

Grasp and Path-Relinking for Coalition Structure Generation

Nicola Di Mauro ^C

Department of Computer Science

University of Bari

ndm@di.uniba.it

Stefano Ferilli

Department of Computer Science

University of Bari

ferilli@di.uniba.it

Teresa M.A. Basile

Department of Computer Science

University of Bari

basile@di.uniba.it

Floriana Esposito

Department of Computer Science

University of Bari

esposito@di.uniba.it

Abstract. In Artificial Intelligence with Coalition Structure Generation (CSG) one refers to those cooperative complex problems that require to find an optimal partition (maximizing a social welfare) of a set of entities involved in a system. The solution of the CSG problem finds applications in many fields such as Machine Learning (set covering machines, clustering), Data Mining (decision tree, discretization), Graph Theory, Natural Language Processing (aggregation), Semantic Web (service composition), and Bioinformatics. The problem of finding the optimal coalition structure is NP-complete. In this paper we present a greedy adaptive search procedure (GRASP) with path-relinking to efficiently search the space of coalition structures. Experiments and comparisons to other algorithms prove the validity of the proposed method in solving this hard combinatorial problem.

Keywords: Coalition Structure Generation, Stochastic-Local-Search, GRASP procedure, Path-Relinking

1. Introduction

An active area of research in Artificial Intelligence regards methods and algorithms to solve complex problems that require to find an optimal partition (maximizing a social welfare) of a set of entities involved in a system into exhaustive and disjoint coalitions. This problem has been studied a lot in the

Address for correspondence: Department of Computer Science, University of Bari

^CCorresponding author

area of multi-agent systems (MASs) where it is named coalition structure generation (CSG) problem (equivalent to the complete set partitioning problem). In particular it is interesting to find coalition structures maximizing the sum of the values of the coalitions, that represent the maximum payoff the agents belonging to the coalition can jointly receive through cooperation. A coalition structure is defined as a partition of the agents involved in a system into disjoint coalitions. The problem of finding the optimal coalition structure is \mathcal{NP} -complete [6, 25].

Coalition generation shares a similar structure with a number of common problems in theoretical computer science and artificial intelligence, such as in combinatorial auctions; in job shop scheduling, Machine Learning, Data Mining, Graph Theory, Natural Language Processing, Semantic Web, and in Bioinformatics. In this paper we will use the term coalition structure generation as a general term to refer to all these grouping problems.

Sometimes there is a time limit for finding a solution, the agents must be reactive and they should act as fast as possible. Hence for the specific task of CSG it is necessary to have approximation algorithms able to quickly find solutions that are within a specific factor of an optimal solution. The goal of this paper is to propose a new algorithm for the CSG problem able to quickly find a near optimal solution.

The problem of CSG has been studied in the context of characteristic function games (CFGs) in which the value of each coalition is given by a characteristic function, and the values of a coalition structure are obtained by summing the values of the coalitions it contains. The problem of coalition structure generation is \mathcal{NP} -hard, indeed as proved in [25], given n the number of agents, the number of possible coalition structures that can be generated is $O(n^n)$ and $\omega(n^{n/2})$. Moreover, in order to establish any bound from the optimal, any algorithm must search at least $2^n - 1$ coalition structures. The CSG process can be viewed as being composed of three activities [25]: a) *coalition structure generation*, corresponding to the process of generating coalitions such that agents within each coalition coordinate their activities, but agents do not coordinate between coalitions. This means partitioning the set of agents into exhaustive and disjoint coalitions. This partition is called a coalition structure (CS); b) *optimization*: solving the optimization problem of each coalition. This means pooling the tasks and resources of the agents in the coalition, and solving this joint problem; and c) *payoff distribution*: dividing the value of the generated solution among agents. Albeit independent of each other, these activities have some interactions. For example, the coalition that an agent wants to join depends on the portion of the value that the agent would be allocated in each potential coalition. This paper focuses on the coalition structure generation in settings where there are too many coalition structures to enumerate and evaluate due to costly or bounded computation and limited time. Instead, agents have to select a subset of coalition structures on which focusing their search.

In this paper we extend the work presented in [3] by adopting a stochastic local search (SLS) procedure [9], named GRASP [4] improved with path-relinking [7], to solve the problem of coalition structure generation in CFGs. The main advantage of using a stochastic local search is to avoid exploring an exponential number of coalition structures providing a near optimal solution. Our algorithm does not provide guarantees about finding the global optimal solution. In particular the questions we would like to pose are: **Q1**) can the metaheuristic GRASP with path-relinking be used as a valuable anytime solution for the CSG problem? In many cases, as in CSG, it is necessary to terminate the algorithm prior to completion due to time limits and to reactivity requirements. In this situation, it is possible to adopt anytime algorithms (i.e. algorithms that may be terminated prior to completion, returning an approximation of the correct answer) whose quality depends on the amount of computation; **Q2**) can the metaheuristic GRASP with path-relinking be adopted for the CSG problem to find optimal solutions faster than the

state of the art exact algorithms? In case of optimization combinatorial problems, stochastic local search algorithms have been proved to be very efficient in finding near optimal solution [9]. In many cases, they outperformed the deterministic algorithms in computing the optimal solution.

The paper is organized as follows: Section 2 introduces some applications sharing a common structure with the CSG problem, Section 3 presents basic concepts regarding the CSG problem, and Section 4 reports the related works on the problem. In Sections 5 and 6 the metaheuristic GRASP and its extension with path-relinking applied to the CSG problem is presented. Section 7 shows some implementation details and Section 9 concludes the paper.

2. Applications

In this section we report some common *grouping problems* in theoretical computer science and artificial intelligence that share a similar structure with the CSG problem.

2.1. Learning Bayesian prototype trees

In [18] the authors present a method to learn Bayesian networks, and specifically Bayesian prototype trees, assuming that the data form clusters of similar vectors. The data can be partitioned into clusters and the maximum likelihood estimates of the partitioning can be computed. A data partition determines the corresponding Bayesian prototype tree model. Hence, given a training data set \mathcal{D} consisting of N data vectors, the authors propose to find the optimal Bayesian prototype tree by finding the optimal partition vector among the N^N different vectors.

2.2. Cluster ensemble problem with graph partitioning

Clustering and graph partitioning are two strongly related concepts. Clustering is a data analysis technique adopted in statistics, data mining, and machine learning communities, and consists in partitioning a set of instances into a given number of groups by optimizing an objective function. Recently, *cluster ensemble* techniques [27] improve clustering performance by generating multiple partitions of the given data set and then combining them to form a superior clustering solution. [27, 5] propose a graph partitioning formulation for cluster ensembles. Given a data set $x = \{X_i, \dots, X_n\}$, a cluster ensemble is a set of clustering solutions, represented as $C = \{C_1, \dots, C_R\}$, where R is the ensemble size. Each clustering solution C_i is a partition of the data set into K_i disjoint clusters. The graph partitioning problem consists in partitioning a weighted graph G into K parts by finding K disjoint clusters of its vertices. G is characterized by the set V of vertices and by a non negative and symmetric similarity matrix W characterizing the similarity between each pair of vertices. The cut of a partition P is defined as $Cut(P, W) = \sum w(i, j)$, where i and j are vertices not belonging to the same cluster. The general goal of graph partitioning is to find a partition that maximizes the cut.

2.3. Aggregation for Natural Language Generation

Aggregation represents a main component of natural language generation systems. The task is to merge two or more linguistic structures into a single sentence. In [2] the authors presented an automatic tool for performing the semantic grouping task by formalizing it as a CSG problem, where each coalition

corresponds to a sentence. The strength of the proposed approach lies in its ability to capture global partitioning constraints by performing collective inference over local pairwise assignments. Pairwise constraints capture the semantic compatibility between pairs of linguistic structures at a local level. The global task is to search a semantic grouping that maximally agrees with the pairwise preferences while simultaneously satisfying constraints on the partitioning as a whole.

2.4. Feature set partitioning and privacy-preserving data mining

The concept of feature set partitioning [24, 23] has been proposed in order to improve the accuracy of supervised learning algorithms. Feature set partitioning methods consist of decomposing the original set of features into several subsets and then combining the classifications made up by all classifier models trained for each subset. Privacy-preserving data mining [28] is a new research area that focuses on preventing privacy violations that might arise during data mining operations. To reach this goal many techniques modify the original datasets in order to preserve privacy even after the mining process is activated ensuring minimal data loss and obtaining qualitative data mining results. In [15] the authors presented a technique considering anonymization for classification through feature set partitioning. Given a learner I , a combination method C , and a training set S with input feature set A , the goal is to find an optimal partitioning of the input feature set A into w mutually exclusive subsets. Optimality is defined in terms of minimization of the generalization error of the classifier combined using the method C .

3. Definitions

In this section we present the basic notions about the CSG problem.

Definition 3.1. (Coalition)

Given a set $N = \{a_1, a_2, \dots, a_n\}$ of n agents ($|N| = n$), called the *grand coalition*, a *coalition* S is a non-empty subset of N , $\emptyset \neq S \subseteq N$.

Definition 3.2. (Coalition Structure)

A *coalition structure* (CS), or *collection*, $\mathcal{C} = \{C_1, C_2, \dots, C_k\} \subseteq 2^N$ is a partition of the set N , and k is its size, i.e. $\forall i, j : C_i \cap C_j = \emptyset$ and $\cup_{i=1}^k C_i = N$.

Given $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$, we define $\cup \mathcal{C} \triangleq \cup_{i=1}^k C_i$. We will denote the collection of all coalition structures of N as $\mathcal{M}(N)$.

As reported in [1], assuming a comparison relation \triangleright , $\mathcal{A} \triangleright \mathcal{B}$ means that the way \mathcal{A} partitions N , where $N = \cup \mathcal{A} = \cup \mathcal{B}$, is preferable to the way \mathcal{B} partitions N .

In this paper we consider the following rules that allow us to transform partitions of the grand coalition. Given a CS $\mathcal{C} = \{C_1, C_2, \dots, C_t\}$:

SPLIT: $C \rightarrow C \setminus \{C_i\} \cup \{C_k, C_h\}$, where $C_k \cup C_h = C_i$, with $C_k, C_h \neq \emptyset$;

MERGE: $C \rightarrow C \setminus \{C_i, C_j\}_{i \neq j} \cup \{C_k\}$, where $C_k = C_i \cup C_j$;

SHIFT: $C \rightarrow C \setminus \{C_i, C_j\}_{i \neq j} \cup \{C'_i, C'_j\}$, where $C'_i = C_i \setminus \{a_i\}$ and $C'_j = C_j \cup \{a_i\}$, with $a_i \in C_i$.

