Figure4 Plot Code

Overview

This doc shows reproducible plot code to generate the Figure 4 result in manuscript *Community Detection with Heterogeneous Block Covariance Model.*

Install packages

Install the HBCM package from github and other helper packages.

```
# install helper package
packages_helper <- c(
  'devtools', "parallel",
  "foreach", "doParallel",
  "kernlab", "matrixcalc",
  "tidyverse", "ggpubr",
  "scales", "ggforce",
  "patchwork"
  )
install.packages(packages_helper, repos = "http://cran.us.r-project.org")

# install hbcm package
devtools::install_github("xiangli2pro/hbcm")

# load all required packages
lapply(c('hbcm', packages_helper), require, char = TRUE)

# set random seed to reproduce the result
set.seed(2022)
```

Gene expression data

Gene expression data used in the manuscript is already loaded once the HBCM package is loaded. Execute the command below to see details in the Help panel on how data is processed.

```
?hbcm::data_gene_sample
?hbcm::data_gene_group
```

Assume there are 5 cluster groups, use HBCM to estimate the clusters of the genes.

```
# assume K=5
centers <- 5

# spectral cluster estimation
spec_labels <- rSpecc(abs(cov(data_gene_sample)), centers = centers)$.Data

# HBCM cluster estimation
hbcm_res <- hbcm::heterogbcm(scale(data_gene_sample, center = TRUE, scale = FALSE),
  centers = centers,
  tol = 1e-3, iter = 100, iter_init = 3,
  labels = spec_labels,
  verbose = FALSE
)
```

The paper that generated the gene expression data has a recommended cluster assignment for the genes, which is used as truth to compare with the estimated clusters.

```
# get recommended cluster label
matchProbe <- match(names(data_gene_sample), data_gene_group$Probe)
matchTF <- data_gene_group$TF_regulator[matchProbe]
uniTFs <- unique(matchTF)

# convert to numerical label
trueLabel <- rep(0, ncol(data_gene_sample))
for (i in c(1:length(uniTFs))) {
  trueLabel[which(matchTF == uniTFs[i])] <- i
}
table(trueLabel)
```

Use the adjusted rand index (adjRand) to evaluate the estimated cluster performance.

```
# adjRand of recommended label VS HBCM
matchLabel(trueLabel, hbcm_res$cluster)$adjRand
```

[1] 0.1876328

Make heat-map plot

Rearrange the genes by their HBCM estimated clusters and make heatmap plot. Note that the black line around each block is added intentionally to visually distinguish the clusters.

```
hbcm::colMat_heatMap(
  affMatrix = cor(data_gene_sample),
  centers,
  labels = hbcm_res$cluster,
  margin = 0.5, midpoint = 0, limit = c(-1,1), size = 0.2,
  legendName = "Correlation", title = "HeatMap of correlation by HBCM groups")
```


HeatMap of correlation by HBCM groups
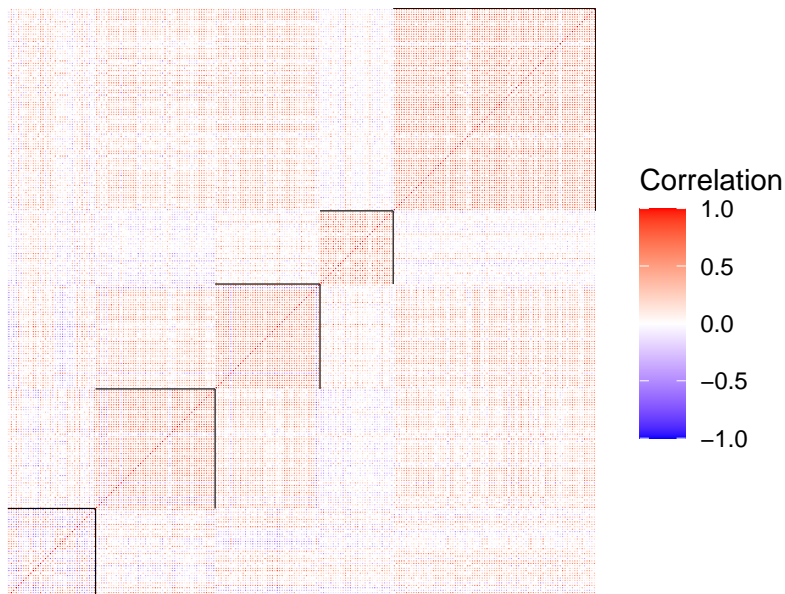
```
# ggsave("plots/heatmap_HBCM.png", dpi = 300)
```