIB instances explored during a run of IRACE.

From the results of each elite configuration, we chose two promising candidates, for a training time of  $-85$  hours of CPU-time, and  $-74$  hours for test:

- $\theta_1^{irace}$ :  $-q: -m: -solr: -d: -O=-2 -S -glb=2$ 
  - Better 38 times on the objective value than the baseline, against 23 worse,
  - Better 14 times on the execution time than the baseline, against 7 worse,
  - Two more instances solved to optimality : *lipa40a*, *had18*,
  - Faster by more than 3,000 seconds in cumulative time.
- $\theta_2^{irace}$ :  $-q=1000000 -m=1 -solr: -d: -O=-6 -S -glb=1$

- Better 41 times on the objective value than the baseline, against 21 times worse,
- Better 14 times on the execution time than the baseline, against 6 times worse,
- One more instance solved to optimality : *had18*,
- Faster by more than 2,000s in cumulative time.

As those configurations were identified on a limited part of the instances, an execution was then done on the full set of instances to allow a complete comparison.

Two other configurations were also added. These configurations were obtained by running all possible configurations (768) on all the 132 instances for a CPU time limit of 1 minute (total configuration time of  $\approx 1,500$  hours) and keeping the best *oracle* configuration  $\theta_1^{oracle1min}$  (*-q: -m=2 -solr: -d: -O=-1 -S -glb=2*) if we consider equally the number of solved instances and the cumulative time, and the best one  $\theta_2^{oracle1min}$  (*-q: -m=2 -solr: -d: -O=-6 -S -glb=1*) considering average objective gap while penalizing the lack of solution found with a 10% objective gap.

Figure 4 compares those four potentially good configurations to the baseline (in blue). We can see that all the learnt configurations gave better results in number of solved instances and cumulative total time.

In Table 5 we report optimality gap distribution ( $\frac{ub-best}{best+1}$ , 100% if no solution found), number of solved instances, and number of instances with at least one solution found. The best configurations are found by IRACE.

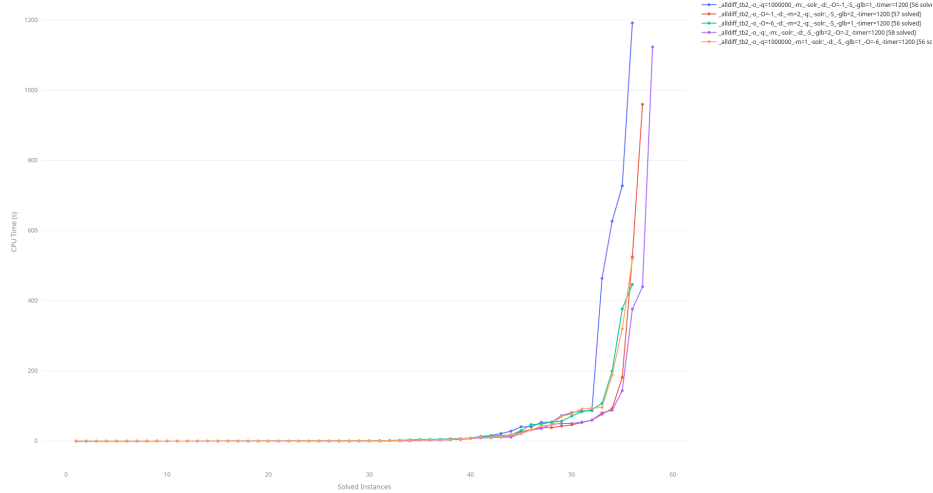


Fig. 4: Cactus plot of handcrafted/elite/oracle configurations on the original model for QAPLIB benchmark.

| Gap                     | mean        | std         | Nb. solved instances | Nb. feasible instances |
|-------------------------|-------------|-------------|----------------------|------------------------|
| $\theta^{def}$          | 1.56        | 2.61        | 56                   | <b>132</b>             |
| $\theta_1^{irace}$      | 2.04        | 8.79        | <b>58</b>            | 131                    |
| $\theta_2^{irace}$      | <b>1.31</b> | <b>2.02</b> | 56                   | <b>132</b>             |
| $\theta_1^{oraclelmin}$ | 2.00        | 8.78        | 57                   | 131                    |
| $\theta_2^{oraclelmin}$ | 1.34        | 2.22        | 56                   | <b>132</b>             |

Table 5: Optimality gap percentages and number of solved/feasible instances on the original model for QAPLIB.