As in common practice [25, 19], we consider coalition structure generation in *characteristic function games* (CFGs). A CFG is a pair (N, v) , where $N = \{a_1, a_2, \dots, a_n\}$ and v is a function $v : 2^N \rightarrow \mathbb{R}$. Given \mathcal{C} a coalition structure, $v(\mathcal{C}) = \sum_{C_i \in \mathcal{C}} v(C_i)$, where $v(C_i)$ is the value of the coalition C_i . Intuitively, $v(C_i)$ represents the maximum payoff the members of C_i can jointly receive by cooperating. As in [25], we assume that $v(C_i) \geq 0$. In case of negative values, it is possible to normalize the coalition values, obtaining a game strategically equivalent to the original game [10], by subtracting a lower bound value from all coalition values. For CFGs the comparison relation on coalition structures is induced in a canonical way, $\mathcal{A} \triangleright \mathcal{B} \iff v(\mathcal{A}) \triangleright v(\mathcal{B})$.

It is possible to prove the following results:

Lemma 3.3. ([11])

Given two coalition structures $\mathcal{A}, \mathcal{B} \in \mathcal{M}(N)$, with $\mathcal{A} \neq \mathcal{B}$, then \mathcal{A} can be transformed into \mathcal{B} by doing at most $n - 1$ applications of the SPLIT or MERGE rules.

Lemma 3.4. ([11])

Given two coalition structures $\mathcal{A}, \mathcal{B} \in \mathcal{M}(N)$, with $\mathcal{A} \neq \mathcal{B}$, then \mathcal{A} can be transformed into \mathcal{B} by doing at most $n - 1$ applications of the SHIFT rule.

Now we can define the coalition structure problem as follows.

Definition 3.5. (Coalition structure generation problem)

Given a set of agents A , the *coalition structure generation* problem consists in maximizing the social welfare of the agents by finding a coalition structure $C^* = \arg \max_{C \in \mathcal{M}(A)} v(C)$.

Formally, a CSG problem may be formulated as a set partitioning problems (SPP). Let $I = \{1, \dots, m\}$ be a set of objects, and let $\{P_1, \dots, P_n\}$ be a collection of subsets of I , with a cost $c_j \in \mathbb{R}^+$ associated to each subset P_j . Given a $n \times m$ binary matrix $A = \{a_{ij}\}$, where $a_{ij} = 1$ if $i \in P_j$ and $a_{ij} = 0$ otherwise, let $\hat{\mathcal{J}}$ be a solution of SPP represented as the n -dimensional vector $\vec{x} = \langle x_1, \dots, x_n \rangle$ of binary decision variables. An integer programming formulation of the set partitioning problem is

$$z(x) = \max \sum_{j=1}^n c_j x_j$$

$$\text{subject to } \sum_{j=1}^n a_{ij} x_j = 1, i = 1 \dots m.$$

Given n agents, the size of the input to a CSG algorithm is exponential, since it contains the values $v(\cdot)$ associated to each of the $(2^n - 1)$ possible coalitions. Furthermore, the number of coalition structures grows as the number of agents increases and corresponds to $\sum_{i=1}^n Z(n, i)$, where $Z(n, i)$, also known as the Stirling number of the second kind, is the number of coalition structures with i coalitions, and may be computed using the following recurrence: $Z(n, i) = iZ(n - 1, i) + Z(n - 1, i - 1)$, where $Z(n, n) = Z(n, 1) = 1$. As proved in [25], the number of coalition structures is $O(n^n)$ and $\omega(n^{n/2})$, and hence an exhaustive enumeration becomes prohibitive.

In this paper we focus on games that are neither *superadditive* nor *subadditive* for which the problem of coalition structure generation is computationally complex. Indeed, for superadditive games where $v(S \cup T) \geq v(S) + v(T)$ (meaning any two disjoint coalitions are better off by merging together), and

for subadditive games where $v(S \cup T) < v(S) + v(T)$ for all disjoint coalitions $S, T \subseteq A$, the problem of coalition structure generation is trivial. In particular, in superadditive games, the agents are better off forming the grand coalition where all agents operate together ($C^* = \{A\}$), while in subadditive games, the agents are better off by operating alone ($C^* = \{\{a_1\}, \{a_2\}, \dots, \{a_n\}\}$).

Instances of the CSG problem have been defined using the following distributions, as proposed in [14, 20], for the values of the characteristic function v :

- Uniform (U): $v(C) \sim U(a, b)$ where $a = 0$ and $b = 1$;
- Uniform Scaled (US): $v(C) \sim |C| \cdot U(a, b)$ where $a = 0$ and $b = 1$;
- Normal (N): $v(C) \sim N(\mu, \sigma^2)$ where $\mu = 1$ and $\sigma = 0.1$;
- Normal Scaled (NS): $v(C) \sim |C| \cdot N(\mu, \sigma^2)$ where $\mu = 1$ and $\sigma = 0.1$;
- Normally Distributed (ND): $v(C) \sim N(\mu, \sigma^2)$ where $\mu = |C|$ and $\sigma = \sqrt{|C|}$.

Figure 1 plots the coalition structures' values according to the previous five distributions of the characteristic function for 10 agents. Each graph plots on the x-axis the value of the coalition structures whose cardinality is represented by a point on the y-axis. As we can see, it seems easy to find optimal solutions in the case of normal and uniform distributions, while it becomes more complicated for the case of scaled distributions. For the normal distribution, the optimal solution belongs to the less populated region corresponding to CSs with none or ten coalitions. The same scenario arises for the uniform distribution, although here the region containing the optimal solution is more populated than in the previous case. The CFGs with scaled distributions are very hard to solve since the optimal solution may belong to very populated regions. More formally, given n agents, let \mathbf{c} be the random variable of the value of a CS with k coalitions c_i . For each distribution the expected value of the \mathbf{c} variable may be computed as follows: $\mathbb{E}_U(\mathbf{c}) = \sum_i^k \mathbb{E}(c_i) = k/2$ (maximum when $k = n$, many coalitions); $\mathbb{E}_{US}(\mathbf{c}) = \sum_i^k |c_i|/2$ (maximum with few coalitions where it is more likely to assign a high value to each one); $\mathbb{E}_N(\mathbf{c}) = \sum_i^k 1 = k$ (maximum when $k = n$); $\mathbb{E}_{NS}(\mathbf{c}) = \sum_i^k |C_i|$ (maximum with few coalitions); $\mathbb{E}_{ND}(\mathbf{c}) = \sum_i^k |C_i|$ (maximum with few coalitions).

4. Related Work

Previous works on CSG can be broadly divided into two main categories: exact algorithms that return an optimal solution, and approximate algorithms that find an approximate solution with limited resources.

A deterministic algorithm must systematically explore the search space of candidate solutions. One of the first algorithms returning an optimal solution is the dynamic programming algorithm (DP) proposed in [29] for the set partitioning problem. This algorithm is polynomial in the size of the input ($2^n - 1$) and it runs in $O(3^n)$ time, which is significantly less than an exhaustive enumeration ($O(n^n)$). However, DP is not an anytime algorithm, and has tough memory requirements. Indeed, for each coalition C it computes the tables $t_1(C)$ and $t_2(C)$. It computes all the possible splits of coalition C , assigning to $t_1(C)$ the best split and to $t_2(C)$ its value. In [19] the authors proposed an improved version of the DP algorithm (IDP) performing fewer operations and requiring less memory than DP. IDP, as shown by the authors, is considered one of the fastest available exact algorithms in the literature for computing an optimal solution.

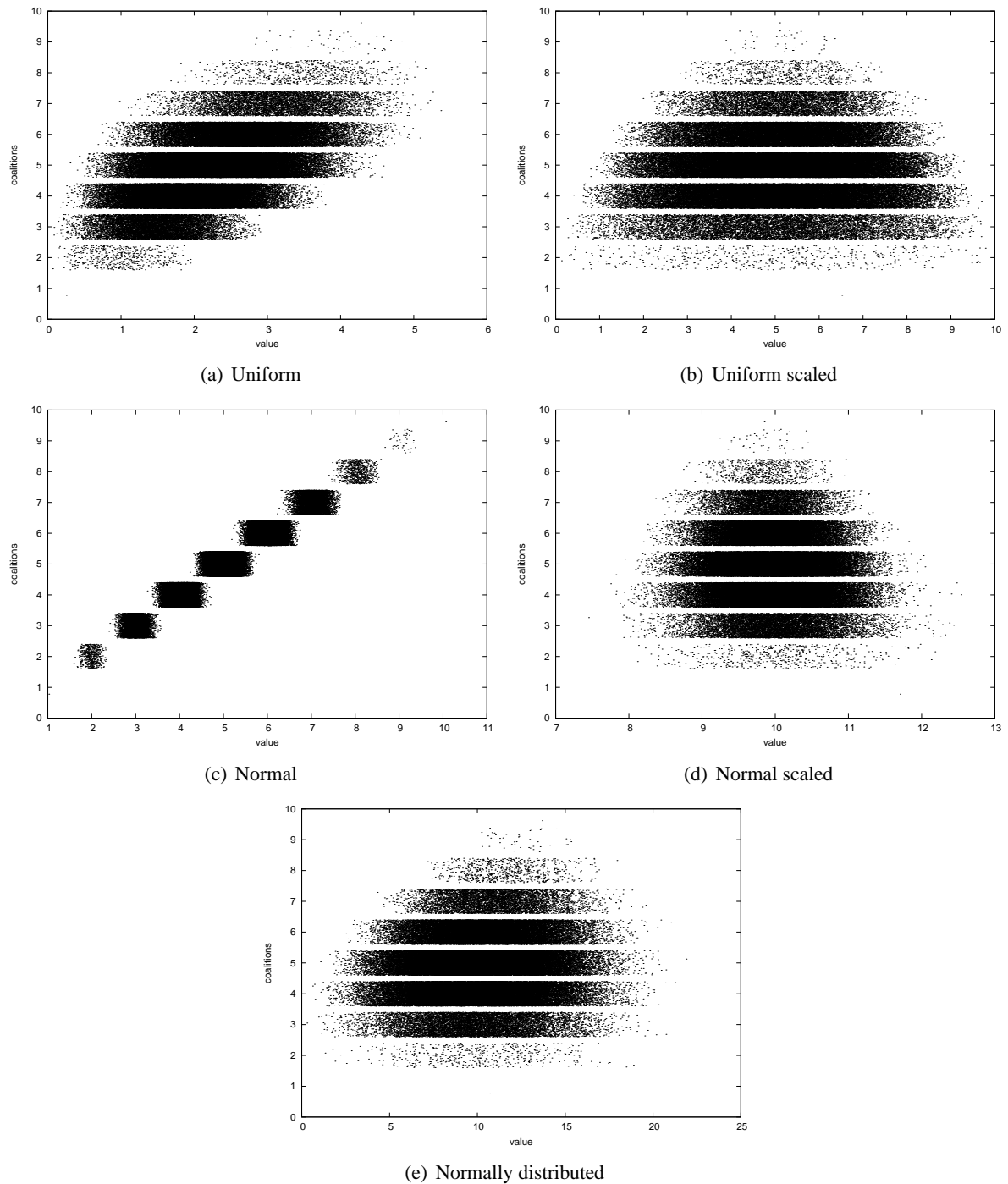


Figure 1. Plots of the coalition structures' values according to the Uniform, Uniform scaled, Normal, Normal scaled, and Normally Distributed distributions of the characteristic function for 10 agents.

