Example

This document provide some examples of using the functions in `mixedSCORE.R` for clustering analysis. See J. Jin, Ke, and Luo (2017) for the details of the method.

First source the code

```
source('mixedSCORE.R')
```

The main function is `mixedSCORE`

Input parameters:

- `A`: n-by-n adjacency matrix of the network
- `K`: number of clusters
- `verbose`: (optional) boolean, whether to generate messages, by default is False.

Outputs: a list containing

- `R`: n-by-(k-1) ratio matrix
- `L`: selected L by the vertex hunting algorithm
- `centers`: L cluster centers
- `vertices`: K vertices selected from L centers
- `memberships`: n-by-K matrix, the memberships of the n nodes
- `degrees`: a vector of length n, estimated degrees of each node
- `puritys`: a vector of length n, estimated purity (the maximum of memberships on K clusters) of each node
- `major.labels`: the hard clustering labels

Here we use the citee network data from Ji and Jin (2016) as an example, which is included as `citee.RData` in the repository. The citee network has 1790 nodes, where each node represents an author, and two nodes are connected if the two authors were once cited together.

```
load('citee.RData')
dim(citee)
```

```
## [1] 1790 1790
```

```
ms.out = mixedSCORE(citee, K = 3)
names(ms.out)
```

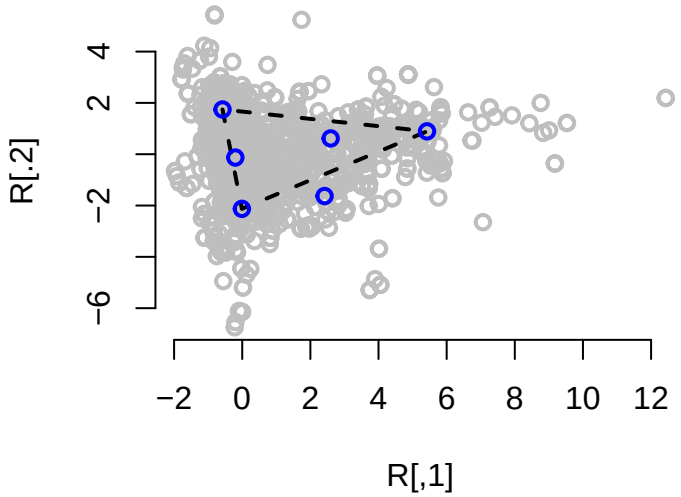
```
## [1] "R"            "L"            "vertices"     "centers"
## [5] "memberships"  "degrees"      "puritys"      "major.labels"
```

```
ms.out$L
```

```
## [1] 6
```

Plot the L vertices and the selected K cluster centers on top of the first two ratios

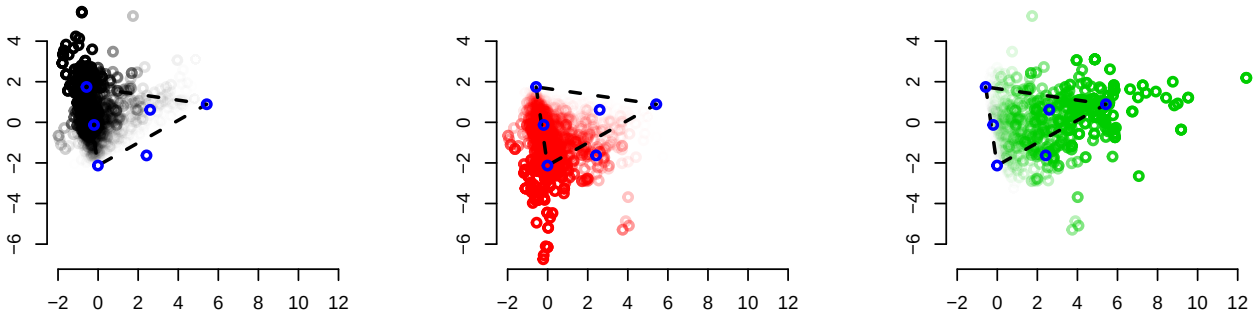
```
plot(ms.out$R, col='grey', lwd = 2, xlab = 'R[,1]', ylab = 'R[.2]',bty="n")
lines(ms.out$vertices[c(1,2,3,1),1], ms.out$vertices[c(1,2,3,1),2],
      lty = 2, lwd = 2, col = 'black')
points(ms.out$centers, lwd = 2, col = 'blue')
```



The membership for each of three clusters

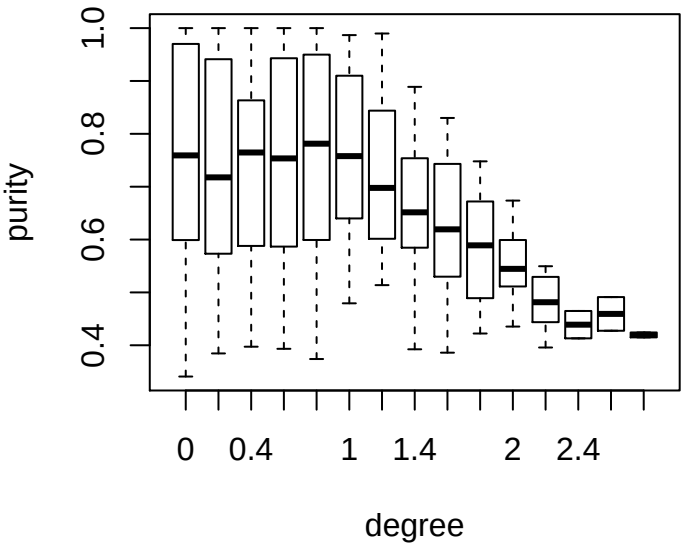
```
load('citee.RData')
# dim(citee) [1] 1790 1790
par(mfrow = c(1,3))
for (i in 1:3){

    plot(ms.out$R, col=scales::alpha(i, ms.out$memberships[,i]^2),
         lwd = 2, bty="n", xlab = '',ylab = '')
    lines(ms.out$vertices[c(1,2,3,1),1], ms.out$vertices[c(1,2,3,1),2],
      lty = 2, lwd = 2, col = 'black')
    points(ms.out$centers, lwd = 2, col = 'blue')
}
```



Plot the purity versus degree of nodes

```
boxplot(ms.out$puritys ~as.factor(round(ms.out$degrees*5)/5),
        bty = 'n', xlab = 'degree', ylab = 'purity')
```



Reference

Ji, P. S., and J. S. Jin. 2016. "Coauthorship and Citation Networks for Statisticians." Journal Article. *Annals of Applied Statistics* 10 (4): 1779–1812. doi:10.1214/15-Aoas896.

Jin, Jiashun, Zheng Tracy Ke, and Shengming Luo. 2017. "Estimating Network Memberships by Simplex Vertex Hunting." *arXiv Preprint arXiv:1708.07852*.