Package 'CollessLike'

July 17, 2017

Type Package

Title Methods to compute distribution and percentile of balance indices of phylogenetic trees

Version 0.0.67

Date 2017-03-10

Author Arnau Mir, Francesc Rossello, Lucia Rotger

Maintainer Lucia Rotger <lucia.rotger@uib.es>

Description Methods to compute distribution and percentile of balance indices (Colless-like, Sackin and cophenetic indices) of phylogenetic tree. It is also possible to generate random trees according to the alpha-gamma model.

License GPL (>= 2)

LazyData true

Depends R (>= 3.3.0)

Imports ape, igraph

RoxygenNote 6.0.1

Encoding UTF-8

NeedsCompilation no

R topics documented:

`CollessLike-package` *Methods to compute distribution and percentile of balance indices of phylogenetic trees*

Description

Methods to compute distribution and percentile of balance indices (Colless-like, Sackin and cophenetic indices) of phylogenetic tree. It is also possible to generate random trees according to the alpha-gamma model.

Details

Package:	CollessLike
Type:	Package
Title:	Methods to compute distribution and percentile of balance indices of phylogenetic trees
Version:	0.0.67
Date:	2017-03-10
Author:	Arnau Mir, Francesc Rossello, Lucia Rotger
Maintainer:	Lucia Rotger <lucia.rotger@uib.es>
Description:	Methods to compute distribution and percentile of balance indices (Colless-like, Sackin and cophenetic i
License:	GPL (>= 2)
LazyData:	true
Depends:	R (>= 3.3.0)
Imports:	ape, igraph
RoxygenNote:	6.0.1
Encoding:	UTF-8

Index of help topics:

Author(s)

Arnau Mir, Francesc Rossello, Lucia Rotger

Maintainer: Lucia Rotger <lucia.rotger@uib.es>

References

Chen, B., Ford, D., Winkel, M., A new family of Markov branching trees: the alpha-gamma model. *Electr. J. Probab*. **14** (2009), 400-430. MR2480547

A. Mir, F. Rossello, L.Rotger, A Colless-like balance index for multifurcating phylogenetic trees.

A. Mir, F. Rossello, L.Rotger, A new balance index for phylogenetic trees. Math. Biosc. 241 (2013).

M. J. Sackin, "Good" and "bad" phenograms. Sys. Zool, 21 (1972), 225-226.

Examples

```
# An example of a tree generated by the alpha-gamma model (igraph object)
 a.g.tree = a.g.model(5,0.5,0.3)
# To compute its percentile in the distribution
# distribution(a.g.tree,0.5,0.3,db.path=getwd())

# It is possible to do a new simulation of indices
# ('Repetitions' set as 100 for a fast example)
  indices.data = indices.simulation(5,0.5,0.3,100)
  distribution(a.g.tree,0.5,0.3,set.indices=indices.data)
```

`a.g.model` *Generates a random tree*

Description

Generates a random tree according to the alpha-gamma model

Usage

```
a.g.model(n, alpha, gamma)
```

Arguments

`n`	the number of leaves in the tree.
`alpha`	parametrer of the alpha-gamma model, between 0 and 1.
`gamma`	parametrer of the alpha-gamma model, between 0 and alpha.

Value

An igraph object that is the ford's tree created.

Author(s)

Lucia Rotger

References

Chen, B., Ford, D., Winkel, M., A new family of Markov branching trees: the alpha-gamma model. *Electr. J. Probab*. **14** (2009), 400-430. MR2480547

Examples

```
tree = a.g.model(10,0.8,0.1)
# plot(tree,layout=layout.reingold.tilford(tree,root=which(degree(tree,mode="in")==0)))

tree = a.g.model(5,0.5,0.3)
# plot(tree,layout=layout.reingold.tilford(tree,root=which(degree(tree,mode="in")==0)))
```

`balance.indices` *Computes 3 balance index of a tree*

Description

Computes Colles-like, sackin and cophenetic indices of a phylogenetic tree.

Usage

```
balance.indices(tree, norm = FALSE)
```

Arguments

`tree`	a single phylogenetic tree. It can be introduced as a string in the Newick format, as a "phylo" object (`ape` package) or as an "igraph" object (`igraph` package).
`norm`	a logical object indicating if the indices should been normalized or not.

Details

The Colless-like index is the generalization of the Colless' index for non-binary trees.

The Sackin's index is computed as the sum of the number of ancestors for each leave of the tree.

Cophenetic index is computed as the sum of the depths of the least common ancestor (LCA) of every pair of leaves.

Value

A numeric vector with the three computed indices of the tree: `Colless-like`, `Sackin` and `Cophenetic` values.

