A Hybrid Data Mining GRASP with Path-Relinking

Hugo Barbalho[1], **Isabel Rosseti**[1], **Simone L. Martins**[1], **Alexandre Plastino**[1]

[1]Departamento de Ciência de Computação – Universidade Federal Fluminense (UFF)
Niterói – RJ – Brazil

{hbarbalho,rosseti,simone,plastino}@ic.uff.br

Abstract. *The exploration of hybrid metaheuristics — combination of metaheuristics with concepts and processes from other research areas — has been an important trend in combinatorial optimization research. In this work, we developed a hybrid version of the GRASP metaheuristic which incorporates the path-relinking procedure — a memory-based intensification strategy — and a data mining module. Computational experiments showed that employing the combination of path-relinking and data mining allowed GRASP to find better results in less computational time. Another contribution of this work is the application of the path-relinking hybrid proposal for the 2-path network design problem, which improved the state-of-the-art solutions for this problem.*

Keywords*: GRASP, path-relinking, data mining*

Resumo. *A exploração de metaheurísticas híbridas — combinação de metaheurísticas com conceitos e processos de outras áreas — vem sendo uma importante linha de pesquisa em otimização combinatória. Nesse trabalho, desenvolvemos uma versão híbrida da metaheurística GRASP que incorpora a técnica de reconexão por caminhos e um módulo de mineração de dados. Experimentos computacionais mostraram que a combinação da técnica de reconexão por caminhos com mineração de dados contribuiu para que o GRASP encontrasse soluções melhores em um menor tempo computacional. Outra contribuição desse trabalho é a aplicação dessa proposta híbrida ao problema de síntese de redes a 2 caminhos, o que proporcionou melhores soluções para esse problema.*

Keywords*: GRASP, reconexão por caminhos, mineração de dados*

1. Introduction.

Metaheuristics represent an important class of approximate techniques for solving hard combinatorial optimization problems, for which the use of exact methods is impractical. A trend in metaheuristics research is the exploration of hybrid metaheuristics. One kind of such hybrid methods results from the combination of concepts and strategies behind two or more classic metaheuristics. Another kind corresponds to metaheuristics combined with concepts and processes from other research areas responsible for improving the original method. An instance of the latter case is the hybrid version of the GRASP metaheuristic that incorporates a data mining process, called DM-GRASP (Data Mining GRASP) [7].

The GRASP (Greedy Randomized Adaptive Search Procedures) metaheuristic, since it was proposed, has been successfully applied to solve many optimization problems[5].

Data mining refers to the automatic extraction of knowledge from datasets [6]. The extracted knowledge, expressed in terms of patterns or rules, represents important features of the dataset at hand. The hybridization of GRASP with a data mining process was first applied to the set packing problem [8]. The basic hypothesis was that patterns found in good quality solutions could be used to guide the search, leading to a more effective exploration of the solution space. The resulting method, the DM-GRASP metaheuristic, achieved promising results not only in terms of solution quality but also in terms of execution time required to obtain good solutions. Afterwards, the method was evaluated on three other applications: the maximum diversity problem, the server replication for reliable multicast problem [7] and the p-median problem [1], and the results were equally successful.

The first contribution of this work is to show that not only the traditional GRASP metaheuristic but also GRASP procedures improved with the path-relinking heuristic [11] — a memory-based intensification mechanism — can benefit from the incorporation of a data mining procedure to extract patterns of sub-optimal solutions in order to guide more efficiently the search for better solutions.

In this work, we present two path-relinking hybrid strategies, called DM-GRASP-PR and MDM-GRASP-PR, which combine a data mining procedure into the GRASP with path-relinking, and show that these strategies can improve the solution quality and computational time of the original GRASP with path-relinking.

The second contribution is the application of the path-relinking hybrid proposals to solve the 2-path network design problem (2PNDP). This problem has shown to be NP-hard and many applications of this problem can be found in the design of communication networks. GRASP procedures with path-relinking have achieved excellent results for this problem [2]. The computational experiments conducted in this work show that the implemented path-relinking hybrid strategies were able to improve the state-of-the-art solutions for the 2PNDP.