## 7 Conclusion

Taking ideas from CP (singleton consistency, global cardinality constraint) and OR (Gilmore-Lawler bound, linear and semi-assignment problems), we significantly improved a CFN solver on quadratic assignment problems, reaching state-of-the-art OR results in a few cases (mostly on asymmetric instances such as *bur26* and *lipa*, and also *tai64c*). In the future, we will explore other soft local consistency methods [10, 34] and problem formulation [1] to get stronger bounds. From our preliminary results on static solver configuration, we will investigate dynamic configurations using portfolios.

**Supplementary Materials.** Detailed results (csv files) are available at: <https://web-genobioinfo.toulouse.inrae.fr/~degivry/cpaior2026/cpaior2026supp.zip>.

**Acknowledgments.** This work has benefited from French grants managed by the National Research Agency under the "Investissements d’Avenir" program with the references ANR-18-EURE-0021, ANR-23-IACL-0002, and ANR-24-CE23-3429-03.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Adams, W.P., Guignard, M., Hahn, P.M., Hightower, W.L.: A level-2 reformulation–linearization technique bound for the quadratic assignment problem. *European Journal of Operational Research* **180**(3), 983–996 (2007)
2. Allouche, D., de Givry, S., Katsirelos, G., Schiex, T., Zytnicki, M.: Anytime Hybrid Best-First Search with Tree Decomposition for Weighted CSP. In: *Proc. of CP-15*. pp. 12–28. Cork, Ireland (2015)
3. Barták, R., Erben, R.: A new algorithm for singleton arc consistency. In: *FLAIRS*. pp. 257–262 (2004)
4. Beldjilali, A., Montalbano, P., Allouche, D., Katsirelos, G., de Givry, S.: Parallel Hybrid Best-First Search. In: *Proc. of CP-22*. vol. 235, pp. 7:1–7:10. Haifa, Israel (2022)
5. Bessiere, C., Cardon, S., Debruyne, R., Lecoutre, C.: Efficient algorithms for singleton arc consistency. *Constraints* **16**(1), 25–53 (2011)

6. Bessiere, C., Debruyne, R.: Theoretical analysis of singleton arc consistency and its extensions. *Artificial Intelligence* **172**(1), 29–41 (2008)
7. Boussemart, F., Hemery, F., Lecoutre, C., Sais, L.: Boosting systematic search by weighting constraints. In: *ECAI*. vol. 16, p. 146 (2004)
8. Burkard, R., Dell’Amico, M., Martello, S.: *Assignment problems: revised reprint*. SIAM (2012)
9. Claus, G., Cambazard, H., Jost, V.: Analysis of reduced costs filtering for alldifferent and minimum weight alldifferent global constraints. In: *Proc. of 24th European Conference on Artificial Intelligence*. vol. 325, pp. 323–330. Santiago de Compostela, Spain (2020)
10. Cooper, M., de Givry, S., Sanchez, M., Schiex, T., Zytnicki, M., Werner, T.: Soft arc consistency revisited. *Artificial Intelligence Journal* **174**, 449–478 (2010)
11. Cooper, M.C., de Givry, S., Schiex, T.: Valued constraint satisfaction problems. In: *A Guided Tour of Artificial Intelligence Research: Volume II: AI Algorithms*, pp. 185–207. Springer (2020)
12. De Givry, S., Heras, F., Zytnicki, M., Larrosa, J.: Existential arc consistency: Getting closer to full arc consistency in weighted csps. In: *IJCAI*. vol. 5, pp. 84–89 (2005)
13. Debruyne, R.: Optimal and suboptimal singleton arc consistency algorithms. In: *Ijcai* (2005)
14. Debruyne, R., Bessière, C.: Some practicable filtering techniques for the constraint satisfaction problem. In: *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, IJCAI 97, Nagoya, Japan, August 23–29*. pp. 412–417 (1997)
15. Demirović, E., Chu, G., Stuckey, P.J.: Solution-based phase saving for CP: A value-selection heuristic to simulate local search behavior in complete solvers. In: *International Conference on Principles and Practice of Constraint Programming*. pp. 99–108. Springer (2018)
16. Dlask, T., Werner, T., de Givry, S.: Bounds on weighted csps using constraint propagation and super-reparametrizations. In: *Proc. of CP-21. Montpellier, France* (2021)
17. Dlask, T., Werner, T., de Givry, S.: Super-Reparametrizations of Weighted CSPs: Properties and Optimization Perspective. *Constraints* **28**, 277–319 (2023)
18. Erdoğan, G., Tansel, B.: A branch-and-cut algorithm for quadratic assignment problems based on linearizations. *Computers & Operations Research* **34**(4), 1085–1106 (2007)
19. Fischetti, M., Monaci, M., Salvagnin, D.: Three ideas for the quadratic assignment problem. *Operations research* **60**(4), 954–964 (2012)
20. Frieze, A.M., Yadegar, J.: On the quadratic assignment problem. *Discret. Appl. Math.* **5**(1), 89–98 (1983)
21. Gilmore, P.C.: Optimal and suboptimal algorithms for the quadratic assignment problem. *Journal of the society for industrial and applied mathematics* **10**(2), 305–313 (1962)
22. Hurley, B., O’Sullivan, B., Allouche, D., Katsirelos, G., Schiex, T., Zytnicki, M., de Givry, S.: Multi-language evaluation of exact solvers in graphical model discrete optimization. *Constraints An Int. J.* **21**(3), 413–434 (2016)
23. Jean-Charles, R.: Generalized arc consistency for global cardinality constraint. *American Association for Artificial Intelligence (AAAI 1996)* pp. 209–215 (1996)
24. Jonker, R., Volgenant, A.: A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing* **38**, 325–340 (1987)

