

Package ‘multigraphr’

October 13, 2022

Title Probability Models and Statistical Analysis of Random
Multigraphs

Version 0.1.0

Description Methods and models for analysing multigraphs as introduced by Shafie (2015) <[doi:10.21307/joss-2019-011](https://doi.org/10.21307/joss-2019-011)>, including methods to study local and global properties <[doi:10.1080/0022250X.2016.1219732](https://doi.org/10.1080/0022250X.2016.1219732)> and goodness of fit tests.

Depends R (>= 3.2.0)

URL <https://github.com/termehs/multigraphr>

BugReports <https://github.com/termehs/multigraphr/issues>

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.2

Suggests rmarkdown, knitr, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Termeh Shafie [aut, cre]

Maintainer Termeh Shafie <termeh.shafie@manchester.ac.uk>

Repository CRAN

Date/Publication 2021-09-16 18:50:02 UTC

R topics documented:

get_degree_seq	2
get_edge_assignment_probs	3
get_edge_multip_seq	4
gof_sim	5
gof_stats	7
gof_test	9
iea_model	11
nsumk	13
rsm_model	14

get_degree_seq	<i>Degree sequence of a multigraph</i>
----------------	--

Description

Finds the degree sequence of the adjacency matrix of an observed graph or multigraph

Usage

```
get_degree_seq(adj, type = "multigraph")
```

Arguments

adj	matrix of integers representing graph adjacency matrix.
type	equals 'graph' if adjacency matrix is for graphs (default), equals 'multigraph' if it is the equivalence of the adjacency matrix for multigraphs (with matrix diagonal representing edge loops double counted).

Details

Gives the degree sequence of the adjacency matrix of an observed graph or multigraph. Note that the matrix diagonal should be even if the matrix represents a multigraph.

Value

Vector of integers representing the degree sequence of a (multi)graph.

Author(s)

Termeh Shafie

References

Shafie, T. (2015). A Multigraph Approach to Social Network Analysis. *Journal of Social Structure*, 16.

Shafie, T. (2016). Analyzing Local and Global Properties of Multigraphs. *The Journal of Mathematical Sociology*, 40(4), 239-264.

Examples

```
# Adjacency matrix for undirected network with 3 nodes
A <- matrix(c(0, 1, 2,
             1, 2, 1,
             2, 1, 2), nrow=3, ncol=3)

# If A represents a graph
get_degree_seq(adj = A, type = 'graph')

# If A represents a multigraph
get_degree_seq(adj = A, type = 'multigraph')
```

```
get_edge_assignment_probs
```

Edge assignment probabilities under IEA model given fixed degrees

Description

Calculates the edge assignment probabilities given specified degree sequence under the two ways in which the RSM model can be approximated by the IEA model:

- the IEAS (independent edge assignment of stubs) model,
- the ISA (independent stub assignment) model.

Usage

```
get_edge_assignment_probs(m, deg.seq, model)
```

Arguments

<code>m</code>	integer giving number of edges in multigraph.
<code>deg.seq</code>	vector of integers with the sum equal to $2m$ representing the degree sequence of the multigraph.
<code>model</code>	character string, either 'IEAS' or 'ISA'.

Details

The IEAS and ISA edge assignment probabilities to possible vertex pairs are calculated given a fixed degree sequence `deg.seq` under the IEAS model, and `deg.seq/2m` under the ISA model.

Number of possible vertex pair sites (and thus the length of the edge assignment sequence) is given by $(n + 1)n/2$ where n is number of vertices.

Value

A numeric vector representing the edge assignment probabilities to all possible vertex pair sites. The number of vertex pair sites is given by $n(n + 1)/2$.

Author(s)

Termeh Shafie

References

Shafie, T. (2015). A Multigraph Approach to Social Network Analysis. *Journal of Social Structure*, 16.

Shafie, T. (2016). Analyzing Local and Global Properties of Multigraphs. *The Journal of Mathematical Sociology*, 40(4), 239-264.