The remaining of this paper is organized as follows. In Section 2, we review the main concepts and the structure of both GRASP metaheuristic and path-relinking strategy. In Section 3, we present the hybrid strategy DM-GRASP-PR developed for the 2PNDP and compare the computational results obtained by this strategy and the traditional GRASP with path-relinking. In Section 4, the strategy MDM-GRASP-PR is described and computational results are presented comparing the DM-GRASP-PR and the MDM-GRASP-PR strategies. Finally, in Section 5, concluding remarks are made.

2. GRASP with path-relinking

GRASP [10] is a metaheuristic already applied successfully to many optimization problems. The first phase of a GRASP iteration is the construction phase, in which a complete solution is built. Since this solution is not guaranteed to be locally optimal, a local search is performed in the second phase. This iterative process is repeated until a termination criterion is met and the result is the best solution found over all iterations.

In the construction phase, the initial solution is the empty set. The components not in the solution are ranked according to a greedy function. The better ranked components form a list and one component is randomly selected from this list and incorporated into the current solution. This process is repeated until the partial solution is completely built. Then, the solution obtained in the construction phase becomes the starting point for the local search phase, in which the neighborhood of the solution is explored.

The GRASP metaheuristic is a memoryless method, because all iterations are independent and no information about the solutions is passed from one iteration to another.Path-relinking [4] is a technique developed to explore possible trajectories connecting high quality solutions obtained by heuristics. The objective of introducing path-relinking to a pure GRASP is to retain previous good solutions and use them as guides in the search of new good solutions [11].

Path-relinking is applied to a pair of solutions $\{s_i, s_g\}$ by starting from the initial solution s_i and gradually incorporating attributes from the guide solution s_g to s_i, until s_i becomes equal to s_g. To use path-relinking within GRASP [11], an elite set is maintained, in which good solutions found in previous GRASP iterations are stored. In this work, path-relinking is performed after each GRASP iteration using a solution from the elite set and a local optimum obtained after the GRASP local search.

3. The Hybrid DM-GRASP-PR Proposal

In this section, we describe the 2-path network design problem and the GRASP with path-relinking procedure developed in [2] to solve this problem. Then we present the DM-GRASP-PR heuristic, which is a hybrid version of the GRASP metaheuristic with path-relinking presented in [2] incorporated with a data mining process.

Let $G = (V, E)$ be a connected undirected graph, where V is the set of nodes and E is the set of edges. A k-path between nodes $s, t \in V$ is a sequence of at most k edges connecting them. Given a non-negative weight function $w : E \rightarrow R_+$ associated with the edges of G and a set D of pairs of origin-destination nodes, the *2-path network design problem* (2PNDP) consists in finding a minimum weighted subset of edges $E' \subseteq E$ containing a 2-path between every origin-destination pair in D. The decision version of the 2PNDP has been proved to be NP-complete by Dahl and Johannessen [3].

3.1. GRASP-PR for 2PNDP

The construction phase of the GRASP with path-relinking heuristic for the 2PNDP algorithm starts with the computation from scratch of a solution x using edge weights w' that are initially equal to the original weights w. The procedure is performed until a 2-path has been computed for every origin-destination pair.

Each solution x may be viewed as a collection of $|D|$ 2-paths. Given any solution x, its neighbor solutions x' may be obtained by replacing any 2-path in x by another 2-path between the same origin-destination pair. The local search phase attempts to improve the solutions built greedily during the construction phase.

In each iteration, the path-relinking is applied to the solution obtained by local search and to a randomly selected solution from the elite pool P twice (one using the latter as the starting solution and the other using the former). The locally optimal solution obtained by local search and the best solutions found along each relinking trajectory are considered as candidates for insertion into P. A solution is inserted in the pool if it is different from all solutions of the pool and its cost is better than the cost of the worst solution of the pool.

3.2. DM-GRASP-PR heuristic

The DM-GRASP is composed of two phases. The first one is called the elite set generation phase, which consists of executing n pure GRASP iterations. The d best obtained solutions compose the elite set. After this first phase, the data mining process is applied to extract patterns from the elite set. The patterns to be mined are sets of elements that frequently appear in solutions from the elite set. This extraction of patterns characterizes a frequent itemset mining application [6]. A frequent itemset mined with support s represents a set of elements that occur in $s\%$ of the elite solutions.

