A Clustering Approach using Weighted Similarity Majority Margins

Raymond Bisdorff[1], Patrick Meyer[23], and Alexandru-Liviu Olteanu[123]

[1] CSC/ILIAS, FSTC, University of Luxembourg
[2] Institut Télécom, Télécom Bretagne, UMR CNRS 3192 Lab-STICC, Technopôle Brest Iroise, CS 83818, 29238 Brest Cedex 3, France
[3] Université Européenne de Bretagne

Abstract. We propose a meta-heuristic algorithm for clustering objects that are described on multiple incommensurable attributes defined on different scale types. We make use of a bipolar-valued dual similarity-dissimilarity relation and perform the clustering process by first finding a set of cluster cores and then building a final partition by adding the objects left out to a core in a way which best fits the initial bipolar-valued similarity relation.

1 Introduction

Clustering is defined as the unsupervised process of grouping objects that are similar and separating those that are not. Unlike classification, clustering has no a priori information regarding the groups to which to assign the objects. It is widely used in many fields like artificial intelligence, information technology, image processing, biology, psychology, marketing and others. Due to the large range of applications and different requirements many clustering algorithms have been developed. Jain [15] gives a thorough presentation of many clustering methods and classifies them into partitioning [20, 19], hierarchical [12, 14, 28], density-based [2, 29], grid-based [1, 26] and model-based methods [8, 18]. New graph-based methods have also been developed in the emerging field of community detection [9, 24, 25]. Fortunato [10] covers many of the latest ones.

In this paper we present the GAMMA-S method (a Grouping Approach using weighted Majority MArgins on Similarities) for clustering objects that are described by multiple incommensurable attributes on nominal, ordinal and/or cardinal scales. We draw inspiration from the bipolar-valued outranking approach proposed by [4–6] for dealing with multiple criteria decision aid problems. As such, we assume that the data is extracted in a prior stage, such that each attribute has a clear meaning and expresses a distinct viewpoint for a human agent. Also, this agent has a clear view on the importance of each attribute when he compares two objects and what can be considered as a discriminating difference in their evaluations. For this we first characterize pairwise global similarity statements by balancing marginal similarity and dissimilarity situations observed at attribute level in order to get majority margins, i.e. a bipolar-valued similarity

graph. Good maximal cliques in this graph, with respect to a fitness measure, are chosen as cluster cores and then expanded to form a complete partition. As the enumeration of all the maximal cliques is well known to be potentially exponential [22], we develop a special meta-heuristic for dealing with the first step. The aim of our method is to achieve a partition that will minimize the differences between the original similarity relation and the relation that is implied by the clustering result.

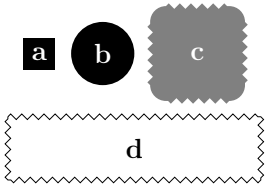
2 Dual similarity-dissimilarity modelling

To illustrate the relational concepts of similarity and dissimilarity we first present a small didactic problem.

Let us consider in Figure 1 a set of objects $\{a, b, c, d\}$ that are described by four attributes, one cardinal and three ordinal. We may notice that objects a,

Fig. 1. Objects' evaluations on the four attributes (left) and their schematic representation (right, smooth (resp. zigzagged) lines representing the smooth (resp. rough) texture).

Attributes	a	b	c	d
Size	2 cm	3 cm	4 cm	20 cm
Texture	Smooth	Smooth	Rough	Rough
Color	Black	Black	Gray	White
Shape	Square	Circle	Rounded Square	Rectangle



b and c are quite small, while d is significantly larger. On the second attribute a and b, as well as c and d have the same texture. On the color attribute we notice some objects are dark, and some are light or we could consider each color level to be different. This can be perceived differently by anyone who looks at these objects. On the last attribute, we have the shapes of each object, and we could consider that object a is different from b but similar to the rest, object b is similar to both a and c but different from d and c is also different from d. None of these objects are similar on all attributes, therefore we could consider two objects to be similar overall if they have similar evaluations on a majority of attributes. For example objects a and b have close evaluations on three out of four attributes, therefore they are considered to be globally *similar*. Objects c and d have also three attributes out of four on which they are similar. But on the first attribute, they show a very large difference in evaluations (4 cm compared to 20 cm). Here, we would rather like to say that *we are not sure* if they are similar or not.

