Multi-view Banded Spectral Clustering (mvBSC)

```
# devtools::install_github("celehs/mvBSC")
```

```
library(mvBSC)
library(data.table)
```

```
va_cosK <- readRDS(paste0("data/va_I00-I25_cosineMat.rds"))
bio_cosK <- readRDS(paste0("data/biobank_I00-I25_cosineMat.rds"))
this.R <- readRDS(paste0("data/I00-I25_distR_wt_avg_1.rds"))
icd.info <- readRDS("data/rollable_new_icd_info_20190130.rds")
codes_in_use <- colnames(bio_cosK)
codes_with_phecode <- codes_in_use[!is.na(icd.info[codes_in_use]$PheCode)]
pheCodes <- icd.info[codes_with_phecode]$PheCode
Z0 <- get_Z(codes_with_phecode, pheCodes)
delta0 <- min(apply(this.R, 1, max)) / 2
delta0
```

```
## [1] 24.5
```

```
initial <- mvbsc_fit(
  codes = codes_with_phecode, # rownames(this.R),
  distance =  this.R,
  similarity = list(va_cosK, bio_cosK),
  ncluster = 20,
  weights = c(va.wt = 0.5, bio.wt = 0.5),
  delta = delta0,
  band = 4,
  seed = 123)
initial
```

```
## $cluster
##   394.9   395.9   396.0   396.1   396.2   396.3   396.8   396.9   397.0
##       2       2      18      18      18      18      18      18      18
##   397.1   397.9  398.99   401.0   401.1   401.9  402.00  402.01  402.10
##      14      18      14       4       4       4      20      20      20
##  402.11  402.90  402.91  403.00  403.01  403.10  403.11  403.90  403.91
##      20      20      20       7       7       8       8       8       7
##  404.00  404.01  404.02  404.03  404.10  404.11  404.12  404.13  404.90
##       7       7       7       8       8       8       7       7       8
##  404.91  404.92  404.93  405.01  405.09  405.11  405.19  405.91  405.99
##       8       7       8      15      15      15       4      15      15
##  410.00  410.01  410.02  410.10  410.11  410.12  410.20  410.21  410.22
##      13      13      13      13      13      13      12      12      12
##  410.30  410.31  410.32  410.40  410.41  410.42  410.50  410.51  410.52
##      12      12      12      12      12      12       4       4      12
##  410.60  410.61  410.62  410.70  410.71  410.72  410.80  410.81  410.82
##       4      12      12       4      13      13      13      12      13
##  410.90  410.91  410.92   411.1  411.89   413.0   413.9  414.00  414.01
##      12      12      12      13       1      13      13      11      11
##  414.02  414.03  414.04  414.05  414.06  414.07  414.10  414.19   414.8
##       1       1       1       1      11      11       1       1      11
##   414.9   429.2  429.79    I00.   I01.0   I01.1   I01.2   I01.8   I01.9
##      11       1       4       4       9       9       4       9       4
```

```
##   I02.0   I02.9   I05.0   I05.1   I05.2   I05.8   I05.9   I06.0   I06.1
##       4       4       2       2       2       2       2       2       2
##   I06.2   I06.8   I06.9   I07.1   I07.2   I07.8   I07.9   I08.0   I08.1
##       2       2       2      10      10      10      10      10      10
##   I08.2   I08.3   I08.8   I08.9   I09.1   I09.2  I09.81  I09.89   I09.9
##      10      10      10       4       4       4      14      14      14
##    I10.   I11.0   I11.9   I12.0   I12.9   I13.0  I13.10  I13.11   I13.2
##       4       4       4      17       3       3       3      17      17
##   I15.0   I15.1   I15.2   I15.8   I15.9   I20.0   I20.1   I20.8   I20.9
##       6       6       6       6       6       9      13       9       9
##  I21.01  I21.02  I21.09  I21.11  I21.19  I21.21  I21.29   I21.3   I21.4
##      16      16      16      16      16      16      16      16       5
##   I22.0   I22.1   I22.2   I22.8   I22.9   I23.1   I23.2   I23.6   I23.7
##      16      16       5       5       5       5       5       5       5
##   I23.8   I24.0   I24.1   I24.8   I24.9  I25.10   I25.2   I25.3  I25.41
##       5       4       4      19      19      19      19      11      19
##  I25.42   I25.5   I25.6 I25.810 I25.811 I25.812  I25.82  I25.83  I25.84
##      19      19      11      19      19       4      19      19      19
##  I25.89   I25.9
##      19      19
##
## $cluster_info
##    cluster size max_dist
## 1       14    5     0.19
## 2       12   16     0.39
## 3        3    3     1.40
## 4       11    8     1.60
## 5       15    5     1.60
## 6        6    5     1.80
## 7       20    6     1.80
## 8        1    8     3.01
## 9       16   10     3.19
## 10      19   14     3.20
## 11       7    9     3.40
## 12       8    9     3.40
## 13      17    3     3.40
## 14      13   14     3.80
## 15      10    9     4.40
## 16       2   12     4.80
## 17      18    8     5.47
## 18       5    9     6.80
## 19       9    6    78.80
## 20       4   23    96.61
```

```
cluster0 <- subset(initial$cluster_info, max_dist > delta0)$cluster
cluster0
```

```
## [1] 9 4
```

```
regroup <- vector("list", length(cluster0))
# names(regroup) <- paste0("initial_", cluster0)
```

```
for (i in 1:length(cluster0)) {
  tmp <- names(initial$cluster[initial$cluster == cluster0[i]])
  for (k in 2:(length(tmp) - 1)) {
```

```
    try <- mvbsc_fit(
      codes = tmp,
      distance =  this.R,
      similarity = list(va_cosK, bio_cosK),
      ncluster = k,
      weights = c(va.wt = 0.5, bio.wt = 0.5),
      delta = delta0,
      band = 4,
      seed = 123)
    if (all(try$cluster_info$max_dist <= delta0)) break
  }
  regroup[[i]] <- try
}
regroup
```

```
## [[1]]
## [[1]]$cluster
## I01.0 I01.1 I01.8 I20.0 I20.8 I20.9
##     1     1     1     2     2     2
##
## [[1]]$cluster_info
##   cluster size max_dist
## 1       1    3      1.6
## 2       2    3      1.8
##
##
## [[2]]
## [[2]]$cluster
##   401.0   401.1   401.9  405.19  410.50  410.51  410.60  410.70  429.79
##       4       9       4      13       3       3       3       3      10
##    I00.   I01.2   I01.9   I02.0   I02.9   I08.9   I09.1   I09.2    I10.
##      11       1       7       7       2       7       7       6       5
##   I11.0   I11.9   I24.0   I24.1 I25.812
##      12      12      10      10       8
##
## [[2]]$cluster_info
##    cluster size max_dist
## 1        1    1     0.00
## 2        2    1     0.00
## 3        4    2     0.00
## 4        5    1     0.00
## 5        6    1     0.00
## 6        8    1     0.00
## 7        9    1     0.00
## 8       11    1     0.00
## 9       13    1     0.00
## 10       3    4     0.31
## 11      12    2     1.80
## 12      10    3     3.20
## 13       7    4    22.40
```