Next, the second phase, called hybrid phase, is performed. Another n slightly different GRASP iterations are executed. In these n iterations, an adapted construction phase starts building a solution guided by a pattern selected from the set of mined patterns. Initially, all elements of the selected pattern are inserted into the partial solution, from which a complete solution will be built executing the standard construction procedure. This way, all constructed solutions will contain the elements of the selected pattern.

In this work, we developed the hybrid procedure DM-GRASP-PR, which incorporates a data mining procedure to the GRASP with path-relinking heuristic (GRASP-PR), in order to show that not only the traditional GRASP metaheuristic but also GRASP procedures improved with the path-relinking heuristic — a memory-based intensification mechanism — can benefit from the incorporation of a data mining procedure.

The useful patterns to be mined are sets of edges that commonly appear in sub-optimal solutions of the 2PNDP. A frequent itemset mined from the elite set with support s represents a set of edges that occur in $s\%$ of the elite solutions. A frequent itemset is called maximal if it has no superset that is also frequent. In order to avoid mining frequent itemsets which are subset of one another, in the DM-GRASP-PR proposal for the 2PNDP, we decided to extract only maximal frequent itemset.

The adapted construction algorithm is quite similar to the GRASP construction phase code with the difference that, we try to construct a 2-path between a pair (a, b) using only the edges from the pattern or the edges already used which had their weight modified to 0. If a 2-path was not found using just these edges, we compute a 2-path starting from the partial solution found so far and using all edges from E.

3.3. Computational Results for DM-GRASP-PR

In this section, the results obtained for GRASP-PR and DM-GRASP-PR are compared. We generated 25 instances similar to the ones generated in [2]. The instances are complete graphs with $|V| \in \{100, 200, 300, 400, 500\}$. The edge costs were randomly generated from the uniform distribution on the interval (0, 10] and $10 \times |V|$ origin-destination pairs were randomly chosen. The algorithms were implemented in C and compiled with gcc 4.4.1. The tests were performed on a 2.4 GHz Intel Core 2 Quad CPU Q6600 with 3 Gbytes of RAM, running Linux Kernel 2.6.24. Both GRASP-PR and DM-GRASP-PR were run 10 times with a different random seed in each run. Each strategy executed 1000 iterations. After having conducted some tuning experiments, we set some parameter values: (d) and (t) were set to 10, and (s) was set to 2.

In Table 1, the results related to the solution quality and computational time are shown. The instances are associated to groups according to the number of vertices $|V|$. The first column presents the group identifier of the instance ax, where $x = |V|$. The second and third columns present the deviation value of the average cost obtained by GRASP-PR and DM-GRASP-PR. The deviation value is computed as follows:

$$dev = \frac{(HeuristicCost - BestCost)}{BestCost} \times 100, \tag{1}$$

where $HeuristicCost$ is the average cost obtained by the heuristic technique and the $BestCost$ is the optimal or best known value for the working instance. The last column shows the percentage difference between the strategies average times, obtained for 10 runs.

The proposed DM-GRASP-PR obtained the best cost values and the best average cost values for all instances. These results show that the proposed DM-GRASP-PR strategy was able to improve all results obtained by GRASP with path-relinking. For all instances, the execution times for DM-GRASP-PR were smaller. The last line of the table presents the average of the percentage differences. We can observe that, on average, DM-GRASP-PR was 20.23% faster than GRASP-PR.

Table 1. GRASP-PR and DM-GRASP-PR quality and time results

Instance group	GRASP-PR	DM-GRASP-PR	Time
a100	0.49	0.0	15.44
a200	0.60	0.0	19.59
a300	0.65	0.0	22.21
a400	0.60	0.0	23.09
a500	0.76	0.0	22.98
Average	0.62	0.0	20.23

There are two main reasons for the faster behavior of DM-GRASP-PR. First, the computational effort of the adapted construction phase is smaller than the original construction, since a smaller set of edges is processed to find a 2-path for each pair. Second, the use of patterns leads to the construction of better solutions which will be input for the local search. This incurs in less effort taken to converge to a local optimal solution.

4. The hybrid MDM-GRASP-PR proposal

In the proposed hybrid DM-GRASP-PR, the data mining procedure is executed just once and at the middle point of the whole process. Although the obtained results were satisfactory, we believe that mining more than once, and as soon as the elite set is stable and good enough, can improve the original DM-GRASP framework. Based on this hypothesis, in this work we also propose and evaluate another version of the DM-GRASP for the 2PNDP, called MDM-GRASP-PR (Multi Data Mining GRASP-PR).