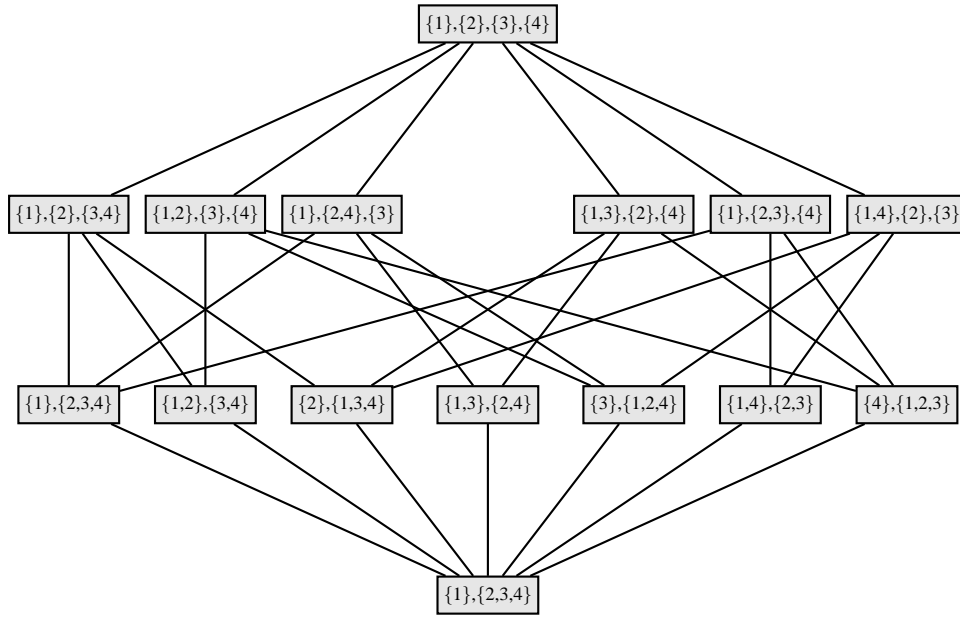


Figure 2. Coalition structure graph for a 4-agent game.

Given a coalition C , $|C| = n$, the number of splits of C into two coalitions C_1 and C_2 , with $|C_1| = s_1$ and $|C_2| = s_2$ is computed as follows

$$S(s_1, s_2) = \begin{cases} \mathbf{C}(s_1 + s_2, s_2)/2 & \text{if } s_1 = s_2 \\ \mathbf{C}(s_1 + s_2, s_2) & \text{otherwise} \end{cases}$$

where $\mathbf{C}(n, k) = \binom{n}{k}$ is the binomial coefficient, i.e. the number of k -combinations from a set made up of n elements. Now, the total number of splits computed by DP is

$$\mathbf{S}_{DP} = \sum_{s=1}^n \mathbf{C}(n, s) \sum_{k=\lceil s/2 \rceil}^{s-1} S(s-k, k),$$

while those computed by IDP are

$$\mathbf{S}_{IDP} = \sum_{s=1}^n \mathbf{C}(n, s) \left(\sum_{k=\lceil s/2 \rceil}^{s-1} S(s-k, k) \mathbf{1}_{\{k \leq n-s \vee s=n\}} \right),$$

where $\mathbf{1}_{\{k \leq n-s \vee s=n\}}$ is 1 if $k \leq n-s$ or $s = n$, 0 otherwise.

Neither DP nor IDP are anytime algorithms, they cannot be stopped before their normal termination. In [25], Sandholm et al. have presented the first anytime algorithm, sketched in Algorithm 1, that can be stopped to obtain a solution (not guaranteed to be optimal) within a given time limit. When not stopped before completion it returns the optimal solution. The CSG process can be viewed as a search in a coalition structure graph as reported in Figure 2. One desideratum is the ability to guarantee that the CS

is within a worst case bound from optimal, i.e. that searching through a subset N of coalition structures, $k = \min\{k'\}$ where $k' \geq \frac{V(S^*)}{V(S_N^*)}$ is finite, and as small as possible, where S^* is the best CS and S_N^* is the best CS that has been seen in the subset N . In [25] it was proved that: a) to bound k , it suffices to search the lowest two levels of the coalition structure graph (with this search, the bound $k = n$, and the number of nodes searched is 2^{n-1}); b) this bound is tight; and c) no other search algorithm can establish any bound k while searching only 2^{n-1} nodes or less.

Algorithm 1 Sandholm et al. algorithm

1. Search the bottom two levels of the coalition structures graph.
 2. Continue with a breadth-first search from the top of the graph as long as there is time left, or until the entire graph has been searched (this occurs when this breadth-first search completes level 3 of the graph, i.e. depth $n-3$).
 3. Return the coalition structure that has the highest welfare among those seen so far.
-

A new anytime algorithm has been proposed in [20], named IP, whose idea is to partition the space of possible solutions into sub-spaces such that it is possible to compute upper and lower bounds on the values of the best CSs they contain. Then, these bounds are used to prune all the sub-spaces that cannot contain the optimal solution. Finally, the algorithm searches through the remaining sub-spaces adopting a branch-and-bound technique avoiding to examine all the solutions within the searched sub-spaces. IP can be used to find optimal coalition structures avoiding to search most of the search space. As reported in [20], IP finds optimal solutions much faster than any previous algorithm designed for this purpose.

As regards the approximate algorithms, [26] proposed a solution based on a genetic algorithm, which performs well when there is some regularity in the search space. Indeed, the authors assume, in order to apply their algorithm, that the value of a coalition depends on other coalitions in the CS, making the algorithm not well suited for the general case. A new solution [11] is based on a Simulated Annealing algorithm [12], a widely used stochastic local search method. At each iteration the algorithm selects a random neighbor solution s' of a CS s . The search proceeds with an adjacent CS s' of the original CS s if s' yields a better social welfare than s . Otherwise, the search is continued with s' with probability $e^{(V(s')-V(s))/t}$, where t is the temperature parameter that decreases according to the annealing schedule $t = \alpha t$.

5. GRASP for the CSG problem

The resource limits posed by some intelligent systems, such as the time for finding a solution, require to have approximation algorithms able to quickly find solutions that are within a specific factor of an optimal solution. In this section we firstly present the anytime algorithm for CSG proposed in [3] with some improvements, and then its extension with path-relinking.

A method to find high-quality solutions for a combinatorial problem is a two-step approach consisting of a greedy construction phase followed by a perturbative¹ local search [9]. Namely, the greedy

¹A perturbative local search changes candidate solutions by modifying one or more of the corresponding solution components.

Algorithm 2 GRASP CSG

Require: v : the characteristic function;
 A : the set of n agents;
maxIter: maximum number of iterations;
neighOp: neighborhood operator;
riiSteps: max non improving search steps for the RII procedure;
wp: RII walk probability

Ensure: solution $\hat{C} \in \mathcal{M}(A)$

- 1: $\hat{C} = \emptyset, v(\hat{C}) = -\infty$
- 2: iter = 0
- 3: **while** iter < maxIter **do**
- 4: $\alpha = \text{rand}(0,1)$;
- 5: $C = \emptyset; i = 0$
- 6: */* construction */*
- 7: **while** $i < n$ **do**
- 8: $\mathcal{S} = \{C' | C' = \text{add}(C, A)\}$
- 9: $\bar{s} = \max\{v(T) | T \in \mathcal{C}\}$
- 10: $\underline{s} = \min\{v(T) | T \in \mathcal{C}\}$
- 11: $\text{RCL} = \{C' \in \mathcal{S} | v(C') \geq \underline{s} + \alpha(\bar{s} - \underline{s})\}$
- 12: randomly select an element C from RCL
- 13: $i \leftarrow i + 1$
- 14: */* local search */*
- 15: $C = \text{RandomisedIterativeImprovement}(C, \text{wp}, \text{riiSteps}, \text{neighOP})$
- 16: **if** $v(C) > v(\hat{C})$ **then**
- 17: $\hat{C} = C$
- 18: iter = iter + 1
- 19: **return** \hat{C}

construction method starts the process from an empty candidate solution and at each construction step adds the best ranked component according to a heuristic selection function. Successively, a perturbative local search algorithm is used to improve the candidate solution thus obtained. Advantages of this search method, over other stochastic local search algorithms, are the much better solution quality and fewer perturbative improvement steps to reach the local optimum. Greedy Randomized Adaptive Search Procedures (GRASP) [4] solve the problem of the limited number of different candidate solutions generated by a greedy construction search methods by randomizing the construction method. GRASP is an iterative process, in which each iteration consists of a construction phase, producing a feasible solution, and a local search phase, finding a local optimum in the neighborhood of the constructed solution. The best overall solution is returned.

Algorithm 2 reports the outline of the GRASP procedure for the CSG problem, denoted in the following with GRASP. In each iteration, it computes a solution C by using a randomized constructive search procedure and then applies a local search procedure to C yielding an improved solution. The main procedure is made up of two components: a constructive phase (lines 7-13) and a local search phase (line 15). The constructive search algorithm used in GRASP iteratively adds a solution component by randomly se-

lecting it, according to a uniform distribution, from a set, named *restricted candidate list* (RCL), of highly ranked solution components with respect to a greedy function $g : C \rightarrow \mathbb{R}$. The probabilistic component of GRASP is characterized by randomly choosing one of the best candidates in the RCL. In our case the greedy function g corresponds to the characteristic function v presented in Section 3. In particular, given v , the heuristic function, and \mathcal{C} , the set of feasible solution components, $\underline{s} = \min\{v(C) | C \in \mathcal{C}\}$ and $\bar{s} = \max\{v(C) | C \in \mathcal{C}\}$ are computed. Then the RCL is defined by including in it all the components C such that $v(C) \geq \underline{s} + \alpha(\bar{s} - \underline{s})$. The parameter α controls the amount of greediness and randomness. A value $\alpha = 1$ corresponds to a greedy construction procedure, while $\alpha = 0$ produces a random construction. As reported in [17], GRASP with a fixed nonzero RCL parameter α is not asymptotically convergent to a global optimum. The solution to make the algorithm asymptotically globally convergent, could be to randomly select the parameter value from the continuous interval $[0, 1]$ at the beginning of each iteration and to use this value during the entire iteration, as we implemented in GRASP.

Given a set of nonempty subsets of n agents $A, C = \{C_1, C_2, \dots, C_t\}$, such that $C_i \cap C_j \neq \emptyset$ and $\cup C \subset A$, the function $add(C, A)$ used in the construction phase returns a refinement C' obtained from C using one of the following operators:

1. $C' \rightarrow C \setminus \{C_i\} \cup \{C'_i\}$ where $C'_i = C_i \cup \{a_i\}$ and $a_i \notin \cup C$, or
2. $C' \rightarrow C \cup \{C_i\}$ where $C_i = \{a_i\}$ and $a_i \notin \cup C$.

