Package: mlsbm (via r-universe)

September 6, 2024

Title Efficient Estimation of Bayesian SBMs & MLSBMs

Version 0.99.4

Description Fit Bayesian stochastic block models (SBMs) and multi-level stochastic block models (MLSBMs) using efficient Gibbs sampling implemented in 'Rcpp'. The models assume symmetric, non-reflexive graphs (no self-loops) with unweighted, binary edges. Data are input as a symmetric binary adjacency matrix (SBMs), or list of such matrices (MLSBMs).

License GPL (>= 2)

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.2

LinkingTo Rcpp, RcppArmadillo

Imports Rcpp, igraph, stats, bluster, tidyr, tidyselect, rlang, dplyr, ggplot2, ggraph, patchwork

Depends R (>= 2.10)

Repository https://carter-allen.r-universe.dev

RemoteUrl https://github.com/carter-allen/mlsbm

RemoteRef HEAD

RemoteSha f42504b709a62ad54559242d950fbeb2e9d88cd7

Contents

AL 2
col_summarize 2
fit_mlsbm 3
fit_sbm 4
get_scores 5
mean_CRI 5
mlsbm 6

plot_connectivity_matrix 6
plot_connectivity_network 7
remap_canonical2 7
sample_mlsbm 8
sample_sbm 9

Index **10**

`AL` *Simulated 3-layer network data*

Description

A data set containing 3 layers of undirected, symmetric adjacency matrices simulated from an SBM with 3 true clusters

Usage

```
AL
```

Format

A list of length 3

`col_summarize` *The col_summarize function*

Description

Function to quickly return credible intervals

Usage

```
col_summarize(MAT, dig = 2, level = 0.95)
```

Arguments

`MAT`	A matrix
`dig`	Number of digits to round estimates and CrIs to
`level`	Confidence level

Value

A character vector of posterior estimates and intervals

Examples

```
M <- matrix(rnorm(1000),ncol = 4)
col_summarize(M)
```

fit_mlsbm *R/Rcpp function for fitting multilevel stochastic block model*

Description

This function allows you to fit multilevel stochastic block models.

Usage

```
fit_mlsbm(
  A,
  K,
  z_init = NULL,
  a0 = 2,
  b10 = 1,
  b20 = 1,
  n_iter = 1000,
  burn = 100,
  verbose = FALSE,
  r = 1.2
)
```

Arguments

`A` An adjacency list of length L, the number of levels. Each level contains an n x n symmetric adjacency matrix.

`K` The number of clusters specified a priori.

`z_init` Initialized cluster indicators. If NULL, will initialize automatically with Louvain algorithm.

`a0` Dirichlet prior parameter for cluster sizes for clusters 1,...,K.

`b10` Beta distribution prior paramter for community connectivity.

`b20` Beta distribution prior parameter for community connectivity.

`n_iter` The number of total MCMC iterations to run.

`burn` The number of burn-in MCMC iterations to discard. The number of saved iterations will be n_iter - burn.

`verbose` Whether to print a progress bar to track MCMC progress. Defaults to true.

`r` Resolution parameter for Louvain initialization. Sould be >= 0 and higher values give a larger number of smaller clusters.

Value

A list of MCMC samples, including the MAP estimate of cluster indicators (z)

Examples

```
data(AL)
# increase n_iter in practice
fit <- fit_mlsbm(AL,3,n_iter = 100)
```

fit_sbm *R/Rcpp function for fitting single level stochastic block model*

Description

This function allows you to fit single level stochastic block models.

Usage

```
fit_sbm(
  A,
  K,
  z_init = NULL,
  a0 = 1,
  b10 = 2,
  b20 = 2,
  n_iter = 1000,
  burn = 100,
  verbose = FALSE,
  r = 1.2
)
```

Arguments

A	An n x n symmetric adjacency matrix.
K	The number of clusters specified a priori.
z_init	Initialized cluster indicators. If NULL, will initialize automatically with Louvain algorithm.
a0	Dirichlet prior parameter for cluster sizes for clusters 1,...,K.
b10	Beta distribution prior paramter for community connectivity.
b20	Beta distribution prior parameter for community connectivity.
n_iter	The number of total MCMC iterations to run.
burn	The number of burn-in MCMC iterations to discard. The number of saved iterations will be n_iter - burn.
verbose	Whether to print a progress bar to track MCMC progress. Defaults to true.
r	Resolution parameter for Louvain initialization. Sould be >= 0 and higher values give a larger number of smaller clusters.

