# Finding communities in the graph of efficient solutions of the bi-objective {0,1}-knapsack problem

Carlos Gomes da Silva(1,2) and João Clímaco(2,3)
(1) School of Technology and Management, Polytechnic Institute of Leiria
Morro do Lena, Alto Vieiro, 2401-951 Leiria, Portugal
Phone: +351 244 820 300, Fax: +351 244 820 310
E-mail: cgsilva@estg.ipleiria.pt
(2) INESC-Coimbra
Rua Antero de Quental, 199
3000-033 Coimbra, Portugal
Phone: +351 239 851 040, Fax: +351 239 824 692
E-mail: jclimaco@inescc.pt
(3) Faculty of Economics, University of Coimbra

**Abstract:** In this paper it is studied the graph generated by the efficient solutions of the bi-objective {0,1}-knapsack problem, trying to decompose it into meaningful communities. The communities can give a better understanding about the entire set of efficient solutions and can disclose the development of new search algorithms for finding efficient solutions. Three types of random instances are used and some interpretation contexts are created.

**Key-Words:** Networks, {0,1}-multiobjective problems, community

# 1 Introduction

In a network (or a graph), by community it is usually meant a subset of nodes highly linked among them, comparing to the links to the rest of the network (Newman and Girvan, 2004 and Donetti and Munõz, 2004). These modules in the network can be useful for better understanding the entire network, especially when it has an intricately complex structure. Identifying such modules can also enable the detection of substructures with particular functions in the system represented by the networks. An additional interest is that the original network can thus be summarized by the interconnection of these fewer components and by this way becomes easier to be analyzed. The interest of identifying communities is present in several diverse areas such as computer science, sociology, biochemistry, taxonomy and the World Wide Web (Newman and Girvan, 2004; Donetti and Muñoz, 2004; Fortunato, 2010). In the business domain, for instance, knowing the communities structures can be helpful for defining client market policies (Du et al., 2007).

This paper is about finding and analysing communities in the set of efficient solutions (that is, the feasible solutions for which there are no other feasible solutions that can improve the value of one criterion without degrading the value of the other) of the bi-objective {0,1}-knapsack problem and aims to bring new insights about the structure of the efficient solutions, namely how they are organized and its relation to the objective space and some known properties. This knowledge could, hopefully, be useful for developing new methods for discovering efficient solutions for this and other combinatorial problems. As in many other areas, it is expected, for instance, that the communities can simplify the original problem, in this case, of finding efficient solutions.

In multiobjective linear programming problems the efficient solutions set is already perfectly characterized, benefiting from the connectedness property, which assures that an efficient solution can be obtained from another one by performing an efficient pivoting (Steuer, 1986). In general, for {0,1}-multiobjective linear problems the above results do not hold (Ehrgott and Klamroth, 1997; Gorski *et al.*, 2006). As a result, the computation of the efficient solutions set of {0,1}-multiobjective linear problems has been a difficult task. In general, only small-medium size instances can be exactly solved by the available traditional exact methods such as dynamic programming or branch-and-bound based algorithms.

A considerable effort has been given to the development of strengthened formulations of sub problems used to find efficient solutions, aiming at using efficiently the well-known properties of the linear programming theory. Despite these efforts, a large amount of computational time is still required, and, for many problems, large size instances remain unsolved in practice.

Heuristics methods face similar problems, since these methods usually pay a significant price in terms of the accuracy of the approximate set of solutions to the exact set of solutions. In some of these type of methods, the improvement of its performance is searched by defining and exploring different neighborhoods of solutions found during the search process (Paquete *et al.*, 2007 and Beausoleil *et al.*, 2009, for example) or known characteristics found in samples of efficient solutions (Gomes da Silva *et al.*, 2006, for example). Another line of research consists of studying the composition of the set of efficient solutions in order to find some sort of rule or constancy which could be used to propose some alternative resolution procedures. In Gomes da Silva *et al.* (2009) this line of research is followed concerning the bi-objective {0,1}-knapsack problem. Converting the efficient solutions into a connected graph, it was found, for example, that the graph of efficient solutions share some properties of the small-world phenomenon: a small shortest path between any pair of nodes, a high clustering coefficient, and a node degree (that is, number of links of a node), approximated by a power law distribution. This research is further explored in the present paper by studying partition of the graphs.