2.1 Pairwise similarity and dissimilarity statements

Let $X = \{x, y, z, ...\}$ denote a set of n objects. Each object $x \in X$ is described on a set $I = \{i, j, k, ...\}$ of m attributes of nominal, ordinal and/or cardinal type, where the actual evaluation x_i may be encoded without loss of generality in the real interval $[m_i, M_i]$ ($m_i < M_i \in \mathbb{R}$). The attributes may not all be of the same significance for assessing the global similarity between the objects. Therefore we assign to the attributes normalized weights $w_i \in [0, 1]$ s.t. $\sum_{i \in I} w_i = 1$, which can be given by the human agent and depend on his knowledge of the problem and his perception of the importance of each attribute in the comparison of the objects.

In order to characterize the *marginal pairwise similarity* and *marginal pairwise dissimilarity* relations between two alternatives x and y of X for each attribute i of I , we use the functions $s_i, d_i : X \times X \to \{-1, 0, 1\}$ defined as follows:

$$s_i(x,y) := \begin{cases} +1 & \text{, if } |x_i - y_i| \leqslant \sigma_i; \\ -1 & \text{, if } |x_i - y_i| \geqslant \delta_i; \\ 0 & \text{, otherwise.} \end{cases} \qquad d_i(x,y) := \begin{cases} -1 & \text{, if } |x_i - y_i| \leqslant \sigma_i; \\ +1 & \text{, if } |x_i - y_i| \geqslant \delta_i; \\ 0 & \text{, otherwise.} \end{cases} \tag{1}$$

where $0 \leq \sigma_i < \delta_i \leq M_i - m_i, \forall i \in I$ denote marginal similarity and dissimilarity discrimination thresholds. These thresholds are parameters which can be fixed by the human agent according to his a priori knowledge on the data and may be constant and/or proportional to the values taken by the objects being compared. If $s_i(x, y) = +1$ (resp. $d_i(x, y) = +1$) we conclude that x and y are similar (resp. dissimilar) on attribute i. If $s_i(x, y) = -1$ (resp. $d_i(x, y) = -1$) we conclude that x and y are not similar (resp. not dissimilar) on attribute i. When $s_i(x, y) = 0$ (resp. $d_i(x, y) = 0$) we are in doubt whether x and y are, on attribute i, to be considered similar or not similar (resp. dissimilar or not dissimilar). Missing values are also handled by giving an indeterminate $s_i(x, y) = 0$, as we cannot state anything regarding this comparison.

The *weighted similarity* and *weighted dissimilarity* relations between x and y, aggregating all marginal similarity statements and all dissimilarity statements are characterized via the functions $ws, wd : X \times X \to [-1, 1]$ defined as follows:

$$ws(x,y) := \sum_{i \in I} w_i \cdot s_i(x,y) \qquad wd(x,y) := \sum_{i \in I} w_i \cdot d_i(x,y) \tag{2}$$

Again, if $0 < ws(x, y) \leqslant 1$ (resp. $0 < wd(x, y) \leqslant 1$) we may assume that it is more sure than not that x is similar (resp. dissimilar) to y; if $-1 \leqslant ws(x, y) < 0$ ($-1 \leqslant wd(x, y) < 0$) we may assume that it is more sure that x is not similar (not dissimilar) to y than the opposite; if, however, $ws(x, y) = 0$ (resp. $wd(x, y) = 0$) we are in doubt whether object x is similar (resp. dissimilar) to object y or not.

Property: The weighted dissimilarity is the *negation* of the weighted similarity relation: $wd = -ws$.

2.2 Taking into account strong dissimilarities

In some cases two objects may be similar on most of the attributes but show a very strong dissimilarity on some other attribute. In this case the objects cannot be considered overall similar or dissimilar. To model this *indeterminate* situation, we define a *marginal strong dissimilarity* relation between objects x and y with the help of function $sd_i : X \times X \to \{0, 1\}$ as follows:

$$sd_i(x, y) := \begin{cases} 1 & \text{, if } |x_i - y_i| \geq \delta_i^+; \\ 0 & \text{, otherwise.} \end{cases} \tag{3}$$

where δ_i^+ is such that $\delta_i < \delta_i^+ \leqslant M_i - m_i$ and represents a strong dissimilarity threshold. Again, this threshold is given by the human agent, in accordance with his experience concerning the underlying problem. If $sd_i(x, y) = 1$ (resp. $sd_i(x, y) = 0$) we conclude that x and y are *strongly dissimilar* (resp. not strongly dissimilar) on attribute i.

