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 = rownames(this.R),
  distance =  this.R,
  similarity = list(va_cosK, bio_cosK),
  ncluster = 30,
  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
##      28      28      25      25      25      25       4       4      28
##   397.1   397.9  398.99   401.0   401.1   401.9  402.00  402.01  402.10
##       4      25       4      10      25      10      21      10      21
##  402.11  402.90  402.91  403.00  403.01  403.10  403.11  403.90  403.91
##      21      21      10       9       9      14      14      14      10
##  404.00  404.01  404.02  404.03  404.10  404.11  404.12  404.13  404.90
##       9       9       9       9       9      14       9      30      14
##  404.91  404.92  404.93  405.01  405.09  405.11  405.19  405.91  405.99
##      14       9       9      29      29      29      25      29      29
##  410.00  410.01  410.02  410.10  410.11  410.12  410.20  410.21  410.22
##       3       3       3       3       3       3      19      16      16
##  410.30  410.31  410.32  410.40  410.41  410.42  410.50  410.51  410.52
##      16      16      16       2       2       2      19      19       2
##  410.60  410.61  410.62  410.70  410.71  410.72  410.80  410.81  410.82
##      19      16       2      19       5       5       3       2       5
##  410.90  410.91  410.92   411.1  411.89   413.0   413.9  414.00  414.01
##       2       2       2       5      26       5       5      23      23
##  414.02  414.03  414.04  414.05  414.06  414.07  414.10  414.19   414.8
##      26      26      26      26      23      23      26      26      26
##   414.9   429.2  429.79    I00.   I01.0   I01.1   I01.2   I01.8   I01.9
##      23      11       7      25      17      17      17      17      17
```

```
##   I02.0   I02.9   I05.0   I05.1   I05.2   I05.8   I05.9   I06.0   I06.1
##      25      25       1       1       1      15      15       1      28
##   I06.2   I06.8   I06.9   I07.1   I07.2   I07.8   I07.9   I08.0   I08.1
##       1      15      15      27      27      27      27      27      27
##   I08.2   I08.3   I08.8   I08.9   I09.1   I09.2  I09.81  I09.89   I09.9
##      27      27      27      27      25      25       4       4       4
##    I10.   I11.0   I11.9   I12.0   I12.9   I13.0  I13.10  I13.11   I13.2
##      25      25      25      30      18      18      18      30      30
##   I15.0   I15.1   I15.2   I15.8   I15.9   I20.0   I20.1   I20.8   I20.9
##      24      24      24      24      24       8       5       8       8
##  I21.01  I21.02  I21.09  I21.11  I21.19  I21.21  I21.29   I21.3   I21.4
##      20       8       8      20      20       8       8       8       8
##   I22.0   I22.1   I22.2   I22.8   I22.9   I23.1   I23.2   I23.6   I23.7
##      20      20       8       7       7       7       7       7       7
##   I23.8   I24.0   I24.1   I24.8   I24.9  I25.10 I25.110 I25.111 I25.118
##       7       7       7       6      12       6      13       6       6
## I25.119   I25.2   I25.3  I25.41  I25.42   I25.5   I25.6 I25.700 I25.701
##       6       6      12      11      12       6      12      13      22
## I25.708 I25.709 I25.710 I25.711 I25.718 I25.719 I25.720 I25.721 I25.728
##      11      13      13      22      13      22      13      22      22
## I25.729 I25.739 I25.750 I25.758 I25.759 I25.760 I25.769 I25.790 I25.791
##      13      22      22      11      22      13      22      13      13
## I25.798 I25.799 I25.810 I25.811 I25.812  I25.82  I25.83  I25.84  I25.89
##      22      22      13      11      25      13       6       6       6
##   I25.9
##       6
##
## $cluster_info
##    cluster size max_dist
## 1       22   11     0.09
## 2       16    6     0.28
## 3        2    9     0.31
## 4        3    7     0.40
## 5       19    5     0.51
## 6       18    3     1.40
## 7       11    5     1.41
## 8       13   12     1.41
## 9       23    5     1.60
## 10      29    5     1.60
## 11      17    5     1.80
## 12      21    4     1.80
## 13      24    5     1.80
## 14      12    4     2.40
## 15      26    8     3.01
## 16      20    5     3.19
## 17       4    7     3.20
## 18       6   11     3.20
## 19      14    6     3.20
## 20      15    4     3.20
## 21       1    5     3.40
## 22       9   10     3.40
## 23      30    4     3.40
## 24       5    7     3.80
## 25       7   10     4.60
```

```
## 26      27   10     4.60
## 27      28    4     5.13
## 28      10    5     6.00
## 29       8   10     6.40
## 30      25   16    96.61
```

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

```
## [1] 25
```

```
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
```

```
## $initial_25
## $initial_25$cluster
##   396.0   396.1   396.2   396.3   397.9   401.1  405.19    I00.   I02.0
##       6       6       6       6       6       3       7       8       6
##   I02.9   I09.1   I09.2    I10.   I11.0   I11.9 I25.812
##      10       1       2       5       4       4       9
##
## $initial_25$cluster_info
##    cluster size max_dist
## 1        1    1      0.0
## 2        2    1      0.0
## 3        3    1      0.0
## 4        5    1      0.0
## 5        7    1      0.0
## 6        8    1      0.0
## 7        9    1      0.0
## 8       10    1      0.0
## 9        4    2      1.8
## 10       6    6     21.2
```