Starting from the empty set, in the first iteration all the coalitions containing exactly one agent are considered and the best one is selected for further specialization. At iteration i , the working set of coalitions C is refined by trying to add an agent to one of the coalitions in C or a new coalition containing the new agent is added to C .

5.1. Local search: Randomized Iterative Improvement

To improve the solution generated by the construction phase, a local search is used. It works by iteratively replacing the current solution with a better solution taken from its neighborhood while there is a better solution in the neighborhood. In order to build the neighborhood of a coalition structure C we adopted the previously reported operators SPLIT, MERGE and SHIFT, leading to the following two neighborhood relations:

- $N_{s/m}(\mathcal{C}) = \{C' \in S | s' \in \text{SPLIT}(s) \cup \text{MERGE}(s)\}$
- $N_s(\mathcal{C}) = \{C' \in S | s' \in \text{SHIFT}(s)\}$

In particular, as a local search procedure in GRASP we used a Randomized Iterative Improvement (RII) technique [9], as reported in Algorithm 3. The algorithm starts from the solutions obtained in the constructive phase of GRASP, and then tries to improve the current candidate solution with respect to v . RII uses a parameter $w_p \in [0, 1]$, called walk probability, that corresponds to the probability of performing a random walk step (line 12) instead of an improvement step (lines 14-23). The uninformed random walk randomly selects a solution from the complete neighborhood (line 12). The improvement step randomly selects one of the strictly improving neighbors $I(\mathcal{C})$ (line 16) or a minimally worsening neighbor if the set $I(\mathcal{C})$ is empty (line 18). The search is terminated when a given number of search steps (steps) has been performed without achieving any improvement (line 4).

Algorithm 3 Randomized Iterative Improvement (RII)

Require: \mathcal{C} : candidate solution;
 neighOP: neighborhood operator;
 steps: max non improving search steps;
 wp: walk probability

Ensure: $\hat{\mathcal{C}}$: candidate solution

- 1: improvingSteps = 0
- 2: bestValue = $v(\mathcal{C})$
- 3: $\hat{\mathcal{C}} = \mathcal{C}$
- 4: **while** improvingSteps < steps **do**
- 5: improvingSteps++
- 6: **if** neighOP == SPLIT/MERGE **then**
- 7: compute the neighborhood $N(\mathcal{C}) = N_{s/m}(\mathcal{C})$ of \mathcal{C}
- 8: **else**
- 9: compute the neighborhood $N(\mathcal{C}) = N_s(\mathcal{C})$ of \mathcal{C}
- 10: $u = \text{rand}([0,1])$
- 11: **if** $u \geq \text{wp}$ **then**
- 12: randomly select \mathcal{C}' from $N(\mathcal{C})$
- 13: **else**
- 14: $I(\mathcal{C}) = \{\mathcal{S} \in N(\mathcal{C}) \mid v(\mathcal{S}) < v(\mathcal{C})\}$
- 15: **if** $I(\mathcal{C}) \neq \emptyset$ **then**
- 16: randomly select \mathcal{C}' from $I(\mathcal{C})$
- 17: **else**
- 18: select \mathcal{C}' from $N(\mathcal{C})$ such that $\forall \mathcal{S} \in N(\mathcal{C}) : v(\mathcal{C}') < v(\mathcal{S})$
- 19: $\mathcal{C} = \mathcal{C}'$
- 20: **if** $v(\mathcal{C}) < \text{bestValue}$ **then**
- 21: bestValue = $v(\mathcal{C})$
- 22: improvingSteps = 0
- 23: $\hat{\mathcal{C}} = \mathcal{C}$
- 24: **return** $\hat{\mathcal{C}}$

5.2. GRASP evaluation

Stochastic Local Search algorithms are typically incomplete when applied to a given instance of an optimization problem and the time required for finding a solution may be considered as a random variable [9]. Given an optimization SLS algorithm S for an optimization problem Π and a soluble instance $\pi \in \Pi$, let $P(T_{S,\pi} \leq t, Q_{S,\pi} \leq q)$ denote the probability that S applied to π finds a solution of quality less than or equal to q in time less than or equal to t . The *run-time distribution* (RTD) of S on the specific instance π is the probability distribution of the bivariate random variable $(T_{S,\pi}, Q_{S,\pi})$, characterized by the *run-time distribution function* $rtd : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow [0, 1]$ defined as $rtd(t, q) = P(T_{S,\pi} \leq t, Q_{S,\pi} \leq q)$ [9]. To empirically measure RTDs, let k be the total number of runs performed with a cutoff time t' , and let $k' < k$ be the number of successful runs (i.e., runs during which a solution was found). Let $rt(j)$ denote the run-time for the j th entry in the list of successful runs, ordered by increasing run-times. The

cumulative empirical RTD is then defined by $\hat{P}(T \leq t) := \#\{j | rt(j) \leq t\} / k$.

A CPU time measurement is always based on specific implementations and run-time environments. It is often more appropriate to measure run-time in a way that abstracts from these factors. This can be done using *operation counts*, reflecting the number of operations that are considered to contribute significantly towards an algorithm's performance. Run-time measurements corresponding to actual CPU times and abstract run-times measured in operation counts may be distinguished by referring to the latter as *run-lengths*. We refer to RTDs obtained from run-times measured in terms of operation counts as *run-length distributions* or RLDs [9].

In order to evaluate the proposed algorithms, we implemented them in C language and included the corresponding source code in the ELK system². ELK includes also our implementation of the algorithm proposed by Sandholm et al. in [25], DP [29], IDP [19], and IP [20].

All the following experimental results about the behavior of both GRASP and its extension with path-relinking are obtained executing the algorithms included in ELK on a PC with an Intel(R) Core(TM) i5 CPU 670 @ 3.47GHz and 8GB of RAM, running GNU/Linux kernel 2.6.32-25-server.

The first evaluation, whose results are reported in Figure 3 and Table 1, regards the behavior of GRASP adopting the SPLIT/MERGE ($N_{s/m}$) or the SHIFT (N_s) neighborhood relation in the local search phase. We set the number of agents to 15, the walk probability of the RII procedure to 0.7, and a cutoff run-length to 10^7 operations. In particular, for each instance of the problem we computed the solution quality obtained with GRASP, computed as the ratio between the optimal solution value and the GRASP best solution value. The cutoff run-length limited the number of operations of the GRASP algorithm. In particular, GRASP ends either when the solution quality is 1 or when the number of computed operations is greater than the cutoff run-length. The operations taken into account are the sum of the visited nodes during the construction and local search phase. Hence, `maxiter` is set to $+\infty$ and the GRASP stop criterion is based on the solution quality and number of operations.

We generated 100 problem instances for each distribution type (Uniform, Uniform scaled, Normal, Normal scaled and Normally distributed) of the characteristic function. For each instance 10 different runs of the GRASP algorithm were executed. Figure 3 plots the graphs of the RLD for GRASP about each distribution of the characteristic function. Each graph reports the curves corresponding to the cumulative empirical run-length distribution when the local search uses the SPLIT/MERGE and SHIFT neighborhood operators. As we can see the SPLIT/MERGE neighborhood operator is more robust than the SHIFT operator and permits GRASP to find good solutions more quickly.

An insight of the experiment is reported in Table 1, where some descriptive statistics for the RLDs shown in Figure 3 are indicated. The first column reports the adopted distribution for the characteristic function (Uniform (U), Uniform scaled (US), Normal (N), Normal scaled (NS), and Normally distributed (ND)); the second column indicates the neighborhood operator used; `mean`, `min`, `max` and `stddev` indicate, respectively, the mean, the minimum, the maximum and the standard deviation of the number of operations over the 1000 runs (10 different GRASP execution for problem instance); `vc` denotes the *variational coefficient* ($vc = stddev/mean$); $q_{0.75}/q_{0.25}$ is a quantile ratio; and `#opt` is the number of runs in which GRASP found the optimal solution within the cutoff run-length operations. From these statistics the improvement obtained adopting the SPLIT/MERGE operator in the local search phase becomes more evident.

²ELK is a system including many algorithms for the CSG problem whose source code is publicly available at <http://www.di.uniba.it/~ndm/elk/>.

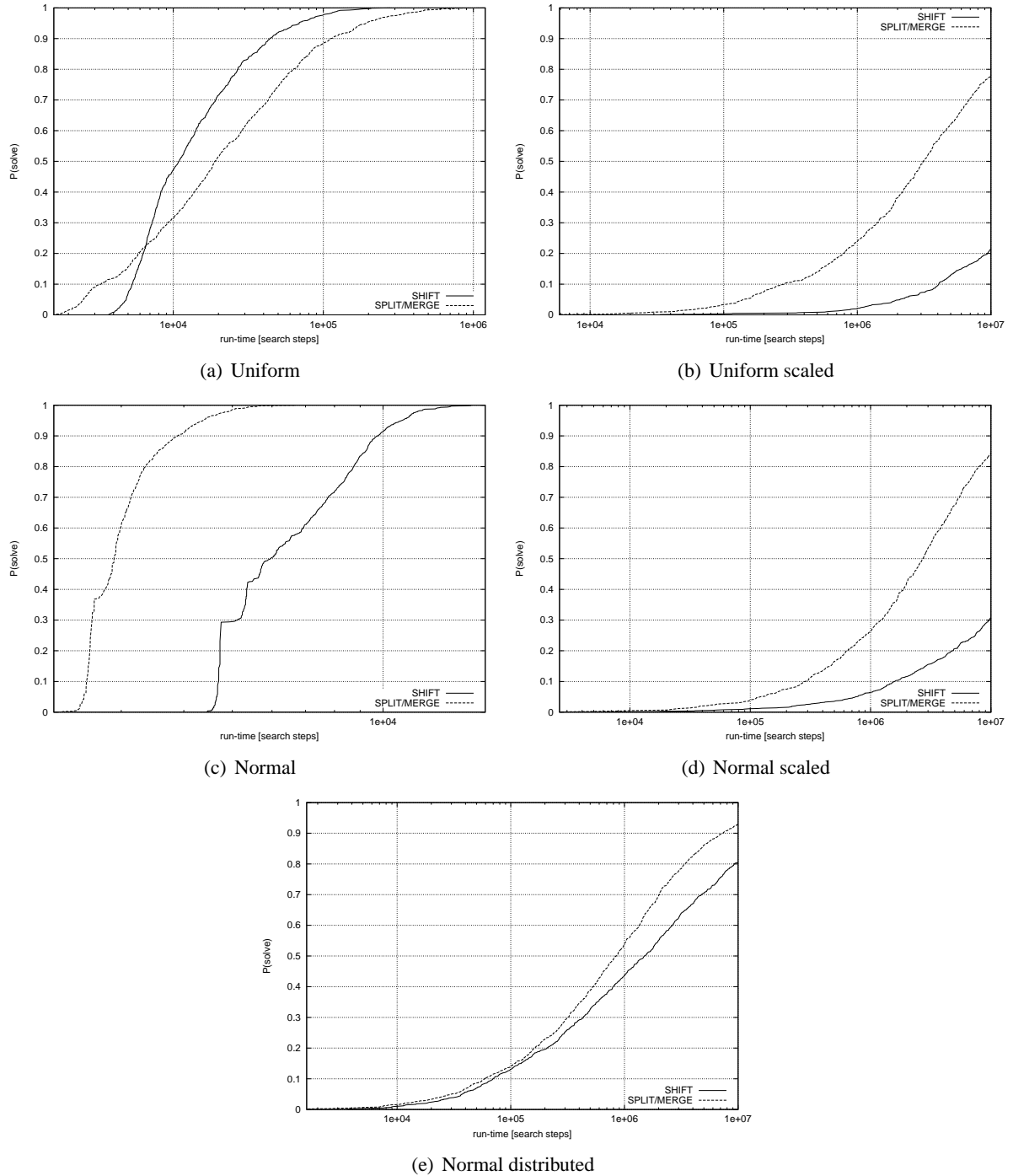


Figure 3. Semi log-plot of RLDs for GRASP, with SHIFT and SPLIT/MERGE neighbourhood operators, applied to 100 Uniform, Uniform scaled, Normal, Normal scaled, Normal distributed CS instances, based on 10 runs per instance.