See Also

[get_degree_seq](#), [iea_model](#)

Examples

```
# Under the IEAS model with 10 possible vertex pair sites (4 vertices)
get_edge_assignment_probs(m = 8, deg.seq = c(4,4,4,4), model = 'IEAS')

# Under the ISA model with 21 possible vertex pair sites (6 vertices)
get_edge_assignment_probs(m = 10, deg.seq = c(8,4,2,2,2,2), model = 'ISA')
```

get_edge_multip_seq *Edge multiplicity sequences of multigraphs given fixed degrees*

Description

Given a degree sequence, this function finds all unique multigraphs represented by their edge multiplicity sequences.

Usage

```
get_edge_multip_seq(deg.seq)
```

Arguments

deg.seq vector of integers with the sum equal to $2m$ representing the degree sequence of the multigraph.

Details

Multigraphs are represented by their edge multiplicity sequence \mathbf{M} with elements $M(i,j)$, denoting edge multiplicity at possible vertex pair sites (i,j) ordered according to $(1,1) < (1,2) < \dots < (1,n) < (2,2) < (2,3) < \dots < (n,n)$, where n is number of nodes.

Only practical for small multigraphs.

Value

All unique edge multiplicity sequences as rows in a data frame. Each row in the data frame represents a unique multigraph given the degree sequence.

Author(s)

Termeh Shafie

References

Shafie, T. (2015). A Multigraph Approach to Social Network Analysis. *Journal of Social Structure*, 16.

Shafie, T. (2016). Analyzing Local and Global Properties of Multigraphs. *The Journal of Mathematical Sociology*, 40(4), 239-264.

See Also

[get_degree_seq](#)

Examples

```
# Adjacency matrix for undirected network with 3 nodes
A <- matrix(c(0, 1, 2,
             1, 2, 1,
             2, 1, 2), nrow=3, ncol=3)
deg <- get_degree_seq(A, 'multigraph')
get_edge_multip_seq(deg)
```

gof_sim

Goodness of fit test simulations for random multigraph models

Description

Goodness of fits test simulations of specified multigraph models using Pearson (S) and information divergence (A) test statistics under the random stub matching (RSM) and the independent edge assignments (IEA) model, where the latter is either independent edge assignments of stubs (IEAS) or independent stub assignment (ISA).

These can be used to check the reliability of the tests by examining the exact probability distributions of the test statistics and their fit to their asymptotic χ^2 -distributions. Only practical for small multigraphs as exact distributions are calculated.

Usage

```
gof_sim(m, model, deg.mod, hyp, deg.hyp)
```

Arguments

<code>m</code>	integer giving number of edges in multigraph.
<code>model</code>	character string representing assumed model, either 'RSM', 'IEAS' or 'ISA'.
<code>deg.mod</code>	vector of integers with the sum equal to $2m$ representing the degree sequence of the multigraph under specified model.
<code>hyp</code>	character string representing the hypothesized null model, either 'IEAS' or 'ISA'.
<code>deg.hyp</code>	vector of integers with the sum equal to $2m$ representing the hypothetical degree sequence of the multigraph under the null model: - if <code>hyp = 'IEAS'</code> , then simple IEAS hypothesis with fully specified degree sequence <code>deg.hyp</code> - if <code>hyp = 'ISA'</code> , then simple ISA hypothesis with with fully specified stub assignment probabilities <code>deg.hyp/2m</code> - if <code>hyp = 'IEAS'</code> and <code>deg.hyp = 0</code> , then composite IEAS hypothesis with edge multiplicity sequence estimated from data - if <code>hyp = 'ISA'</code> and <code>deg.hyp = 0</code> , then composite ISA hypothesis with edge multiplicity sequence estimated from data

Details

The tests are performed using goodness-of-fit measures between simulated edge multiplicity sequence of a multigraph according to an RSM or IEA model, and the expected multiplicity sequence according to a simple or composite IEA hypothesis.

Test statistics of Pearson type (S) and of information divergence (A) type are used and summary of tests given these two statistics are given as output. The adjusted statistics and χ^2 -distributions are useful for better power calculations.

