Belgian Journal of Operations Research, Statistics and Computer Science, Vol 29, n° 3

SOME SILHOUETTE-BASED GRAPHICS FOR CLUSTERING INTERPRETATION

E. TRAUWAERT, P. ROUSSEEUW and L. KAUFMAN

V.U.B
Centrum S.T.O.O.
Pleinlaan 2
B-1050 Brussels

ABSTRACT

Silhouettes were developed as a graphical display for nonhierarchical cluster analysis. They are based on the ratio between the tightness of a cluster and its separation from other clusters. A possible extension is to represent for each object both these characteristics in a two dimensional graph.

The same technique can also be used with fuzzy clustering, making use directly of the fuzzy membership functions to measure the tightness of the links of each object with its principal cluster and its neighbour.

1° Introduction

Visual representation has always been an important means of communication. Nowadays many other mathematical tools, such as analytical formulas and computers, are at the disposal of the researcher to describe phenomena in a precise way. However, graphical representation still possesses a very suggestive power that no other mathematical description is able to provide. The reason is that a graph yields a global view of the phenomena together with all the relations between its parts. This is clearly an advantage over most formal mathematical models.

No wonder that for cluster analysis, which is sometimes defined as the art of discovering groups in data, graphical representation is a much cherished tool. It may even be the main tool in examples where all objects can be represented in a two-dimensional space. In multidimensional situations clustering algorithms are necessary, but graphs are still very helpful to illustrate the results and to reveal some features which may be the start for a further investigation.

In hierarchical clustering, dendrograms [see e.g. ref. 1 to 5] represent the relations between the partitions at different levels, the merging sequence, and the level of each partition.

For nonhierarchical clustering, a representation by means of silhouettes was recently proposed by Rousseeuw [6]. Silhouettes are based on the ratio between the distances of an object to its own cluster and to its neighbour cluster.

In the present note, silhouettes will be extended in two directions: a two-dimensional representation for each object (Section 3) and a modification for fuzzy clusters, either as a one-dimensional (Section 4) or as a two-dimensional graph (Section 5). Some further considerations and conclusions are given in Section 6.

2° Recalling silhouettes

Silhouettes were developed by Rousseeuw [6] to evaluate the quality of a clustering allocation, independently of the clustering technique that was used. Only two streams of information are needed: the partition of the objects into a number of clusters (at least two) and the matrix of proximities between all objects.

The silhouettes are then defined as follows (we restrict ourselves to dissimilarities, although one could also use a collection of similarities between objects):

-let D(i,j) be the dissimilarity between objects i and j;

-let a(i) be the average dissimilarity of object i, which has been allocated to cluster A, to all other objects of the same cluster:

$$a(i) = \frac{\sum_j D(i,j)}{n_A - 1}$$

with $j \in A$ and n_A = number of objects in A. It is assumed that $n_A > 1$.

-let d(i,C) be the average dissimilarity of object i of cluster A to all objects of any cluster C, different from A; hence

$$d(i,C) = \frac{\sum_j D(i,j)}{n_C}$$

with $j \in C$ and n_C = number of objects in C.

-let b(i) be the minimum over all clusters C of d(i,C), corresponding to the neighbour cluster B (see Figure 1).

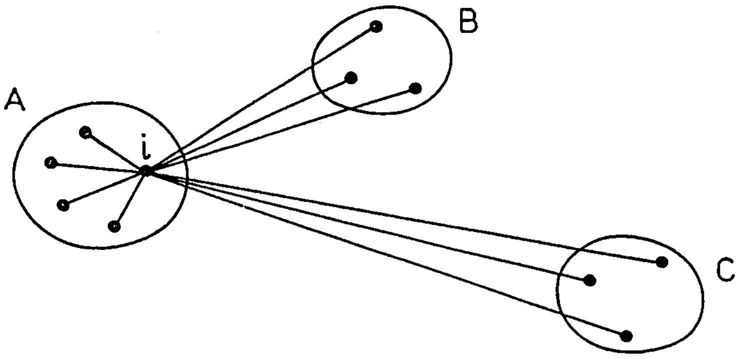


Fig.1: An illustration of the elements involved in the computation of s(i), where the object i belongs to cluster A (from [6]).

-let, for $n_A > 1$,

$$s(i) = 1 - \frac{a(i)}{b(i)} \qquad \text{if } a(i) \leq b(i)$$

$$= \frac{b(i)}{a(i)} - 1 \qquad \text{if } a(i) \geq b(i) \qquad (1)$$

for $n_A = 1$, $s(i) = 0$ by convention.

