

Package ‘snowboot’

August 29, 2016

Type Package

Title Bootstrap Methods for Network Inference

Version 0.5.1

Date 2016-3-23

Author Leticia Ramirez-Ramirez, Kusha Nezafati, Vyacheslav Lyubchich, Yulia Gel

Maintainer Kusha Nezafati <kusha.nezafati@gmail.com>

Description Functions for analysis of network objects, which are imported or simulated by the package. The non-parametric methods of analysis center around snowball and bootstrap sampling.

Depends R (>= 3.2.2)

License GPL-3

LazyData TRUE

Imports VGAM, igraph, stats, Rcpp

RoxygenNote 5.0.1

LinkingTo Rcpp

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-04-04 08:37:21

R topics documented:

artificial_networks	2
bootCI	3
bootdeg	4
BparametersEst	5
cross_validation	6
igraph_to_network	8
local.network.MR.new5	8
LSMI	9
Oempdegreedistrib	10
sample_about_one_seed	12

summary_net	13
vertboot	14

Index	15
--------------	-----------

artificial_networks	<i>200 Simulated Networks of order 2000 with Polylogarithmic (0.1, 2) Degree Distributions</i>
---------------------	--

Description

A list called "networks" containing 200 network objects of order 2000. These networks were simulated using the polylogarithmic (aka Gutenberg-Richter law) degree distribution (Newman et al., 2001; Newman, 2002) with parameters $\delta = 0.1$ and $\lambda = 2$ as see in the following equations:

$$f(k) = k^{-\delta} e^{-k/\lambda} / Li_{\delta}(e^{-1/\lambda})$$

$$Li = \sum_{j=1}^{\infty} z^{-j} / j^{\delta}$$

where $\lambda > 0$. Please see reference below for details (Thompson et al, 2016).

Usage

```
artificial_networks
```

Format

a list containing 200 network objects. Each network object is a list with three elements:

edges The edgelist of the network. A two column matrix where each row is an edge.

degree The degree sequence of the network, which is an integer vector of length n.

n The network order. The order for every network is 2000.

References

Thompson, M. E., Ramirez Ramirez, L. L., Lyubchich, V. and Gel, Y. R. (2015), Using the bootstrap for statistical inference on random graphs. Can J Statistics. doi: 10.1002/cjs.11271

bootCI *Build Bootstrap Confidence Intervals for \hat{p}_k*

Description

The function will build bootstrap confidence intervals for the bootstrap estimate of and μ with a lower-bound of 0.025 and an upper-bound of 0.975.

Usage

```
bootCI(outBootdeg, bootstrap_mean = T, lower_bound = 0.025,
       upper_bound = 0.975)
```

Arguments

outBootdeg A list that is the output of [bootdeg](#)

bootstrap_mean A Boolean option to return the bootstrap confidence interval for the mean.

lower_bound The lower quantile for the bootstrap confidence intervals.

upper_bound The upper quantile for the bootstrap confidence intervals.

Value

A list of two elements

p_k_CI This a list of length `length(outBootdeg$num.sam)`, one element per LSMI. Each element contains three sets of bootstrap confidence intervals for \hat{p}_k^* corresponding to the three estimation methods. See [bootdeg](#) for more on the three estimation methods.

mean_CI This a list of length `length(outBootdeg$num.sam)`, one element per LSMI. Each element contains three sets of bootstrap confidence intervals for $\hat{\mu}$ corresponding to the three estimation methods. See [bootdeg](#) for more on the three estimation methods.

References

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics*, 1-26.

Thompson, M. E., Ramirez Ramirez, L. L., Lyubchich, V. and Gel, Y. R. (2015), Using the bootstrap for statistical inference on random graphs. *Can J Statistics*. doi: 10.1002/cjs.11271

Examples

```
net <- artificial_networks[[1]]
sam.out <- Oempdegreedistrib(net = net, n.seeds = 40, n.neigh = 1, num.sam = 1)
outBootdeg <- bootdeg(sam.out = sam.out, n.boot = 50)
a <- bootCI(outBootdeg)
```

bootdeg *Bootstrapping Empirical Degree Distribution*

Description

This function delivers a bootstrap estimate of network degree distribution based on a LSMI sample. Default is one bootstrap replication.

