

# 250518-Spectral-Clustering

2025-05-18

## Define Spectral Clustering Function

```
## Compute the Similarity Matrix with Gaussian Radial Kernel
S <- function(X) {
  exp(-0.5 * as.matrix(dist(X)))
}

## Weighted adjacency matrix with 10 nearest neighbors
knn = function(values, K){
  index = order(values, decreasing = TRUE)[(1:K) + 1]
  values[-index] = NA
  return(values)
}

## Compute the weight Matrix
W <- function(S, type = "Full", extra = NA) {
  if (type == "Soft") {
    lambda <- extra$lambda
    W <- S - lambda
    W[W < 0 ] <- 0

    ## Epsilon neighborhood
  } else if (type == "Epsilon") {
    S > epsilon

    ## Fully Connected Graph
  } else {
    W <- apply(S, 2, function(x) knn(x, K = 10))
    W = (W + t(W)) / 2
    W[is.na(W)] = 0
    W
  }
}

## Compute the Laplacian
L <- function(W) {
  D <- diag(rowSums(W))
  D - W
}

## Pick the K smallest eigenvalues of L and corresponding eigenvectors
bottomK <- function(L, K) {
```

```

n <- nrow(L)
E <- eigen(L)
U <- E$vectors[, (n-K+1):n]

U
}

## K-mean Cluster rows of U
KM <- function(U, K) {
  kmeans(U, centers = K, iter.max = 100, nstart = 10)$cluster
}

## Wrapper Function
spectralClust <- function(X, K, cancerType = "All", extra = NA) {
  s <- S(X) ## Similarity
  w <- W(s, "Full", extra) ## Weight
  l <- L(w) ## Laplacian
  u <- bottomK(l, K) ## Eigen values
  clustVec <- KM(u, K)

  par(mfrow = c(1,2))
  ## Plot first two eigenvectors, colored by cluster
  library(ggplot2)
  U2 <- data.frame(cbind(u[,1:2], clustVec))
  colnames(U2) <- c("U1", "U2", "C")

  plot(U2[,1], U2[,2], col = factor(clustVec),
       main = "Clustering in top 2 eigenvectors",
       xlab = "U1", ylab = "U2")

  ## Plot Graph Representation
  library(igraph)
  G = graph_from_adjacency_matrix(w, mode="undirected", weighted = TRUE)
  layout_circle <- layout_in_circle(G)
  plot(G, vertex.size=3, vertex.label=NA,
       layout = layout_circle,
       main = paste("Spectral Graph for ", cancerType))

  mtext(cancerType,
       side = 3, line = -2, outer = TRUE, cex = 1.5)

  clustVec
}

```

## Test Trial using Class Example

```

## Generate Data
set.seed(226)
n = 100
r1 = 0.3
r2 = 1
r3 = 1.8

```

```

X1 = matrix(rnorm(n * 2), n, 2)
X1 = X1/sqrt(rowSums(X1^2)) * r1
X2 = matrix(rnorm(n * 2), n, 2)
X2 = X2/sqrt(rowSums(X2^2)) * r2
X3 = matrix(rnorm(n * 2), n, 2)
X3 = X3/sqrt(rowSums(X3^2)) * r3
E = matrix(rnorm(3*n * 2, 0, 0.1), 3*n, 2)
X = rbind(X1, X2, X3) + E
X = scale(X, scale = F, center = T)
colnames(X) = c("x1", "x2")
K <- 3

specClusters <- spectralClust(X, K, "All")

```

```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

```
## Warning: package 'igraph' was built under R version 4.4.3
```

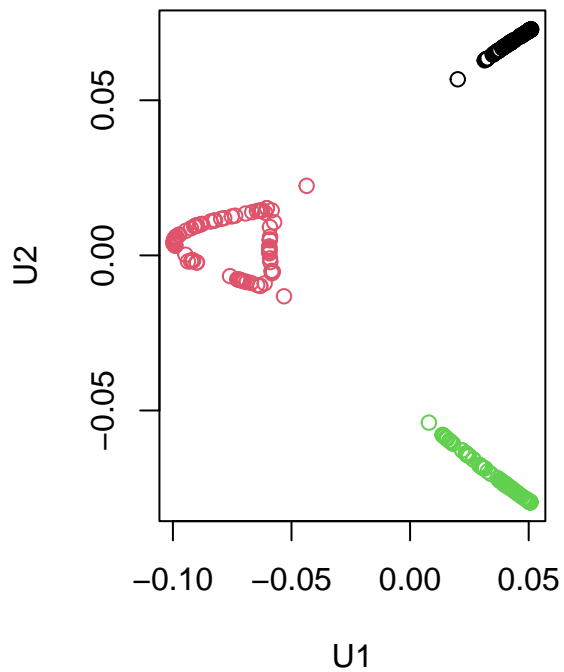
```
##
## Attaching package: 'igraph'
```