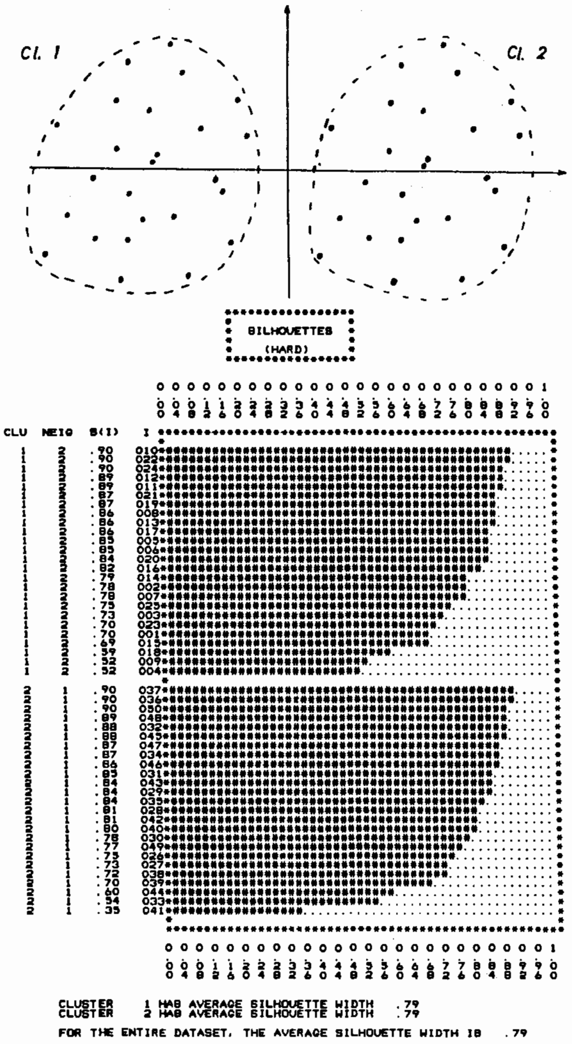
It can be seen that always

$$-1 \leq s(i) \leq 1. \qquad (2)$$

An s(i) near +1 means that the object i has a small average dissimilarity to objects of the same cluster and a high average dissimilarity to the neighbour cluster, and hence to all other clusters. A value near -1 expresses the opposite.

Having computed s(i) for each object of the data set, it is now possible to draw the silhouette of each cluster. For each object of that cluster, one draws a horizontal line with length proportional to s(i), pointing to the right whenever s(i) is positive and to the left otherwise (although this last part of the representation can be deleted as it is of less interest). All these lines are drawn below each other in decreasing order of magnitude. Each cluster has its own silhouette, the height of which is proportional to its number of objects whereas the width expresses its relative tightness.

Fig. 2 and 3 illustrate this technique on a set of objects consisting of two "natural" clusters. In Fig. 2 the natural clusters have effectively been found by some clustering technique. Because the clusters are fairly symmetric, so are both silhouettes. The largest values of s(i) correspond to objects at



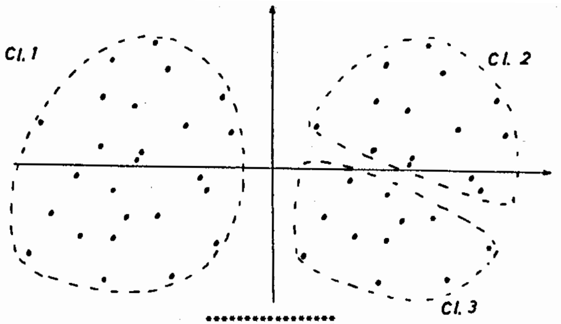
```
SILHOUETTES
  (HARD)

CLU  NEIG  S(I)   I
 1    2    .90   010
 1    2    .90   022
 1    2    .90   024
 1    2    .89   012
 1    2    .89   011
 1    2    .87   021
 1    2    .87   019
 1    2    .86   008
 1    2    .86   013
 1    2    .86   017
 1    2    .85   005
 1    2    .85   006
 1    2    .84   020
 1    2    .82   016
 1    2    .79   014
 1    2    .78   002
 1    2    .78   007
 1    2    .75   025
 1    2    .73   003
 1    2    .70   023
 1    2    .70   001
 1    2    .69   015
 1    2    .59   018
 1    2    .52   009
 1    2    .52   004

 2    1    .90   037
 2    1    .90   036
 2    1    .90   050
 2    1    .89   048
 2    1    .88   032
 2    1    .88   045
 2    1    .87   047
 2    1    .87   034
 2    1    .86   046
 2    1    .85   031
 2    1    .84   043
 2    1    .84   029
 2    1    .84   035
 2    1    .81   028
 2    1    .81   042
 2    1    .80   040
 2    1    .78   030
 2    1    .77   049
 2    1    .75   026
 2    1    .73   027
 2    1    .72   038
 2    1    .70   039
 2    1    .60   044
 2    1    .54   033
 2    1    .35   041

CLUSTER   1 HAS AVERAGE SILHOUETTE WIDTH   .79
CLUSTER   2 HAS AVERAGE SILHOUETTE WIDTH   .79

FOR THE ENTIRE DATASET, THE AVERAGE SILHOUETTE WIDTH IS   .79
```

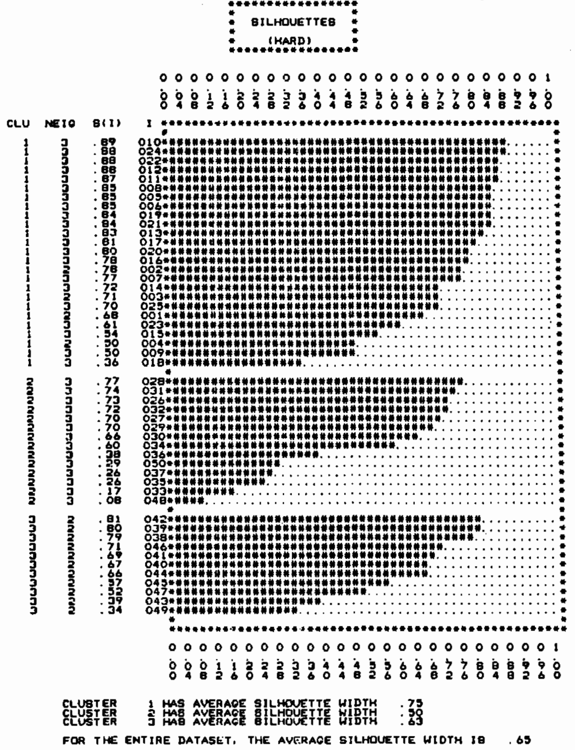
Fig.2: Silhouettes of basic model: 2 clusters.

