Constrained–Divisive Algorithm

Input: (data set D = {d1,...,dn}, must-links Con=⊆ D × D, cannot-links Con≠ ⊆ D × D)

0. Request the user to select a distance metric for computing dissimilarity.

Variant 1: Euclidean distance

Variant 2: Pearson metric

Variant 3: Spearman metric

Variant 4: Manhattan distance

1. Let C be the initial cluster group: C = {D}.

2. Select the cluster Cm ∈ C with the largest dissimilarity between any two of its objects. Divide Cm following (3) to (7).

3. Select the element sz with the highest average dissimilarity to all other elements and, if clustering the first level, with an unresolved cannot-link. sz initiates splinter group S. If clustering the first level:

∀d ∈ Cm, s ∈ S: If (di, sj ) ∈ Con=, move di to S.

4. ∀d still ∈ Cm, s ∈ S: Compute the difference of distances: Diff(i)=[averagedistance(di, dj)] − [averagedistance(di, sj)]. Select the element dh with the greatest difference, Diff(h). If clustering the first level: If ∃sj ∈ S : (dh, sj) ∈ Con≠, set Diff(h) = 0. If Diff(h) > 0, move dh to the splinter group S.

5. Repeat (4) and (5) until all differences Diff(h) are negative.

6. Now, the original cluster is split into two clusters. One is the splinter group S, and the other is formed by the remaining elements in Cm.

7. Iterate between (2) and (7) until all clusters {C1 ...Ck} contain only a single element.