The remainder of the paper is organized as follows. Section 2 presents the bi-objective {0,1}-knapsack problem and how the graph of efficient solutions is built. In Section 3, it is presented a summary of possible methods for finding the communities. Section 4 concerns the computational experiments. Section 5 concludes the paper.

## 2 The bi-objective {0,1}-knapsack problem and the graph of efficient solutions

This paper is about the graphs of efficient solutions of the well-known bi-objective {0,1}-knapsack problem, which can be mathematically formulated as:

$$\begin{aligned} &\max\ z_1(x_1,...,x_j,...,x_n) = \sum_{j=1}^{n} c_j^1 x_j \\ &\max\ z_2(x_1,...,x_j,...,x_n) = \sum_{j=1}^{n} c_j^2 x_j \\ &s.t.: \\ &\sum_{j=1}^{n} w_j x_j \leq W \\ &x_j \in \{0,1\}, j = 1,...,n \end{aligned} \quad (1)$$

where $c_j^i$ represents the value of item $j$ on criterion $i, i = 1,2, x_j = 1$ if item $j$ $(j = 1,...,n)$ is included in the knapsack and $x_j = 0$ otherwise, $w_j$ is the weight of item $j$ and $W$ is the overall knapsack capacity.

Combining a simple mathematical structure with a hard computational resolution complexity, this problem is one of the most preferred for testing new search algorithms in the multiobjective combinatorial optimization field.

Till now, there was not drawn any efficient process for finding the complete set of efficient solutions for medium-large size instances.

The efficient solutions of general random instances of the problem have no specific characteristics. Nevertheless, by imposing additional constraints on the parameters, the solutions have some particular features (Gomes da Silva *et al.*, 2004): 1) if the weights of the items $(w_j)$ are all equal, then all the efficient solutions have the same cardinality, $|x|$, that is, the same number of included items; 2) if $c_j^1 + c_j^2$ is equal for all the items, then the sum of the criteria values of efficient solutions is proportional to its cardinality, leading to a decision space organized by lines such that $z_1(x) + z_2(x) = k\,|x|$. These properties will be used to interpret the meaning of some found communities in the graphs of efficient solutions, built like as in Gomes da Silva *et al.* (2009). In such graphs, a node $n_i$ represents the efficient solution $x^i$ and the edge between nodes $n_i$ and $n_j$ exists if and only if $d(n_i, n_j) = \sum_{k=1}^{n} \left| x_k^i - x_k^j \right| \leq \gamma$. The value $\gamma$ is a constant previously defined, named neighborhood radius, increased from the value 2 until a connected graph is achieved.

## 3 Methods for finding communities

The problem of identifying communities is a hard combinatorial problem (Newman and Girvan, 2004), being available many methods. A recent review of some methods for finding communities can be found in Fortunato (2010). The reviewed methods are aggregated into divisive algorithms, modularity based methods, spectral algorithms, dynamic algorithms, methods based on statistical inference and alternative methods. The recent methods try to go beyond the assumptions of the previous ones, incorporating different concepts of community, intending to improve the performance of the methods over different sets of testing graphs or assure enough flexibility, for instance, by incorporating heuristic procedures, to analyze graphs with a very large number of nodes and edges. For example, the modularity optimization implicit in the method of Newman and Girvan (2004), and one of the most used techniques, was put into question in the identification of small modules in the graphs, depending on the degree of interdependency of the modules, that is, this method has a resolution limit (Fortunato and Barthélemy, 2007). However, it is an optimal method for a specific type of graphs, such as the ones with communities of similar size and nodes degree (Rosvall and Bergstrom, 2007).

In Lancichinetti and Fortunato (2009) it is presented a comparative study of the most significant methods considered representative of the most interesting ideas and techniques developed in the last years.