the extremities of the set; the smallest values characterize objects near the interface between the clusters. The largest value is 0.90 for both clusters, and the smallest is 0.52 for the first cluster and 0.35 for the second. One can also calculate an average silhouette width for each cluster and for the entire data set; in our example all these values happen to be 0.79.

If a partition into three groups is performed (Fig.3), the first cluster remains unchanged whereas the second is split up in two parts. The silhouette of the first cluster is very similar to the one in Figure 2: not only the general shape is similar, but also the ordering of the objects. The s(i) values become slightly smaller because b(i), the average dissimilarity to the objects of the nearest of the other two clusters, is usually less than the average dissimilarity to the big cluster in Figure 2. This yields an average silhouette width of 0.75, as compared with 0.79 in Figure 2.

As for the two "half" clusters, the changes are of course more striking. Although for each object i the value a(i) is decreased, at the same time b(i) becomes smaller still, so $s(i)=1-a(i)/b(i)$ decreases. This results in an average silhouette width of 0.50 for cluster 2 and 0.63 for cluster 3, as compared with 0.79 in Figure 2. The overall average silhouette width of all three clusters is 0.65, or about 20% less than in the case of two clusters. Therefore, the overall average silhouette width gives some indication about the "best" number of clusters.





```
SILHOUETTES
  (HARD)

                         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
                         . . . . . . . . . . . . . . . . . . . . . . . . . .
                         0 0 0 1 1 2 2 2 3 3 4 4 4 5 5 6 6 6 7 7 8 8 8 9 9 0
                         0 4 8 2 6 0 4 8 2 6 0 4 8 2 6 0 4 8 2 6 0 4 8 2 6 0
CLU  NEIG  S(I)     I
 1    3    .89     010
 1    3    .88     024
 1    3    .88     022
 1    3    .88     012
 1    3    .87     011
 1    3    .85     008
 1    3    .85     005
 1    3    .85     006
 1    3    .84     019
 1    3    .84     021
 1    3    .83     013
 1    3    .81     017
 1    3    .80     020
 1    3    .78     016
 1    2    .78     002
 1    3    .77     007
 1    3    .72     014
 1    2    .71     003
 1    3    .70     025
 1    2    .68     001
 1    3    .61     023
 1    3    .54     015
 1    2    .50     004
 1    3    .50     009
 1    3    .36     018

 2    3    .77     028
 2    3    .74     031
 2    3    .73     026
 2    3    .72     032
 2    3    .70     027
 2    3    .70     029
 2    3    .66     030
 2    3    .60     034
 2    3    .38     036
 2    3    .29     050
 2    3    .26     037
 2    3    .26     035
 2    3    .17     033
 2    3    .08     048

 3    2    .81     042
 3    2    .80     039
 3    2    .79     038
 3    2    .71     046
 3    2    .69     041
 3    2    .67     040
 3    2    .66     044
 3    2    .57     045
 3    2    .52     047
 3    2    .39     043
 3    2    .34     049

                         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
                         . . . . . . . . . . . . . . . . . . . . . . . . . .
                         0 0 0 1 1 2 2 2 3 3 4 4 4 5 5 6 6 6 7 7 8 8 8 9 9 0
                         0 4 8 2 6 0 4 8 2 6 0 4 8 2 6 0 4 8 2 6 0 4 8 2 6 0

CLUSTER    1 HAS AVERAGE SILHOUETTE WIDTH   .75
CLUSTER    2 HAS AVERAGE SILHOUETTE WIDTH   .50
CLUSTER    3 HAS AVERAGE SILHOUETTE WIDTH   .63

FOR THE ENTIRE DATASET, THE AVERAGE SILHOUETTE WIDTH IS   .65
```

Fig.3: Silhouettes of basic model: 3 clusters.

3° Unfolding silhouettes in two dimensions

Silhouettes are based on the evaluation of two functions for each object:

the "tightness" a(i)

the "separation" b(i).

Instead of calculating the ratio of these two functions, it is also possible to simply plot these functions in a two-dimensional graph, using, say, a(i) for the x-axis and b(i) for the y-axis.

