Package 'SyNet'

November 16, 2011

Type Package

Title Inference and Analysis of Sympatry Networks

Version 2.0

Date 2011-11-11

Author Daniel A. Dos Santos

Maintainer Daniel A. Dos Santos <pseudalopex_79@yahoo.com>

Depends tkrplot, tcltk

Description Infers sympatry matrices from distributional data and analyzes them in order to identify groups of species cohesively connected.

License GPL

R topics documented:

SyNet-package	*Inference and Analysis of Sympatry Networks*

Description

Historical Biogeography focuses basically on sympatry patterns among species. **SyNet** extracts hypotheses of sympatry from available distributional evidence and integrates them into weighted and binary sympatry networks. **SyNet** tests the adequacy of networks to be segregated into groups of species cohesively sympatric (units of co-occurrence). It identifies units of co-occurrence from the more inclusive network where they are embedded. NAM algorithm iteratively removes intermediary species that are linked to those units otherwise disconnected. Here, you can do among other things: i) to analyze the spatial affinity between sets of records and ii) to explore dynamically the cleavogram associated to the analysis of sympatry networks. You can track many new ideas on numerical classification throughout the examples of functions. Remarkably, geographical data associated to any network partitioning can be exported to KML files which can be opened via Google Earth.

Details

Package:	SyNet
Type:	Package
Version:	2.0
Date:	2011-11-11
License:	GPL

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Dos Santos D.A., Fernandez H.R., Cuezzo M.G., Dominguez E. 2008. *Sympatry Inference and Network Analysis in Biogeography.* Systematic Biology 57:432-448.

Dos Santos D.A., Cuezzo M.G., Reynaga M.C., Dominguez E. 2011. *Sympatry Inference and Network Analysis in Biogeography*. Systematic Biology (in press)

Examples

```
data(sciobius2x2)
# Derive a sympatry matrix from distributional table.
# Entries equal to 1 mean co-extensive sympatry.
infer <- gridinfer(dntable = sciobius2x2)
# Run NAM method on the previously inferred sympatry network
outnam <- nam(infer)
# Extract the sympatry network with elements optimally rearranged
# according to the NAM sequence of splitting events.
sm <- outnam$mt
```

```
# Plot the original network
forgraph <- seq(0, by = 2*pi/47, length = 47)
xcoord <- cos(forgraph)
ycoord <- sin(forgraph)
plot(xcoord, ycoord, main = "SCIOBIUS Sympatry Network",
     xlab = "", ylab = "", pch = 19, col = 2, cex = 1.3, axes = FALSE)
text (xcoord*1.05, ycoord*1.05, outnam$leaves)
for (i in 1:46)
  for (j in (i+1):47)
    if(sm[i,j] > 0) lines (c(xcoord[i], xcoord[j]), c(ycoord[i], ycoord[j]))
mtext("Subnetwork 0 (Original network)", side = 1, line = 2, col = 4)
# Next, we define a control window to go from initial network to the last sub-network.
# Here, you can visualize the intrinsic dynamic associated to the iterative removal
# of intermediary species.
subnet <- 0
onClick <- function(action) {
  if(action == 1) subnet <<- 0
  if(action == 2) subnet <<- pmax(0, subnet - 1)
  if(action == 3) subnet <<- pmin(outnam$nsub - 1, subnet + 1)
  if(action == 4) subnet <<- outnam$nsub - 1
  notremoved <- outnam$LastNet >= subnet
  plot(xcoord, ycoord, main = "SCIOBIUS Sympatry Network",
       xlab = "", ylab = "", type = "n", axes = FALSE)
  points(xcoord[notremoved], ycoord[notremoved], pch = 19, col = 2)
  text (xcoord[notremoved]*1.05, ycoord[notremoved]*1.05, outnam$leaves[notremoved])
  for (i in 1:(outnam$nsp - 1)){
     if(!notremoved[i]) next
     for (j in (i+1):outnam$nsp) {
        if(!notremoved[j]) next
        if(sm[i,j] > 0) lines (c(xcoord[i], xcoord[j]), c(ycoord[i], ycoord[j]))
     }
  }
  mtext(paste("Subnetwork", subnet), side = 1, line = 2, col = 4)
}
## Not run:
tt <- tktoplevel()
tkwm.title(tt, "Evolution of NAM ...")
but1 <- tkbutton(tt, text = "<<", command = function(...) onClick(1), padx = 20)
but2 <- tkbutton(tt, text = "<", command = function(...) onClick(2), padx = 20)
but3 <- tkbutton(tt, text = ">", command = function(...) onClick(3), padx = 20)
but4 <- tkbutton(tt, text = ">>", command = function(...) onClick(4), padx = 20)
tkgrid(tklabel(tt, text = "*** Change Sub-Network ***", font = "Times 16", foreground = "blue"),
       columnspan = 4)
tkgrid(but1, but2, but3, but4)
#Finally, type the following command to see the cleavogram called 'outnam':
cleavogram()

## End(Not run)
```