Dist	Op	mean	min	max	stddev	vc	$q_{0.75}/q_{0.25}$	#opt
U	S	20318.4	3698	279817	26048.4	1.28	3.46	1000
	S/M	48909.2	1658	1127280	89177.1	1.82	6.64	1000
US	S	8844716.9	33292	10006969	2565704.5	0.29	1.00	213
	S/M	4422351.4	5958	10138012	3719544.2	0.84	7.76	781
N	S	6737.1	4438	15017	2134.0	0.32	1.76	1000
	S/M	3057.6	2280	6683	600.5	0.20	1.23	1000
NS	S	8100474.3	6249	10006889	3299610.4	0.41	1.34	307
	S/M	3923760.0	2681	10076751	3496925.3	0.89	7.22	845
ND	S	3461246.2	3637	10006795	3846373.4	1.11	22.78	803
	S/M	2131138.6	1633	10004863	2903255.4	1.36	10.65	929

Table 1. Descriptive statistics for the RLDs shown in Figure 3; $vc = stddev/mean$ denotes the variation coefficient, and $q_{0.75}/q_{0.25}$ the quantile ratio, where q_x denotes the x -quantile.

A second experiment was run in order to evaluate the impact of the walk probability value in the Randomized Iterative Improvement procedure adopted as local search. The corresponding results are plotted in Figure 4. As in the previous experiment we run GRASP 10 times over 100 problem instances for distribution and taking fixed the neighborhood operator to SPLIT/MERGE. The best values are obtained with a walk probability equal to 0.6 or 0.7. There is an increasing improvement for values ranging from 0.2 to 0.6/0.7; then the quality of the found solutions starts decreasing for values ranging from 0.7 to 0.95.

6. GRASP with path-relinking for the CSG problem

Path-relinking is an intensification strategy, proposed in [7], that explores trajectories connecting *elite* solutions obtained by tabu search [8] or scatter search [21].

Given a set of elite solutions, paths among elite solutions in the solution space are generated and traversed hoping to visit better solutions. Paths are generated taking into account the moves incorporating attributes of the guiding solution into the current one. Algorithm 4 reports the path-relinking procedure applied to a pair of solutions x_s (starting solution) and x_t (target solution), assuming that $f(x_s) < f(x_t)$, where f is the heuristic function computing the solution's value.

The algorithm iteratively computes the symmetric difference $\Delta(x, x_t)$ between the current solution x and the target one x_t corresponding to the set of moves needed to reach x_t from x . At each step, the algorithm considers all possible moves $m \in \Delta(x, x_t)$ and selects the one whose result is the least cost solution, i.e. the one which minimizes $f(x \oplus m)$, where $x \oplus m$ is the solution resulting from applying move m to solution x . The best move m^* is made, producing solution $x \oplus m^*$. The algorithm terminates when x_t is reached, i.e. when $\Delta(x, x_t) = \emptyset$, and returns the best solution x^* obtained during the iterations.

Given two elite solutions a and b , some of the alternatives to relink a and b [22, 21] considered in this paper are:

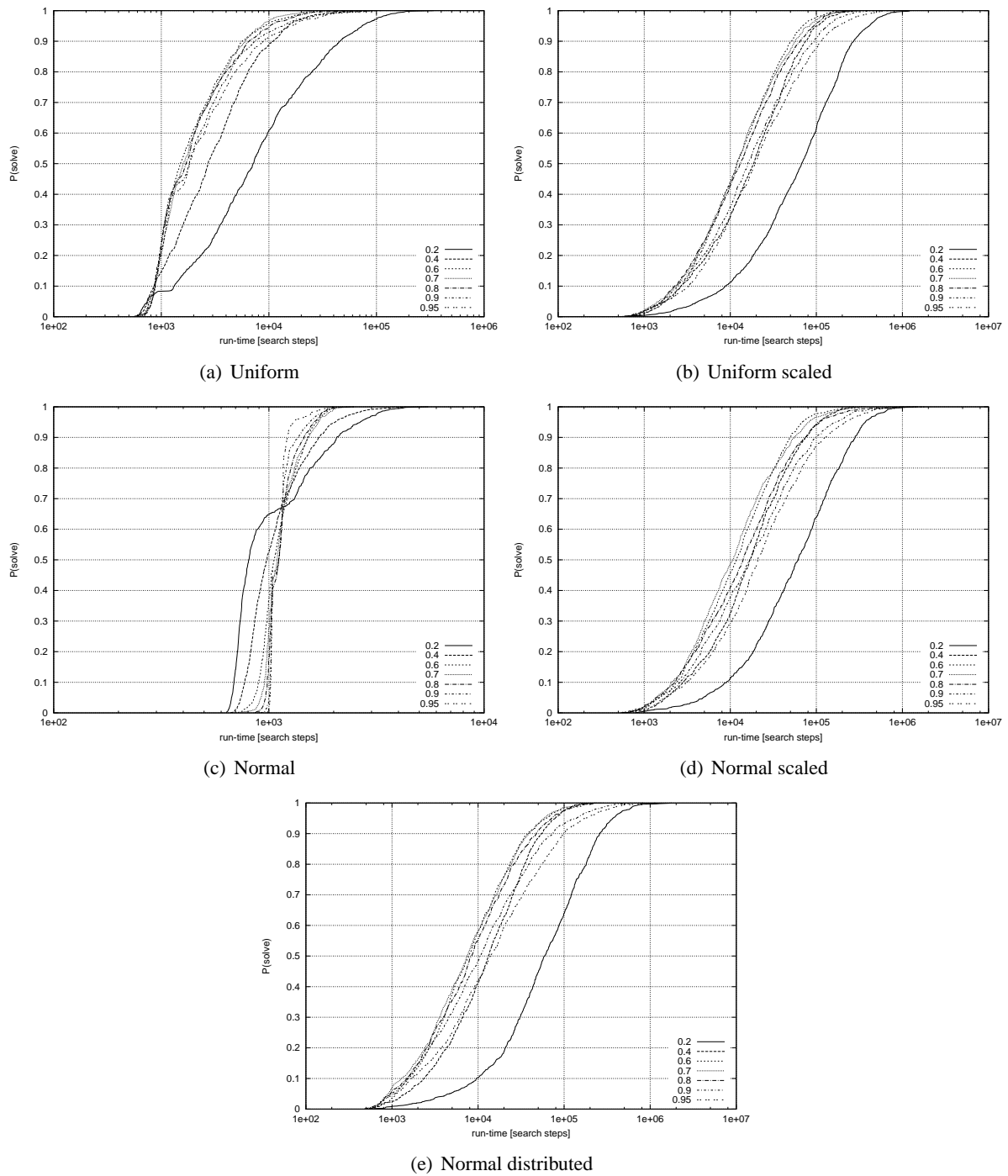


Figure 4. Semi log-plot of RLDs for GRASP with SPLIT/MERGE neighborhood operator, applied to 100 Uniform, Uniform scaled, Normal, Normal scaled, Normal distributed CS instances, based on 10 runs per instance with different values of the walk probability for the RII procedure.

Algorithm 4 PATH-RELINKING**Require:** x_s and x_t : starting and target solution such that $f(x_s) < f(x_t)$ **Ensure:** best solution x^* in path from x_s to x_t

```

1:  $x = x_s$ 
2:  $\mathcal{D} = \Delta(x, x_t)$ 
3: while  $\mathcal{D} \neq \emptyset$  do
4:   select  $m^* \in \mathcal{D}$  such that  $f(x \oplus m^*) < f(x \oplus m_i)$  for all  $m^* \neq m_i \in \mathcal{D}$ 
5:    $\mathcal{D} = \mathcal{D} \setminus \{m^*\}$ 
6:    $x = x \oplus m^*$ 
7:   if  $f(x) < f^*$  then
8:      $f^* = f(x)$ 
9:      $x^* = x$ 
10: return  $x^*$ 

```

- *forward relink*: using $x_s = \min_{a,b}\{f(a), f(b)\}$ and $x_t = \max_{a,b}\{f(a), f(b)\}$;
- *backward relinking*: adopting $x_s = \max_{a,b}\{f(a), f(b)\}$ and $x_t = \min_{a,b}\{f(a), f(b)\}$;
- *back and forward relinking*: both different forward and backward trajectories are explored.

Path-relinking represents a major enhancement to the basic GRASP procedure, leading to significant improvements in solution time and quality, firstly proposed in [13]. The path-relinking intensification strategy adopted in this paper is applied to each local optimum obtained after the local search phase.

The algorithm adopts a pool of `maxElite` elite solutions that is originally empty. Then, each locally optimal solution obtained by a local search is considered as a candidate to be inserted into the pool if it is different from every solution currently contained in the pool. The strategy adopted in this paper is the following. If the pool already contains `maxElite` solutions, then the candidate is inserted into the pool if it is better than the worst of them, that is then removed from the pool. If the pool is not full, the candidate is simply inserted.

Algorithm 5 reports GRASP with path-relinking for the CSG problem, referred in the following as GRASP+PR, where a new step to the construction and local search phase is added. The path-relinking algorithm is applied to the solution returned by local search and to all the solutions from the pool. Improving solutions along the trajectories are considered as candidates for insertion into the pool.

