Package ‘ergmclust’

February 1, 2021

Type Package
Title ERGM-Based Network Clustering
Version 1.0.0
Date 2021-01-24
Description Implements clustering and estimates parameters in Exponential-Family Random Graph Models (ERGMs) for static undirected and directed networks, developed in Vu et. al. (2013) <https://projecteuclid.org/euclid.aoas/1372338477>.
Encoding UTF-8
License GPL-2
Imports Rcpp (>= 1.0.1), MASS, lda, quadprog, igraph, viridis
LinkingTo Rcpp, RcppArmadillo
Depends R (>= 2.10)
URL https://sites.psu.edu/sldm/netclust/
NeedsCompilation yes
Author Amal Agarwal [aut, cre],
Kevin H. Lee [aut],
Lingzhou Xue [aut, ths],
Anna Yinqi Zhang [com]
Maintainer Amal Agarwal <amalag.19@gmail.com>
Repository CRAN
Date/Publication 2021-02-01 08:50:02 UTC

R topics documented:

  ergmclust-package ............................................................ 2
  armsnet ................................................................. 3
  ergmclust .............................................................. 3
  ergmclust.ICL ............................................................ 5
  ergmclust.plot .......................................................... 7
  tradenet ................................................................. 9

Index 10
Description

Clustering and estimation of parameters in ERGMs for static undirected and directed networks with inference based on VEM algorithm.

Details

The ergmclust package is an R implementation that serves as an estimation framework for static binary networks, in both undirected and directed cases. Its main functions include ergmclust for clustering and parameter estimation, ergmclust.ICL for model selection, and ergmclust.plot for visualizing the clustered network. The package is based on VEM algorithm (Vu et al., 2013) and works well with both simulated and real-world data.

Author(s)

Authors: Amal Agarwal [aut, cre], Kevin Lee [aut], Lingzhou Xue [aut, cre], Anna Yinqi Zhang [cre]
Maintainer: Amal Agarwal <amalag19@gmail.com>

References

Agarwal, A. and Xue, L. (2019) Model-Based Clustering of Nonparametric Weighted Networks With Application to Water Pollution Analysis, Technometrics, to appear

https://ieeexplore.ieee.org/document/865189


https://arxiv.org/abs/1712.07325

https://projecteuclid.org/euclid.aos/1372338477
armsnet

Arms Trade Network Data in 2003.

Description
The directed network on all transfers of major conventional weapons internationally. We define the edges as $y_{ij} = 1$, if the volume of international transfers of arms, measured by Trend Indicator Value (TIV) from country $i$ to country $j$ exceeds 1 million dollars, and $y_{ij} = 0$ otherwise.

Usage
data(armsnet)

Format
The format is a $69 \times 69$ network adjacency matrix.

Source
https://www.sipri.org/databases/armstransfers

References

Examples
data(armsnet)

---

ergmclust

Model-Based Clustering of Large Networks Through ERGMs.

Description
Model-based clustering and cluster-specific parameter estimation through the mixed membership Exponential-Family Random Graph Models (ERGMs) using Variational Expectation-Maximization algorithm.

Usage
ergmclust(adjmat, K, directed = FALSE, thresh = 1e-06, iter.max = 200, coef.init = NULL)
Arguments

adjmat
An object of class matrix of dimension (N x N) containing the adjacency matrix, where N is the number of nodes in the network.

K
Number of clusters in the mixed membership Exponential-Family Random Graph Models (ERGMs).

directed
If TRUE, the network is supposed to be directed (and therefore adjmat must be asymmetric in general). By default, this is set as FALSE.

thresh
Optional user-supplied convergence threshold for relative error in the objective in Variational Expectation-Maximization (VEM) algorithm. The default value is set as 1e-06.

iter.max
The maximum number of iterations after which the algorithm is terminated. The default value is set as 200.

coef.init
Optional user-supplied network canonical parameter vector (K-dimensional). Default is NULL and ergmclust chooses a random perturbation around zero vector.

Details

ergmclust is an R implementation for the model-based clustering through the mixed membership Exponential-Family Random Graph Models (ERGMs) with undirected and directed network data. It uses the Variational Expectation-Maximization algorithm to solve the approximate maximum likelihood estimation.

Value

Returns a list of ergmclust object. Each object of class ergmclust is a list with the following components:

coefficients
An object of class vector of size (K x 1) containing the canonical network parameters in Exponential-Family Random Graph Models (ERGMs).

probability
An object of class matrix of size (N x K) containing the mixed membership probabilities of the model for N nodes distributed in K clusters.

clust.labels
An object of class vector of size (N x 1) containing the cluster membership labels in \{1, ..., K\} for N nodes.

ICL
Integrated Classification Likelihood (ICL) score calculated from completed data log-likelihood and penalty terms.

Author(s)

Authors: Amal Agarwal [aut, cre], David R. Hunter [aut], Kevin Lee [aut], Lingzhou Xue [aut, cre], Anna Yinqi Zhang [cre]

Maintainer: Amal Agarwal <amalag.19@gmail.com>
References

Agarwal, A. and Xue, L. (2019) Model-Based Clustering of Nonparametric Weighted Networks With Application to Water Pollution Analysis, Technometrics, to appear


https://arxiv.org/abs/1712.07325

https://projecteuclid.org/euclid.aos/1372338477

Examples

```r
## undirected network:
data(tradenet)
## clustering and estimation for K = 2 groups
ergmclust(adjmat = tradenet, K = 2, directed = FALSE,
thresh = 1e-06, iter.max = 120, coef.init = NULL)

## directed network:
data(armsnet)
## clustering and estimation for K = 2 groups
ergmclust(adjmat = armsnet, K = 2, directed = TRUE,
thresh = 1e-06, iter.max = 120, coef.init = NULL)
```

Description

Model-based clustering and cluster-specific parameter estimation through the mixed membership Exponential-Family Random Graph Models (ERGMs) for the different number of clusters. Model selection is based on maximum value of Integrated Classification Likelihood (ICL).