Author(s)

Lucia Rotger

References

A. Mir, F. Rossello, L.Rotger, A Colless-like balance index for multifurcating phylogenetic trees.

A. Mir, F. Rossello, L.Rotger, A new balance index for phylogenetic trees. Math. Biosc. 241 (2013).

M. J. Sackin, "Good" and "bad" phenograms. Sys. Zool, 21 (1972), 225-226.

Examples

```
#From a newick string
balance.indices("(1,2,3,4,5);")
balance.indices("(1,(2,(3,(4,5))));")

#From a phylo object
require(ape)
random.tree = rtree(5,rooted=TRUE)
balance.indices(random.tree)

#An example of a tree generated by the alpha-gamma model (igraph object)
a.g.tree = a.g.model(5,0.5,0.3)
balance.indices(a.g.tree)

#All of them can be normalized (between 0 and 1)
balance.indices("(1,2,3,4,5);",norm=TRUE)
balance.indices("(1,(2,(3,(4,5))));",norm=TRUE)
balance.indices(random.tree,norm=TRUE)
balance.indices(a.g.tree,norm=TRUE)
```

---

colless.like.index *Computes the Colless-like index of a tree*

Description

Computes the Colless-like balance index of a phylogenetic tree.

Usage

```
colless.like.index(tree, f.size = "ln", diss = "MDM", norm = FALSE)
```

Arguments

`tree`	a single phylogenetic tree. It can be introduced as a string in the Newick format, as a "phylo" object (`ape` package) or as an "igraph" object (`igraph` package).
`f.size`	by default is `f(n)=ln(n+e)`. It can be set as "ln" (default) or "exp" (`f(n)=exp(n)`). It aso can be a function defined by user, in this case the index can not be normalized.
`diss`	by default the dissimilarity is `MDM`. It can also be set as "var" (the sample variance) or "sd" (the sample standard deviation). It aso can be a function defined by user, in this case the index can not be normalized.
`norm`	a logical object indicating if the indices should been normalized or not.

Details

The Colless-like index is the generalization of the Colless' index for non-binary trees.

By default, the f-size function is `f(n)=exp(n)` and the dissimilarity is the mean deviation from the median (MDM). It is possible to change them by specifying it with the parameters `f.size` and `diss`, with "exp" the f-size would be `f(n)=exp(n)`, and with "var" (or "sd") the dissimilarity would be the sample variance (or the sample standard deviation). It is also possible to set a new function for both parameters, see "References".

Value

A numeric value.

Author(s)

Lucia Rotger

References

A. Mir, F. Rossello, L.Rotger, A Colless-like balance index for multifurcating phylogenetic trees.

Examples

```
#' #From a newick string
colless.like.index("(1,2,3,4,5);")
colless.like.index("(1,(2,(3,(4,5))));")

#From a phylo object
require(ape)
random.tree = rtree(5,rooted=TRUE)
colless.like.index(random.tree)

#An example of a tree generated by the alpha-gamma model (igraph object)
a.g.tree = a.g.model(5,0.5,0.3)
colless.like.index(a.g.tree)

#All of them can be normalized (between 0 and 1)
colless.like.index("(1,2,3,4,5);",norm=TRUE)
colless.like.index("(1,(2,(3,(4,5))));",norm=TRUE)
colless.like.index(random.tree,norm=TRUE)
colless.like.index(a.g.tree,norm=TRUE)

#Changing f-size
colless.like.index(a.g.tree,f.size="exp")

#Changing dissimilarity
colless.like.index(a.g.tree,diss="var")
colless.like.index(a.g.tree,diss="sd")

#Changing both
colless.like.index(a.g.tree,f.size="exp",diss="var")
colless.like.index(a.g.tree,f.size="exp",diss="sd")
```

---

cophen.index *Computes the cophenetic index of a tree*

Description

Computes the cophenetic index of balance of a phylogenetic tree.

Usage

```
cophen.index(tree, norm = FALSE)
```

Arguments

tree	a single phylogenetic tree. It can be introduced as a string in the Newick format, as a "phylo" object (ape package) or as an "igraph" object (igraph package).
norm	a logical object indicating if the indices should been normalized or not.

Details

The cophenetic index is computed as the sum of the depths of the least common ancestor (LCA) of every pair of leaves.

Value

A numeric value.