We consider that two objects x and y of X, described on a set I of attributes, are *overall similar*, denoted $(x \,\mathrm{S}\, y)$, if

1. a weighted majority of the attributes in I validates a similarity situation between x and y and,
2. there is no marginal strong dissimilarity situation observed between x and y.

We formally characterize the *overall similarity* and *overall dissimilarity* relations by functions $s, d : X \times X \to [-1, 1]$ as follows:

$$s(x, y) := \bigvee \big(ws(x, y), -sd_1(x, y), \cdots, -sd_m(x, y)\big) \tag{4}$$

$$d(x, y) := \bigvee \big(wd(x, y), sd_1(x, y), \cdots, sd_m(x, y)\big) \tag{5}$$

where, for $q \in \mathbb{N}_0$, the epistemic disjunction operator $\bigvee : [-1, 1]^q \to [-1, 1]$ is defined as follows:

$$\bigvee (p_1, p_2, \ldots, p_q) := \begin{cases} \max(p_1, p_2, \ldots, p_q) & \text{, if } p_i \geqslant 0,\ \forall i \in \{1 \ldots q\}; \\ \min(p_1, p_2 \ldots, p_q) & \text{, if } p_i \leqslant 0,\ \forall i \in \{1 \ldots q\}; \\ 0 & \text{, otherwise.} \end{cases} \tag{6}$$

For two given alternatives x and y of X, if $ws(x, y) > 0$ and no marginal strong dissimilarity has been detected, $s(x, y) = ws(x, y)$ and both alternatives are considered as overall similar. If $ws(x, y) > 0$ and a strong dissimilarity is detected we do not state that x and y are overall similar or not, and $s(x, y) = 0$. If $ws(x, y) < 0$ and, a strong dissimilarity is observed, then x and y are certainly not overall similar and $s(x, y) = -1$. Finally, if $ws(x, y) = 0$ is observed conjointly with a strong dissimilarity, we will conclude that x and y are indeed not overall similar and $s(x, y)$ is put to -1.

Property: The overall dissimilarity represents the *negation* of the overall similarity:
$d = -s$.

Following this property, we can now say that two objects which are not similar according to this caracterization can be called dissimilar.

2.3 The Condorcet similarity graph

We call a *Condorcet similarity graph*, denoted $G(X, s^*)$, the three-valued graph associated with the bipolar-valued similarity relation s, where X denotes the set of nodes and function $s^* : X \times X \to \{-1, 0, 1\}$, named *crisp similarity relation*, weights its set of edges as follows:

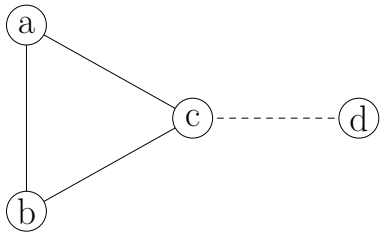
$$s^*(x,y) := \begin{cases} +1 & \text{, if } s(x,y) > 0; \\ -1 & \text{, if } s(x,y) < 0; \\ 0 & \text{, otherwise.} \end{cases} \tag{7}$$

Figure 2 represents on the left the encoding of the attributes on real scales and the corresponding thresholds for the example defined at the beginning of this section. On the right we have the bipolar-valued overall similarity relation s. Notice that a,b and c are more similar than not to each other, whereas d is surely dissimilar both from a and b. Besides, d and c appear to be neither similar nor dissimilar. The corresponding Condorcet similarity graph is shown below. Edges valued by -1 are not represented and the zero-valued one is dashed. As a Condorcet similarity graph is always reflexive, we do not represent the loops on the nodes.

Fig. 2. Encoding of the attributes (left) and bipolar-valued similarity relation with associated Condorcet similarity graph (right)

	Size	Texture	Color	Shape
w	1	1	1	1
a	2	0	0	1
b	3	0	0	3
c	4	1	1	2
d	20	1	2	0
σ	2	0	1	1
δ	4	1	2	2
δ^+	10	-	-	-

s	**a**	**b**	**c**	**d**
a	1.00	**0.50**	**0.50**	-1.00
b	**0.50**	1.00	**0.50**	-1.00
c	**0.50**	**0.50**	1.00	**0.00**
d	-1.00	-1.00	**0.00**	1.00



3 Definition of the clusters

Ideally, a cluster would have all the objects inside it similar to each other and dissimilar from the rest. In graph theory this may be modeled by a maximal

clique, however, we would also need the maximal clique to be totally disconnected from the rest of the graph, which on real data will very rarely be the case. Also there may generally exist a very large number of such maximal cliques, many overlapping one with the other. Moon and Moser have shown that, in the worst case, the number of maximal cliques in a graph can be exponential [22].

