**library("clValid")**

**x <- read.csv(“D:/Essential protein prediction/input/Priority \_based\_Similarity\_measure/Ess\_data.csv”, header=TRUE)**

**x1<-x**

**x1$Protein<-NULL**

**intern <- clValid(x1, 2:10, clMethods = c("hierarchical", "kmeans", "diana", "fanny", "pam", "sota", "clara", "model"), validation = "internal")**

**summary(intern)**

Clustering Methods:

hierarchical kmeans diana fanny pam sota clara model

Cluster sizes:

2 3 4 5 6 7 8 9 10

Validation Measures:

2 3 4 5 6 7 8 9 10

hierarchical

Connectivity 4.2869 8.2766 10.0500 12.5500 19.0770 19.5687 23.4206 25.4206 29.1127

Dunn 0.4214 0.0215 0.0403 0.0403 0.0403 0.0181 0.0188 0.0188 0.0316

Silhouette 0.9831 0.9701 0.9619 0.9617 0.9612 0.9424 0.9418 0.9416 0.9307

kmeans

Connectivity 4.4091 6.6377 10.7357 13.2357 16.4302 20.9996 27.8889 29.8889 31.8532

Dunn 0.0096 0.0220 0.0064 0.0064 0.0026 0.0026 0.0034 0.0034 0.0014

Silhouette 0.9717 0.9683 0.9491 0.9486 0.9351 0.9307 0.9304 0.9302 0.9133

diana

Connectivity 4.3607 8.6476 10.4210 12.9210 18.1484 25.0377 28.8897 32.5817 38.1341

Dunn 0.0062 0.0132 0.0259 0.0271 0.0035 0.0037 0.0041 0.0068 0.0073

Silhouette 0.9720 0.9699 0.9620 0.9617 0.9412 0.9407 0.9403 0.9317 0.9314

fanny

Connectivity 5.2484 10.5036 19.7306 14.2115 23.0940 31.2663 34.1246 43.7802 48.7857

Dunn 0.0004 0.0001 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

Silhouette 0.9255 0.8779 0.8509 0.7720 0.7602 0.7435 0.7179 0.7154 0.6896

pam

Connectivity 6.0742 10.9627 13.4357 13.5036 24.8575 27.0095 27.4873 36.0802 38.0187

Dunn 0.0008 0.0003 0.0005 0.0002 0.0000 0.0001 0.0001 0.0001 0.0001

Silhouette 0.9671 0.9277 0.9252 0.8914 0.8786 0.8794 0.8255 0.8257 0.8266

sota

Connectivity 2.8647 6.9556 11.3083 16.3056 23.4948 27.4817 29.9817 31.7552 35.7476

Dunn 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

Silhouette 0.4705 0.5768 0.6076 0.6121 0.6118 0.6126 0.6124 0.6119 0.6113

clara

Connectivity 4.7409 10.4290 17.4155 20.8560 32.2925 32.4226 30.7683 32.6683 43.8103

Dunn 0.0016 0.0001 0.0001 0.0001 0.0001 0.0001 0.0000 0.0000 0.0000

Silhouette 0.9605 0.8862 0.9289 0.8660 0.8824 0.8724 0.8696 0.8710 0.8712

model

Connectivity 1121.1052 4508.7960 5584.9615 6314.1532 6663.7718 6587.7278 6964.2119 6943.3877 7343.3079

Dunn 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

Silhouette 0.7645 0.1427 0.1188 0.0946 0.1341 0.1171 0.0224 0.0470 -0.0180

***Optimal Scores:***

***Score Method Clusters***

***Connectivity 2.8647 sota 2***

***Dunn 0.4214 hierarchical 2***

***Silhouette 0.9831 hierarchical 2***

**op <- par(no.readonly = TRUE)**

**par(mfrow = c(2, 2), mar = c(4, 4, 3, 1))**

**plot(intern, legend = FALSE)**

**plot(nClusters(intern), measures(intern, "Dunn")[, , 1], type = "n",axes=F,xlab="",ylab="")**

**legend("center", clusterMethods(intern), col = 1:8, lty = 1:8,pch=paste(1:8))**

