

**Input :** A set of tuples  $\mathbf{T}$ ,

- a clustering  $\mathcal{C} = \{c_1, c_2, \dots, c_k\}$  of  $\mathbf{T}$ ,  
where  $c_i$  is the identifier of cluster  $i$
- a distance measure  $d$ .

**Output :** For every tuple  $\mathbf{t}$  in  $\mathbf{T}$ , a probability  $prob(\mathbf{t})$ .

**Main Procedure :**

- (Step 1) For  $i = 1 \dots k$ :
  - \* compute cluster representative  $rep_i$  for  $c_i$   
by merging all the tuples that belong to it.
  - \* initialize sum of distances for  $c_i$ ,  $S(c_i) = 0$ .
- (Step 2) For each tuple  $\mathbf{t} \in \mathbf{T}$  that belongs to  $c_i$ :
  - \* compute  $d_{\mathbf{t}} = d(\mathbf{t}, rep_i)$ , the distance of  $\mathbf{t}$   
to the representative of its cluster.
  - \* Add  $d_{\mathbf{t}}$  to  $S(c_i)$ .
- (Step 3) For each tuple  $\mathbf{t} \in \mathbf{T}$  that belongs to  $c_i$ :
  - \* compute similarity  $s_{\mathbf{t}} = 1 - \frac{d_{\mathbf{t}}}{S(c_i)}$ .
  - \*  $prob(\mathbf{t}) = 1.0$  if  $|c_i| = 1$ , or  
 $prob(\mathbf{t}) = \frac{s_{\mathbf{t}}}{|c_i|-1}$  otherwise.