Usage

```
bootdeg(sam.out, num.sam = sam.out$num.sam, n.boot = 1)
```

Arguments

sam.out	A list that is the output of Oempdegreedistrib .
num.sam	A vector of integers containing the numeric ids of the LSMI samples when sam.out\$num.sam is greater than one. When num.sam is an integer, N, LSMI from 1 to N are taken from the input sam.out.
n.boot	A positive integer number, the number of bootstrap replications.

Value

A list consisting of:

values	A list of length num.sam where each element is a vector containing the unique degree values sampled in each LSMI.
empd	A list of length three where each element contains a different estimate of degree distribution: empd.w.p0s - weighted bootstrap with a proportion of isolated nodes p0 being estimated by simple random sampling of bootstrapped seeds; empd.nw.p0sEkb - non-weighted bootstrap with a proportion of isolated nodes p0 being estimated by simple random sampling of bootstrapped seeds; empd.nw.p0sEks - ignore (see Thompson et al. for details)
num.sam	Numeric indices corresponding to LSMI samples used for bootstrap.
n.boot	The same object as input argument n.boot.
n.neigh	The number of waves carried out by the snowball sample. This is the same value from sam.out\$n.neigh.
seeds1	A matrix of dimension length(num.sam) x n.seeds with the numeric seed ids. Each row corresponds to one LSMI. The rows are present in the same order as the ids in num.sam.
nodes_of_LSMI	A list of length length(num.sam) where each element is vector containing the numeric ids of the nodes sampled using the respective LSMI. The elements are present in the same order as the ids in num.sam. Note: nodes_of_LSMI is unreported when n.neigh equals zero.

References

- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics*, 1-26.
- Thompson, M. E., Ramirez Ramirez, L. L., Lyubchich, V. and Gel, Y. R. (2015), Using the bootstrap for statistical inference on random graphs. *Can J Statistics*. doi: 10.1002/cjs.11271

Examples

```
net <- artificial_networks[[1]]
sam.out <- Oempdegredistrib(net = net, n.seeds = 40, n.neigh = 1, num.sam = 1)
a <- bootdeg(sam.out = sam.out, n.boot = 50)
```

BparametersEst

Summary of the Bootstrap Degree Distribution

Description

This function provides summary statistics of a bootstrap degree distribution.

Usage

```
BparametersEst(outBootdeg)
```

Arguments

outBootdeg A list that is the output of `bootdeg`

Value

A list consisting of:

- | | |
|-----------|---|
| mean | An array of dimension $c(\text{length}(\text{outBootdeg}\$num.\text{sam}), \text{outBootdeg}\$n.\text{boot}, 3)$. The last dimension, of 3, is for the three different methods of obtaining the empirical degree distribution from <code>outBootdeg\$empd</code> (see output <code>empd</code> from <code>bootdeg</code> for details). The (i,j,k) -th element in the array is an estimate of mean degree for the i -th LSMI sample, j -th bootstrap replication, and k -th empirical distribution from <code>outBootdeg\$empd</code> . |
| quartiles | An array of dimension $c(\text{length}(\text{outBootdeg}\$num.\text{sam}), 3, \text{outBootdeg}\$n.\text{boot}, 3)$. The last dimension, of 3, is for the three different methods of estimation from <code>outBootdeg\$empd</code> (see output <code>empd</code> from <code>bootdeg</code> for details). The second dimension, of 3, corresponds to the quartiles (.25, .5, .75). The (i,j,k,l) -th element in the array is an estimate of j -th quartile for the i -th LSMI sample, k -th bootstrap replication, and l -th empirical distribution from <code>outBootdeg\$empd</code> . |
| rfreq | An array of dimension $c(\text{length}(\text{outBootdeg}\$num.\text{sam}), 5, \text{outBootdeg}\$n.\text{boot}, 3)$. The last dimension, of 3, is for the three different methods of estimation from <code>outBootdeg\$empd</code> (see output <code>empd</code> from <code>bootdeg</code> for details). The second dimension, of 5, corresponds to degree values: 0, 1, 2, 3, 4. The (i,j,k,l) -th element in the array is the proportion of nodes with degree j in the i -th LSMI sample, k -th bootstrap replication, and l -th empirical distribution from <code>outBootdeg\$empd</code> . |