6.1. GRASP with path-relinking evaluation

The first part of the evaluation of GRASP+PR for CSG regarded its effectiveness by varying the relinking strategy and the size of the pool of elite solutions. In the first experiment we investigated the GRASP+PR effectiveness by varying the relinking strategy. Figure 5 plots the graphs corresponding to the RLDs for GRASP+PR adopting the SPLIT/MERGE neighborhood operator and different relinking strategies. The setup of the experiment is the same as that used in the previous GRASP evaluations, reported in Section 5.2: 100 problem instances for each distribution and 10 GRASP+PR runs per instance; the number of agents was set to 15, the walk probability of RII to 0.7, and the cutoff run-length to 10^7 operations. For each instance the solution quality obtained with GRASP+PR has been computed as the ratio between the optimal solution value and the GRASP+PR best solution value. As we can see from the graphs reported

Algorithm 5 GRASP+PR

Require: v : the characteristic function;
 A : the set of n agents;
maxIter: maximum number of iterations;
neighOP: neighborhood operator;
maxElite: max pool dimension;
riiSteps: max non improving search steps for the RII procedure;
wp: RII walk probability

Ensure: solution $\hat{C} \in \mathcal{M}(A)$

- 1: iter = 0
- 2: $P = \emptyset$
- 3: **while** iter < maxiter **do**
- 4: $C = \text{GreedyRandomizedConstruction}()$ /* lines 4-13 of the Alg. 2 */
- 5: $C = \text{RandomizedIterativeImprovement}(C, wp, riiSteps, \text{neighOP})$
- 6: **if** iter \geq 1 **then**
- 7: **for all** $x \in P$ **do**
- 8: determine which (C or x) is the initial and which is the target
- 9: $x_p = \text{PathRelinking}(x_s, x_t)$
- 10: update the elite set with x_p
- 11: **if** $v(x_p) > v(\hat{C})$ **then**
- 12: $\hat{C} = x_p$
- 13: update the elite set P with C
- 14: **else**
- 15: insert C into the elite set P
- 16: iter = iter + 1
- 17: **return** \hat{C}

in Figure 5, the FORWARD/BACKWARD strategy is the most robust for all distributions, while the FORWARD strategy seems to be the less beneficial one. Table 2 reports a descriptive statistics for the search space nodes visited by GRASP+PR with a FORWARD relinking strategy, where #nodes is the total number of visited nodes; #construction, #local and #relink are, respectively, the number of nodes visited in the construction, local search and relinking phase; iter represents the mean value of the iterations required by the GRASP+PR algorithm to end a single run.

In the second experiment on GRASP+PR, we evaluated its effectiveness by varying the size of the pool of elite solutions. Adopting the same setting as for the last experiment, we fixed the FORWARD relinking strategy and let the maxElite parameter range on the values {10, 50, 100, 250, 500}. Figure 6 plots the obtained RLDs showing that, for each distribution, the adoption of a large maxiter value allows GRASP+PR to quickly find the best solution.

The third experiment compared the behavior of GRASP to that of GRASP+PR by considering solution qualities and the runtime performances. The experimental setting is the same: 100 problem instances for each distribution and 10 GRASP+PR runs per instance; the number of agents was set to 15, the walk probability of RII to 0.7, the cutoff run-length to 10^7 operations, the relinking strategy was FORWARD, the neighborhood operator was SPLIT/MERGE, and the pool size of the elite solutions to 10. Figure 7

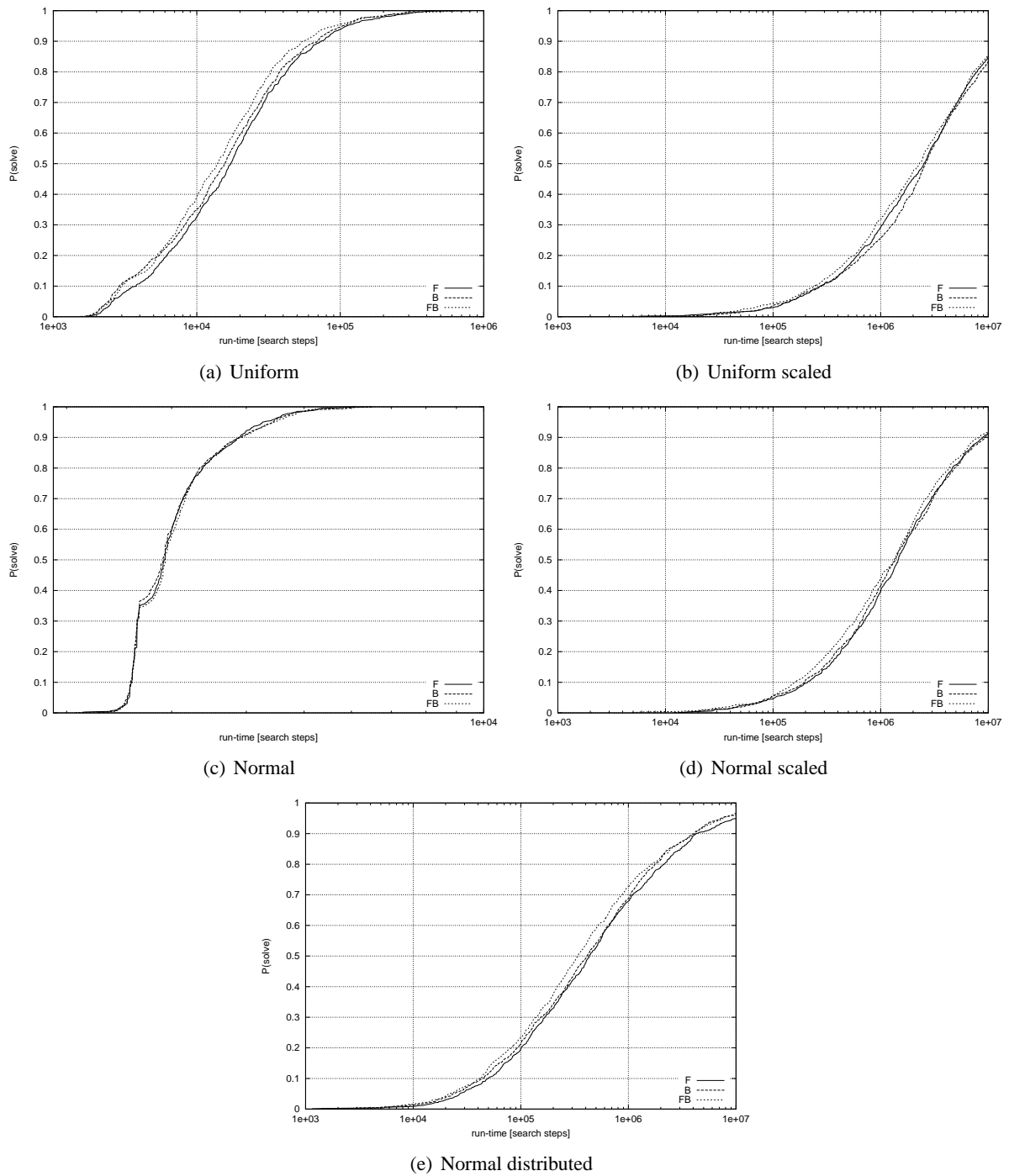


Figure 5. Semi log-plot of RLDs for GRASP-PR, with SPLIT/MERGE neighborhood operator and FORWARD (F), BACKWARD (B), and FORWARD-BACKWARD (FB) relinking strategy, applied to 100 Uniform, Uniform scaled, Normal, Normal scaled, Normal distributed CS instances, based on 10 runs per instance.

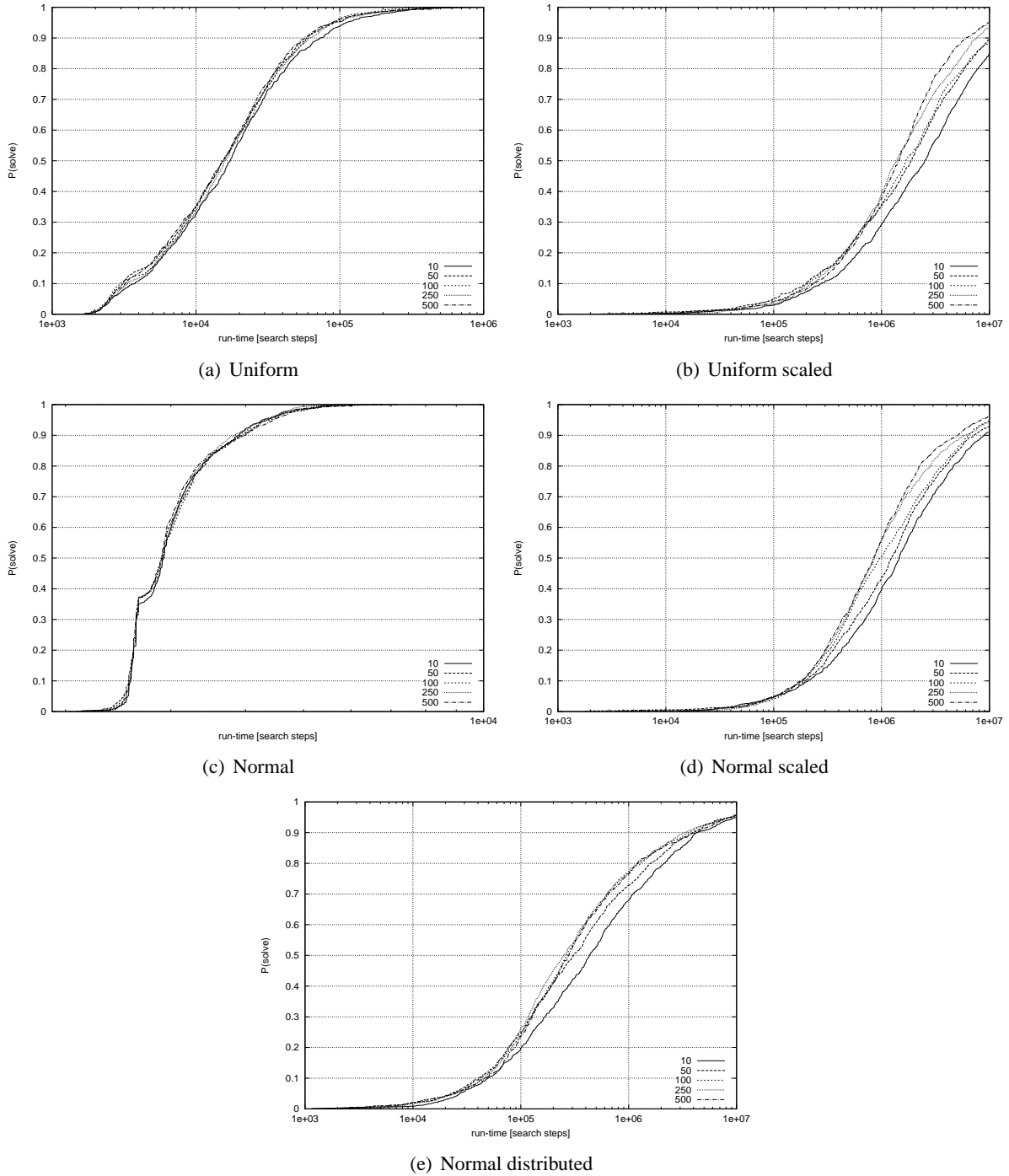


Figure 6. Semi log-plot of RLDs for GRASP+PR with SPLIT/MERGE neighbourhood operator and FORWARD relinking strategy, applied to 100 Uniform, Uniform scaled, Normal, Normal scaled, Normal distributed CS instances, based on 10 runs per instance. Curves refer to different values for the size of the pool of elite solutions.