Value

Output is generated from function `gof_stats`:

<code>test.summary</code>	Expected value and variances of test statistics (<code>stat</code>), critical values (<code>cv</code>) according to asymptotic χ^2 -distribution and according to cdf's of test statistics, significance level (<code>alpha</code>) according to asymptotic χ^2 distribution, power of tests ($P(\text{stat} > \text{cv})$), critical values and power according to the distributions of test statistics ($\text{cv}(\text{stat})$ and $P(\text{Stat} > \text{cv}(\text{Stat}))$).
<code>degrees.of.freedom</code>	Degrees of freedom for tests performed.
<code>probS</code>	Probability distributions of Pearson statistic S .
<code>probA</code>	Probability distributions of information divergence statistic A .
<code>adjusted.stats</code>	Expected values and variances for adjusted test statistics, preferred adjusted statistics.
<code>adjusted.chi2</code>	Degrees of freedom for adjusted χ^2 -distribution.
<code>power.apx</code>	Power approximations according to adjusted statistics.

Author(s)

Termeh Shafie

References

Shafie, T. (2015). A Multigraph Approach to Social Network Analysis. *Journal of Social Structure*, 16.

Shafie, T. (2016). Analyzing Local and Global Properties of Multigraphs. *The Journal of Mathematical Sociology*, 40(4), 239-264.

See Also

[gof_stats](#), [get_edge_assignment_probs](#), [nsumk](#), [rsm_model](#)

Examples

```
# Testing a simple IEAS hypothesis with degree sequence [6,6,6] against
# an IEAS model with degree sequence [14,2,2] on a multigraph with n = 3 nodes and m = 9 edges
deg.mod <- c(14,2,2)
deg.hyp <- c(6,6,6)
test1 <- gof_sim(9, 'IEAS', deg.mod, 'IEAS', deg.hyp)

# Non-null distributions (pdf's and cdf's) of test statistics S and A are given by
test1$probS
test1$probA

# Testing a simple ISA hypothesis with degree sequence [6,6,6] against
# an IEAS model with degree sequence [12,3,3] on a multigraph with n = 3 nodes and m = 9 edges
deg.mod <- c(12,3,3)
deg.hyp <- c(6,6,6)
test2 <- gof_sim(9, 'IEAS', deg.mod, 'ISA', deg.hyp)

# Testing a composite IEAS hypothesis against
# an RSM model with degree sequence [6,6,6,6] on a multigraph with n = 4 nodes and m = 20 edges.
deg.mod <- c(6,6,6,6)
test3 <- gof_sim(12, 'RSM', deg.mod, 'IEAS', deg.hyp = 0)
```

gof_stats

Exact probability distributions and moments of goodness of fit statistics

Description

Goodness of fit between two specified edge multiplicity sequences (e.g. observed vs. expected). Pearson (S) and information divergence (A) tests statistics are used and the exact distribution of these statistics, their asymptotic χ^2 -distributions, and their first two central moments are calculated using this function. Only practical for small multigraphs.

Usage

```
gof_stats(m, dof, m.seq, prob.mg, Q.seq)
```

Arguments

m	integer giving number of edges in multigraph.
dof	integer giving degrees of freedom of test performed.
m.seq	matrix of integers, each row representing the edge multiplicity sequence of a multigraph (which correspond to observed values).
prob.mg	numerical vector representing a given probability distribution of multigraphs/edge multiplicity sequences in m.seq.
Q.seq	a numeric vector representing the hypothetical edge assignment probabilities to all possible vertex pair sites (from which expected values are calculate).

Details

The tests are performed using goodness-of-fit measures between two edge multiplicity sequences (e.g. observed vs. expected).

Test statistics of Pearson type (S) and of information divergence (A) type are used and summary of tests given these two statistics are given as output. The adjusted statistics and χ^2 -distributions are useful for better power calculations.