The main idea of this proposal is to execute the mining process: (a) as soon as the elite set becomes stable — which means that no change in the elite set occurs throughout a given number of iterations — and (b) whenever the elite set has been changed and again has become stable. We hypothesize that mining more than once will explore the gradual evolution of the elite set and allow the extraction of refined patterns.

4.1. Computational Results

In this section, we report the computational results obtained for the proposed MDM-GRASP-PR strategy. The 2PNDP instances are the same used in the previous section. The MDM-GRASP-PR was also run 10 times with a different random seed in each run. The number of executed iterations were also 1000. We performed some experiments using three values for the parameter used to define if the elite set is stable: 1%, 3% and 5% of the total number of iterations. We adopted 1% as this value provided the best cost values.

Since, in the previous analysis, the DM-GRASP-PR outperformed GRASP-PR, we decided to compare the MDM-GRASP-PR only with the DM-GRASP-PR strategy. In Table 2, the results related to quality and computational time are shown. MDM-GRASP-PR found 18 better results for best values and DM-GRASP-PR found four. MDM-GRASP-PR found 24 better results for average values and DM-GRASP-PR just one. These results show that the MDM-GRASP-PR proposal was able to improve the results obtained by DM-GRASP-PR.

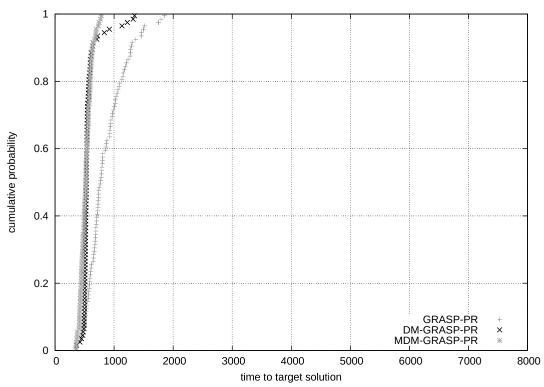
We observed that the DM-GRASP-PR was faster in 18 instances and MDM-GRASP-PR was faster in seven instances. However, we can observe that MDM-GRASP-PR was, on average, just 1.34% slower than DM-GRASP-PR which is not very significant in terms of the heuristic performance. We conclude that both path-relinking hybrid proposals had a similar behavior in terms of computational time.

Table 2. DM-GRASP-PR and MDM-GRASP-PR quality results

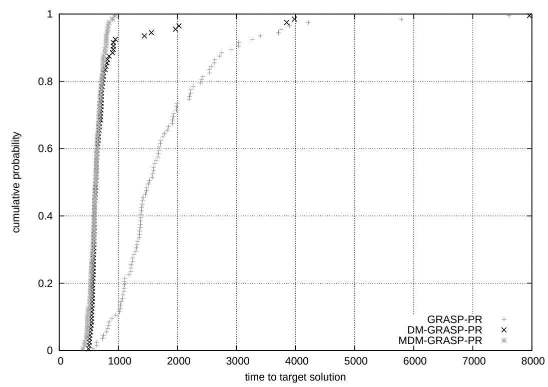
Instance Group	DM-GRASP-PR	MDM-GRASP-PR	Time
a100	0.16	0.0	-3.76
a200	0.19	0.0	-1.17
a300	0.17	0.0	-0.47
a400	0.51	0.0	1.93
a500	0.03	0.004	-3.21
Average	0.21	0.0008	-1.34

In order to verify whether or not the differences of mean values obtained by the strategies presented in Tables 1 and 2 are statistically significant, we employed the unpaired Student's t-test technique. By comparing DM-GRASP-PR with GRASP-PR, we verified that DM-GRASP-PR found better results for all 25 instances and 19 of them are statistically significant, considering a p-value less than 0.01. When comparing MDM-GRASP-PR with GRASP-PR, we verified that MDM-GRASP-PR found better results for all 25 instances and 21 of them are statistically significant. These results show the superiority of the data mining strategies, mainly the good behavior of the MDM-GRASP-PR.

Figures 1(a) and 1(b) show another comparison between the three strategies, based on $Time\text{-}to\text{-}target$ (TTT) plots [9], which are used to analyze the behavior of randomized algorithms. These plots basically show the cumulative probability distributions of running times, i.e., $p(\text{computational\_time} < x)$ vs. x.