As both a(i) and b(i) are always positive, only the first quadrant of the (x,y)-space is used. Looking for the relation between the s(i) values and the (a,b)-plot, it can be observed

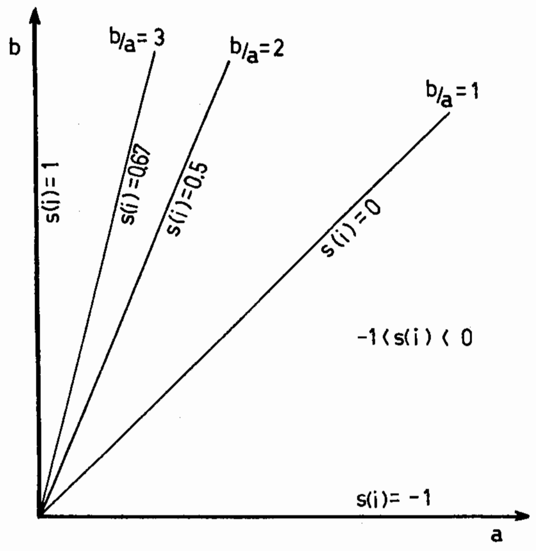


Fig.4: Relation between separation/tightness and silhouettes.

that all objects with the same s(i) values lie on a straight line, starting from the origin and satisfying one of the following equations:

$$b(i) = (1+s(i))\ a(i) \quad \text{if } -1 \leq s(i) \leq 0 \qquad (3)$$

$$b(i) = \frac{1}{1-s(i)}\ a(i) \quad \text{if } 0 \leq s(i) \leq 1 \qquad (4)$$

From these equations it can be seen that objects with s(i)=-1 will be represented by points on the a-axis. Objects with s(i)=0 correspond to the equation b(i)=a(i), and will be represented by points on the 45° line. Objects with negative s(i) will lie below that line, whereas objects with positive s(i) lie above it. Objects with s(i)=1 end up on the b-axis. These relations are represented in Fig. 4. It should be observed that a plot can be drawn for all the objects of a data set as well as for the objects of each cluster separately.

Fig. 5 and 6 show these plots for the example with two "natural" clusters discussed in the previous section. Fig. 5 is very typical of a good clustering allocation. The plots show a rather narrow concentration of the tightness a(i) and a much larger dispersion of the separation, with most objects having a b(i)/a(i) ratio larger than two. The only object with b(i)/a(i) smaller than two is located near both clusters. It almost forms a bridge between them, as can be deduced from the fact that a(i) has one of the largest and b(i) one of the smallest values.

In the three clusters case (Fig.6) things are clearly different. The first cluster still resembles that of the former case, but the two remaining clusters have much smaller values of b(i), which in turn are much nearer to the a(i) values. This could be a first indication that these clusters should not have been separated.

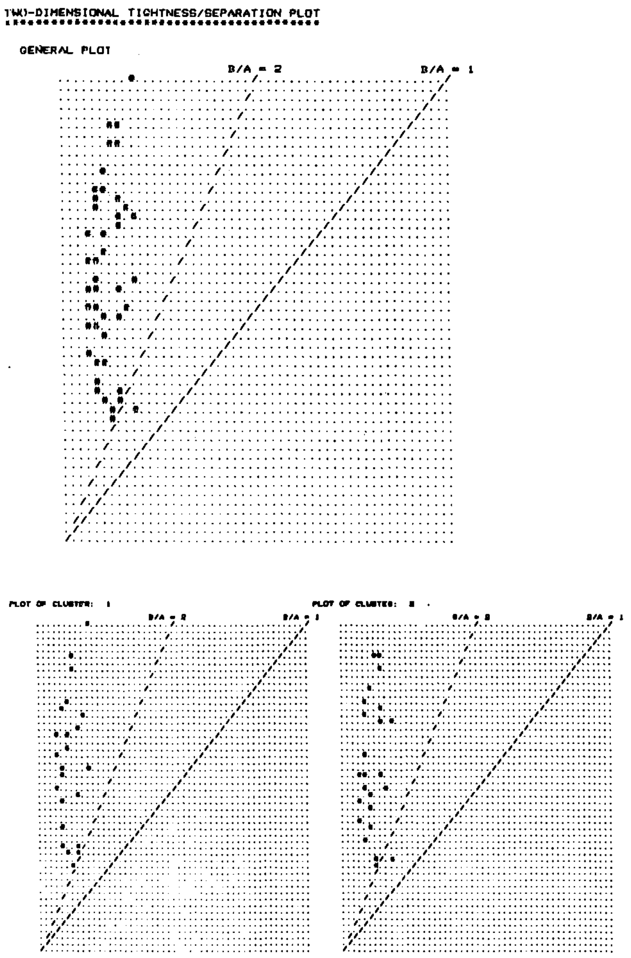


Fig.5: Basic model: two-dimensional hard representation of 2 clusters.