Considering benchmarks compatible with general systems, where the degree of the nodes follows a power law distribution (Lancichinetti and Fortunato, 2009) tested the performance of several methods. They concluded that the method by Rosvall and Bergstorom (2009), referred to as Infomap, is among the most competitive ones.

This method, with a heuristic nature, is based on the flow of information between the nodes of the graph, which is evaluated by considering random walks in the graph. In short Rosvall and Bergstrom characterize their approach as follows: "We use the probability flows of random walks on a network as a proxy for information flows in the real system and decompose the network into modules by compressing the description of the probability flow. The result is a map that both simplifies and highlights the regularities in the structure and their relationships".

The problem of community discovery is thus associated with the problem of finding the minimum length of information required to characterize the graph, based on the information entropy conceptions. In summary, the method aims at optimizing a function describing the length of that flow, $L(M)$, given by (Rosvall *et al.*, 2009):

$$L(M) = q_{\curvearrowright} H(Q) + \sum_{i=1}^{m} p_{\circlearrowright}^{i} H\left(P^{i}\right) \tag{2}$$

where $M$ is a partition of the nodes, $q_{\curvearrowright} = \sum_{i=1}^{m} q_{i\curvearrowright}$ is the probability that the random walk changes modules on a given step, $q_{i\curvearrowright}$ is the fraction of movements that can occur within module $i$, $H(Q)$ is the entropy of changing modules, given by $-\sum_{i=1}^{m} \frac{q_{i\curvearrowright}}{\sum_{j=1}^{m} q_{j\curvearrowright}} p_{\circlearrowright}^{i} \log\left(\frac{q_{i\curvearrowright}}{\sum_{j=1}^{m} q_{j\curvearrowright}}\right)$, $p_{\circlearrowright}^{i} = q_{i\curvearrowright} + \sum_{\alpha \in i} p_{\alpha}$ is the fraction of within module movements in module $i$ plus the probability of exiting module $i$, $m$ is the number of modules that define the partition $M$, and $H\left(P^{i}\right) = -\frac{q_{i\curvearrowright}}{q_{i\curvearrowright} + \sum_{\beta \in i} p_{\beta}} \log\left(\frac{q_{i\curvearrowright}}{q_{i\curvearrowright} + \sum_{\beta \in i} p_{\beta}}\right) - \sum_{\alpha \in i} \frac{p_{\alpha}}{q_{i\curvearrowright} + \sum_{\beta \in i} p_{\beta}} \log\left(\sum_{\alpha \in i} \frac{p_{\alpha}}{q_{i\curvearrowright} + \sum_{\beta \in i} p_{\beta}}\right)$, meaning the entropy inside the module $i$,

The optimality of the length of the flow is treated heuristically by a greedy search and a simulated annealing method.

As described in Rosvall *et al.* (2009), the main procedure of the method consists of considering each node as a module, and then, in a random sequence, it is verified the benefits for the objective function of joining nodes into modules. The process is repeated until no improvement is possible. Two additional routines are used to increase the accuracy of the partition. In the first, each module is considered as a network and is clustered as above. With the resulting clusters, the main procedure is applied. In the second, each node is candidate to represent its module and the main procedure is applied to the network formed by these nodes. The objective is to try to move single nodes into different modules.

In the following experiments, we also applied this method. Nevertheless, it is also convenient to refer that applying other methods, different communities can be found. This is a subjective issue and there is no universal best method.

Another difficult question is about the interpretation of the communities found. In some cases, namely in biological networks, the communities have a functional specificity, which can be used to interpret them. This is not the case in the majority of the systems. Du *et al.* (2007) considered interpreting the communities by using the information that can be extracted from the nodes defining a community. For each node of a community, a set of attribute can be defined, indeed are inherent to them and the frequency of some of those attributes common to central nodes (for instance, the nodes with the highest degree) and non-central nodes, are used to name the community. For example, it was found that the three largest communities in a telecommunication call network (Du *et al.*, 2007) consist of people with similar attributes: close consumption levels, similar ages or live in the same area. The interpretation of the communities requires thus additional information regarding its nodes.