---

acsh   *Average Cost for Spatial Homogenization*

Description

Produces a weighted matrix of spatial association between species from points of occurrence.

Usage

```
acsh(dotdata)
```

Arguments

dotdata   Object of class 'dotdata'

Details

This function performs pairwise comparisons between species point sets. It calculates the average displacement required to achieve homopatry between species. If two species share all their records they are said to be homopatric (etymologically, with the same territory).

Given two sets of points A and B, the measure ACSH as described in Dos Santos et al. (2011) corresponds to the weighted average across the nearest interspecific distances between the elements of A and B. The weight assigned to each element is proportional to the mean length of MST arcs incident to it. Close affinity between distributions is suggested when they deliver an ACSH score near to zero.

Value

Returns a matrix that expresses the dissimilarity between species point sets via non-negative real values.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Dos Santos D.A., Cuezzo M.G., Reynaga M.C., Dominguez E. 2011. *Towards a Dynamic Analysis of Weighted Networks in Biogeography.* Systematic Biology (in press).

See Also

Objects of class 'dotdata' are created via `procdnpoint`.

Examples

```
data(mayflynz)
aux <- procdnpoint(mayflynz) #Pre-processing
head(acsh(aux)) #Generates the dissimilarity matrix and displays some of its values
```

---

acshgral   *Generalized ACSH score*

Description

Returns the ACSH score for some set of species of any cardinality.

Usage

```
acshgral(dotdata, species)
```

Arguments

dotdata	Object of class 'dotdata'
species	Vector of indices specifying the target group of species

Details

This function generalizes the notion of the ACSH score for pairs of species. It may consider simultaneously many elements from a pool of species. For a given universe of records, the overall ACSH averages the longest interspecific gap measured at each point. The interspecific gap corresponds to the shortest distance between a single point and a given set of species points.

I have implemented a procedure of agglomerative hierarchical clustering based on this measure. I think that results are very promising in the context of marked spatial point processes. This procedure advances on the classification problem stated by Mane et al. (2006). Basically, the idea consists of iteratively merging two items into a larger inclusive cluster so that the generalized acsh score is minimal across the totality of feasible merging scenarios.

Value

A single non-negative real number that denotes the cost (in terms of imaginary spatial displacement) to achieve complete overlap between distributions.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Dos Santos D.A., Cuezzo M.G., Reynaga M.C., Dominguez E. 2011. *Towards a Dynamic Analysis of Weighted Networks in Biogeography.* Systematic Biology (in press).

Mane S., Kang J., Shekhar S., Srivastava J., Murray C., Pusey A. 2006. *Identifying Clusters in Marked Spatial Point Processes: A Summary of Results.* Technical Report 06-006. Department of Computer Science and Engineering, University of Minnesota.

See Also

The function acsh addresses pairwise comparisons between species point sets.