```
## The following object is masked _by_ '.GlobalEnv':
##
## knn
```

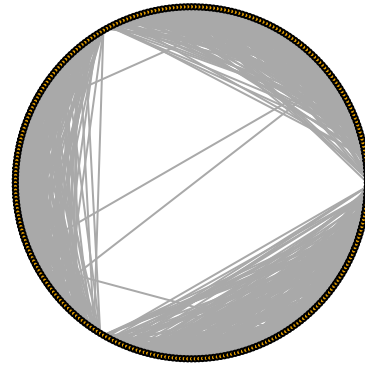
```
## The following objects are masked from 'package:stats':
##
## decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
## union
```

## Clustering in top 2 eigenvectors<sup>All</sup>

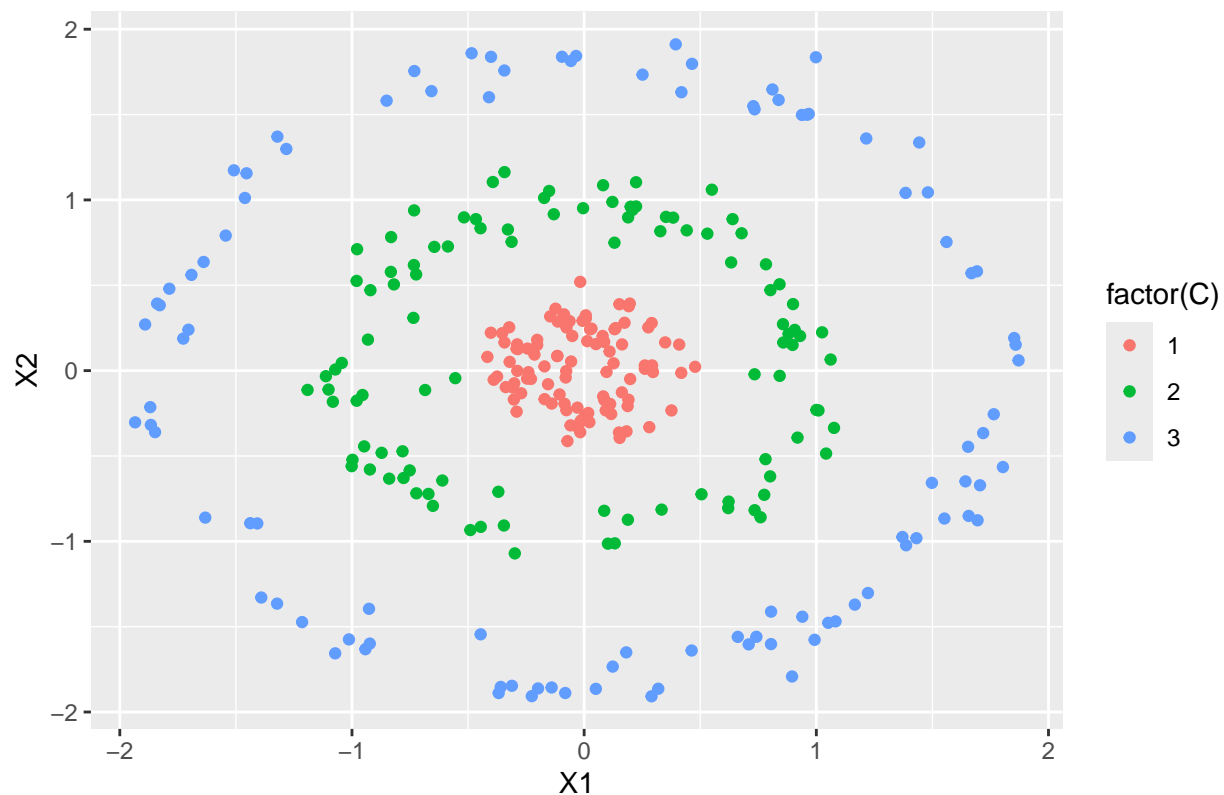


## Spectral Graph for All