Author(s)

Lucia Rotger

References

A. Mir, F. Rossello, L.Rotger, A new balance index for phylogenetic trees. Math. Biosc. 241 (2013).

Examples

```
#' #From a newick string
cophen.index("(1,2,3,4,5);")
cophen.index("(1,(2,(3,(4,5))));")

#From a phylo object
require(ape)
random.tree = rtree(5,rooted=TRUE)
cophen.index(random.tree)

#An example of a tree generated by the alpha-gamma model (igraph object)
a.g.tree = a.g.model(5,0.5,0.3)
cophen.index(a.g.tree)

#All of them can be normalized (between 0 and 1)
cophen.index("(1,2,3,4,5);",norm=TRUE)
cophen.index("(1,(2,(3,(4,5))));",norm=TRUE)
cophen.index(random.tree,norm=TRUE)
cophen.index(a.g.tree,norm=TRUE)
```

---

distribution — *Computes the percentile in the alpha-gamma distribution of the indices of the given tree*

Description

Computes the percentile in the alpha-gamma distribution of the indices (Colless-like, Sackin and Cophenetic) of the given tree. It also plots all the three distributions highlighting where are the indices of the given tree or it plots a percentile plot with the percentiles of the tree.

Usage

```
distribution(tree, alpha = NA, gamma = NA, set.indices = NULL,
  new.simulation = FALSE, repetitions = 1000,
  legend.location = "topright", cex = 0.75, percentile.plot = FALSE,
  db.path = getwd())
```

Arguments

tree	a single phylogenetic tree. It can be introduced as a string in the Newick format, as a "phylo" object (ape package) or as an "igraph" object (igraph package).
alpha	parametrer of the alpha-gamma model, between 0 and 1.
gamma	parametrer of the alpha-gamma model, between 0 and alpha.
set.indices	if NULL(default) the indices data is taken from stored data or from a new simulated data (See "Details"). If not, it must be a 3-column data.frame with the three balance indices (Colles-like, Sackin, Cophenetic). See indices.simulation.
new.simulation	if FALSE(default) the indices data it could be from a data.frame introduced by the user or a data.frame from our database. If it is TRUE, a new indices data set is computed. See indices.simulation.
repetitions	the number of trees to generate in case a new simulation is done.
legend.location	location where the legend is going to be placed. See "Details".
cex	expansion factor of the legend. See "Details".
percentile.plot	if TRUE plots the percentile plot of the indices. If it is FALSE(default) , then a distribution plot is represented.
db.path	by default is the actual working directory. It should be changed if the data base is going to be used and it is located in a different directory.

Details

Two plots are available: one with the acumulated percentiles of the indices (percentile.plot=FALSE), and the other with the distribution (percentile.plot=TRUE).

The stored data available has been calculated for a number of leaves between 3 and 50. For each of them, the parameters are set as alpha in {0,0.1,0.2,...,1} and gamma in {0,0.1,0.2,...,alpha}. If the introduced parameters are not in the list, a new computation is done with them and a new dataset of trees is generated, and computed its indices. The number of trees generated can be modified by the parameter repetitions (see indices.simulation for more information). This computation may take some time, therefore you can computate it separately with indices.simulation, save its value and then call this function by setting it as the parameter set.indices.

The legend is placed with the graphics function legend(), so its location can be specified by setting legend.position to a single keyword from the list "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center". The expansion factor for the legend is controlled by the parameter cex, by default cex=1. See legend.

Value

A numeric vector with the three percentiles.

Author(s)

Lucia Rotger

References

Chen, B., Ford, D., Winkel, M., A new family of Markov branching trees: the alpha-gamma model. *Electr. J. Probab*. **14** (2009), 400-430. MR2480547

A. Mir, F. Rossello, L.Rotger, A Colless-like balance index for multifurcating phylogenetic trees.

A. Mir, F. Rossello, L.Rotger, A new balance index for phylogenetic trees. Math. Biosc. 241 (2013).

M. J. Sackin, "Good" and "bad" phenograms. Sys. Zool, 21 (1972), 225-226.

See Also

legend, indices.simulation, balance.indices

Examples

```
#If it is need, to specify the location of the database
#folder=".../CollesLikeDataBase/"
##If not,
folder=getwd()
## Different ways to introduce the tree
#From a newick string
distribution("(1,2,3,4,5);",0.5,0.3,db.path=folder)
distribution("(1,(2,(3,(4,5))));",0.5,0.3,db.path=folder)

#From a phylo object
require(ape)
random.tree = rtree(5,rooted=TRUE)
distribution(random.tree,0.5,0.3,db.path=folder)

#An example of a tree generated by the alpha-gamma model (igraph object)
a.g.tree = a.g.model(5,0.5,0.3)
distribution(a.g.tree,0.5,0.3,db.path=folder)

## Different indices data
# From our data base
distribution(a.g.tree,0.5,0.3,db.path=folder)

# From a data.frame generated by 'indices.simulation'
# ('Repetitions' set as 10 for a fast example)
indices.data = indices.simulation(5,0.5,0.3,10)
distribution(a.g.tree,0.5,0.3,set.indices=indices.data)

# Allow the function to do a new generation of data and compute their indices
distribution(a.g.tree,0.5,0.3,new.simulation=TRUE,repetitions=10)
# WARNING! it might take a long time, it depends on the parameters
# 'n' (number of leaves) and 'repetition' (number of repetitions)
```

`indices.simulation` *Generates random trees and computes their balance indices*

Description

Generates a list of trees according to the introduced parameters for the alpha-gamma model. Then, this 3 balance index are calulated: Colless-like, Sackin and Cophenetic.