deciles	An array of dimension $c(\text{length}(\text{outBootdeg}\$\text{num.sam}), 9, \text{outBootdeg}\$\text{n.boot}, 3)$. The last dimension, of 3, is for the three different methods of estimation from <code>outBootdeg\$empd</code> (see output <code>empd</code> from <code>bootdeg</code> for details.). The second dimension, of 9, corresponds to the deciles (.1, .2, ..., .9). The (i,j,k,l)-th element in the array is an estimate of j-th decile for the i-th LSMI sample, k-th bootstrap replication, and l-th empirical distribution from <code>outBootdeg\$empd</code> .
num.sam	Numeric indices corresponding to LSMI samples used for bootstrap. See value <code>num.sam</code> from <code>bootdeg</code> .
seeds1	A matrix of dimension $\text{length}(\text{num.sam}) \times \text{n.seeds}$ with the numeric seed ids. Each row corresponds to one LSMI. The rows are present in the same order as the ids in <code>num.sam</code> . See value <code>seeds1</code> from <code>bootdeg</code> .
nodes_of_LSMI	A list of length $\text{length}(\text{num.sam})$ where each element is vector containing the numeric ids of the nodes sampled using the respective LSMI. The elements are present in the same order as the ids in <code>num.sam</code> . Note: <code>nodes_of_LSMI</code> is unreported when <code>n.neigh</code> equals zero. See value <code>nodes_of_LSMI</code> from <code>bootdeg</code> .

References

- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics*, 1-26.
- Thompson, M. E., Ramirez Ramirez, L. L., Lyubchich, V. and Gel, Y. R. (2015), Using the bootstrap for statistical inference on random graphs. *Can J Statistics*. doi: 10.1002/cjs.11271

Examples

```
net <- artificial_networks[[1]]
sam.out <- Oempdegreedistrib(net = net, n.seeds = 40, n.neigh = 1, num.sam = 1)
outBootdeg <- bootdeg(sam.out = sam.out, n.boot = 50)
a <- BparametersEst(outBootdeg)
```

`cross_validation` *A function that uses cross-validation to select seed-wave combination for estimation of a degree's frequency.*

Description

The function's inputs are a network, a vector of possible seed sample-sizes, a vector of possible waves, and a few tuning parameters. The output will contain the best seed-wave combination for each degree and the width of the 95 percent bootstrap confidence intervals at each degree for the best seed-wave combination.

Usage

```
cross_validation(network, n.seeds, n.neigh, n.boot, kmax, proxyRep = 19,
  proxyOrder = 30)
```

Arguments

network	A network object that is list containing: edges The edgelist of the network. A two column matrix where each row is an edge. degree The degree sequence of the network, which is an integer vector of length n. n The network order. The object can be created by <code>local.network.MR.new5</code> or it can be imported.
n.seeds	A numeric vector for the different sample sizes of seed to use in cross-validation.
n.neigh	A numeric vector for the different waves to use in cross-validation.
n.boot	The number of bootstrap sample.
kmax	The largest degree to preform cross-validation on.
proxyRep	The number of time to sample a proxy. Default is 19.
proxyOrder	The size of the proxy sample. Default is 30.

Value

A list consisting of

`selected_seed_wave`

A list of 3 matrices (one per estimation method. See supporting documentation [bootdeg](#)). Each matrix provides the best seed-wave combinations (obtained via cross-validation) for the respective estimation method.

`selected_seed_wave`

A list of 3 matrices (one per estimation method. See supporting documentation [bootdeg](#)). Each matrix provides the 95 percent bootstrap confidence intervals for the estimated degree frequency using the best seed-wave combinations (see above).

Note

Only one LSMI per seed-wave combination is currently supported.

References

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics*, 1-26.

