

# Package ‘ergmclust’

January 8, 2022

**Type** Package  
**Title** ERGM-Based Network Clustering  
**Version** 2.0.1  
**Date** 2022-01-08  
**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, locfit  
**LinkingTo** Rcpp, RcppArmadillo  
**Depends** R (>= 2.10)  
**URL** <https://sites.psu.edu/sldm/netclust/>

## R topics documented:

ergmclust-package . . . . .	1
armsnet . . . . .	2
ergmclust . . . . .	3
ergmclust.ICL . . . . .	5
ergmclust.plot . . . . .	7
tradenet . . . . .	8
<b>Index</b>	<b>9</b>

---

ergmclust-package	<i>ERGM-based network clustering</i>
-------------------	--------------------------------------

---

## 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 <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://amstat.tandfonline.com/doi/abs/10.1080/00401706.2019.1623076>

Biernacki, C., Celeux, G., and Govaert, G. (2000) *Assessing a mixture model for clustering with the integrated completed likelihood*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22(7), 719-725

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

Blei, D. M. , Kucukelbir, A., and McAuliffe, J. D. (2017), *Variational Inference: A Review for Statisticians*, *Journal of the American Statistical Association*, Vol. 112(518), 859-877

<https://www.tandfonline.com/doi/full/10.1080/01621459.2017.1285773>

Daudin, J. J., Picard, F., and Robin, S. (2008) *A Mixture Model for Random Graphs*, *Statistics and Computing*, Vol. 18(2), 173-183

<https://link.springer.com/article/10.1007/s11222-007-9046-7>

Lee, K. H., Xue, L., and Hunter, D. R. (2017) *Model-Based Clustering of Time-Evolving Networks through Temporal Exponential-Family Random Graph Models*, *Journal of Multivariate Analysis*, to appear

<https://arxiv.org/abs/1712.07325>

Vu D. Q., Hunter, D. R., and Schweinberger, M. (2013) *Model-based Clustering of Large Networks*, *The Annals of Applied Statistics*, Vol. 7(2), 1010-1039

<https://projecteuclid.org/euclid.aoas/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**

Akerman, A., & Seim, A. L. (2014) *The global arms trade network 1950–2007*, Journal of Comparative Economics, Vol. 42(3), 535-551

<https://www.sciencedirect.com/journal/journal-of-comparative-economics/vol/42/issue/3>

**Examples**

```
data(armsnet)
```

---

ergmclust	<i>Model-Based Clustering of Large Networks Through ERGMs.</i>
-----------	--

---

**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, weighted = FALSE, thresh = 1e-06,
iter.max = 200, coef.init = NULL, wtmat = 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 is must be asymmetric in general). By default, this is set as FALSE.
weighted	If TRUE, the network is supposed to be weighted. 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	ergmclust chooses the default value as a random perturbation around K-dim zero vector; default is NULL.
wtmat	An object of class matrix of dimension (N x N) containing the weight matrix, where N is the number of nodes in the network; default is NULL.

## 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 $(K \times 1)$ containing the canonical network parameters in Exponential-Family Random Graph Models (ERGMS).
probability	An object of class matrix of size $(N \times K)$ containing the mixed membership probabilities of the model for $N$ nodes distributed in $K$ clusters.
clust.labels	An object of class vector of size $(N \times 1)$ containing the cluster membership labels in $\{1, \dots, K\}$ 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://amstat.tandfonline.com/doi/abs/10.1080/00401706.2019.1623076>

Blei, D. M. , Kucukelbir, A., and McAuliffe, J. D. (2017), *Variational Inference: A Review for Statisticians*, Journal of the American Statistical Association, Vol. 112(518), 859-877

<https://www.tandfonline.com/doi/full/10.1080/01621459.2017.1285773>

Lee, K. H., Xue, L., and Hunter, D. R. (2017) *Model-Based Clustering of Time-Evolving Networks through Temporal Exponential-Family Random Graph Models*, Journal of Multivariate Analysis, to appear

<https://arxiv.org/abs/1712.07325>

Vu D. Q., Hunter, D. R., and Schweinberger, M. (2013) *Model-based Clustering of Large Networks*, The Annals of Applied Statistics, Vol. 7(2), 1010-1039

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

## Examples

```
## 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)
```

ergmclust.ICL

*Model Selection Based On Integrated Classification Likelihood.*

## 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

```
ergmclust.ICL(adjmat, Kmax = 5, directed = FALSE, weighted = FALSE, thresh = 1e-06,
iter.max = 200, coef.init = NULL, wtmat = 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 is must be asymmetric in general). By default, this is set as FALSE.
weighted	If TRUE, the network is supposed to be weighted. 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	ergmclust chooses the default value as a random perturbation around K-dim zero vector; default is NULL.
wtmat	An object of class matrix of dimension (N x N) containing the weight matrix, where N is the number of nodes in the network; default is NULL.

## 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**

Biernacki, C., Celeux, G., and Govaert, G. (2000) *Assessing a mixture model for clustering with the integrated completed likelihood*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22(7), 719-725

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

Daudin, J. J., Picard, F., and Robin, S. (2008) *A Mixture Model for Random Graphs*, Statistics and Computing, Vol. 18(2), 173-183

<https://link.springer.com/article/10.1007/s11222-007-9046-7>

**Examples**

```
## 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)
```

---

ergmclust.plot

*Visualization For Model-Based Clustering of Large Networks.*

---

## 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

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

## Arguments

<code>adjmat</code>	An object of class matrix of dimension (N x N) containing the adjacency matrix, where N is the number of nodes in the network.
<code>K</code>	Number of clusters in the mixed membership Exponential-Family Random Graph Models (ERGMs).
<code>directed</code>	If TRUE, the network is supposed to be directed (and therefore <code>adjmat</code> is must be asymmetric in general). By default, this is set as FALSE.
<code>thresh</code>	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.
<code>iter.max</code>	The maximum number of iterations after which the algorithm is terminated. The default value is set as 200.
<code>coef.init</code>	<code>ergmclust</code> chooses the default value as a random perturbation around K-dim zero vector; default is NULL.
<code>node.labels</code>	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

Vu D. Q., Hunter, D. R., and Schweinberger, M. (2013) *Model-based Clustering of Large Networks*, The Annals of Applied Statistics, Vol. 7(2), 1010-1039

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

## Examples

```
## 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)
```

---

tradenet

---

*International Trade Network Data in 1991.*


---

## Description

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.

## Usage

```
data(tradenet)
```

## Format

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

## Source

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

## References

Westveld, A. H. and Hoff, P. D. (2011) *A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict*, The Annals of Applied Statistics 5(2A), 843–872

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

## Examples

```
data(tradenet)
```



# Index

armsnet, [2](#)

ergmclust, [3](#)

ergmclust-package, [1](#)

ergmclust.ICL, [5](#)

ergmclust.plot, [7](#)

tradenet, [8](#)