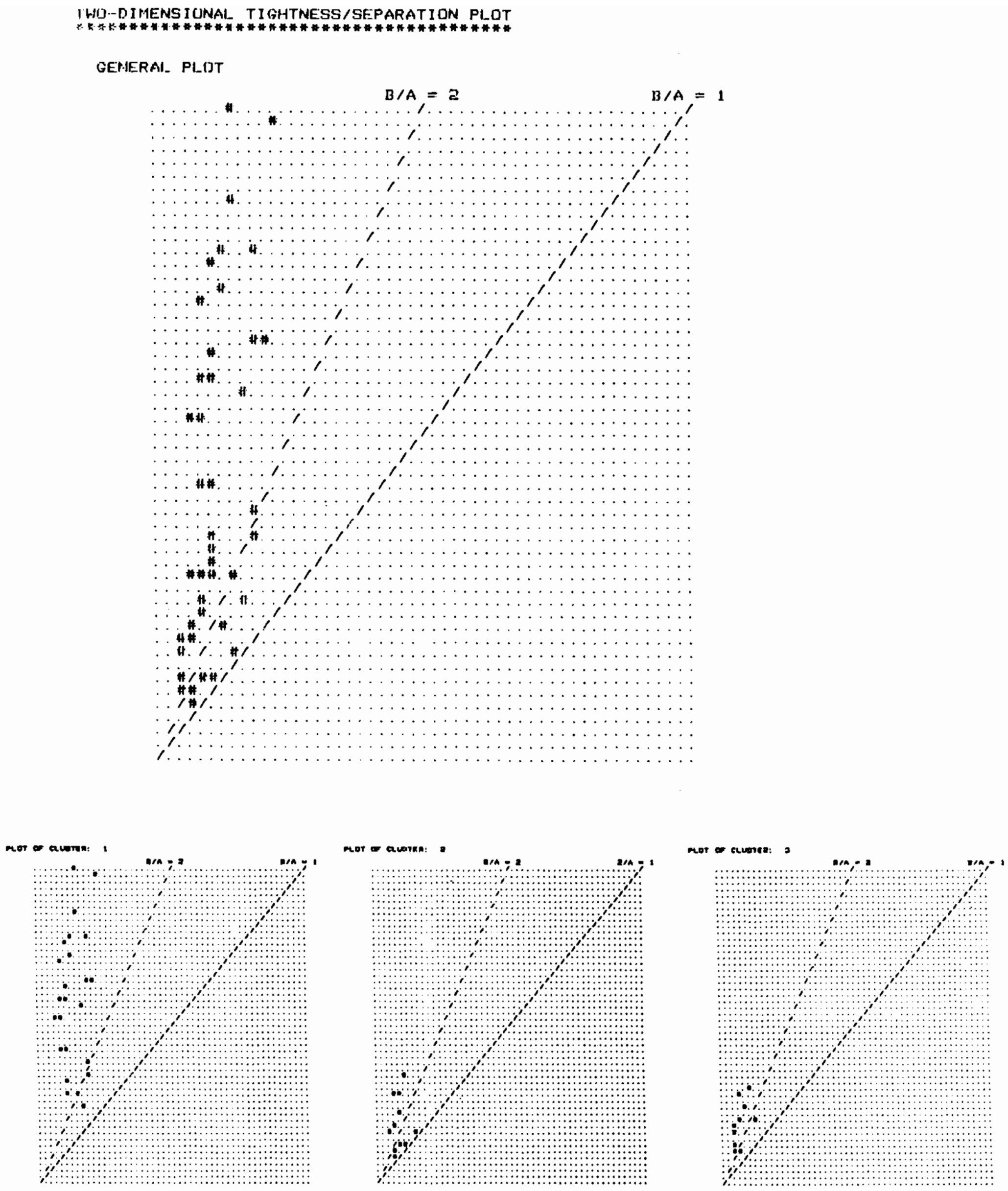


Fig.6: Basic model: two-dimensional hard representation of 3 clusters.

4° Using fuzzy membership functions

The goal of fuzzy clustering is to express, for each object, its relative membership to each cluster. Most fuzzy clustering algorithms [see e.g. ref.7] make use of average dissimilarities. By definition, the sum of membership values of each object to all clusters always equals one. It is also customary to consider the nearest hard classification, allocating each object to the cluster for which its fuzzy membership is largest. Therefore it is possible to define new "tightness" and "separation" factors based on membership functions, keeping in mind that the latter reflect similarity rather than dissimilarity:

$$a(i) = 1 - u^*(i) \quad \text{with } u^*(i) = u(t_o,i) = \max_{t} u(t,i) \qquad (5)$$

$$b(i) = 1 - u^{**}(i) \quad \text{with } u^{**}(i) = \max_{t \neq t_o} u(t,i) \qquad (6)$$

in which the membership functions must satisfy the relations:

$$u(t,i) \geq 0 \qquad \text{for all } i \text{ and } t$$

$$\sum_{t} u(t,i) = 1 \qquad \text{for all } i. \qquad (7)$$

From (5) and (6) we see that

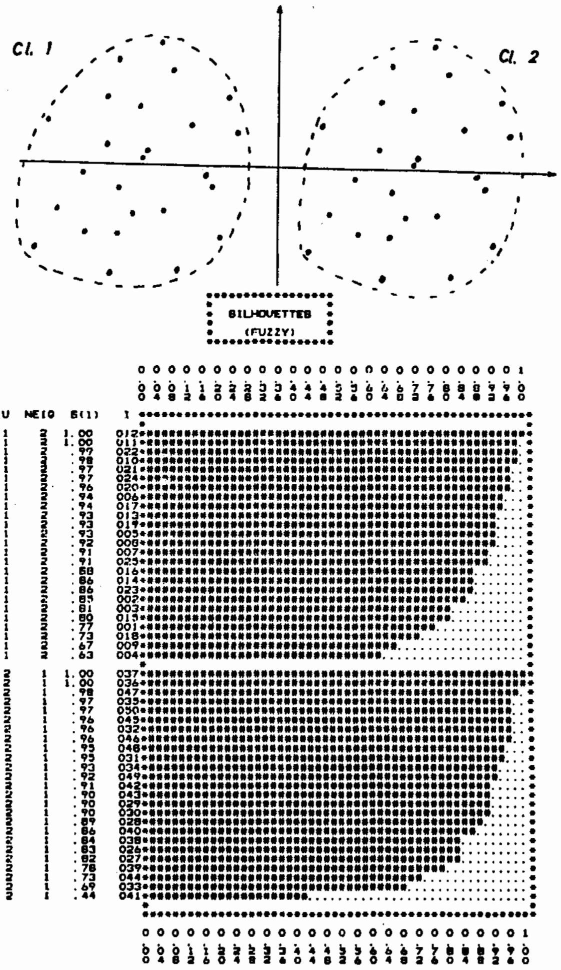
$$u^*(i) \geq u^{**}(i) \qquad (8)$$

and hence we always have

$$a(i) \leq b(i) \qquad (9)$$

resulting in $0 \leq s(i) \leq 1$, excluding the possibility of negative $s(i)$. Apart from this last aspect, the $s(i)$ behave similarly to what was seen in section 2. This is confirmed by Fig.7 which shows the fuzzy silhouette plot of the two-cluster example of that section: the general shape is very similar to that of Fig.2.

The only difference is that the fuzzy s(i) are generally a bit larger than the hard s(i) (which, of course, depends very much on the actual fuzzy algorithm used).



```
CLU  NEIG  S(I)    I
 1    2    1.00   012
 1    2    1.00   011
 1    2     .97   022
 1    2     .98   010
 1    2     .97   021
 1    2     .97   024
 1    2     .96   020
 1    2     .94   006
 1    2     .94   017
 1    2     .93   013
 1    2     .93   019
 1    2     .93   005
 1    2     .92   008
 1    2     .91   007
 1    2     .91   025
 1    2     .88   016
 1    2     .86   014
 1    2     .86   023
 1    2     .85   002
 1    2     .81   003
 1    2     .80   015
 1    2     .77   001
 1    2     .73   018
 1    2     .67   009
 1    2     .63   004

 2    1    1.00   037
 2    1    1.00   036
 2    1     .98   047
 2    1     .97   035
 2    1     .97   050
 2    1     .96   045
 2    1     .96   032
 2    1     .96   046
 2    1     .95   048
 2    1     .95   031
 2    1     .93   034
 2    1     .92   049
 2    1     .91   042
 2    1     .90   043
 2    1     .90   029
 2    1     .90   030
 2    1     .89   028
 2    1     .86   040
 2    1     .84   038
 2    1     .83   026
 2    1     .82   027
 2    1     .78   039
 2    1     .73   044
 2    1     .69   033
 2    1     .44   041

CLUSTER    1 HAS AVERAGE SILHOUETTE WIDTH   .89
CLUSTER    2 HAS AVERAGE SILHOUETTE WIDTH   .88
FOR THE ENTIRE DATASET, THE AVERAGE SILHOUETTE WIDTH IS   .88
```

Fig.7: Fuzzy silhouettes of basic model: 2 clusters.

5° A two-dimensional plot with fuzzy membership functions

As in the case of the original silhouette, it is also possible to unfold the fuzzy membership function in a two-dimensional plot. Compared to section 3, there are two main differences:

1° due to relation (9) all points will lie above the 45° line;

2° relation (7) induces a series of constraints which were absent in the hard approach. As we will see, these depend on the number of clusters that is considered;

a) for 2 clusters, relation (7) becomes

$$u^{\circ}(i)+u^{\circ\circ}(i) = 1$$

and through (5) and (6) we find

$$a(i)+b(i) = 1. \qquad (10)$$

This relation means that all objects in a two-cluster system will be represented on the straight line going from (1,0) to (0,1) (see Fig.8).

b) for 3 clusters, relation (7) becomes

$$u^{\circ}(i) + u^{\circ\circ}(i) + u(t,i) = 1$$

$$\text{or} \quad u^{\circ}(i) + u^{\circ\circ}(i) \leq 1$$

which through (5) and (6) gives

$$a(i) + b(i) \geq 1 \tag{11}$$

and as $u^{\circ\circ}(i) \geq u(t,i)$ through (6) we also have

$$u^{\circ}(i) + 2u^{\circ\circ}(i) \geq 1. \tag{12}$$

Using (5) and (6) this yields

$$1-a(i) \quad + 2(1-b(i)) \geq 1$$

so

$$b(i) \leq 1 - \tfrac{1}{2}\, a(i). \tag{13}$$

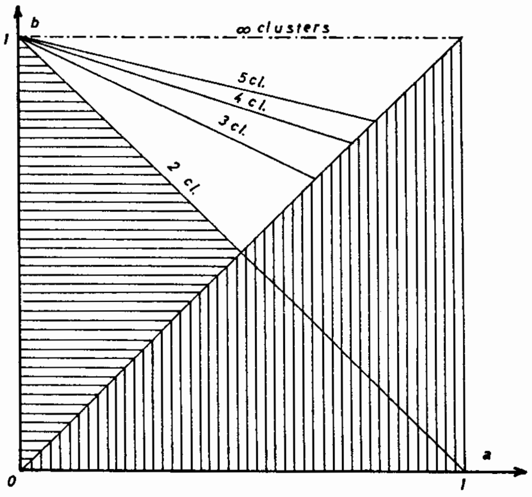


Fig.8: Two-dimensional plot with feasibility regions as function of number of fuzzy clusters.