Thompson, M. E., Ramirez Ramirez, L. L., Lyubchich, V. and Gel, Y. R. (2015), Using the bootstrap for statistical inference on random graphs. *Can J Statistics*. doi: 10.1002/cjs.11271

Examples

```
net <- artificial_networks[[1]]
a <- cross_validation(network = net, n.seeds = c(10, 20, 30), n.neigh = c(1, 2),
  n.boot = 200, kmax = 30)
```

igraph_to_network *Create a "Network" Object from an igraph Object*

Description

This function will convert an igraph object to an object compatible with snowboot functions.

Usage

```
igraph_to_network(in_graph)
```

Arguments

`in_graph` An igraph object. To create igraph objects from field data, see [graph_from_edgelist](#), [graph_from_data_frame](#), [graph_from_adjacency_matrix](#), or [read_graph](#).

Value

A list that contain elements:

<code>edges</code>	The edgelist of the network. A two column matrix where each row is an edge.
<code>degree</code>	The degree sequence of the network, which is an integer vector of length <code>n</code> .
<code>n</code>	The network order.

References

<http://igraph.org/>

Examples

```
hex_ring <- igraph::make_ring(6, directed = FALSE, mutual = FALSE, circular = TRUE)
net <- igraph_to_network(hex_ring)
```

local.network.MR.new5 *Construct Artificial Networks*

Description

This function constructs an artificial network from a given distribution. Only 11 distributions are available.

Usage

```
local.network.MR.new5(n, distrib, param = NULL, degree = NULL,
  take.p = 0.05)
```


Arguments

<code>n</code>	The number of nodes in the desired network.
<code>distrib</code>	An atomic character representing the desired degree distribution. User may choose from 11 available distributions: "fixed", "pois", "zt pois", "geom", "nbinom", "ztgeom", "poly.log", "logarithmic", "power.law", "full" (fully connected), or "none" (no element connected).
<code>param</code>	The distribution parameters. If the function is "fixed", <code>param</code> is a vector of degrees.
<code>degree</code>	An optional vector of degrees that must be of length <code>n</code> . The default is <code>degree = NULL</code> .
<code>take.p</code>	A number between 0 and 1 representing the proportion to take for elimination with each iteration.

Value

A list consisting of

<code>edges</code>	The edgelist of the network. A two column matrix where each row is an edge.
<code>degree</code>	The degree sequence of the network, which is an integer vector of length <code>n</code> .
<code>degree.left</code>	A vector of length <code>n</code> that should be all zeroes.
<code>n</code>	The network order. The order for every network is 2000.

Examples

```
a <- local.network.MR.new5(1000, "poly.log", c(2, 13))
```

 LSMI

Snowball sampling with multiple inclusion.

Description

The function will conduct snowball sampling.

Usage

```
LSMI(net, n.seeds = 10, n.neigh = 1, seeds = NULL)
```

Arguments

<code>net</code>	A network object that is list containing: edges The edgelist of the network. A two column matrix where each row is an edge. degree The degree sequence of the network, which is an integer vector of length <code>n</code> . n The network order.
------------------	---

The object can be created by `local.network.MR.new5` or it can be imported.

<code>n.seeds</code>	A number of seeds in the snowball sample. It must be a positive integer.
<code>n.neigh</code>	A number of waves to be sampled around each seed in LSMI. For example, <code>n.neigh = 0</code> corresponds to seeds only, and <code>n.neigh = 1</code> corresponds to sampling seeds and their first neighbors). Note that the algorithm allows for multiple inclusions.
<code>seeds</code>	A matrix of dimension <code>num.sam x n.seeds</code> containing the numeric ids of the seeds to initiate sampling. Each row of the matrix corresponds to one LSMI sample. Note that this is an optional parameter. WARNING: As of now, this feature is only supported when parameter <code>n.neigh</code> is greater than zero.

Value

A list containing the following elements:

<code>seeds</code>	A numeric a vector containing the numeric ids of sampled seeds.
<code>sampleN</code>	A numeric vector containing ids of the nodes from the snowball sampling and the intial seeds' ids. This vector may have duplicates, since the algorithm allows for multiple inclusions.
<code>unodes</code>	A list of length <code>n.seeds</code> where each element is a numeric vector containing the seed's id and the unique ids of all nodes that were snowball sampled from that seed using <code>sample_about_one_seed</code> (one vector per seed).
<code>nodes.waves</code>	A list of length <code>n.seeds</code> where each element is a list of length <code>n.neigh</code> (Note: these lists are the output object <code>\$nodes.waves</code> from <code>sample_about_one_seed</code>) that contains vectors of numeric id's of the nodes reached in each respective wave from the respective seed.