Dis.	#nodes	#construction	#local	#relink	iter
U	33109.5	6745.0 (20.4%)	25578.4 (77.2%)	786.0 (2.4%)	13.2
US	3789510.5	363175.4 (9.6%)	3334008.3 (88.0%)	92326.8 (2.4%)	893.3
N	3061.9	605.2 (19.8%)	2456.7 (80.2%)	0 (0.0%)	1
NS	2732970.0	350673.3 (12.8%)	2268367.5 (83.0%)	113929.2 (4.17%)	853.0
ND	1528595.9	301562.6 (19.7%)	1105793.2 (72.3%)	121240.1 (7.9%)	699.0

Table 2. Descriptive statistics for the search space nodes visited by GRASP+PR with FORWARD relinking strategy and SPLIT/MERGE neighborhood operator.

Dist.	Alg.	mean	min	max	stddev	vc	$q_{0.75}/q_{0.25}$	#opt
U	G	48909.2	1658	1127280	89177.1	1.82	6.64	1000
	G+PR	33109.5	1697	826931	55954.2	1.69	4.49	1000
US	G	4422351.4	5958	10138012	3719544.2	0.84	7.76	781
	G+PR	3789510.5	5579	10130711	3500039.9	0.92	7.45	848
N	G	3057.6	2280	6683	600.5	0.20	1.23	1000
	G+PR	3061.9	2123	6115	585.4	0.19	1.23	1000
NS	G	3923760.0	2681	10076751	3496925.3	0.89	7.22	845
	G+PR	2732969.9	13193	10014158	3048544.5	1.11	6.81	911
ND	G	2131138.6	1633	10004863	2903255.4	1.36	10.65	929
	G+PR	1528595.9	1144	10002688	2548228.5	1.67	12.00	949

Table 3. Descriptive statistics for the RLDs shown in Figure 7; $vc = stddev/mean$ denotes the variation coefficient, and $q_{0.75}/q_{0.25}$ the quantile ratio, where q_x denotes the x -quantile.

plots the obtained RLDs and Table 3 reports the corresponding descriptive statistics.

Finally, even if GRASP+PR is not a complete algorithm returning the optimal solution, we compared it to the two best performing algorithm able to return the optimal solution. Given different numbers of agents, ranging from 10 to 18, we compared GRASP+PR to IDP and IP reporting the time required to find a solution whose value is at least the 95% (resp. 99%) of the optimal coalition structure value. Since GRASP+PR is a stochastic algorithm, for each problem the corresponding value was obtained by running it 10 times and then averaging the values for each run. As reported in Figure 8, where the time in seconds is plotted in a log scale, GRASP+PR outperforms both IDP and IP for non-scaled distributions. Its effectiveness is comparable to that of IDP and IP on scaled distributions. In particular, considering the problems with 18 agents (rightmost points in the graphs) GRASP+PR is always able to obtain a solution whose value is at least 95% of the optimal coalition structure value with a runtime lesser than that required by both IDP and IP. Increasing the accuracy of the required coalition structure value from 95% to 99% the runtime of GRASP+PR is at least comparable to, or better than, that required by IDP and IP. These results prove that for very hard problems GRASP+PR is a valuable choice to find near optimal solutions.

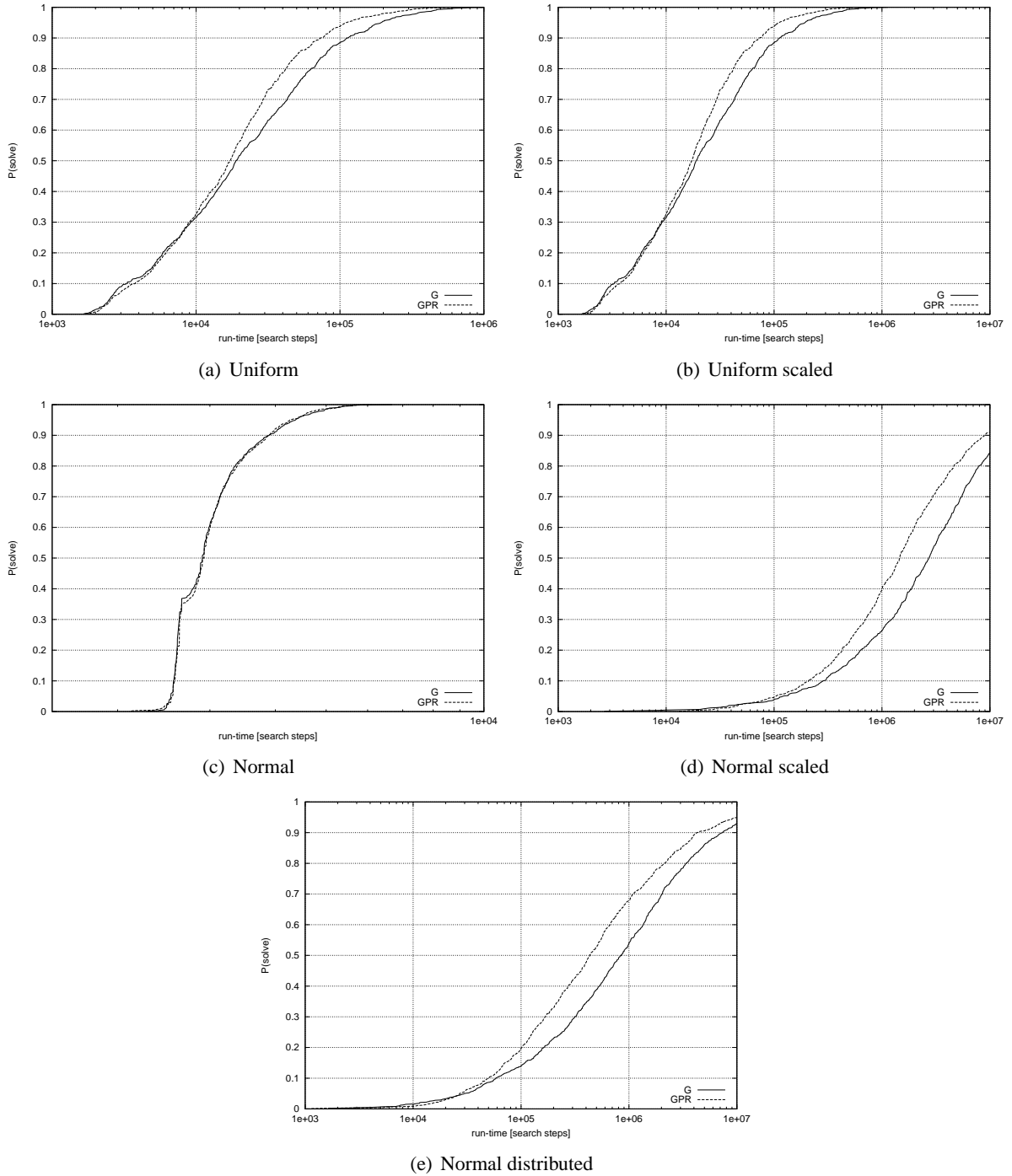


Figure 7. Semi log-plot of RLDs for GRASP and GRASP+PR with FORWARD relinking strategy and SPLIT/MERGE neighborhood operator, applied to 100 Uniform, Uniform scaled, Normal, Normal scaled, Normal distributed CS instances, based on 10 runs per instance.

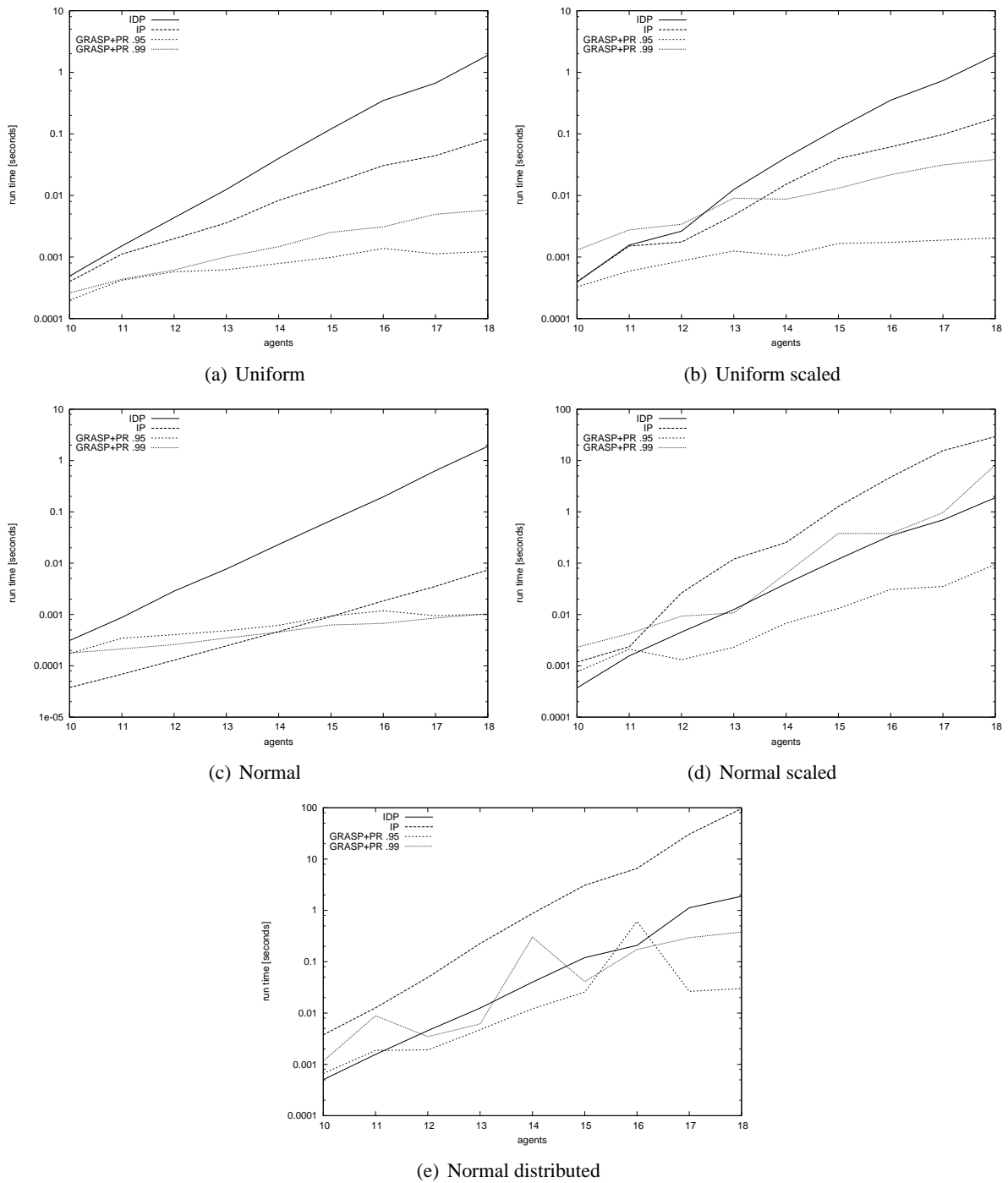


Figure 8. Semi log-plot of run times for GRASP-PR, with SPLIT/MERGE operator and FORWARD relink type, and IP applied to 10 Uniform, Uniform scaled, Normal, Normal scaled, Normal distributed CS instances, based on 10 runs per instance.