Therefore, in a first stage, we propose to select in the Condorcet similarity graph $G(X, s^*)$ the *best* set of maximal cliques on the $+1$ arcs (thus containing objects that are, on a majority, similar to each other), which may be considered as cluster cores. In a second stage, we expand these cores into clusters by adding objects that are well connected to them in such a way that we try to maximize the number similarity arcs inside each cluster, and minimize the number of similarity arcs between the clusters.

Let us introduce several fitness measures we will need in the algorithmic approach. Given a Condorcet similarity graph $G(X, s^*)$ and a set $K \subseteq X$ of objects, we define, for each x of X the crisp similarity majority margin $smm^* : X \times \mathcal{P}(X) \to [-n, n]$ towards the set K:

$$smm^*(x, K) := \sum_{y \in K} s^*(x, y). \tag{8}$$

A large positive value of $smm^*(x, K)$ would show that x is similar to the set K in a consistent manner. A large negative value would mean that x mostly dissimilar from K.

We define the profile of a set K by the set of all similarity majority margins for $x \in X$.

We will consider a cluster to have a strong profile if it contains strongly positive and/or negative similarity majority margins and reflect this using the core fitness function $f_C^* : \mathcal{P}(X) \to [-n^2, n^2]$ defined as:

$$f_C^*(K) := \sum_{x \in X} |smm^*(x, K)|. \tag{9}$$

In Figures 3 and 4 we show how two possible cluster cores could be characterised. The examples show a set of 10 objects $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and a possible cluster core $\{1, 2, 3, 4, 5\}$. On the left we have the crisp similarity relation between the core and all the objects in the example and its similarity majority margins. On the right we show a representation of the Condorcet Similarity Graph.

In the first image $\{1, 2, 3, 4, 5\}$ is well connected to objects 6 and 7, and very well disconnected from the rest of the objects. Objects 6 and 7 are connected to 4 out of 5 objects in the maximal clique and can be added later to eventually form a cluster. This can be regarded as a good cluster core. In the second image the same maximal clique is not consistently connected to all the objects outside. Each of them is connected to either 2 or 3 objects inside the core, and this is reflected in the similarity majority margins which take low positive and negative values. This is regarded as a weaker cluster core.

Fig. 3. Possible good cluster core $K = \{1, 2, 3, 4, 5\}$

$\mathbf{s^*}$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	+1	+1	+1	+1	+1	+1	+1	-1	-1	-1
2	+1	+1	+1	+1	+1	+1	-1	-1	-1	-1
3	+1	+1	+1	+1	+1	-1	+1	-1	-1	-1
4	+1	+1	+1	+1	+1	+1	+1	-1	-1	-1
5	+1	+1	+1	+1	+1	+1	+1	-1	-1	-1
$\mathbf{smm^*}$	**+5**	**+5**	**+5**	**+5**	**+5**	+3	+3	**-5**	**-5**	**-5**

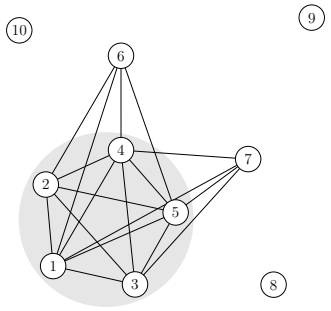
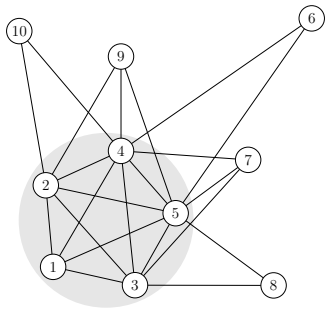


Fig. 4. Possible bad cluster core $K = \{1, 2, 3, 4, 5\}$

$\mathbf{s^*}$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	+1	+1	+1	+1	+1	-1	-1	-1	-1	-1
2	+1	+1	+1	+1	+1	-1	-1	-1	+1	+1
3	+1	+1	+1	+1	+1	-1	+1	+1	-1	-1
4	+1	+1	+1	+1	+1	+1	+1	-1	+1	+1
5	+1	+1	+1	+1	+1	+1	+1	+1	+1	-1
$\mathbf{smm^*}$	**+5**	**+5**	**+5**	**+5**	**+5**	-1	+1	-1	+1	-1



In order to achieve a partitioning of the entire dataset we need to detect well separated maximal cliques that correspond to local maxima of our fitness measure. To find these local maxima, we define the neighborhood of a maximal clique as all the maximal cliques that contain at least one object from it.