References

Thompson, M. E., Ramirez Ramirez, L. L., Lyubchich, V. and Gel, Y. R. (2015), Using the bootstrap for statistical inference on random graphs. *Can J Statistics*. doi: 10.1002/cjs.11271

Examples

```
net <- artificial_networks[[1]]
a <- LSMI(net, n.seeds = 20, n.neigh = 2)
```

Oempdegreedistrib	<i>Obtaining an Empirical Network Degree Distribution from Labeled Snowball Sampling with Multiple Inclusion (LSMI).</i>
-------------------	--

Description

Oempdegreedistrib is used to obtain the empirical network degree distribution from labeled snowball sampling with multiple inclusion (LSMI).

Usage

```
Oempdegreedistrib(net, n.seeds, n.neigh, num.sam = 1, seeds = NULL)
```

Arguments

net A network object that is list containing:
edges The edgelist of the network. A two column matrix where each row is an edge.
degree The degree sequence of the network, which is an integer vector of length `n`.
n The network order.
The object can be created by `local.network.MR.new5` or it can be imported.

n.seeds A number of seeds in the snowball sample. It must be a positive integer.

n.neigh A number of waves to be sampled around each seed in LSMI. For example, `n.neigh = 0` corresponds to seeds only, and `n.neigh = 1` corresponds to sampling seeds and their first neighbors). Note that the algorithm allows for multiple inclusions.

num.sam A number for the LSMI repetitions. Default value is one.

seeds A matrix of dimension `num.sam x n.seeds` containing the numeric ids of the seeds to initiate sampling. Each row of the matrix corresponds to one LSMI sample. Note that this is an optional parameter. **WARNING:** As of now, this feature is only supported when parameter `n.neigh` is greater than zero.

Value

A list consisting of

samples A list of length `num.sam` where each element is a list containing three tables: the frequency of degrees sampled from seeds, non-seeds including duplicated nodes, and non-seeds without duplications.

values A list of length `num.sam` where each element is a vector containing the unique degree values sampled in each LSMI.

Oemp A list of length `num.sam` where each element is a list containing two tables based on different methods to estimating the empirical distribution from the network sample (One list per LSMI).

num.sam A number for the LSMI repetitions.

val.seeds A list of length `num.sam` where each element is a vector of unique degree values sampled solely from seeds (One vector per LSMI).

val.nonseed a list of length `num.sam` where each element is a vector of unique degree values sampled solely from non-seeds (One vector per LSMI). Note: This item is unreported when `n.neigh` equals zero.

n.seeds the number of seeds in the snowball sample.

n.neigh the number of waves carried out by the snowball sample. See input argument for details.

p0.real	proportion of nodes from the network with degree zero. Note: p0.real is unreported when n.neigh equals zero.
p0.seeds	a list of length num.sam where each element is the proportion of seeds with degree zero. (One element per LSMI).
ekseed	a list of length num.sam where each element is the sample mean of the seeds. Note that This is unreported when n.neigh equals zero.
seeds1	a matrix of dimension num.sam x n.seeds with the numeric seed ids. Each row corresponds to one LSMI.
nodes_of_LSMI	a list of length num.sam where each element is vector containing the numeric ids of the nodes sampled using LSMI (One element per LSMI). Note: nodes_of_LSMI is unreported when n.neigh equals zero.

References

Thompson, M. E., Ramirez Ramirez, L. L., Lyubchich, V. and Gel, Y. R. (2015), Using the bootstrap for statistical inference on random graphs. Can J Statistics. doi: 10.1002/cjs.11271

Examples

```
net <- artificial_networks[[1]]
sam.out <- Oempdegreedistrib(net = net, n.seeds = 40, n.neigh = 1, num.sam = 1)
```

sample_about_one_seed *Snowball Sampling with Multiple Inclusion Around a Single Seed*

Description

This function performs snowball sampling with multiple inclusions (LSMI) around a single seed.