```
ggplot(cbind(X, "C" = specClusters)) +  
  geom_point(aes(x1, x2, color = factor(C))) +  
  xlab("X1") +  
  ylab("X2") +  
  ggtitle("Clustering of genes in the the first two eigenvectors' space")
```

## Clustering of genes in the the first two eigenvectors' space



## Shared Gene Networks

```
library(tidyverse)
```

```
## Warning: package 'purrr' was built under R version 4.4.3
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.4      v readr      2.1.5
```

```
## v forcats    1.0.0      v stringr    1.5.1
```

```
## v lubridate  1.9.4      v tibble     3.2.1
```

```
## v purrr      1.0.4      v tidyr      1.3.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x lubridate::%--%()      masks igraph::%--%()
```

```
## x dplyr::as_data_frame() masks tibble::as_data_frame(), igraph::as_data_frame()
```

```
## x purrr::compose()       masks igraph::compose()
```

```
## x tidyr::crossing()      masks igraph::crossing()
```

```
## x dplyr::filter()        masks stats::filter()
```

```
## x dplyr::lag()           masks stats::lag()
```

```
## x purrr::simplify()      masks igraph::simplify()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
seCount <- readRDS(file = "Common_pan_cancer_hyper_bins_adjusted_and_normalized_cnt_in_SE.RDS")
dim(seCount)
```

```
## [1] 24418 378
```

```
sampleInfo <- read.csv("Common_pan_cancer_hyper_bins_adjusted_cnt_in_SE_samples_info.csv")
cancerTypes <- sampleInfo$cancer_type[sampleInfo$sample_id %in% colnames(seCount)]
```

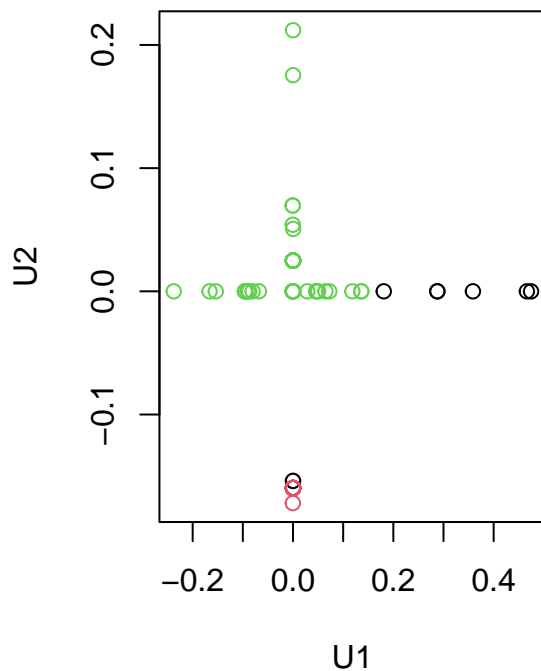
```
## Load gene numbers that are important in PCA
```

```
load("impGenes.RData")
```

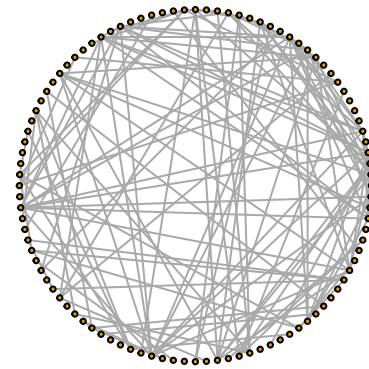
```
X <- seCount[impGenes, ]
```