Let us define now the fitness an alternative x would have as part of a cluster K through function $f^* : X \times \mathcal{P}(X) \to [-n^2, n^2]$ as:

$$f^*(x, K) := \sum_{y \in X} s^*(x, y) \cdot smm^*(x, K). \tag{10}$$

If x is mostly similar to K and compares to the rest of the objects in X mostly the same as the objects in K then $f^*(x, K)$ will be high.

Finally we define the fitness of a partition, with respect to the crisp similarity relation, as the outcome of the clustering method though $f_P^* : \mathcal{O}(X) \to [-n^2, n^2]$, where $\mathcal{O}$ is the set of all possible partitions of X:

$$f_P^*(\mathcal{K}) := \sum_{K \in \mathcal{K}} \sum_{x,y \in K} s^*(x, y) + \sum_{K_1 \neq K_2 \in \mathcal{K}} \sum_{x \in K_1, y \in K_2} -s^*(x, y). \tag{11}$$

As the clustering result will be a partition, we wish to maximize this fitness function and therefore will use it as the criterion to be optimized.

4 Clustering Algorithm

The exact algorithmic approach to find the best partition would be to enumerate all of them and select the one that maximizes the fitness function defined above. However, this approach is not feasible even for small problems. The number of all partitions for a problem of size n is given by the Bell number for which an upper bound of $\left(\frac{0.792 \cdot n}{ln(n+1)}\right)^n$ has been recently given in [3].

Therefore we structure our algorithmic approach in the following four steps:

1. Elicit the thresholds and weights on each attribute from a human agent.
2. Construct the bipolar-valued similarity relation and its associated Condorcet similarity graph according to this preferential information.
3. Find the cluster cores.
4. Expand the cores and achieve a complete partition.

The *first step* is not fully covered in this paper, however this parameter elicitation step is crucial and is one of the distinctive features of our approach. We are currently exploring this step thoroughly. It is inspired from preference elicitation techniques used in multiple criteria decision aid to obtain the preferences of the decision maker (see [23] for a large review of such methods). It should be based both on a direct induction of the parameters and an indirect one, trying to exploit holistic judgements of the human actor on the similarity or dissimilarity of some objects well-known to him.

In the direct elicitation process, the human actor is asked to assign numerical values to the various discrimination thresholds and the importance weights of the attributes, according to his expertise or his perception of the underlying data. Via an indirect process, this information can at any time be complemented by overall judgements of the human actor on some objects. They can be, among others, of the following forms :

- I consider that objects a and b are overall similar (or dissimilar);
- I consider that objects a and b are more similar (or dissimilar) than c and d are.

These judgements are then included as linear constraints on the overall similarity caracteristic functions in a linear program whose goal is to determine a set of discrimination thresholds and importance weights in accordance with these inputs (see [21] for a similar approach in multiple criteria decision aid). Note again that this step should be carried over only a small sample of the original dataset.

The *second step* is straight-forward and derives from the definitions in the above section.

In the *third step*, we may use two resolution strategies: exact enumeration of all the maximal cliques and selection of the fittest ones as potential cluster cores, or a population-based metaheuristic approach.

For the exact approach we use the Bron-Kerbosch algorithm [7], with the pivot point improvement from Koch [17]. We then evaluate the fitness of each maximal clique and compute the neighbourhood matrix from which we retrieve the maximal cliques that are the local maxima of the fitness function. As previously mentioned, the number of maximal cliques in a graph can be exponential [22], making the use of exact approaches for large or even medium clustering problems rapidly intractable.

To overcome this operational problem, we use a population-based metaheuristic close in structure to evolutionary strategies [27]. As such, the metaheuristic contains 4 steps: initialization, selection, reproduction and replacement. Each individual in the population is a maximal clique in the Condorcet similarity graph. Our aim is to discover all maximal cliques that are local maxima of our fitness measure.

In the *initialization step* we, first, iteratively generate maximal cliques that do not overlap with each other. After each object has been covered by at least one maximal clique, the rest of the population is then generated randomly.

The *selection step* has a large number of potential variations. We have opted after several tests for the rank-based roulette wheel method.

The *reproduction step* is based on a mutation operator specifically designed for maximal cliques. The maximal clique that will generate a new individual in the population is incrementally stripped with a given probability of its objects and then grown by adding other objects until the property of maximality is reached. The generated population is of equal size with the old one.