In this paper, for interpreting the set of efficient solutions we propose taking into account 1)

| | Type 1 | | | | Type 2 | | | |
|---|---|---|---|---|---|---|---|---|
| Statistics | Nodes | Edges | Av.Degree | $\gamma$ | Nodes | Edges | Av.Degree | $\gamma$ |
| *Max* | 177 | 1504 | 10,5 | 5 | 538 | 3853 | 7,2 | 2 |
| *Min* | 73 | 307 | 4,1 | 4 | 225 | 1281 | 5,4 | 2 |
| *Average* | 124,9 | 804,9 | 6,5 | 4,3 | 326,5 | 1978,0 | 6,0 | 2 |
| *Std* | 27,9 | 303,3 | 2,1 | 0,5 | 64,1 | 522,1 | 0,4 | 0 |

Table 1: Graphs structure - Type 1 and Type 2

the objective space; 2) the known characteristics of the efficient solutions (cardinality and the sum of the values of the objective functions).

## 4 Empirical evidence

In the experiments below three types of instances are considered, with the coefficients randomly generated from the uniform distribution within the range [1,100], $U(1, 100)$:

Type 1: $c_j^1, c_j^2, w_j \sim U(1, 100), j = 1, ..., n$ (uncorrelated instances)

Type 2: $c_j^1, c_j^2 \sim U(1, 100), w_j = 100, j = 1, ..., n$ (uncorrelated criteria functions and constant weight)

Type 3: $c_j^1, w_j \sim U(1, 100), c_j^2 = 101 - c_j^1, j = 1, ..., n$ (uncorrelated criteria and weight functions and strongly correlated criteria functions)

In the experiments, the number of variables was set to 100 ($n = 100$) for Type 1 and 2 instances, and 50 ($n = 50$) for Type 3 instances. Thirty instances for types 1 and 2 were used and 14 instances for Type 3. The Infomap method (Section 3) is used to find the communities. The results about the number of communities refer to the best value of the criteria function of that method, after 100 repetitions and where the computational seed was arbitrarily set to 1000.

### 4.1 The graphs of efficient solutions

The obtained graphs of efficient solutions are described in tables 1 and 2, where it is presented some statistics about the number of nodes, edges and the average degree of a node and the radius used to connect the graphs. Note that the number of nodes corresponds to the number of efficient solutions of each instance.

As can be seen, instances Type 1 are more regular in terms of number of nodes and instances Type 2 are more stable in terms of the average degree. In the opposite side, are the instances Type 3, where the number of nodes, number of edges and average degree varies significantly. The configuration of the graphs is also different among the three types of instances. Indeed, the graphs generated by instances Type 1 have a lower number of nodes and edges compared

| | Type 3 | | | |
|---|---|---|---|---|
| Statistics | Nodes | Edges | Av.Degree | $\gamma$ |
| *Max* | 20996 | 1348043 | 317,6 | 6 |
| *Min* | 851 | 19140,0 | 9,3 | 2,0 |
| *Average* | 4891,4 | 432073,8 | 112,6 | 3,8 |
| *Std* | 5224,7 | 428212,9 | 106,2 | 1,1 |

Table 2: Graphs structure - Type 3

with the graphs generated by instances Type 2. And the same when comparing the graphs generated by instances Type 2 and Type 3. The instances Type 3 have a significantly higher number of efficient solutions. This is due to the existence of several alternative solutions for each non-dominated solution.

It can also be referred to that the graphs of instances Type 1 and 3 are connected with a similar average neighborhood radius, larger than the required for connecting solutions of Type 2. Additionally, please note that despite the fact that the neighborhood radius is equal to 2 for all the random instances Type 2, this is not a general property (Gorshi *et al.*, 2006).

## 4.2 The number of communities

The number of communities, the average percentage distribution of the nodes among communities, and the average number of communities that contain at least 70% and 90% of the efficient solutions, are presented in Table 3.