Value

test.summary	Expected value and variances of test statistics (stat), critical values (cv) according to asymptotic χ^2 -distribution and according to cdf's of test statistics, significance level (alpha) according to asymptotic χ^2 distribution, power of tests ($P(\text{stat} > \text{cv})$), critical values and power according to the distributions of test statistics ($\text{cv}(\text{stat})$ and $P(\text{Stat} > \text{cv}(\text{Stat}))$).
degrees.of.freedom	Degrees of freedom for tests performed.
probS	Probability distributions of Pearson statistic S .
probA	Probability distributions of information divergence statistic A .
adjusted.stats	Expected values and variances for adjusted test statistics, preferred adjusted statistics.
adjusted.chi2	Degrees of freedom for adjusted χ^2 -distribution.
power.apx	Power approximations according to adjusted statistics.

Author(s)

Termeh Shafie

References

Shafie, T. (2015). A Multigraph Approach to Social Network Analysis. *Journal of Social Structure*, 16.

Shafie, T. (2016). Analyzing Local and Global Properties of Multigraphs. *The Journal of Mathematical Sociology*, 40(4), 239-264.

See Also

[gof_sim](#), [get_edge_assignment_probs](#), [nsumk](#)

Examples

```
# Generate a set of edge multiplicity sequences (random multigraphs) and
# its probability distribution using rsm_model() with degree sequence [4,4,6,6]
rsm <- rsm_model(deg.seq = c(4,4,6,6))
mg <- as.matrix(rsm$m.seq)
mg.p <- rsm$prob.dists[, 1]

# Generate edge assignment probabilities from which the second set of
# edge multiplicity sequences is generated from using the iea_model()
deg.f <- (4*5)/2 - 1
eap <- get_edge_assignment_probs(m = 10,
                                deg.seq = c(4,4,6,6), model = 'IEAS')

# Perform the test
test <- gof_stats(m = 10, dof = deg.f,
                  m.seq = mg, prob.mg = mg.p, eap)
```

gof_test

Goodness of fit tests for random multigraph models

Description

Goodness of fit tests between an observed edge multiplicity sequence and an expected edge multiplicity sequence according to specified RSM or IEA hypotheses using Pearson (S) and information divergence (A) tests statistics.

Usage

```
gof_test(adj, type, hyp, deg.hyp, dof)
```

Arguments

adj	matrix of integer representing graph adjacency matrix.
type	equals 'graph' if adjacency matrix is for graphs (default), equals 'multigraph' if it is the equivalence of the adjacency matrix for multigraphs (with matrix diagonal representing loops double counted).
hyp	character string representing the null model, either 'IEAS' or 'ISA'.
deg.hyp	vector of integers with the sum equal to to $2m$ representing the degree sequence of the multigraph under the null model: - if hyp = 'IEAS', then simple IEAS hypothesis with fully specified degree sequence deg.hyp - if hyp = 'ISA', then simple ISA hypothesis with with fully specified stub assignment probabilities deg.hyp/2m - if hyp = 'IEAS' and deg.hyp = 0, then composite IEAS hypothesis with edge multiplicity sequence estimated from data - if hyp = 'ISA' and deg.hyp = 0, then composite ISA hypothesis with edge multiplicity sequence estimated from data
dof	integer giving degrees of freedom of test, $r-1$ for simple hypotheses and $r-n$ for composite hypotheses where $r = n(n+1)/2$

Details

This function can be used to test whether there is a significant difference between observed multigraph and the expected multiplicity sequence according to a simple or composite IEA hypothesis.

Test statistics of Pearson (S) and of information divergence (A) type are used and test summary based on these two statistics are given as output.

p -values indicate whether we have sufficient evidence to reject the null that there is no significant difference between the observed and expected edge multiplicity sequence.

Value

summary	Data frame including observed values of test statistics S and A , degrees of freedom for the asymptotic χ^2 -distributions of tests statistics, and p -values for the tests performed.
---------	---

Author(s)

Termeh Shafie

References

Shafie, T. (2015). A Multigraph Approach to Social Network Analysis. *Journal of Social Structure*, 16.

Shafie, T. (2016). Analyzing Local and Global Properties of Multigraphs. *The Journal of Mathematical Sociology*, 40(4), 239-264.