In the *replacement step*, all maximal cliques in the current population that are local maxima of the fitness measure, based however on the limited exploration of their neighborhoods that has been done at previous iterations, are kept in the new population. The rest of the individuals to be kept are selected at random, in order to maintain a good exploration of the search space.

The *last step* orders all the objects that were not included in a core based on their best fitness to be added to a core. The majority margins heuristic, in fact, tells us how many relations are in accordance with the decision to add the object to a particular core, therefore iteratively taking the best pair of object and core and adding that object to the core is well justified considering our goals to extract a partition that is in most accordance to the original similarity relation.

5 Results

We would like to present some results on the 2010-11 Times Higher Education World University Ranking data [**?**]. The dataset contains 199 universities evaluated on 13 separate indicators designed to capture a broad range of activities, from teaching and research to knowledge transfer. These elements are brought

together into five attributes on scales from 0 to 100: Teaching environment (T), International mix (I), Industry income (Ind), Research (R) and Citations (C).

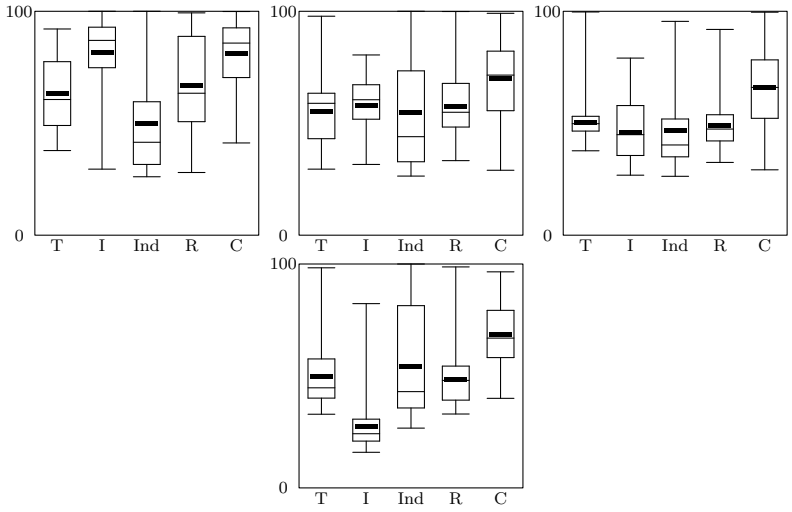
We do not wish to rank the universities, as it could be easily misunderstood, but to find clusters of similar universities according to two persons' viewpoints. Let's consider person A to be a student who is more interested in the teaching environment and international side of the universities, and let person B be a postdoc who is more interested in looking at the research environment characterising each university. For each person we will find the set of clusters which make most sense according to their viewpoints.

Due to the fact that each attribute is constructed and brought to the same scale, we take as thresholds $\sigma = 5\%$, $\delta = 10\%$, $\delta^+ = 50\%$ of the scales range for each attribute. We then select a set of weights in accordance with each person's point of view as seen in Table 1.

Table 1. Weights extracted to reflect two persons' viewpoints)