On average, for instances Type 1, less than four communities contain at least 70% of the nodes (solutions) of the graph, with approximately five communities containing at least 90% of the nodes. It can be seen, that the two largest communities contain more than 50% of the nodes. For instances Type 1, the number of communities varies from 3 to 11, being around 6 the average number. The instances Type 2 have a more stable number of communities, varying between 7 and 11. The average is 8 communities. For these instances, achieving 70% and 90% of the nodes requires more communities: 5 and 7 respectively. The graphs related to instances Type 3 are quite diverse. The number of communities goes from 2 till 70, being the average equal to 18. Many of these graphs are organized in such a way, that a single community has at least 70% of the nodes. On average, it is required 8 and 12 communities to incorporate at least 70% and 90% of the nodes.

It was investigated the linear relationship between the number of communities, the number of edges, number of nodes and the average degree of the nodes. Nonetheless, the results showed a weak relationship. Therefore, the number of communities cannot be forecasted from those attributes, at least in a linear form. It seems that the number of communities is grounded in a nonlinear dynamic inside the graphs.

| | | | *Community* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Instances* | *Statistic* | *Communities* | 1 | 2 | 3 | 4 | 5 | 70% | 90% |
| *Type*1 | *Max* | 11 | 56,4 | 38,4 | 27,4 | 18 | 17,2 | 5 | 9 |
| | *Min* | 3 | 17,0 | 15,0 | 13,5 | 7,2 | 3,8 | 2,0 | 3,0 |
| | *Average* | 5,8 | 31,1 | 23,7 | 18,2 | 13,1 | 10,4 | 3,5 | 5,1 |
| | *Std* | 1,9 | 9,9 | 5,9 | 3,4 | 3,1 | 3,1 | 1,0 | 1,5 |
| *Type*2 | *Max* | 11 | 24,7 | 21,7 | 17,5 | 15,5 | 13,7 | 7 | 9 |
| | *Min* | 7 | 14,6 | 13,7 | 11,0 | 9,5 | 8,7 | 4 | 6 |
| | *Average* | 8,8 | 19,5 | 17,0 | 14,6 | 12,3 | 11,0 | 5,2 | 7,3 |
| | *Std* | 1,0 | 2,6 | 2,2 | 1,7 | 1,4 | 1,3 | 0,7 | 0,7 |
| *Type*3 | *Max* | 70 | 99,0 | 22,9 | 22,9 | 13,4 | 6,5 | 26 | 46 |
| | *Min* | 2 | 5,6 | 0,3 | 0,1 | 0,3 | 0,1 | 1 | 1 |
| | *Average* | 17,6 | 51,8 | 9,6 | 5,0 | 5,4 | 4,5 | 7,6 | 12,1 |
| | *Std* | 20,1 | 40,6 | 7,7 | 6,2 | 3,9 | 2,2 | 8,3 | 14,0 |

Table 3: Percentual composition of communities - Type 1 instances

### 4.3 Interpreting the meaning of the communities

How can we interpret the communities found? What do they mean in terms of the problem? Is there any evidence that suggests that the known characteristics about the solutions are related to the communities found?

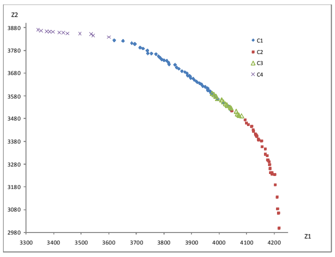

Figure 1: Instance Type 1

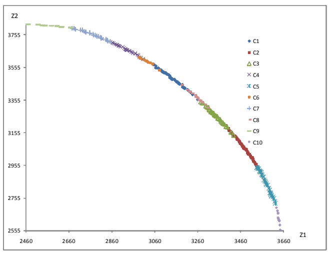

Figure 2: Instance Type 2

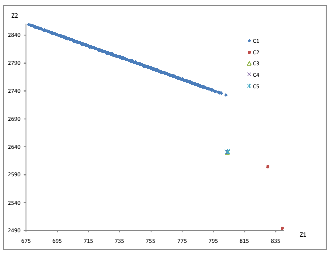

Figure 3: Instance Type 3

As the considered graphs refer to a mathematical problem the interpretation of the communities is not straightforward. We need to have some starting point to interpret the communities.