Usage

```r
ergmclust.ICL(adjmat, Kmax = 5, directed = FALSE,
thresh = 1e-06, iter.max = 200, coef.init = NULL)
```
Arguments

adjmat
An object of class matrix of dimension (N x N) containing the adjacency matrix, where N is the number of nodes in the network.

Kmax
Maximum number of clusters.

directed
If TRUE, the network is supposed to be directed (and therefore adjmat must be asymmetric in general). By default, this is set as FALSE.

thresh
Optional user-supplied convergence threshold for relative error in the objective in Variational Expectation-Maximization (VEM) algorithm. The default value is set as 1e-06.

iter.max
The maximum number of iterations after which the algorithm is terminated. The default value is set as 200.

coef.init
Optional user-supplied network canonical parameter vector (K-dimensional). Default is NULL, and ergmclust chooses a random perturbation around zero vector.

Details

ergmclust.ICL is an R implementation for the model selection for an appropriate number of clusters in the mixed membership Exponential-Family Random Graph Models (ERGMs). The Integrated Classification Likelihood (ICL) was proposed by Biernacki et al. (2000) and Daudin, et. al. (2008) to assess the model-based clustering.

Value

Returns a list of ergmclust object. Each object of class ergmclust is a list with the following components:

Kselect
Optimum number of clusters chosen after model selection through Integrated Classification Likelihood (ICL).

coefficients
An object of class vector of size (Kselect x 1) containing the canonical network parameters of the model.

probability
An object of class matrix of size (N x Kselect) containing the mixed membership probabilities of the model for N nodes distributed in Kselect clusters.

clust.labels
An object of class vector of size (N x 1) containing the cluster membership labels in {1, ..., Kselect} for N nodes.

ICL
Integrated Classification Likelihood (ICL) score calculated from completed data log-likelihood and penalty terms.

Author(s)

Authors: Amal Agarwal [aut, cre], Kevin Lee [aut], Lingzhou Xue [aut, cre], Anna Yinqi Zhang [cre]
Maintainer: Lingzhou Xue <lzxue@psu.edu>
**References**


https://ieeexplore.ieee.org/document/865189


**Examples**

```r
## undirected network:
data(tradenet)
## Model selection for Kmax = 3
ergmclust.ICL(adjmat = tradenet, Kmax = 3, directed = FALSE,
thresh = 1e-06, iter.max = 120, coef.init = NULL)

## directed network:
data(armsnet)
## Model selection for Kmax = 3
ergmclust.ICL(adjmat = armsnet, Kmax = 3, directed = TRUE,
thresh = 1e-06, iter.max = 60, coef.init = NULL)
```

**Description**

Visualization of the network data with the clusters node colors representing different clusters in the Exponential-Family Random Graph Models (ERGMs) clustered network.

**Usage**

```r
ergmclust.plot(adjmat, K, directed = FALSE, thresh = 1e-06,
iter.max = 200, coef.init = NULL, node.labels = NULL)
```

**Arguments**

- **adjmat**: An object of class matrix of dimension (N x N) containing the adjacency matrix, where N is the number of nodes in the network.
- **K**: Number of clusters in the mixed membership Exponential-Family Random Graph Models (ERGMs).
- **directed**: If TRUE, the network is supposed to be directed (and therefore adjmat must be asymmetric in general). By default, this is set as FALSE.
thresh

Optional user-supplied convergence threshold for relative error in the objective in Variational Expectation-Maximization (VEM) algorithm. The default value is set as $1e-06$.

iter.max

The maximum number of iterations after which the algorithm is terminated. The default value is set as 200.

coeff.init

Optional user-supplied network canonical parameter vector ($K$-dimensional); default is NULL, and ergmclust chooses a random perturbation around zero vector.

node.labels

Optional user-supplied network node names character vector ($N$-dimensional); default is NULL.

Details

ergmclust.plot provides the visualization tool for network data clustered through mixed membership Exponential-Family Random Graph Models (ERGMs). The optional argument node.labels could help track the cluster membership of specific nodes.

Value

Returns a plot of network object with colored nodes corresponding to $K$ clusters.

Author(s)

Authors: Amal Agarwal [aut, cre], Kevin Lee [aut], Lingzhou Xue [aut, cre], Anna Yinqi Zhang [cre]

Maintainer: Amal Agarwal <amalag.19@gmail.com>

References


https://projecteuclid.org/euclid.aoas/1372338477

Examples

```r
## undirected network:
data(tradenet)
## Plotting clustered network
ergmclust.plot(adjmat = tradenet, K = 2, directed = FALSE, thresh = 1e-06)

## directed network:
data(armsnet)
## Plotting clustered network
ergmclust.plot(adjmat = armsnet, K = 2, directed = TRUE, thresh = 1e-06)
```
The undirected network on all trade relations internationally among 58 countries. We define the edges as $y_{ij} = 1$, if there is a bilateral trade between country $i$ and $j$, and $y_{ij} = 0$ otherwise.

The format is a $58 \times 58$ network adjacency matrix.

https://projecteuclid.org/euclid.aoas/1310562208#supplemental


https://projecteuclid.org/euclid.aoas/1310562208

Examples

data(tradenet)
Index

armsnet, 3

ergmclust, 3
ergmclust-package, 2
ergmclust.ICL, 5
ergmclust.plot, 7

tradenet, 9