Relations (11) and (12) force all objects in a three-cluster configuration to remain between two straight lines starting from the y-axis at the value b(i)=1 and with slopes -1 and $-\frac{1}{2}$ (see Fig.8).

c) for k clusters, relation (11) is still valid whereas relation (12) becomes

$$u^{\circ}(i) + (k-1)\, u^{\circ\circ}(i) \geq 1 \qquad (14)$$

which upon consideration of (5) and (6) becomes

$$1 - a(i) + (k-1)(1 - b(i)) \geq 1$$

so

$$b(i) \leq 1 - 1/(k-1)\, a(i). \qquad (15)$$

Hence the lower and right hand feasibility limits (11) and (9) remain unchanged whatever the number of clusters; the upper limit starts from the point on the b(i) axis with value 1 and has a negative slope proportional to 1/(k-1) (see Fig.8). This upper limit coincides with the lower limit in the case of only two clusters (k=2) and tends to an horizontal line for an infinite number of clusters (k=OO). It can further be observed that whenever points are represented on the lower limit, i.e. when the sum of a(i) and b(i) is equal to one, these objects have zero membership to all clusters but the principal one and the first neighbour; points represented on the upper limit line corresponding to the number of clusters, indicate that equation (14) has to be considered with an equality sign and hence that

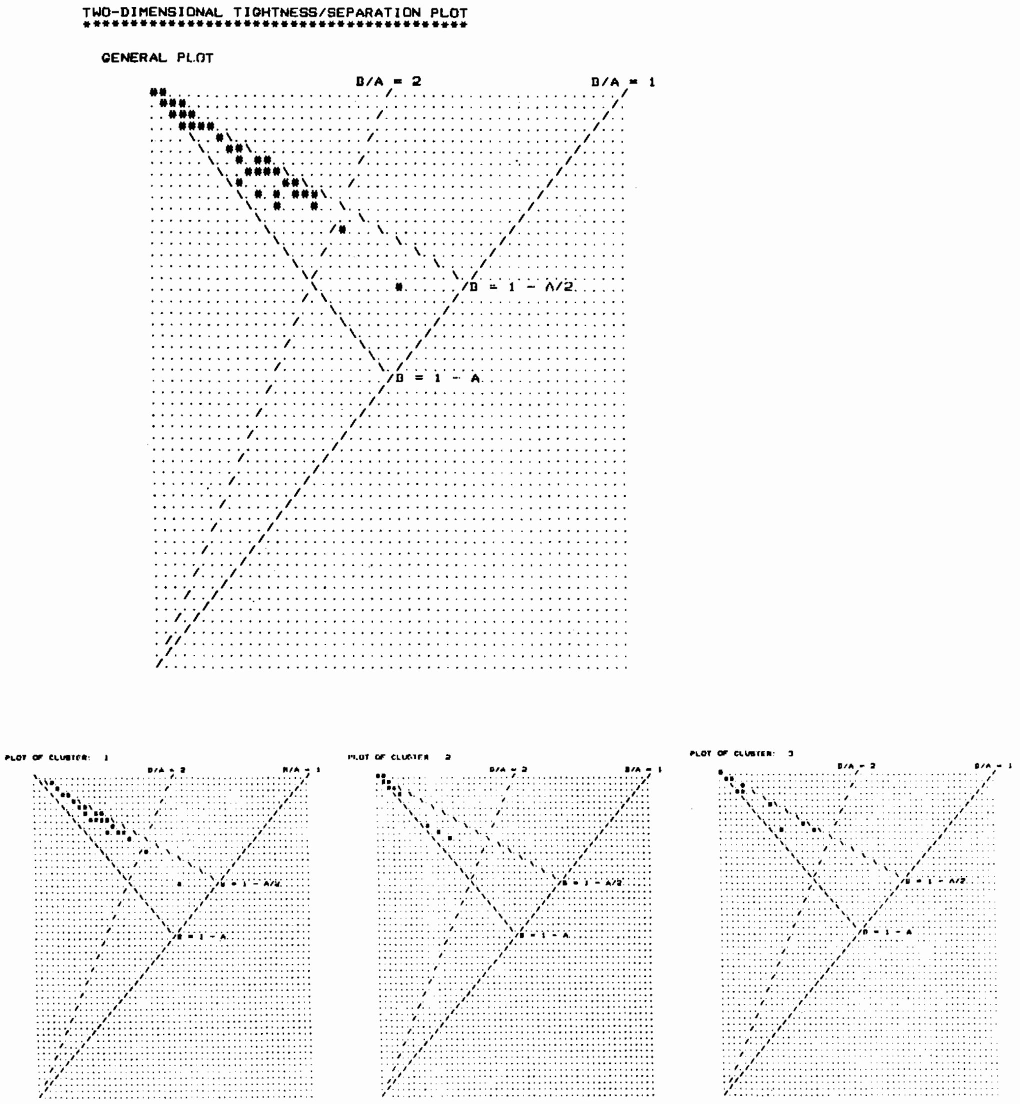


Fig.9: Ruspini's data: two-dimensional fuzzy representation of 3 clusters.

the corresponding object, apart from its membership to its principal cluster, has an equal membership to all the other clusters.

An example is provided by the Ruspini data [8], which contain four rather well-separated clusters. A partitioning into three fuzzy clusters shows two well-characterized clusters and a third one that is not so tight (Fig. 9). The partition in four clusters gives an improved image for all clusters, confirming the existence of four "natural clusters" (Fig.10).

6° Conclusions

Graphical representations are very useful to get a global impression of a clustering. It was shown how silhouettes could be extended to a two-dimensional plot, providing some new information such as a distinction between bridging objects and outliers.