Examples

```
data(mayflynz)
procnz <- procdnpoint(mayflynz)
nsp <- length(procnz$Label)
testmt <- acsh(procnz)
# Following, we will produce a matrix of ACSH scores for pairs of species,
# but now with the generalized form.
calcmt <- matrix(0, nrow = nsp, ncol = nsp)
for(i in 1:(nsp-1))
  for(j in (i+1):nsp)
    calcmt[i,j] <- acshgral(procnz, c(i, j))
all(calcmt + t(calcmt) == testmt)
#The previous statement should be TRUE because the generalized ACSH
#corresponds to the standard pairwise ACSH when
#two species are submitted to the function "acshgral".
################
```

```
###############
# This code makes explicit the above idea of hierarchical clustering based on the minimization
# of the ACSH profile throughout the nested structure of groups formed during the merging process
diag(testmt) <- Inf
newdist <- matrix(-1, nsp, nsp)
notation <- as.character(1:nsp)
classes <- 1:nsp
analyze <- array(TRUE, nsp)
while(max(classes)!= 1) {
  aux <- which.min(testmt)
  items <- sort(c(row(testmt)[aux], col(testmt)[aux]))
  #Update cluster structure and ACSH scores
  ngr <- which(classes %in% items)
  classes[ngr] <- items[1]
  notation[items[1]] <- paste("(", notation[items[1]], " ", notation[items[2]], ")", sep = "")
  newdist[items[1], items[2]] <- newdist[items[2], items[1]] <- testmt[aux]
  analyze[items[2]] <- FALSE
  testmt[items, ] <- testmt[, items] <- Inf
  for(i in which(analyze)) {
    ogr <- which(classes == classes[i])
    if(i == items[1]) next
    testmt[items[1], i] <- testmt[i, items[1]] <- acshgral(procnz, c(ngr, ogr))
  }
}
#Plot the respective dendrogram
tt <- max(newdist) + 1
newdist <- ifelse(newdist == -1, tt, newdist)
plot(hclust(as.dist(newdist), method = "single"), label = procnz$Label,
     cex = 0.7, xlab = "", ylab = "Generalized ACSH")
#Display relationships into a parenthetical notation. You can track the indices of the species in the
#leaves of the following dendrogram
plot(hclust(as.dist(newdist), method = "single"))
cat(notation[1])
```

---

cleavogram	*Cleavogram*

Description

Using a graphical user interface (GUI) this function plots an interactive cleavogram, and allows the
search of connected groups of species under different scenarios of cohesiveness.

Usage

```
cleavogram()
```

Details

The cleavogram is a visual tool for exploring the structural changes in a sympatry network as the
removal of intermediary species proceeds. See the referenced paper for more information.

The main window has three panels. The upper left panel is the cohesiveness manager. Here, the
values of the cohesiveness parameters can be specified. If connected dyads are considered to be
meaningful entities itself, keep transitivity to 0 and eccentricities set 'NONE'. The lower left panel

exhibits in real time the spatial expressions of the current selected branch of the cleavogram. The right panel holds the cleavogram. There are several buttons at the left of the cleavogram to change its general appearance.

A left click on the cleavogram displays a contextual menu to copy/save the plot for further editing actions. The double right click opens a new window to better explore the spatial expression associated to the target branch of the cleavogram. If you press down the left hand mouse button, and while keeping it pressed you move the mouse pointer throughout the cleavogram, branches are highlighted in blue and the geographical panel refreshes accordingly.

There is a combo box with a list of species (i.e. vertices of the network or leaves of the cleavogram) below the cleavogram panel. This enables to traverse the cleavogram from the root to the respective leaf, emphasizing the trajectory with a distinctive line. This is a nice resource to see the distribution of sister taxa over the cleavogram.

Menu Analysis allows to perform flat partitions following different search strategies. It also launches a visualizer of spatial expressions associated to a given flat partition.

Value

This function leaves the operating environment to allow the user access to the data. Flat partitions can be saved into an R object of class 'nampartition' provided of the next elements:

kind	Character. Specifies the kind of spatial data either points or grids.
status	Two-columns data frame. Taxa names are located at the first column. The other column refers to the classification obtained after applying the flat partition.
occupancy	List of records by individual taxon.
coords	Two-column matrix. Values of the spatial coordinates associated to each record of the data set.