Usage

```
sample_about_one_seed(net, seed0, n.neigh = 1)
```

Arguments

net	A network object that is list containing: edges The edgelist of the network. A two column matrix where each row is an edge. degree The degree sequence of the network, which is an integer vector of length n. n The network order. The object can be created by local.network.MR.new5 or it can be imported.
seed0	num. Id of a seed to be sampled around.
n.neigh	A number of waves to be sampled around each seed in LSMI. For example, n.neigh = 0 corresponds to seeds only, and n.neigh = 1 corresponds to sampling seeds and their first neighbors). Note that the algorithm allows for multiple inclusions.

Value

a list containing:

seed	seed0 num. Id of a seed to be sampled around.
sampleN	A vector of numeric ids of the nodes from LSMI along with the original seed. This vector may have duplicates, since the algorithm allows for multiple inclusions.
unode	A vector containing the unique values in \$sampleN.
nodes.waves	A list of length n.neigh containing vectors where each vector reports numeric ids of nodes sampled in a particular wave.

Note

[LSMI](#).

References

Thompson, M. E., Ramirez Ramirez, L. L., Lyubchich, V. and Gel, Y. R. (2015), Using the bootstrap for statistical inference on random graphs. *Can J Statistics*. doi: 10.1002/cjs.11271

Examples

```
net <- artificial_networks[[1]]
a <- sample_about_one_seed(net = net, seed0 = 1, n.neigh = 2)
```

summary_net

Summary of a Network Degree Sequence

Description

This function provides summary statistics of a network degree distribution.

Usage

```
summary_net(net)
```

Arguments

net A network object that is list containing:

- edges** The edgelist of the network. A two column matrix where each row is an edge.
- degree** The degree sequence of the network, which is an integer vector of length n.
- n** The network order.

The object can be created by [local.network.MR.new5](#) or it can be imported.

Value

a list consisting of:

realdd	A vector of length <code>net\$n</code> where each element corresponds to the degree of a node in the network.
rmean	A numeric vector of length one that is the arithmetic mean of <code>realdd</code> . see mean .
rquart	The lower, median, and upper quartiles of degree sequence <code>realdd</code> .
rfreq	A numeric vector of length five where each element corresponds to the proportion of seeds with degree: 0, 1, 2, 3, and 4 (respectively).
rdec	A numeric vector of length nine containing the deciles of of degree sequence <code>realdd</code> in increasing order.

Examples

```
net <- artificial_networks[[1]]
a <- summary_net(net)
```

vertboot

Bootstrapping a network with vertex bootstrap

Description

This function bootstraps the the original network using a vertex bootstrap technique.

Usage

```
vertboot(m1, n.boot)
```

Arguments

<code>m1</code>	An adjacency matrix,the matrix represents the original network.
<code>n.boot</code>	A positive integer number, the number of bootstrap replications.

Value

A list of bootstrapped networks as adjacency matrices.

References

Tom A.B.Snijders., Stephen P.Borgatti. (1999). Non-Parametric Standard Errors and Tests for Network Statistics.

Examples

```
graph_ex <- igraph::graph_from_edgelist(artificial_networks[[1]]$edges)
m1 <- igraph::as_adjacency_matrix(graph_ex)
m1 <- as.matrix(m1)
vertboot_out <- vertboot(m1,20)
```

Index

*Topic **datasets**

- artificial_networks, [2](#)
- artificial_networks, [2](#)
- bootCI, [3](#)
- bootdeg, [3](#), [4](#), [5–7](#)
- BparametersEst, [5](#)
- cross_validation, [6](#)
- graph_from_adjacency_matrix, [8](#)
- graph_from_data_frame, [8](#)
- graph_from_edgelist, [8](#)
- igraph_to_network, [8](#)
- local.network.MR.new5, [7](#), [8](#), [10–13](#)
- LSMI, [9](#), [13](#)
- mean, [14](#)
- Oempdegreedistrib, [4](#), [10](#)
- read_graph, [8](#)
- sample_about_one_seed, [10](#), [12](#)
- summary_net, [13](#)
- vertboot, [14](#)