25. Kennington, J., Wang, Z.: A shortest augmenting path algorithm for the semi-assignment problem. *Operations Research* **40**(1), 178–187 (1992)
26. Koopmans, T.C., Beckmann, M.: Assignment problems and the location of economic activities. *Econometrica: journal of the Econometric Society* pp. 53–76 (1957)
27. Kuhn, H.W.: The hungarian method for the assignment problem. *Naval Research Logistics Quarterly* **2**(1-2), 83–97 (1955)
28. Lecoutre, C., Saïs, L., Tabary, S., Vidal, V.: Reasoning from last conflict(s) in constraint programming. *Artificial Intelligence Journal* **173**, 1592,1614 (2009)
29. Lecoutre, C., Cardon, S.: A greedy approach to establish singleton arc consistency. In: *IJCAI*. vol. 5, pp. 199–204 (2005)
30. Loiola, E.M., De Abreu, N.M.M., Boaventura-Netto, P.O., Hahn, P., Querido, T.: A survey for the quadratic assignment problem. *European journal of operational research* **176**(2), 657–690 (2007)
31. López-Ibáñez, M., Dubois-Lacoste, J., Pérez Cáceres, L., Birattari, M., Stützle, T.: The irace package: Iterated racing for automatic algorithm configuration. *Operations Research Perspectives* **3**, 43–58 (2016). <https://doi.org/https://doi.org/10.1016/j.orp.2016.09.002>
32. Montalbano, P.: Linear Constraints and Conflict-free Learning for Graphical Models. Ph.D. thesis, Université de Toulouse (2023)
33. Montalbano, P., de Givry, S., Katsirelos, G.: Multiple-choice knapsack constraint in graphical models. In: *International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research*. pp. 282–299. Springer (2022)
34. Montalbano, P., de Givry, S., Katsirelos, G.: Virtual arc consistency for linear constraints in cost function networks. In: *Proc. of ICTAI-25*. Athens, Greece (2025)
35. Peng, J., Mittelman, H., Li, X.: A new relaxation framework for quadratic assignment problems based on matrix splitting. *Mathematical Programming Computation* **2**(1), 59–77 (2010)
36. Régim, J.C.: Cost-based arc consistency for global cardinality constraints. *Constraints* **7**(3), 387–405 (2002)
37. Rossi, F., Van Beek, P., Walsh, T.: *Handbook of constraint programming*. Elsevier (2006)
38. Roussel, O.: Controlling a solver execution: the runsolver tool. *JSAT* **7**, 139–144 (11 2011). <https://doi.org/10.3233/SAT190083>
39. Sellmann, M.: An arc-consistency algorithm for the minimum weight all different constraint. In: Hentenryck, P.V. (ed.) *Proc. of 8th International Conference on Principles and Practice of Constraint Programming*, Ithaca, NY, USA, September 9–13. Lecture Notes in Computer Science, vol. 2470, pp. 744–749. Springer (2002)
40. Sewa, G., Allouche, D., de Givry, S., Katsirelos, G., Montalbano, P., Schiex, T.: Assignment problems in cost function networks. In: *Proc. of AAI-26*. Singapore (2026)
41. Trösser, F., De Givry, S., Katsirelos, G.: Relaxation-aware heuristics for exact optimization in graphical models. In: *International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research*. pp. 475–491. Springer (2020)
42. Van Hentenryck, P., Saraswat, V., Deville, Y.: Design, implementation, and evaluation of the constraint language cc (fd). *The Journal of Logic Programming* **37**(1-3), 139–164 (1998)
43. Xia, Y.: Gilmore-lawler bound of quadratic assignment problem. *Frontiers of Mathematics in China* **3**(1), 109–118 (2008)