Note

Once you have set the criteria of cohesiveness, make sure you have filtered the branches of the cleavogram satisfying those criteria. Go to menu Analysis and then Filter by criteria.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Dos Santos D.A., Cuezzo M.G., Reynaga M.C., Dominguez E. 2011. *Towards a Dynamic Analysis of Weighted Networks in Biogeography.* Systematic Biology (in press).

See Also

nam creates objects of class cleavogram that can be opened by the menu Data –> Choose cleavogram ...

Examples

```
## Not run:
#NAM method applied on the example of New Zealand mayflies
data(mayflynz)
dotdata <- procdnpoint(mayflynz)
toposimilar(dotdata) -> toponz
```

```
acsh(dotdata) -> acshnz
reweight(toponz) -> toporew
#The next step consists of obtaining the binary sympatry network, that is to
#create the respective object of class 'dotinference'. This task can be done
#interactively with function dotinfer.
#Here, we will create the required 'dotinference' object by hand. The thresholding
#rule match that used by Dos Santos et al. (2011).
rslt <- c()
rslt$sm <- ifelse(acshnz < 100 & toporew > 0.8, 1, 0)
rslt$Label <- dotdata$Label
rslt$occupancy <- dotdata$occupancy
rslt$coords <- dotdata$coords
rslt$kind <- "points"
class(rslt) <- "dotinference"
#Now, run NAM over the previous created object. Then go to the cleavogram and explore it.
outnz <- nam(rslt)
cleavogram()

## End(Not run)
```

---

creategrid   *Project Point Data into a Rectangular Mesh of OGUs*

Description

Dispalys punctual records into a grid system and creates the respective distributional table. Actions are facilitated by a graphical user interface (GUI).

Usage

```
creategrid(dnpoint)
```

Arguments

dnpoint   An object of class 'dnpoint'.

Details

The grid parameters (cell size and upper leftmost corner) can be manually adjusted. Keep press the left mouse button and move freely on the plotting region to select the upper leftmost corner.

A choropleth of taxonomic richness is next to the graticule/gradicule under manipulation. A brief report of transformations (from points to grid) can be exported to a PDF file.

Value

A data frame with species and rectangular OGUs (operative geographic units) in rows and columns, respectively. The distributional table is saved into the global environment. The presence of a taxon into a given cell of the grid is coded 1, otherwise 0. The coordinates for each cell or OGU are arranged into the first two rows of the table, preceding the species themselves. The coordinates are integer numbers and match the indices of rows and columns in the grid system.

Note

Tobler (1975) is a good starting point to put the geographical reasoning based on grids into an adequate context. Grids are appealing because of their isomorphism with a matrix structure. Unfortunately, this simple method to deal with spatial data takes arbitrary decisions on features like shape, size and placement of cells over the surface. Moreover, it is divorced from the spherical model of the Earth and forces distributions to accommodate their extensions to the rectangular pattern of the grid. This last charge can be analogized with the Greek myth of Procrustes.

Procrustes lived near the city of Eleusis. He captured his victims and took them back to his iron bed and stretched them out until they fit. If the victim was too long, he would cut him off. In any case, he made them fit a pre-established frame. Theseus put an end to this obsession: he made the giant undergo the same 'normalizing' treatment.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Tobler W. 1975. *Cellular geography.* Translation from the conference entitled "Schachbrett Modelle in der Geographie" exposed at "Arbeitskreis fur neue Methoden in der Regionalforschung", Wien.

See Also

See `read.coord` to capture point data from an external txt file

Examples

```
data(mayflynz)
## Not run:
creategrid(mayflynz)

## End(Not run)
```

---

dotinfer   *Inference of a Sympatry Network from Point Data*

Description

Using a Graphical User Interface (GUI), this function explores the pairwise relationships scored by weighted matrices of spatial association. It also helps to choose the cutoff value, creates an interactive network and plots intensity matrices.

Usage

```
dotinfer(dotdata, mtxdata)
```

Arguments

dotdata	Object of class 'dotdata'.
mtxdata	List of weighted matrices based on the previous data. The names for each element of this list are considered as the identity labels of matrices.