See Also

[get_degree_seq](#), [get_edge_assignment_probs](#), [gof_sim](#) to check the reliability of your test

Examples

```
# Adjacency matrix of observed network (multigraph), n = 4 nodes , m = 15 edges
A <- t(matrix(c( 0, 1, 0, 3,
               0, 0, 1, 7,
               0, 1, 0, 3,
               3, 6, 3, 2), nrow= 4, ncol=4))
deg <- get_degree_seq(adj = A, type = 'multigraph')

# Testing a simple IEAS hypothesis with above degree sequence
gof_test(adj = A, type = 'multigraph', hyp = 'IEAS', deg.hyp = deg, dof = 9)

# Testing a composite IEAS hypothesis
gof_test(adj = A, type = 'multigraph', hyp = 'IEAS', deg.hyp = 0, dof = 6)
```

iea_model

*Independent edge assignment model for multigraphs***Description**

Summary of estimated statistics for analyzing global structure of random multigraphs under the independent edge assignment model given observed adjacency matrix.

Two versions of the IEA model are implemented, both of which can be used to approximate the RSM model:

1. independent edge assignment of stubs (IEAS) given an edge probability sequence
2. independent stub assignment (ISA) given a stub probability sequence

Usage

```
iea_model(
  adj,
  type = "multigraph",
  model = "IEAS",
  K = 0,
  apx = FALSE,
  p.seq = NULL
)
```

Arguments

`adj` matrix of integers representing graph adjacency matrix.

`type` equals 'graph' if adjacency matrix is for graphs (default), equals 'multigraph' if it is the equivalence of the adjacency matrix for multigraphs (with matrix diagonal representing loops double counted).

model	character string representing which IEA model: either 'IEAS' (default) or 'ISA'.
K	upper limit for the complexity statistics $R(k)$, $k=(0,1,\dots,K)$, representing the sequence of frequencies of edge multiplicities (default is maximum observed in adjacency matrix).
apx	logical (default = 'FALSE'). if 'TRUE', the IEA model is used to approximate the statistics under the random stub matching model given observed degree sequence.
p.seq	if model = ISA and apx = FALSE, then specify this numerical vector of stub assignment probabilities.

Details

When using the IEAS model:

If the IEAS model is used as an approximation to the RSM model, then the edge assignment probabilities are estimated by using the observed degree sequence. Otherwise, the edge assignment probabilities are estimated by using the observed edge multiplicities (maximum likelihood estimates).

When using the ISA model:

If the ISA model is used as an approximation to the RSM model, then the stub assignment probabilities are estimated by using the observed degree sequence over $2m$. Otherwise, a sequence containing the stub assignment probabilities (for example based on prior belief) should be given as argument p.seq.

Value

nr.multigraphs	Number of unique multigraphs possible.
M	Summary and interval estimates for <i>number of loops</i> (M1) and <i>number of multiple edges</i> (M2).
R	Summary and interval estimates for frequencies of edge multiplicities R_1, R_2, \dots, R_K , where K is a function argument.

Author(s)

Termeh Shafie

References

Shafie, T. (2015). A Multigraph Approach to Social Network Analysis. *Journal of Social Structure*, 16.

Shafie, T. (2016). Analyzing Local and Global Properties of Multigraphs. *The Journal of Mathematical Sociology*, 40(4), 239-264.

See Also

[get_degree_seq](#), [get_edge_multip_seq](#), [iea_model](#)

Examples

```

# Adjacency matrix of a small graph on 3 nodes
A <- matrix(c(1, 1, 0,
             1, 2, 2,
             0, 2, 0),
           nrow = 3, ncol = 3)

# When the IEAS model is used
iea_model(adj = A , type = 'graph', model = 'IEAS', K = 0, apx = FALSE)

# When the IEAS model is used to approximate the RSM model
iea_model(adj = A , type = 'graph', model = 'IEAS', K = 0, apx = TRUE)

# When the ISA model is used to approximate the RSM model
iea_model(adj = A , type = 'graph', model = 'ISA', K = 0, apx = TRUE)

# When the ISA model is used with a pre-specified stub assignment probabilities
iea_model(adj = A , type = 'graph', model = 'ISA', K = 0, apx = FALSE, p.seq = c(1/3, 1/3, 1/3))

```

nsumk

Ordered n-tuples of non-negative integers summing to k

Description

Finds ordered n -tuples of non-integers summing to k . Only practical for $n < 15$.