Let us firstly visualize some examples of the communities found in the previous experiments. Figures 1, 2 and 3 show, in the objective space, the communities of the first instance of Types 1, 2 and 3, respectively, used in the experiments above. For those instances, and as the figures show, there is some clear organization in the objective space. This relation is formally investigated in the following section, as a first context for interpreting the communities. In these cases, the solutions appear to be clustered according to their proximity in the objective space. Additionally, the largest community appears to be placed in the central area of this space and there is some overlapping of the region of some communities, particularly for instances Type 3.

The objective space is the most straightforward way of interpreting the communities, particularly because it has a graphical meaning. This characteristic is in some way distorted in the presence of alternative solutions for the same non-dominated solution. In such cases, the same point in the objective space is associated with different efficient solutions and those belong to possible different communities.

Since some characteristics of the efficient solutions are known, they can also be tested as interpretation contexts. As already referred to, the efficient solutions of types 2 and 3 have some properties, namely the same cardinality (Type 2), and a sum of the criteria values proportional to the cardinality of the solutions (Type 3). Solutions of instances Type 1 have no especial known property. Each of these two characteristics will also be used to interpret the communities.

### 4.3.1 Relation to the objective space

In order to quantify the relationship between the communities and their position in the objective space, we had computed the coefficient of linear correlation between the ranking of the efficient solutions in the objective space and their ranking in the communities. A high coefficient (positive or negative) value means a strong correlation between the two rankings. More formally, the procedure is as follows:

1) The communities are built using the procedure presented in a section above and are named, $C^1, C^2, ...,$ such that, $\min\left\{\psi\left(x^i\right) : x^i \in C^1\right\} \leq \min\left\{\psi\left(x^i\right) : x^i \in C^2\right\} \leq$ ...with $\psi\left(x^i\right)$ being the order of $x^i$ in the objective space (non-decreasing values of $z_1(x)$).

   Within each community, the efficient solutions are also sorted according to non-decreasing values of the community order, $\psi\left(x\right)$.

2) For each efficient solution, $x^i$, two parameters, $u_i$ and $v_i$, are defined. Parameter $u_i$ gives the order of $x^i$ in the objective space and $v_i$ its order in the community.

3) The coefficient of linear correlation between $u$ and $v$, $r_{uv}$, is computed, as a measure of association between the two orders:

   The coefficient $r_{uv}$ is between [-1,1] and $r_{uv} = 1$ or $r_{uv} = -1$ gives an exact linear relation between $u$ and $v$.

As an illustration of this procedure, let us suppose that we have six efficient solutions: $x^1, x^2, x^3, x^4, x^5, x^6$. Besides, also suppose that $z_1\left(x^1\right) \leq z_1\left(x^2\right) \leq z_1\left(x^3\right) \leq z_1\left(x^4\right) \leq z_1\left(x^5\right) \leq z_1\left(x^6\right)$. Additionally, let $C_1 = \left\{x^1, x^2, x^6\right\}, C_2 = \left\{x^3, x^5\right\}, C_3 = \left\{x^4\right\}$. Then, $u_i = i, i = 1, ..., 6; v_1 = 1; v_2 = 2; v_3 = 4; v_4 = 6; v_5 = 5; v_6 = 3$ and $r_{uv} = 0,6$.

In the presence of draws in the ranking, different correlation coefficient can be obtained, being necessary an optimization procedure in order to retrieve the most favorable coefficient value.