```
sClusters <- spectralClust(X, 3, "All")
```

## Clustering in top 2 eigenvectors<sup>All</sup>



## Spectral Graph for All



## Cancer-Specific Gene Networks

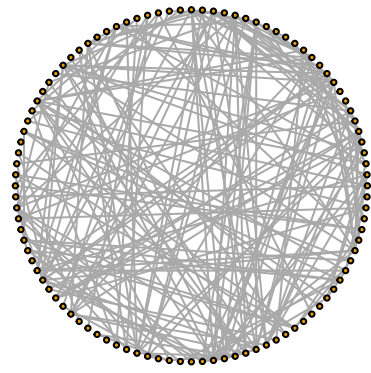
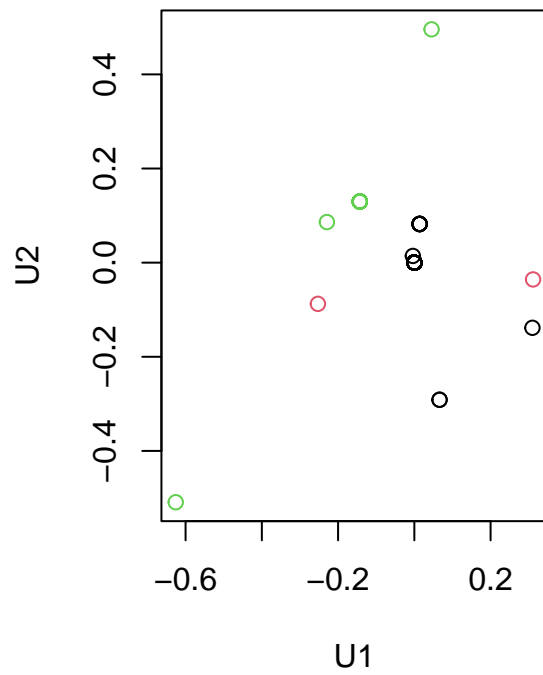
```
cancerFrameList <- split(data.frame(t(X)), cancerTypes)
```