Usage

```
nsumk(n, k)
```

Arguments

n a positive integer.
 k a positive integer.

Details

Useful for finding all possible degree sequences for a network with n nodes and $k/2$ number of edges, or for finding all possible edge multiplicity sequence n that sum up to k number of edges. The number of vertex pair sites (or length of edge multiplicity sequence) for a multigraph with n nodes is given by $n(n + 1)/2$.

Value

A matrix with choose($k+n-1, n-1$) rows and n columns. Each row comprises non-negative integers summing to k .

Author(s)

Termeh Shafie

See Also[gof_sim](#)**Examples**

```
# All possible degree sequences for
# a network with 4 nodes and 5 edges
D <- nsumk(4, 10)

# Remove degree sequences with isolated nodes
D <- D[-which(rowSums(D == 0) > 0), ]

# All edge multiplicity sequences/multigraph with 2 nodes and 4 edges
r <- (2*3)/2 # vertex pair sites (or length of edge multiplicity sequences)
mg <- nsumk(r,4) # number of rows give number of possible multigraphs
```

rsm_model

*Random stub matching model for multigraphs***Description**

Given a specified degree sequence, this function finds all unique multigraphs represented by their edge multiplicity sequences. Different complexity statistics together with their probability distributions and moments are calculated.

Usage

```
rsm_model(deg.seq)
```

Arguments

deg.seq vector of integers representing the degree sequence of a multigraph.

Details

The probability distributions of all unique multigraphs given fixed degree sequence, together with the first two central moments and interval estimates of the statistics $M1 = \text{number of loops}$ and $M2 = \text{number of multiple edges}$, under the RSM model are calculated.

For other structural statistics and for large multigraphs, use the IEA approximation of the RSM model via the function [iea_model](#)

Value

m.seq	possible multigraphs represented by edge multiplicity sequences
prob.dists	probability distribution of the multigraphs/edge multiplicity sequences, and the probability distributions of the statistics <i>number of loops</i> , <i>number of multiple edges</i> , and <i>simple graph</i> (logical) for each multigraph
M	summary of moments and interval estimates for <i>number of loops</i> and <i>number of multiple edges</i> (M1 and M2)).

Author(s)

Termeh Shafie

References

Shafie, T. (2015). A Multigraph Approach to Social Network Analysis. *Journal of Social Structure*, 16.

Shafie, T. (2016). Analyzing Local and Global Properties of Multigraphs. *The Journal of Mathematical Sociology*, 40(4), 239-264.

See Also

[get_degree_seq](#), [get_edge_multip_seq](#), [iea_model](#)

Examples

```
# Given a specified degree sequence
D <- c(2,2,3,3) # degree sequence
mod1 <- rsm_model(D)
mod1$m.seq
mod1$prob.dists
mod1$M

# Given an observed graph
A <- matrix(c(0, 1, 2,
             1, 2, 1,
             2, 1, 2), nrow=3, ncol=3)
D <- get_degree_seq(adj = A, type = 'graph')
mod2 <- rsm_model(D)
```

Index

`get_degree_seq`, [2](#), [4](#), [5](#), [11](#), [12](#), [15](#)
`get_edge_assignment_probs`, [3](#), [7](#), [9](#), [11](#)
`get_edge_multip_seq`, [4](#), [12](#), [15](#)
`gof_sim`, [5](#), [9](#), [11](#), [14](#)
`gof_stats`, [6](#), [7](#), [7](#)
`gof_test`, [9](#)

`iea_model`, [4](#), [11](#), [12](#), [14](#), [15](#)

`nsumk`, [7](#), [9](#), [13](#)

`rsm_model`, [7](#), [14](#)