The coefficients of linear correlation were computed for each network used in the experiments and are presented in Table 4. The results reveal a great consistency in the values of $r_{uv}$ for instances Type 1 and 2, with the standard deviation values being small. The $r_{uv}$ is very high

| *Statistic* | *Type* 1 | *Type* 2 | *Type* 3 |
|---|---|---|---|
| *Max* | 1,0000 | 1,0000 | 1,0000 |
| *Min* | 0,9749 | 0,9930 | 0,0418 |
| *Average* | 0,9946 | 0,9986 | 0,7841 |
| *Std* | 0,0064 | 0,0020 | 0,3220 |

Table 4: Linear correlation between euclidean and community rankings

and on average is higher than 0,99, both for Types 1 and 2. This is a very strong association between the two orders. As a conclusion, it can be said that the efficient solutions are clustered in the objective space. A denser relationship in the graph of efficient solutions is found between nearer solutions in the objective space. The transition among near solutions in the objective space appears to be also easier in the decision space. These features can justify the interest of heuristics that search for efficient solutions on a basis of proximity of the objective values. For instances Type 3, $r_{uv}$ varies significantly. The lowest value was obtained with the instance with 70 communities. The alternative solutions for the same non-dominated solution were spread in different communities, justifying the small value of $r_{uv}$.

The ideal point in the objective space, that is the vector with coordinates equal to the best value of each objective function, is frequently used as a reference point when dealing with multiobjective problems. The Chebyshev distance, which gives the worst case distance among the values of the objective functions, is particularly used to access the distance between non-dominated solutions (images of efficient solutions in the objective space) and the ideal solution and is here also analyzed for the three types of instances.

Considering a non-dominated solution $z^i = \left(z_1^i, z_1^i\right)$ and the ideal solution by $z^* = \left(z_1^*, z_1^*\right)$, the Chebyshev distance between $z^i$ and $z^*$ is given by $\max\left\{z_1^* - z_1^i; z_2^* - z_2^i\right\}$. Computing the autocorrelation coefficient for each instances of types 1, 2 and 3, we obtained also high coefficients, nevertheless smaller than the ones presented in the Table 6: $0,9043; 0,9707$ and $0,66583$, respectively. This shows a weaker association for the composition of the communities, being thus a less relevant interpretation context.

#### 4.3.2 Relation to the characteristics of the efficient solutions

The relevance of the characteristics of the efficient solutions is assessed as above. The coefficient of correlation between the rankings derived from the characteristics and the community order are presented in Table 5. For Type 2, the $r_{uv}$ was not computed since all the efficient solutions have the same cardinality and as a consequence $r_{uv}$ is necessarily equal to 1.

| | Cardinality versus ranking | | | Sum of criteria versus ranking | | |
|---|---|---|---|---|---|---|
| *Statistic* | *Type* 1 | *Type* 2 | *Type* 3 | *Type* 1 | *Type* 2 | *Type* 3 |
| *Max* | 1,0000 | _ | 1,0000 | 0,9826 | 0,9890 | 1,0000 |
| *Min* | 0,6645 | _ | 0,9794 | 0,7298 | 0,8880 | 0,9794 |
| *Average* | 0,8939 | _ | 0,9967 | 0,9077 | 0,9694 | 0,9967 |
| *Std* | 0,0820 | _ | 0,0065 | 0,0591 | 0,0225 | 0,0065 |

Table 5: Linear correlation between cardinality and sum of criteria versus community rankings

The results show that the cardinality reveals to be a worst context for interpreting the organization of the graph of efficient solutions of Type 1. The average value of $r_{uv}$ is much smaller than the one obtained with the Euclidean ranking. For instances Type 3, the cardinality gives higher and consistent correlation values. The communities revealed thus the known influence that cardinality has in definition of the nature of efficient solutions of these instances.

About the sum ranking, the coefficient of correlation is high for both Types 1 and 2, however, smaller than the one obtained with the Euclidean ranking. As can be easily recognized (in Figures 1 and 3 is presented a typical configuration of the efficient solutions in the objective space), the highest sums are generally obtained with the central solutions in the objective space. As the Figures 1 and 2 show the largest communities appear to be also in the central area. The observation of this fact in the instances of the experiments can be pointed to justify a high coefficient of correlation.