Details

The list of weighted matrices in addition to the list of involved taxa are loaded into two different combo boxes. It is very important to identify the right nature of the data (i.e. dissimilarity/similarity). In similarity matrices there is a direct relationship between their scores and the strength of association. On the contrary, in dissimilarity matrices there is an inverse relationship between their scores and the strength of association.

Once you have selected a given matrix of spatial association, its single linkage dendrogram is drawn. On the other hand, after you choose a given focus species, its couples are decreasingly ordered by the strength of spatial association between them. The upper left panel shows the spatial distributions of species pairs under consideration. The current focus species is indicated through red dots, whereas the current couple is represented by blue dots. There are buttons to shift the couples of the focus species, thus the user can move from more to less associated neighbors according to the values dictated by the matrix under analysis.

Thresholds are used to dichotomize a weighted matrix into a binary one. I know this procedure entails loss of information. In order to preserve the largest amount of information, the bottom right panel assists the user in this task. Thus, the strongest relationships found on the matrix are recovered using the notion of stable couples (Gale and Shapley 1962). The stable matchings are assumed to be representative of the strongest links in the underlying weighted network. Then, the distribution of affinity scores across the stable couples is estimated via kernel density estimates. The percentage under the density curve can be used as objective guideline to select the cutoff value.

Two appealing features of this interface can be found at the Display menu : i) manipulations of VAT images and ii) interactive layout of the inferred sympatry network.

The VAT approach presents pairwise association information about the set of objects O = {o_1, ..., o_n} as a square digital image with n*n pixels, after the objects are suitably reordered so that the image is better able to highlight potential cluster structure. VAT operates on the matrix of association between items and transforms each score into a value in the gray tone scale.

The network derived from thresholding is displayed in an interactive graph drawing facility. Here, nodes can be picked out and moved throughout the canvas (in addition to its incident edges). The right click on a given node highlights the target node and its neighbors in red and blue colors, respectively.

Value

An object of class 'dotinference' is created into the working environment. It is a list with components:

sm	An adjacency matrix that suggests (1) or not (0) the occurrence of a sympatric link between species.
Label	Character vector of species labels.
occupancy	List of records occupied by each species.
coords	Spatial coordinates for each record occurring in the data set. They are arranged into a two columns matrix.
kind	Character. Specifies the kind of distributional data, that is "points".

Note

Intensity matrices are produced with the VAT algorithm published by Bezdek and Hathaway (2002). Certainly, improvements on the layout of the network need to be introduced. I hope to work on that in the near future!

Author(s)

Daniel A. Dos Santos, <dadossantos@csnat.unt.edu.ar>

References

Bezdek J.C., Hathaway R. 2002. *VAT: A Tool for Visual Assessment of (Cluster) Tendency.* Proc. Int. Joint Conf. Neural Networks (IJCNN 2002): 2225-2230.

Dos Santos D.A., Cuezzo M.G., Dominguez E., Reynaga M.C. 2011. *Towards a Dynamic Analysis of Weighted Networks in Biogeography.* Systematic Biology (in press).

Gale D., Shapley L. 1962. *College Admissions and the Stability of Marriage.* Amer. Math. Month. 69:9-15.

See Also

Objects of class 'dotdata' result from submitting raw punctual data to the function procdnpoint. Kernel density estimates are calculated with the density function. Stables couples are obtained with the standard arguments available at stablecouple.

Examples

```
data(mayflynz)
aux <- procdnpoint(mayflynz) #Pre-processing of data
toponz <- toposimilar(aux) #Similarity matrix.
acshnz <- acsh(aux) #Dissimilarity matrix.
rewnz <- reweight(toponz) #Similarity matrix.
#Explore the content of previous matrices in addition to the distributions
#of involved species.
## Not run:
dotinfer(aux, list(toponz = toponz, acshnz = acshnz, rewnz = rewnz))

## End(Not run)
```

---

epiphragmophora	*Landsnails of Norhtwestern Argentina*

Description

Provides records for 21 species of the landsnail genus *Epiphragmophora* consisting of 145 records with coordinates in decimal format.

