Robust Clustering

The *robust* clustering algorithm has been designed to take numerous sets of clusters as input and uses these to generate a set of *robust clusters*. A robust cluster must only consist of objects that appear together in *all* the input clusters produced from different clustering algorithms. Firstly, an upper triangular $n \times n$ agreement matrix is generated with each cell containing the number of agreements amongst methods for clustering together the two variables, represented by the indexing row and column indices (see Figure 1). This matrix is then used to cluster variables based upon their cluster agreement (as found in the matrix).

$$\begin{array}{cc}
 & \text{To Gene} \\
\text{From Gene} & \begin{bmatrix}
0 & A_{12} & A_{13} & A_{14} & \cdots & \cdots & \cdots & \cdots & \cdots & A_{1n} \\
\vdots & 0 & A_{23} & A_{24} & \cdots & \cdots & \cdots & \cdots & \cdots & A_{2n} \\
\vdots & \vdots & 0 & A_{34} & \cdots & \cdots & \cdots & \cdots & \cdots & A_{3n} \\
\vdots & \vdots & \vdots & \ddots & \ddots & & & & & \vdots \\
\vdots & \vdots & \vdots & & \ddots & \ddots & & & & \vdots \\
\vdots & \vdots & \vdots & & & \ddots & A_{ij} & & & \vdots \\
\vdots & \vdots & \vdots & & & & \ddots & \ddots & & \vdots \\
\vdots & \vdots & \vdots & & & & & \ddots & \ddots & \vdots \\
\vdots & \vdots & \vdots & & & & & & \ddots & A_{(n-1)n} \\
0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0
\end{bmatrix}
\end{array}$$

Figure 1. The Agreement Matrix, *A*

The algorithm works by taking in the agreement matrix in order to generate a list, *List*, which contains all the pairs where the appropriate cell in the agreement matrix contains a value equal to the number of methods being combined (i.e. full agreement). Starting with an empty set of robust clusters *RC*, where RC_i is the *i*th robust cluster, the first cluster is created containing the elements of the first pair in *List*. Then the pairs in *List* are iterated through and checked to see if one of the members of the current pair is within any of the clusters in *RC*.

If one element of the current pair is found and the other element of the pair is not in the same cluster then the other element is added to that cluster. If neither element of the pair is found in *RC* then a new cluster is added to *RC* containing each element of the pair. When the end of the list is reached, the set of robust clusters, *RC*, is the output. The results of applying robust clustering are summarised in Table 7 and the algorithm follows.

```
Input:  Agreement Matrix ($n \times n$), A; The Number of Clustering Methods
        Set List = all pairs ($x$,$y$) in A, with agreement = number of methods
        Set RC to be an empty list of clusters
        Create a cluster and insert the first pair ($x$,$y$) in List into it
        For $i$ = 2 to size of List
            For $j$ = 1 to number of Clusters in RC
                If $x$ or $y$ of $List_i$ is found within $RC_j$
                    If the other member of the pair $List_i$ is not found in $RC_j$
                        Add the other member to $RC_j$
                    End If
                Else If the other member of the pair $List_i$ is not found in $RC_j$
                    Add a new cluster to RC containing $x$ and $y$
                End If
            End For
        End For
Output: Set of Robust Clusters, RC
```