```
for (cancer in names(cancerFrameList)) {
  paste("Cancer type:", cancer, "\n")
  X <- t(cancerFrameList[[cancer]])
  sClusters <- spectralClust(X, 3, cancer)
}
```

# Bladder Cancer

## Clustering in top 2 eigenvectors

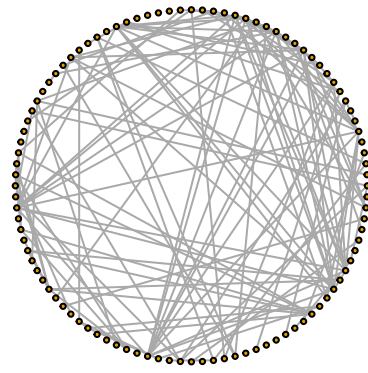
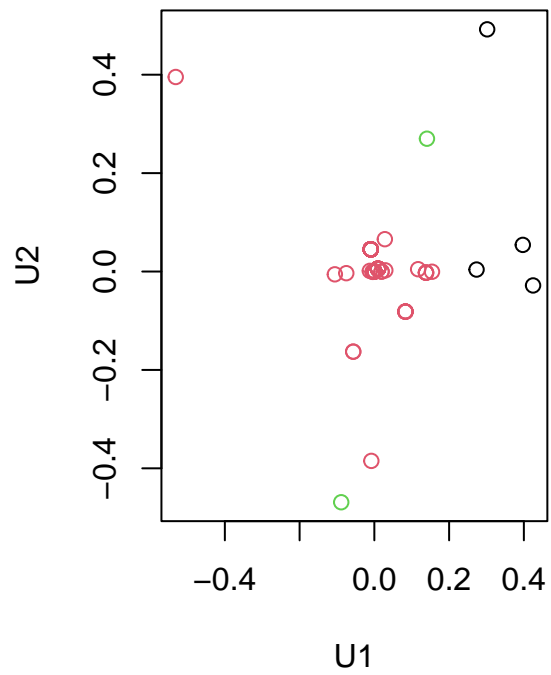
## Spectral Graph for Bladder Canc



# Blood Cancer

## Clustering in top 2 eigenvectors

## Spectral Graph for Blood Cancer

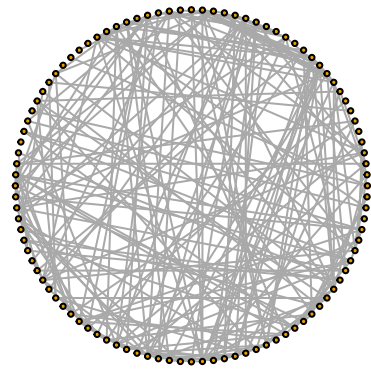
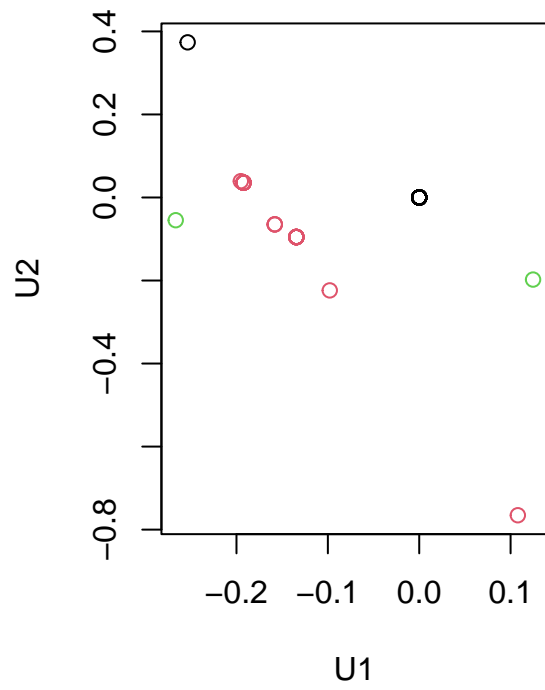




# Breast Cancer

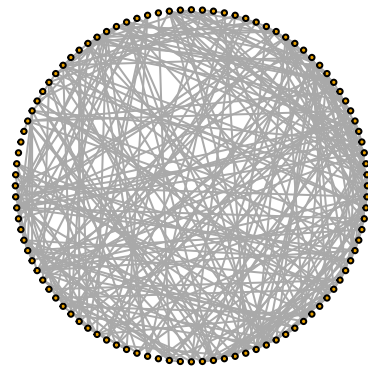
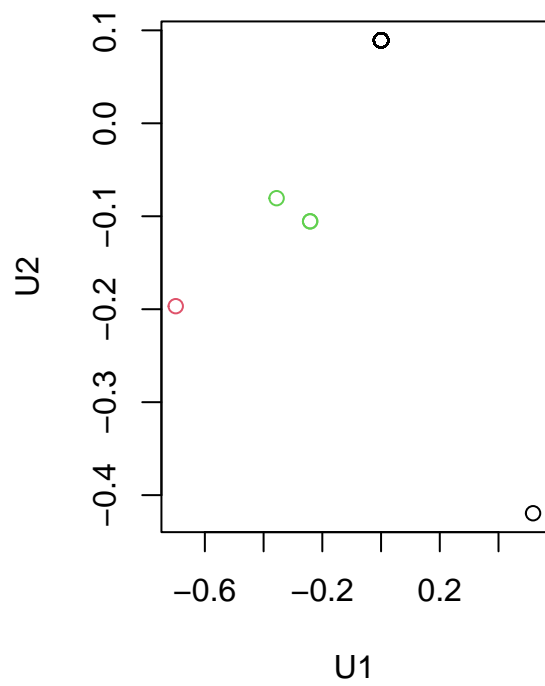
## Clustering in top 2 eigenvectors

## Spectral Graph for Breast Cancer



# Colorectal Cancer

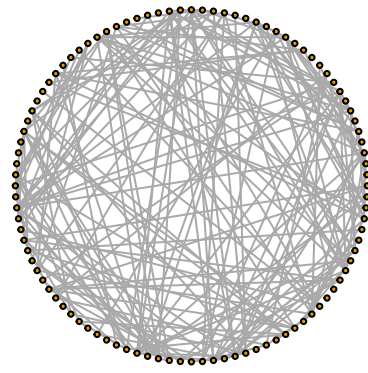