Usage

```
data(epiphragmophora)
```

Format

List of 3 elements: $ Numpoints: int 145 $ Points :'data.frame': 145 obs. of 3 variables: ..$ IDsp : num [1:145] 1 1 1 1 1 1 1 1 1 1 ... ..$ Longitud: num [1:145] -65.5 -65.4 -65.3 -65.2 -65.4 ... ..$ Latitud : num [1:145] -26.8 -26.8 -26.7 -26.6 -26.8 ... $ Label : chr [1:21] "argentina" "cryptomphala" "escoipensis" "guevarai" ... $ Type : chr "geographical" - attr(*, "class")= chr "dnpoint"

Details

Data were previously incorporated to an object of class 'dnpoint'.

Source

M.G. Cuezzo, curator of Malacological Collection of Miguel Lillo Museum, Tucuman, Argentina.

References

Cuezzo M.G. 2006. *Systematic Revision and Cladistic Analysis of Epiphragmophora Doering from Argentina and Southern Bolivia (Gastropoda: Stylommatophora: Xanthonychidae).* Malacologia 49:121-188.

Dos Santos D.A., Fernandez H.R., Cuezzo M.G., Dominguez E. 2008. *Sympatry Inference and Network Analysis in Biogeography.* Systematic Biology 57:432-448.

Examples

```
data(epiphragmophora)
plot(epiphragmophora[[2]][,2:3], main = "Landsnails (Epiphragmophora genus) from Argentina",
     xlab = "Latitude", ylab = "Longitude", asp = 1.3)
```

---

gridinfer   *Sympatry Inference from Grids*

Description

Produces a sympatry matrix from distributional data organized as tables of species vs. pre-defined spatial units.

Usage

```
gridinfer(file = NULL, dntable = NULL, sp_row = TRUE, reciprocity = TRUE,
 criterion = "max", tolerance = sqrt(2), conditioned = TRUE, ...)
```

Arguments

file	Character string naming the ASCII file to read it. The file is read by `read.table` and contains an species-by-grids matrix of presence/absence. Entries must be non-negative and scores higher than zero are interpreted as presence.
dntable	A matrix or data frame object with non-negative entries. It is an species-by-grids distributional table. Entries must be non-negative and scores higher than zero are interpreted as presence. Indices or coordinates for each cell must be arranged in the same table, preceding the species themselves
sp_row	Logical. If `TRUE` rows are interpreted as species and columns as grids. If FALSE, the opposite is considered.
reciprocity	Logical. If `TRUE`, comparisons between species ranges are bidirectional.
criterion	Character string. It should be a valid R function to extract a summary statistic from a profile of proximities.
tolerance	Numeric. Represents the upper threshold to assess co-extensive sympatry.

conditioned	Logical. If TRUE, sympatry between taxa is subordinated to their co-occurrence in at least one cell.
...	Arguments to be passed to the function `read.table`

Details

Species are sympatric if their ranges overlap, whereas they are allopatric if their ranges show spatial disjunction. Inference of sympatry can be addressed over raw distributional data (i. e., dot maps or coordinates of species records) or over distributional tables (i. e., tables of species vs. pre-defined areas that indicates occupancy or not).

In case of punctual data, sympatry is inferred by the interaction of geographical proximity and interpenetration of species point sets. Thus, sympatry is proposed when records are close together, sharing an underlying area of unknown boundary. In case of distributional data based on grids, sympatry is inferred by co-occurrence of species in pre-defined spatial unit (OGUs, operative geographical units).

In the context of grids, the older version of SyNet considered two taxa sympatric if they shared at least a single OGU of occurrence. This relaxed prescription has been changed now.

Each cell has assigned a pair of coordinates. They are the integer indices of the row and column associated to that cell in the grid system. Then, given a pair of species, we calculate the nearest interspecific Euclidean distances among their OGUs. If species A and B occupy OGUs {1,2} and {3,4}, respectively, we should obtain the following vectors of distances.

For species A: vecAB = {min($\mathbf{d}$(1,3), $\mathbf{d}$(1,4)), min($\mathbf{d}$(2,3),$\mathbf{d}$(2,4))}.