Person	T	I	Ind	R	C
A	0.31	0.42	0.10	0.10	0.05
B	0.05	0.10	0.10	0.42	0.31

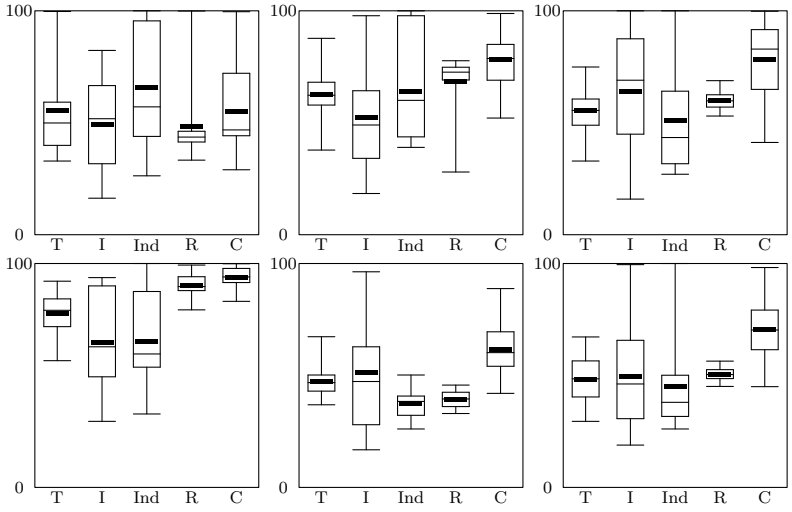
Fig. 5. Person A cluster box plots



For person A we find 4 clusters for which we present the boxplots of the objects inside them in Figure 5. We notice how we have grouped together universities with close evaluations on the second attribute which was deemed as most important for person A. In the first cluster we find universities with high evaluations on the International attribute (I), medium-high values in the second, medium-low in the third, and very low in the last. The Teaching attribute (T) also differentiates well the clusters, with progressively lower evaluations from the first to third clusters, while slightly higher on the last.

In the case of the second point of view we find 6 clusters where universities are grouped together mostly based on the fourth attribute, Research (R), as seen in Figure 6. Also the last attribute, Citations (C), is well defining for each cluster.

Fig. 6. Person B cluster box plots



As we wish that the clustering result would contain clusters of objects that are all similar to each other and dissimilar from the rest, we could model this by a similarity relation that contains $+1$ values between objects inside the same cluster and -1 values between objects in different clusters. To measure the quality of the results we use the fitness of a partition that we defined in Section **??**. We compute this fitness accroding to the similarity relation defined by each point of view that we have modeled. We also compare our results to some well-known algorithms like K-means [19], Single-Link and Complete-Link Agglomerative Hierarchical Clustering [14] and Partitioning Around Medoids [16]. For all of these, we give them as input the number of clusters we have found with our approach.

Table 2. Fitness of clusterings according to two persons' viewpoints)

	Person A POV	Person B POV
GAMMA-S for A	**0.784**	0.707
GAMMA-S for B	0.763	**0.853**
K-Means	0.710	0.698
SL AHC	0.204	0.213
CL AHC	0.682	0.728
PAM	0.704	0.791

We further present some results on a few well-known datasets such as the Iris, Wines and Breast Cancer datasets from the UCI Machine Learning Repository [11]. Each of these datasets comes, based on the analysis of some experts on the corresponding problem, with the set of clusters we should obtain. Therefore a good criterion for evaluation is the *Jaccard Coefficient* [13] in order to measure how close the results of the clustering algorithms are to the desired clusters.

Due to the absence of the experts that proposed the clusters for each problem we have used $\sigma = 10\%$, $\delta = 20\%$ and $\delta^+ = 70\%$ of the value range on each attribute as thresholds and given equal significance to all attributes. We show, however, on the Iris dataset which contains 150 objects defined on 4 attributes, that if such an interaction were possible, we could have extracted a threshold set that would bring the clustering results of GAMMA-S very close to the clusters that were proposed by the experts. Therefore, with $\sigma_1 = 0\%$, $\sigma_2 = 33\%$, $\sigma_3 = 25\%$, $\sigma_4 = 12\%$ of the value ranges of each attribute and the dissimilarity thresholds higher by 1% the Jaccard Coefficient of the GAMMA-S result would be equal to 0.878.

Except for our algorithm, we have given the a priori knowledge regarding how many clusters the outcome should have to the rest. The results come from running every non-deterministic algorithm 100 times over each instance. For the first two datasets, due to their small size, we have used the exact approach of our algorithm.

Table 3. Average results on Jaccard Coefficient (standard deviations in brackets)

Algorithm/Dataset	Iris	Wines	Breast Cancer
K-means	0.529 (0.118)	0.461 (0.100)	0.554 (0.130)
SL AHC	0.589 (-)	0.336 (-)	0.531 (-)
CL AHC	0.622 (-)	0.805 (-)	0.588 (-)
PAM	0.712 (0.007)	0.734 (0.000)	0.622 (0.025)
GAMMA-S	0.525 (-)	0.766 (-)	0.539 (0.028)

We notice overall that *GAMMA-S* performs very well considering the assumptions we make. We also notice that if we would be able to extract prefer-

ential information from a person who wants to cluster this data, we would get results more in accordance with his point of view on the problem. In addition we neither need to provide commensurable cardinal attributes nor an a priori number of clusters.

6 Conclusions and Perspectives

We conclude that our clustering method does indeed give consistent results, however without any requirements on the data, as all kinds of attribute types can be considered. Furthermore, imprecision, uncertainties and even missing values can easily be handled by the similarity relation defined in this article. There are many improvements that could be done to increase the performance of our approach, which will be explored in the future. At the moment we need to explore elicitation techniques for the parameters of the model. We also wish to present more extensive results on datasets on which we could have this interaction with a real person. The complexity of the algorithm should be improved by fine-tuning the meta-heuristic, while the final result our method proposes could be further improved by means of a local search method.