A similar plot can be constructed from fuzzy membership functions. There all points remain within a triangle, of which only the upper boundary is a function of the number of clusters. Moreover, the position of each object within this triangle tells a lot about the clustering characteristics.

As seen from the examples, the above graphs can even be drawn with a plain line printer. This allows the implementation of these graphical representations in almost any computer environment.

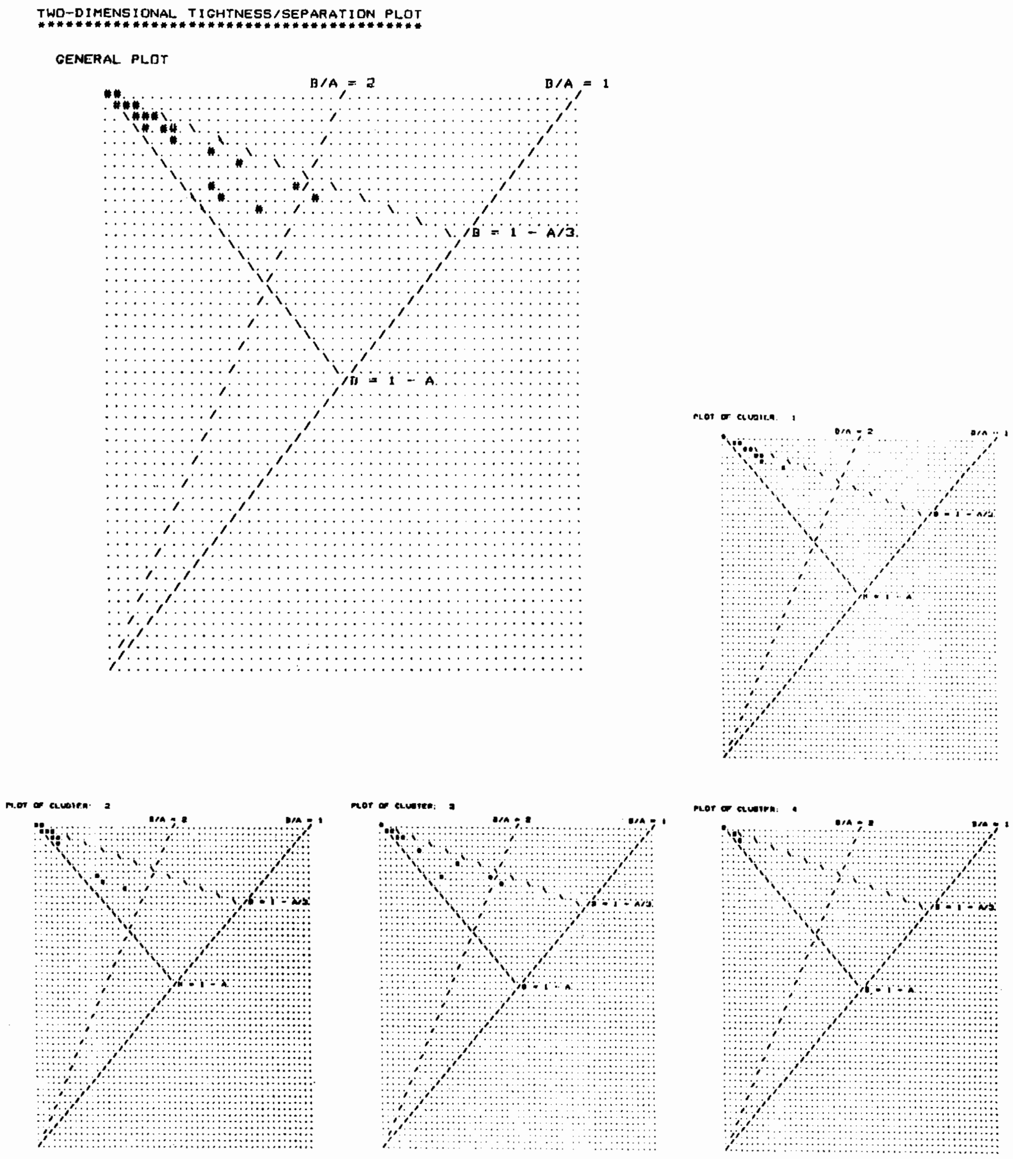


Fig.10: Ruspini's data: two-dimensional fuzzy representation of 4 clusters.

7° References

[1] Massart D.L. and Kaufman L.: The interpretation of analytical chemical data by the use of cluster analysis; J.Wiley & S. (1983).

[2] Everitt B.: Cluster analysis; J.Wiley & S. (1973).

[3] Bock H.H.: Automatische Klassifikation; Vandenhoeck & Ruprecht (1974).

[4] Sneath P.H.A. and Sokal R.R.: Numerical taxonomy: the principles and practice of numerical classification; Freeman, San Francisco (1973).

[5] Lance G.N. and Williams W.T.: A general theory of classificatory sorting strategies; 1. Hierarchical systems; Computer Journal, 9 (1967) pp373-380.

[6] Rousseeuw P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis; Jour. of Comp. & App. Math., 20 (1987) pp53-65.

[7] Bezdek J.C.: Pattern recognition with fuzzy objective function algorithms; Plenum Press, NY (1981).

[8] Ruspini E.: Numerical methods for fuzzy clustering; Information Sciences, 2 (1970) pp319-350.