For species B: vecBA = {min($\mathbf{d}$(3,1), $\mathbf{d}$(3,2)), min($\mathbf{d}$(4,2),$\mathbf{d}$(4,3))}.

In these statements, $\mathbf{d}$ stands for Euclidean distance between a pair of OGUs.

Inference of sympatry is based on the values found on the previous vectors of proximity. Firstly, we need a summary or reference value for each vector, which is dictated by the above argument `criterion`. For example, under the default setting we obtain:

refA = max(vecAB); refB = max(vecBA).

Secondly, those reference statistics are compared against the upper threshold indicated by the argument `tolerance`. Note that the default value, that is sqrt(2), is the maximal distance that contiguous cells may exhibit (corner-to-corner cells).

Now, let suppose that refA < tolerance, but refB > tolerance. They will pass the test of sympatry if the argument `reciprocity` would have been set FALSE. This scenario of asymmetry is expected for nested distributions. If you are interested on co-extensive sympatry, `reciprocity` should be set TRUE and sympatry statement proceeds if both vectors are below the threshold.

Finally, the summary values derived from both vectors (e.g. vecAB, vecBA) may be lower than the threshold, despite the actual list of OGUs do not intersect. In order to avoid considering sympatry when there is no co-occurrence, the argument `conditioned` acts preventively. So, if `conditioned` is TRUE, two species must inhabit a common OGU to be considered candidates for sympatry. In this way, the default setting makes co-occurrence a necessary but not a sufficient condition to postulate meaningful sympatry between taxa.

Value

An object of class `gridinference`, which is a list with components:

sm	An adjacency matrix that reflects the existence (1) or not (0) of a sympatric link between species.
occupancy	List of OGUs occupied by species.

| coords | Coordinates of each cell arranged into a two-columns matrix. |
| kind | Character. Specifies the kind of distributional data, that is "grids". |

Note

Do not forget to provide the coordinates for the OGUs in the same input table, preceding the species themselves. Thus, if species are rows then the first two rows are considered to have the coordinates of the OGUs, and similarly for species arranged into columns.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Dos Santos D.A., Fernandez H.R., Cuezzo M.G., Dominguez E. 2008. *Sympatry Inference and Network Analysis in Biogeography.* Systematic Biology 57:432-448.

Dos Santos D.A., Cuezzo M.G., Reynaga M.C., Dominguez E. 2011. *Towards a Dynamic Analysis of Weighted Networks in Biogeography.* Systematic Biology (in press).

See Also

Objects of class gridinference can be submitted to the function nam.

Examples

```
data(sciobius2x2)
#Do inference and discount reciprocity.
######
aux1 <- gridinfer(dntable = sciobius2x2, reciprocity = FALSE)$sm # Displays the sympatry matrix
#Check that the widespread S. pullus (pu) has here many neighbors because distributions
#are nested inside it. In a network analysis this kind of element will behave as intermediary node.
aux1["pu",] #There are many connections to S. pullus coded 1.
######
#Do inference and force to consider reciprocity in the pairwise comparisons of species ranges.
#That is, study now co-extensive sympatry.
aux2 <- gridinfer(dntable = sciobius2x2, reciprocity = TRUE)$sm # Displays the sympatry matrix
#S. pullus is now an isolated node because there is no other taxa that spreads over the
#surface like S. pullus.
aux2["pu",] #Only the loop is present in this vector of connections for S. pullus.
```

---

mayflynz   *Mayfly Fauna of New Zealand*

Description

List of 679 geographical records associated to the endemic fauna of mayflies in the New Zealand archipelago. This data set belongs to the class dnpoint.