Value

A list of MCMC samples, including the MAP estimate of cluster indicators (z)

Examples

```
data(AL)
fit <- fit_sbm(AL[[1]],3)
```

get_scores *Calculate continuous uncertainty scores*

Description

This function allows you to augment the discrete cell type assignments with continuous propensity and uncertainty scores

Usage

```
get_scores(fit)
```

Arguments

`fit`	A list returned by fit_sbm() or fit_mlsbm()

Value

A list with populated entries C_scores (N x K matrix for cell type propensities) and U_scores (N x 1 vector of uncertainty scores)

mean_CRI *The mean_CRI function*

Description

Simple function to return the mean (95% CrI) for a vector

Usage

```
mean_CRI(y, dig = 2)
```

Arguments

`y`	A numeric vector
`dig`	The number of digits to round to

Value

A string of mean and 95% quantile interval rounded to 'dig'

Examples

```
mean_CRI(rnorm(1000))
```

mlsbm *mypackage: A package for fitting single and multilevel SBMs.*

Description

This package fits Bayesian stochastic block models (SBMs)

mlsbm functions

The mlsbm functions ...

plot_connectivity_matrix *Plot community structure of cell sub-populations as matrix*

Description

This function allows you to visualize the community structure of cell sub-populations in matrix format via the connectivity parameters of the BANYAN model

Usage

```
plot_connectivity_matrix(fit)
```

Arguments

`fit`	A list returned by fit_banyan().

Value

A ggplot object

`plot_connectivity_network` *Plot community structure parameters as a K x K network*

Description

This function allows you to visualize the inferred community structure as a community-community connectivity network

Usage

```
plot_connectivity_network(fit)
```

Arguments

`fit`	A list returned by fit_sbm() or fit_mlsbm()

Value

A ggplot object

`remap_canonical2` *Canonical re-mapping of mixture component labels*

Description

Avoid label switching by re-mapping sampled mixture component labels at each iteration (Peng and Carvhalo 2016).

Usage

```
remap_canonical2(z)
```

Arguments

`z`	A length-n vector of discrete mixture component labels

Value

A length-n vector of mixture component labels re-mapped to a canonical sub-space

Examples

```
# parameters
n <- 10 # number of observations
K <- 3 # number of clusters (mixture components)
pi <- rep(1/K,K) # cluster membership probability
z <- sample(1:K, size = n, replace = TRUE, prob = pi) # cluster indicators
z <- remap_canonical2(z)
```

sample_mlsbm *R/Rcpp function for sampling from a multilevel stochastic block model*

Description

This function allows you to sample a multilevel stochastic block model.

Usage

```
sample_mlsbm(z, P, L)
```

Arguments

z	An n x 1 vector of community labels for each node
P	A K x K symmetric matrix of community connectivity probabilities
L	The number of levels to sample

Value

A list of adjecency matrices – one for each level of the MLSBM

Examples

```
n = 100
K = 3
L = 2
pi = rep(1/K,K)
z = sample(1:K, size = n, replace = TRUE, prob = pi)
p_in = 0.50
p_out = 0.05
P = matrix(p_out, nrow = K, ncol = K)
diag(P) = p_in
AL = sample_mlsbm(z,P,L)
```

`sample_sbm` *R/Rcpp function for sampling from a single level stochastic block model*

Description

This function allows you to sample a single level stochastic block model.

Usage

```
sample_sbm(z, P)
```

Arguments

`z`	An n x 1 vector of community labels for each node
`P`	A K x K symmetric matrix of community connectivity probabilities

Value

An adjacency matrix

Examples

```
n = 100
K = 3
pi = rep(1/K,K)
z = sample(1:K, size = n, replace = TRUE, prob = pi)
p_in = 0.50
p_out = 0.05
P = matrix(p_out, nrow = K, ncol = K)
diag(P) = p_in
A = sample_sbm(z,P)
```

Index

∗ **Bayesian**

fit_mlsbm, 3
fit_sbm, 4
get_scores, 5
plot_connectivity_matrix, 6
plot_connectivity_network, 7

∗ **Credible**

col_summarize, 2

∗ **Gibbs**

fit_mlsbm, 3
fit_sbm, 4
get_scores, 5
plot_connectivity_matrix, 6
plot_connectivity_network, 7

∗ **MCMC**

mean_CRI, 5

∗ **MLSBM**

fit_mlsbm, 3
fit_sbm, 4
get_scores, 5
plot_connectivity_matrix, 6
plot_connectivity_network, 7

∗ **SBM**

fit_mlsbm, 3
fit_sbm, 4
get_scores, 5
plot_connectivity_matrix, 6
plot_connectivity_network, 7
sample_mlsbm, 8
sample_sbm, 9

∗ **datasets**

AL, 2

∗ **expression**

get_scores, 5
plot_connectivity_matrix, 6
plot_connectivity_network, 7

∗ **gene**

get_scores, 5
plot_connectivity_matrix, 6
plot_connectivity_network, 7

∗ **intervals**

col_summarize, 2

∗ **networks**

fit_mlsbm, 3
fit_sbm, 4
get_scores, 5
plot_connectivity_matrix, 6
plot_connectivity_network, 7

∗ **spatial**

get_scores, 5
plot_connectivity_matrix, 6
plot_connectivity_network, 7

AL, 2

col_summarize, 2

fit_mlsbm, 3
fit_sbm, 4

get_scores, 5

mean_CRI, 5
mlsbm, 6

plot_connectivity_matrix, 6
plot_connectivity_network, 7

remap_canonical2, 7

sample_mlsbm, 8
sample_sbm, 9