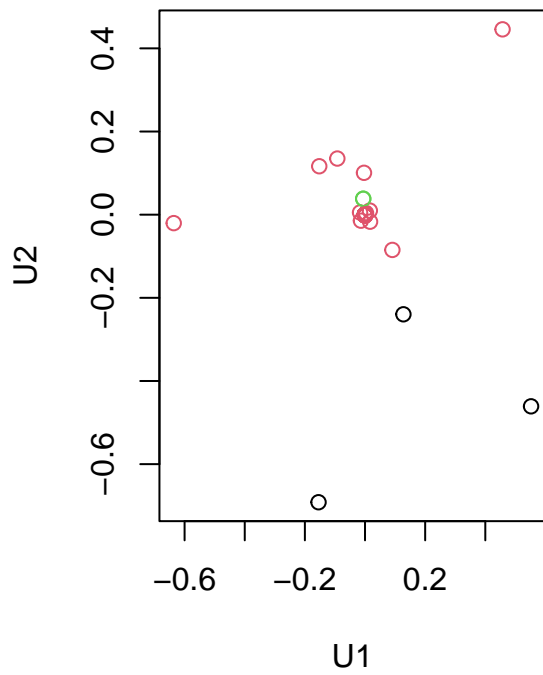
## Clustering in top 2 eigenvectors Spectral Graph for Colorectal Can



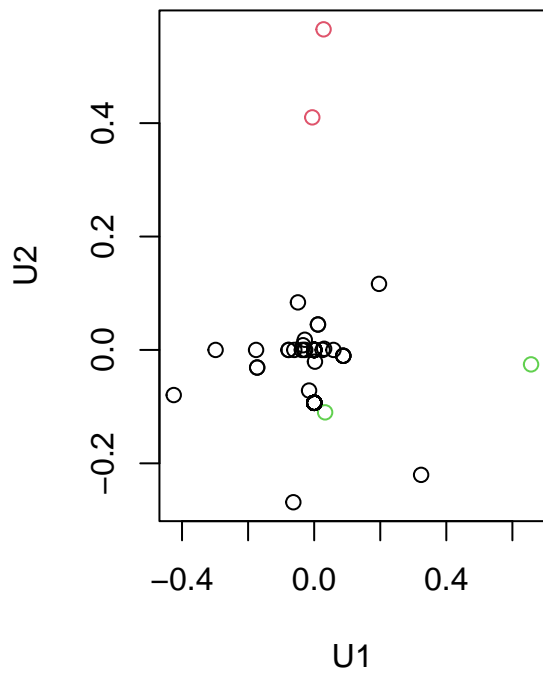
# Lung Cancer

## Clustering in top 2 eigenvectors

## Spectral Graph for Lung Cance

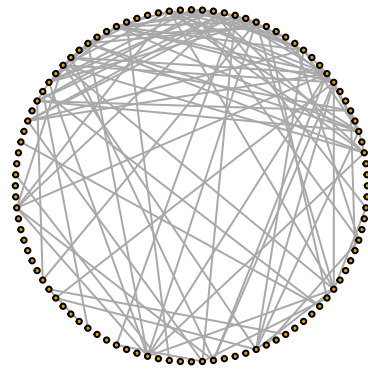


Clustering in top 2 eigenvectors



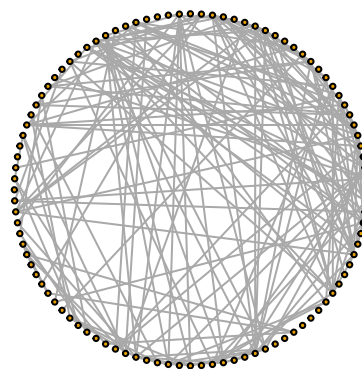
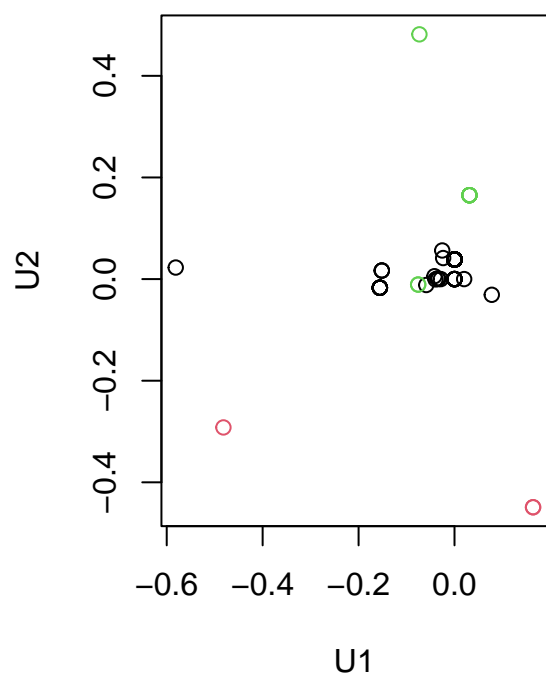
Normal

Spectral Graph for Normal



# Pancreatic Cancer

## Clustering in top 2 eigenvectors Spectral Graph for Pancreatic Can



# Renal Cancer

## Clustering in top 2 eigenvectors

## Spectral Graph for Renal Cance