7. Implementation details

7.1. The characteristic function

Concerning the representation of the characteristic function and the search space, given n agents $N = \{a_1, a_2, \dots, a_n\}$, we recall that the number of possible coalitions is $2^n - 1$. Hence, the characteristic function $v : 2^n \rightarrow \mathbb{R}$ is represented as a vector CF in the following way. Each subset $S \subseteq A$ (coalition) is described as a binary number $c_B = b_1 b_2 \dots b_n$ where each $b_i = 1$ if $a_i \in S$, $b_i = 0$ otherwise. For instance, given $n = 4$, the coalition $\{a_2, a_3\}$ corresponds to the binary number 0110. Now, given the binary representation of a coalition S , its decimal value corresponds to the index in the vector CF where its corresponding value $v(S)$ is stored. This allows to have random access to the values of the characteristic functions in order to efficiently compute the value v of a coalition structure.

7.2. Coalition structure

Given a coalition structure $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$, assuming that the C_i 's are ordered by their smallest elements, a convenient representation of the CS is an integer sequence $d_1 d_2 \dots d_n$ where $d_i = j$, if the agent a_i belongs to coalition C_j . Such sequences are known as *restricted growth sequences* [16] in the combinatorial literature. The binary representation of coalition C_i is $b_1 b_2 \dots b_n$ where $b_j = 0$ if $d_j \neq i$, and $b_j = 1$ otherwise.

For instance, the sequence corresponding to the coalition structure $\mathcal{C} = \{C_1, C_2, C_3\} = \{\{1, 2\}, \{3\}, \{4\}\}$ is 1123. Now, in order to compute $v(\mathcal{C})$, we have to solve the sum $v(C_1) + v(C_2) + v(C_3)$, where C_1 corresponds to the binary number 1100, C_2 corresponds to the binary number 0010, and C_3 corresponds to the binary number 0001. Hence, $v(\mathcal{C}) = v(C_1) + v(C_2) + v(C_3) = \text{CF}[1100_2] + \text{CF}[0010_2] + \text{CF}[0001_2] = \text{CF}[12] + \text{CF}[2] + \text{CF}[1]$, where CF is the vector containing the values of the characteristic function.

8. Feature set partitioning as coalition structure generation: an example

Feature set partitioning [24, 23] is a generalization of the task of feature selection that aims to partition the feature set into subsets of features that are collectively useful. Here we give an illustrative example on how the feature set partitioning problem can be solved using the ELK system.

In a classification problem we have a training set of instances $D = \{(\mathbf{x}_i, y_i)\}_{i=1, \dots, N}$, where $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})$ is the vector of the values assigned to the features A and y_i denotes the class value. The projection of an instance \mathbf{x}_i (resp. a training set S) onto a subset of features $G \subseteq A$ may be denoted as $\pi_G \mathbf{x}_i$ (resp. $\pi_G S$).

Solving the problem of partitioning an input feature set corresponds to finding the best partitioning such that the combination of the classifiers trained on each feature subset has the highest possible accuracy. More formally:

Definition 8.1. (Feature set partitioning)

Given a learning algorithm L , a combination method C , and a training set S defined over a set of input features A and a target feature y , the *feature set partitioning problem* corresponds to finding an optimal partitioning $Z = \{G_1, \dots, G_m\}$ of the input feature set A into m mutually exclusive subsets $G_i \subseteq A$. The optimality is defined in terms of minimization of the generalization error of the induced classifiers $L(G_i)$ combined using method C , where $L(G_i)$ denotes the classifier induced on the projection $\pi_{G_i} S$.

For instance, assuming the naive Bayes as a combination method, a classification of a new instance is based on the product of the conditional probability of the target feature, given the values of the input features in each subset:

$$l_{\text{MAP}}(\mathbf{x}_i) = \arg \max_{c_j} P_{L(S)}(y_i = c_j) \prod_{k=1}^m \frac{P_{L(\pi_{G_k \cup y} S)}(y_i = c_j | \pi_{G_k} \mathbf{x}_i)}{P_{L(S)}(y_i = c_j)}.$$

In terms of coalition structure generation, the problem corresponds to having a characteristic function game (N, v) , where $N = A = \{a_1, a_2, \dots, a_n\}$ are the input features, and v is the function $v : 2^A \rightarrow [0, 1]$ that computes for a subset of the input features $G_i \subseteq A$ the predictive accuracy of a model learned with the projection π_{G_i} . The accuracy may be computed adopting a k-fold cross validation on the projected training set. The only modification that is needed to our algorithm is the construction of the characteristic function v . Now, for each coalition G_i (a subset of the input features) we have to adopt a classifier to compute its predictive accuracy on the training set π_{G_i} .

9. Conclusions

The paper presented an algorithm applicable to cooperative complex problems that require to find an optimal partition, maximizing a social welfare, of a set of entities involved in a system into exhaustive and disjoint coalitions. We present a greedy adaptive search procedure with path-relinking to efficiently search the space of coalition structures of those grouping problems. As reported in the experimental section the proposed algorithm outperforms in some cases the state of the art algorithms in computing optimal coalition structures.

References

- [1] Apt, K. R., Witzel, A.: A Generic Approach To Coalition Formation, *International Game Theory Review*, **11**(3), 2009, 347–367.
- [2] Barzilay, R., Lapata, M.: Aggregation via set partitioning for natural language generation, *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, Association for Computational Linguistics, 2006.
- [3] Di Mauro, N., Basile, T. M. A., Ferilli, S., Esposito, F.: Coalition Structure Generation with GRASP, *Proceedings of the 14th International Conference on Artificial Intelligence: Methodology, Systems, and Applications* (D. Dicheva, D. Dochev, Eds.), 6304, Springer, 2010.
- [4] Feo, T. A., Resende, M. G. C.: Greedy randomized adaptive search procedures, *Journal of Global Optimization*, **6**, 1995, 109–133.
- [5] Fern, X. Z., Brodley, C. E.: Solving cluster ensemble problems by bipartite graph partitioning, *Proceedings of the 21st International Conference on Machine Learning* (C. E. Brodley, Ed.), 2004.
- [6] Garey, M. R., Johnson, D. S.: *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman & Co., 1990.
- [7] Glover, F.: Tabu search and adaptive memory programming Advances, applications and challenges, *Interfaces in Computer Science and Operations Research*, Kluwer, 1996.

- [8] Glover, F., Laguna, M.: *Tabu Search*, Kluwer, 1997.
- [9] Hoos, H., Stützle, T.: *Stochastic Local Search: Foundations & Applications*, Morgan Kaufmann Publishers Inc., 2004.
- [10] Kahan, J. P., Rapoport, A.: *Theories of Coalition Formation*, Lawrence Erlbaum Associates Publisher, 1984.
- [11] Keinänen, H.: Simulated Annealing for Multi-agent Coalition Formation, *Proceedings of 3rd International KES Symposium on Agents and Multi-agent Systems Technologies and Applications*, Springer, 2009.
- [12] Kirkpatrick, S., Gelatt Jr., C. D., Vecchi, M. P.: Optimization by Simulated Annealing, *Science*, **220**(4598), 1983, 671–680.
- [13] Laguna, M., Mart, R.: GRASP and path relinking for 2-layer straight line crossing minimization, *INFORMS Journal on Computing*, **11**, 1999, 44–52.
- [14] Larson, K. S., Sandholm, T. W.: Anytime coalition structure generation: an average case study, *Proceedings of the third annual conference on Autonomous Agents*, ACM, 1999.
- [15] Matatov, N., Rokach, L., Maimon, O.: Privacy-preserving data mining: A feature set partitioning approach, *Information Sciences*, **180**, 2010, 2696–2720.
- [16] Milne, S. C.: Restricted Growth Functions, Rank Row Matchings of Partitions Lattices, and q-Stirling Numbers, *Advances in Mathematics*, **43**, 1982, 173–196.
- [17] Mockus, J., Eddy, E., Mockus, A., Mockus, L., Reklaitis, G. V.: *Bayesian Heuristic Approach to Discrete and Global Optimization*, Kluwer Academic Publishers, 1997.
- [18] Myllymaki, P., Tirri, H.: Constructing Computationally Efficient Bayesian Models via Unsupervised Clustering, *Probabilistic Reasoning and Bayesian Belief Networks* (A. Gammerman, Ed.), Alfred Waller Publishers, 1995.
- [19] Rahwan, T., Jennings, N. R.: An improved dynamic programming algorithm for coalition structure generation, *Proceedings of 7th International Conference on Autonomous Agents and Multiagent Systems*, 2008.
- [20] Rahwan, T., Ramchurn, S. D., Jennings, N. R., Giovannucci, A.: An anytime algorithm for optimal coalition structure generation, *Journal of Artificial Intelligence Research*, **34**(1), 2009, 521–567.
- [21] Resende, M. G., Ribeiro, C. C., Glover, F., Mart, R.: Scatter Search and Path-Relinking: Fundamentals, Advances, and Applications, in: *Handbook of Metaheuristics* (F. S. Hillier, M. Gendreau, J.-Y. Potvin, Eds.), vol. 146 of *International Series in Operations Research & Management Science*, Springer US, 2010, 87–107.
- [22] Resende, M. G. C., Ribeiro, C.: GRASP with path-relinking: Recent advances and applications, *Metaheuristics: Progress as Real Problem Solvers*, Springer-Verlag, 2005.
- [23] Rokach, L.: Decomposition methodology for classification tasks – a meta decomposer framework, *Pattern Analysis and Applications*, **9**(2), 2006, 257–271.
- [24] Rokach, L., Maimon, O.: Theory and Application of Attribute Decomposition, *Proceedings of the First IEEE International Conference on Data Mining*, IEEE Computer Society Press, 2001.
- [25] Sandholm, T., Larson, K., Andersson, M., Shehory, O., Tohmé, F.: Coalition structure generation with worst case guarantees, *Artificial Intelligence*, **111**(1-2), 1999, 209–238.
- [26] Sen, S., Dutta, P. S.: Searching for Optimal Coalition Structures, *Proceedings of the 4th International Conference on Multi-Agent Systems*, IEEE Computer Society, 2000.
- [27] Strehl, A., Ghosh, J.: Cluster ensembles - a knowledge reuse framework for combining multiple partitions, *Journal of Machine Learning Research*, **3**, 2003, 583–617.

- [28] Verykios, V. S., Bertino, E., Fovino, I. N., Provenza, L. P., Saygin, Y., Theodoridis, Y.: State-of-the-art in privacy preserving data mining, *Newsletter ACM SIGMOD Record*, **33**, 2004, 50–57.
- [29] Yeh, D. Y.: A dynamic programming approach to the complete set partitioning problem, *BIT*, **26**(4), 1986, 467–474.