Usage

```
indices.simulation(n, alpha = NA, gamma = NA, repetitions = 1000,
  norm = FALSE)
```

Arguments

`n`	the number of leaves in the tree.
`alpha`	parametrer of the alpha-gamma model, between 0 and 1.
`gamma`	parametrer of the alpha-gamma model, between 0 and alpha.
`repetitions`	the number of trees to generate.
`norm`	a logical object indicating if the indices should been normalized or not.

Details

Given a number of leaves, the function generates a tree with that number of leaves and computates the three index of balance (Colles-like, Sackin and Cophenetic with function `balance.indices`). This is done as many times as it is set by 'repetitions' parameter , and it generates a 3-column data.frame of indices.

The trees are generated according to the alpha-gamma model. This parameters can be specified by `alpha` and `gamma` parameters of the function. The following cases are distinguish:

- `alpha = NA` and `gamma = NA` : All the 66 combinations of `alpha` in { 0, 0.1, 0.2, ... ,0.9, 1 } and `gamma` in { 0, 0.1, ... ,`alpha` } are done.
- `alpha` in [0,1] and `gamma = NA` : Since `alpha` is fixed, all the combinations with that `alpha` and `gamma` in { 0, 0.1, ... ,`alpha` } are done.
- `alpha` in [0,1] and `gamma` in [0,`alpha`] : Both parameters are fixed then, only that combination is done.

Value

A 3-column data.frame with the indices of Colless-like, Sackin and Cophenetic for every generated tree. If more than one data.frame has been generated, then the returned value is a data.frame list (its names specify which alpha and gamma parameters have generated that data.frame, for instance "a0.5g0.3" indicates alpha=0.5 and gamma=0.3).

Author(s)

Lucia Rotger

References

Chen, B., Ford, D., Winkel, M., A new family of Markov branching trees: the alpha-gamma model. *Electr. J. Probab*. **14** (2009), 400-430. MR2480547

A. Mir, F. Rossello, L.Rotger, A Colless-like balance index for multifurcating phylogenetic trees.

A. Mir, F. Rossello, L.Rotger, A new balance index for phylogenetic trees. Math. Biosc. 241 (2013).

M. J. Sackin, "Good" and "bad" phenograms. Sys. Zool, 21 (1972), 225-226.

See Also

balance.indices

Examples

```
#('Repetitions' set as 10 for a fast example)
indices.table = indices.simulation(5,0.5,0.3,repetitions=10)
head(indices.table)

#Normalized indices (between 0 and 1)
indices.table = indices.simulation(5,0.5,0.3,repetitions=10,norm=TRUE)
head(indices.table)

#Without specifying alpha and gamma
indices.list = indices.simulation(5,repetitions=10)
#by default alpha=seq(0,1,0.1) and gamma=seq(0,alpha,0.1), thus
length(indices.list) #=66
#all the elements of the list have a name that identifies its parameters
indices.list$a0.5g0.3
indices.list$a0.7g0.2
```

---

sackin.index *Computes the Sackin index of a tree*

Description

Computes the Sackin index of balance of a phylogenetic tree.

Usage

```
sackin.index(tree, norm = FALSE)
```

Arguments

tree	a single phylogenetic tree. It can be introduced as a string in the Newick format, as a "phylo" object (ape package) or as an "igraph" object (igraph package).
norm	a logical object indicating if the indices should been normalized or not.

Details

The Sackin's index is computed as the sum of the number of ancestors for each leave of the tree.

Value

numeric value.

Author(s)

Lucia Rotger

References

M. J. Sackin, "Good" and "bad" phenograms. Sys. Zool, 21 (1972), 225-226.

Examples

```
#From a newick string
sackin.index("(1,2,3,4,5);")
sackin.index("(1,(2,(3,(4,5))));")

#From a phylo object
require(ape)
random.tree = rtree(5,rooted=TRUE)
sackin.index(random.tree)

#An example of a tree generated by the alpha-gamma model (igraph object)
a.g.tree = a.g.model(5,0.5,0.3)
sackin.index(a.g.tree)

#All of them can be normalized (between 0 and 1)
sackin.index("(1,2,3,4,5);",norm=TRUE)
sackin.index("(1,(2,(3,(4,5))));",norm=TRUE)
sackin.index(random.tree,norm=TRUE)
sackin.index(a.g.tree,norm=TRUE)
```

Index