Interaction-Hierarchical Clustering

This project is published for "Cluster fusion-fission dynamics in the Singapore stock exchange",

by Boon Kin Teh and Siew Ann Cheong. Please refer to the paper for more details, and cite the paper if you are using this code to perform interaction-hierarchical clustering. Thank you.

Additional details can be found in Boon Kin Teh dissertation, “Macroscopic & Mesoscopic dynamics of Financial Market”.

# Code: HierarchicalClustering

**Perform hierarchical clustering (First level clustering)**

**Inputs:**

1) DistanceMatrix: NxN double array, contains distance between each pair of variables (total N variables), it is a symmetry matrix with zeros diagonal entries.

2) ClusteringMethod: string, Method to define the distance between clusters, as follow

complete = Maximum distance

average = Average distance

single = minimum distance

centroid = Euclidean distance

**Outputs:**

[SerialIndex,Result\_A] = A\_HierarchicalClustering(DistanceMatrix,ClusteringMethod)  
1) SerialIndex: 1xN integer array, the dendrogram order obtained for performing hierarchical clustering.  
2) Result\_A: structure, input for B\_DetermineRobustClusters.  
3) Linkage: N-1x3 double array, each row represents the cluster with column 1 index merged with the cluster with column 2 index at distance in column 3.  
 Cluster index, CI: if CI <= N, CI represent the index for the N elements  
 if CI > N, represent clusters merged at row CI-N.

# Code: DetermineRobustClusters

**Determine the robust clusters**

**Inputs:**

1) Result\_A: structure data, result produced by A\_HierarchicalClustering.

2) MustBreakCutoff: double, a cut off such that no robust cluster is obtain with distance about this cut off, define as the distance the similarity/correlation become negative**.**

**Outputs:**

[RobustClusterList,RobustLenght,Result\_B] = B\_DetermineRobustClusters(Result\_A,MustBreakCutoff)  
1) RobustClusterList: LxK interger array, storage for element index of K robust clusters determined, the largest cluster size is L, where zeros indicate non-element.  
2) RobustLenght: double, the robust length identified that give least number of cluster.  
3) Result\_B: structure, input for C\_InteractionHierarchicalClustering.

# Code: InteractionHierarchicalClustering

**Perform Interaction-hierarchical clustering (Second level clustering)**

**Inputs:**

1) Result\_B: structure data, result produced by B\_DetermineRobustClusters.   
2) InteractionDistanceDef : string, Method to define the interaction distance between clusters as follow

average = Average distances (default)

Pct2575 = Average between precentile 25 and 75 distances

median = median distance

**Outputs:**

[RBSerialIndex,Result\_C] = C\_InteractionHierarchicalClustering(Result\_B,InteractionDistanceDef)  
1) RBSerialIndex: 1xN integer array, the dendrogram order obtained for performing interaction-hierarchical clustering.  
2) Result\_C: structure, input for D\_IdentifyClusters.  
3) RBLinkage: K-1x3 double array, each row represents the robust cluster with column 1 index merged with the robust cluster with column 2 index at distance in column 3.  
 Robust Cluster index, RCI: if RCI <= K, CI represent the index for the K robust clusters  
 if RCI > K, represent clusters merged at row RCI-K

# Code: IdentifyClusters

**Identify clusters based on interaction-hierarchical clustering**

**Inputs:**

1) Result\_C: structure data, result produced by C\_InteractionHierarchicalClustering2) CL\_2: Nx1 cell list as Cluster list, each cell represents a cluster with at least one elements.

**Outputs:**

IdentifiedClusterList = D\_IdentifyClusters(Result\_C)  
1) ClusterList: LxK interger array, storage for elemnt index of K identified clusters determined, the largest cluster size is L, where zeros indicate non-element.