The communities can be interpreted as composed of solutions with the same cardinality and similar sum of the objective functions. Nevertheless, this interpretation is weaker than using the Euclidean ranking in the cases of instances Type 1 and 2.

## 5 Conclusions

To the best of our knowledge, this was the first attempt to find communities within efficient solutions, at least with the described methodology.

The paper showed that the organization of the connected network of efficient solutions is related to the organization of the objective space. By using different rankings contexts it was seen that some were better than others, which brings the problem of proposing the most accurate interpretation of the communities, that is, proposing the most adequate set of context attributes. These results also showed that for the same problem, different structures of communities can be found. Probably, in the future, some short-cuts can be incorporated in search heuristic procedures to find efficient solutions.

When a characteristic is known to be present in the efficient solutions we could observe in the experiments that the community procedure that was used was able to define communities of efficient solutions influenced by the presence of such characteristic. This aspect is obviously interesting for discovering new properties from studying the communities.

The evaluation of the efficient graphs of other {0,1}-multiobjective problems, the proposal of new communities interpretation since different problems may have different rules governing the organization of its efficient graph and the construction of search algorithms inspired on the presented results are future research lines.

## 6 References

Beausoleil, R., Montejo R. (2009) A Study with Neighborhood Searches to deal with Multiobjective Unconstrained Permutation Problems, Journal of Industrial and Management Optimization, 5, 2, pp. 193-216.

Du, N., Wu, B., Pei, X., Wang, B., Xu, L. (2007) Community detection in large-scale social networks, International Conference on Knowledge Discovery and Data Mining, Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis.

Donetti, L., Muñoz, M. (2004) Detecting network communities: a new systematic and efficient algorithm, Journal of Statistical Mechanics: Theory and Experiment, 1742-5468/04/P10012.

Ehrgott, M., Klamroth, K (1997). Connectedness of efficient solutions in multiple criteria combinatorial optimization. European Journal of Operational Research, 97, pp. 159–166.

Fortunato, S., Barthélemy, M. (2007) Resolution limit in community detection, PNAS, 104, pp. 36-41.

Fortunato, S. (2010) Community detection in graphs, Physics Reports, 486, pp. 75 174.

Gomes da Silva, C., Clímaco, J., Figueira, J. (2004) Geometrical configuration of the Pareto frontier of bi-criteria {0,1}-knapsack problems. Research Report, 16, Institute for Systems and Computers Engineering, Coimbra, Portugal.

Gomes da Silva, C., Clímaco, J., Figueira, J. (2006) A scatter search method dedicated to the bi-objective knapsack problems, European Journal of Operational Research, 169, 2, pp. 373-391.

Gomes da Silva, C., Clímaco, J., Almeida Filho, A. (2009) The small world of efficient solutions: empirical evidence from the bi-objective {0,1}-knapsack problem, 4OR, 8, 2, pp.195-211.

Gorski, J., Klamroth, K., Ruzika, S. (2006). Connectedness of efficient solutions in multiple objective combinatorial optimization, Technical Report 102, University of Kaiserslautern, Department of Mathematics.

Paquete, L., Schiavinotto, T., Stützle, T. (2007) On local optima in multiobjective combinatorial problems, Annals of Operations Research, 156, pp. 83-97.

Lancichinetti, A., Fortunato, S. (2009) Community detection algorithms: A comparative analysis, Physical Review E, 80, pp. 056117.1-056117.11.

Newman, M., Girvan, M. (2004) Finding and evaluating community structure in networks, Physical Review E, 69, pp. 026113.1- 026113.15.

Rosval, M., Bergstrom, C. (2007) An information-theoretic framework for resolving community structure in complex networks, PNAS, 104, 18, pp. 7327–7331.

Rosval, M., Bergstrom, C. (2009) Maps of random walks on complex networks reveal community structure, PNAS, 105, 4, pp. 1118-1123.

Rosval, M., Axelsson, D., Bergstrom, C. (2009) The map equation, arXiv: 0906.1405v2.