References

1. The Times Higher Education World University Rankings 2010-2011.
2. R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan. Automatic subspace clustering of high dimensional data, 2005.
3. M. Ankerst, M. M. Breunig, H. Kriegel, and J. Sander. Optics: ordering points to identify the clustering structure. In *International Conference on Management of Data*, pages 49–60, 1999.
4. D. Berend and T. Tassa. Improved bounds on bell numbers and on moments of sums of random variables. *Probability and Mathematical Statistics*, 30:185–205, 2010.
5. R. Bisdorff. Logical foundation of fuzzy preferential systems with application to the electre decision aid methods. *Computers & Operations Research*, 27:673–687, 2000.
6. R. Bisdorff. Electre-like clustering from a pairwise fuzzy proximity index. *European Journal of Operational Research*, 138(2):320–331, 2002.
7. R. Bisdorff. On clustering the criteria in an outranking based decision aid approach. In *Modelling, Computation and Optimization in Information Systems and Management Sciences*, pages 409–418. Springer CCIS, 2008.
8. C. Bron and J. Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Commun. ACM*, 16(9):575–577, 1973.
9. P. Cheeseman and J. Stutz. *Bayesian Classification (AutoClass): Theory and Results*, chapter 6, pages 62–83. AAAI Press/MIT Press, 1996.
10. Nan Du, Bin Wu, Xin Pei, Bai Wang, and Liutong Xu. Community detection in large-scale social networks. In *WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 16–25. ACM, 2007.

11. S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75 – 174, 2010.
12. A. Frank and A. Asuncion. UCI machine learning repository, 2010.
13. S. Guha, R. Rastogi, and K. Shim. CURE: an efficient clustering algorithm for large databases. In Laura Haas, Pamela Drew, Ashutosh Tiwary, and Michael Franklin, editors, *SIGMOD '98: Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, pages 73–84. ACM Press, 1998.
14. P. Jaccard. Nouvelles recherches sur la distribution florale. *Bulletin de la Societe Vaudoise de Sciences Naturelles*, 44:223–370, 1908.
15. A. Jain and R. Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.
16. A. Jain, M. Murty, and P. Flynn. Data clustering: A review. *ACM Computing Survey*, 31(3):264–323, 1999.
17. L. Kaufman and J. Rousseeuw, P. *Finding Groups in Data An Introduction to Cluster Analysis*. Wiley Interscience, 1990.
18. I. Koch. Enumerating all connected maximal common subgraphs in two graphs. *Theoretical Computer Science*, 250(1-2):1–30, 2001.
19. T. Kohonen. Self-organising maps. *Information Sciences*, 1995.
20. J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, 1967.
21. G. McLachlan and T. Krishnan. *The EM algorithm and extensions*. Wiley series in probability and statistics. Wiley, 2. ed edition, 2008.
22. P. Meyer, J.-L. Marichal, and R. Bisdorff. Disaggregation of bipolar-valued outranking relations. In Le Thi Hoai An, Pascal Bouvry, and Pham Dinh Tao, editors, *MCO*, volume 14 of *Communications in Computer and Information Science*, pages 204–213. Springer, 2008.
23. J. Moon and L. Moser. On cliques in graphs. *Israel Journal of Mathematics*, 3(1):23–28, 1965.
24. V. Mousseau. Elicitation des préférences pour l'aide multicritére á la décision, 2003.
25. M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 2004.
26. G. Palla, I. Derenyi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.
27. G. Sheikholeslami, S. Chatterjee, and A. Zhang. Wavecluster: A multi-resolution clustering approach for very large spatial databases. In *VLDB'98, Proceedings of 24rd International Conference on Very Large Data Bases*, pages 428–439. Morgan Kaufmann, 1998.
28. E. Talbi. *Metaheuristics - From Design to Implementation.* Wiley, 2009.
29. T. Zhang, R. Ramakrishnan, and M. Livny. BIRCH: an efficient data clustering method for very large databases. In *SIGMOD '96: Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, pages 103–114. ACM Press, 1996.
30. B. Zhou, D. Cheung, and B. Kao. A fast algorithm for density-based clustering in large database. In *PAKDD*, volume 1574 of *Lecture Notes in Computer Science*, pages 338–349. Springer, 1999.