(a) An average target for instance a400-100



(b) A difficult target for instance a400-100

Figure 1. Time-to-target plotting

For the average target, we observe in Figure 1(a) that GRASP-PR behaves worst than the two other strategies, and that the MDM-GRASP-PR behaves better than DM-GRASP-PR. We can see that the probability for MDM-GRASP-PR to reach the average target in 800s is 100%, for DM-GRASP-PR is approximately 95% and for GRASP-PR is approximately 58%. For the difficult target, Figure 1(b) shows that MDM-GRASP-PR behaves better than DM-GRASP-PR and both behave better than GRASP-PR. These plots indicate that MDM-GRASP-PR is able to reach difficult solutions faster than DM-GRASP-PR and much faster than GRASP-PR, demonstrating that mining more than once and when the elite set is stable brings robustness to the hybrid strategy.

5. Conclusions

In this work, we proposed to combine a data mining technique into a GRASP metaheuristic with path-relinking in order to show that not only the traditional GRASP can benefit from using patterns to guide the search, but also GRASP improved with the path-relinking heuristic.

The experimental results showed that the first version of the proposed path-relinking hybrid strategy, called DM-GRASP-PR, was able to obtain better solutions in less computational time than the original GRASP with path-relinking developed to solve the 2-path network design problem.

To explore the gradual evolution of the elite set of solutions and allow the extraction of better and higher-quality patterns, we proposed another version of the path-relinking hybrid strategy, called MDM-GRASP-PR. The conducted experiments showed that the MDM-GRASP-PR obtained even better results than the DM-GRASP-PR.

These results showed that incorporating a data mining technique is effective, not only to memoryless heuristics, but also to methods that use exchange of information about obtained solutions like the path-relinking strategy.

6. Comments

This work is part of a research project on hybrid metaheuristics with data mining. The student has developed, under supervision of both supervisors, both DM-GRASP-PR and MDM-GRASP-PR strategies based on the GRASP-PR, implemented in [2]. An extended version of this paper has been submitted to the special issue *GRASP with Path Relinking* of the Computers and Operation Research Journal.

References

[1] A. Plastino, E. R. Fonseca, R. Fuchshuber, S. L. Martins, A. A. Freitas, M. Luis, and S. Salhi, *A hybrid data mining metaheuristic for the p-median problem*, Proceedings of the SIAM International Conference on Data Mining, pp. 305-316, 2009.

[2] C. C. Ribeiro and I. Rosseti, *Efficient parallel cooperative implementations of GRASP heuristics*, Parallel Computing 33 (2007), pp. 21-35.

[3] G. Dahl and B. Johannessen, *The 2-path network problem*, Networks, 43 (2004), pp. 190–199.

[4] F. Glover, M. Laguna, and R. Martí, *Fundamentals of scatter search and path-relinking*, Control and Cybernetics 19, pp. 653–684, 1977.

[5] P. Festa and M. G. C. Resende, *An annotated bibliography of GRASP Part II: Applications*, International Transactions in Operational Research, 16 (2009), pp. 131–172.

[6] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2^{nd} Ed., Morgan Kaufmann Publishers, 2006.

[7] L. F. Santos, S. L. Martins, and A. Plastino, *Applications of the DM-GRASP heuristic: A survey*, International Transactions in Operational Research, 15 (2008), pp. 387–416.

[8] M. H. F. Ribeiro, V. F. Trindade, A. Plastino, and S. L. Martins, *Hybridization of GRASP metaheuristic with data mining techniques*, Proceedings of the ECAI Workshop on Hybrid Metaheuristics, pp. 69-78, 2004.

[9] R. Aiex, M. G. C. Resende, and C. C. Ribeiro, *TTT plots: a perl program to create time-to-target plots*, Optimization Letters, 4 (2007), pp. 355–366.

[10] M. G. C. Resende and C. C. Ribeiro, *Greedy randomized adaptive search procedures*, Handbook of Metaheuristics, Kluwer Academic Publishers, 2003.

[11] M. G. C. Resende and C. C. Ribeiro, *GRASP with path-relinking: Recent advances and applications*, Metaheuristics: Progress as Real Problem Solvers (T. Ibaraki et al. editors), (2005), 29–63.