Usage

```
data(mayflynz)
```

Format

The format is: List of 4 $ Numpoints: int 679 $ Points :'data.frame': 679 obs. of 3 variables: ..$ IDsp : num [1:679] 1 1 1 1 1 1 1 1 1 1 ... ..$ Longitud: num [1:679] 175 175 175 175 175 ... ..$ Latitud : num [1:679] -41.2 -41.2 -41.2 -41.1 -40.9 ... $ Label : chr [1:40] "Acanthophlebia cruentata" "Arachnocolus phillipsi" "Atalophlebiodes cromwelli" "Aupouriella pohei" ... $ Type : chr "geographical" - attr(*, "class")= chr "dnpoint"

Details

Ephemeroptera (mayflies) are an ancient lineage of insects. Aquatic nymphs are the dominant life history stage. Adults have short lives ranging from a few hours to a few days. All species are endemic to this archipelago and their distribution patterns seem to be driven by landscape changes associated to Pleistocene glaciations. Leptophlebiidae and Nesameletidae are the most specious families and they have been considered in compiling this data set.

Localities given in New Zealand Map Grid projection metric coordinates were converted to the WGS84 datum with the online utility provided by the Land Information New Zealand government department (LINZ, http://www.linz.govt.nz/. Additional records for Zephlebia pirongia were inferred from its dot map in Hitchings (2008a). Random samples without replacement of thirty percent of records previously extracted for *Z. pirongia* were assigned to its co generic homopatric species {*Z. dentata, Z. nebulosa, Z. inconspicua, Z. versicolor*} (Hitchings 2008a).

Source

Hitchings T.R., Staniczek A.H. 2003. *Nesameletidae (Insecta: Ephemeroptera).* Fauna of New Zealand 46:1-72. Lincoln, New Zealand.

Towns D.R., Peters W.L. 1996. *Leptophlebiidae (Insecta: Ephemeroptera).* Fauna of New Zealand 36. Lincoln, New Zealand.

Hitchings T.R. 2008a. *The post glacial distribution of New Zealand mayflies.* In: Hauer F.R, Stanford J.A., Newell R.L., editors. International advances in the ecology, zoogeography and systematics of mayflies and stoneflies. Univ. Calif. Publ. Entomol. 128:89-101.

Hitchings T.R. 2008b. *A new species of Deleatidium (Penniketellum) and the adult of D. (P.) cornutum Towns and Peters (Ephemeroptera: Leptophlebiidae) from New Zealand.* Rec. Canterbury Mus. 22:31-43.

Hitchings T.R. 2009a. *Three new species of Deleatidium (Deleatidium) (Ephemeroptera: Leptophlebiidae) from New Zealand.* Rec. Canterbury Mus. 23: 35-50.

Hitchings T.R. 2009b. *Leptophlebiidae (Ephemeroptera) of the alpine region of the Southern Alps, New Zealand.* Aquat. Insects 31(Supp. 1): 595-601.

Winterbourn, M.J. 2009. *A new genus and species of Leptophlebiidae (Ephemeroptera) from northern New Zealand.* N.Z. J. Zool. 36:423-430.

References

Dos Santos D.A., Cuezzo M.G., Reynaga M.C., Dominguez E. 2011. *Towards a Dynamic Analysis of Weighted Networks in Biogeography.* Systematic Biology (in press).

Examples

```
data(mayflynz)
plot(mayflynz[[2]][,2:3], main = "Mayfly Fauna of New Zealand", xlab = "Latitude",
     ylab = "Longitude", asp = 1.3)
```

mst	*Minimum Spanning Tree*

Description

Identifies the minimum spanning tree connecting the set of points using their spatial distances.

Usage

```
mst(x)
```

Arguments

x	Distance matrix between points

Details

Certainly, a single vector could be used to capture the topology of a MST, indicating the index of the parent node for each child node. However, I have decided to indicate the endpoints for each MST arc into two separate vectors.

Value

This function returns a list with the following elements:

path	Length of the MST in the given metric units.
wght	Mean leangth of MST arcs incident to each vertex. Values are normalized so that they sum to unity.
xy0	Indices for one of the endpoints of MST arcs.
xy1	Indices for the other endpoints correspondingly ordered.

Note

The algorithm to obtain the MST is valid for distance matrices with metric properties. Consequently, both Euclidean and orthodromic distances can be used.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Prim R.C. 1957. *Shortest Connecting Networks and Some Generalizations.* Bell Syst. Tech. J 36:1389-1401.

Examples

```
xy <- matrix(rnorm(100), ncol = 2) # Sample a random set of points
plot(xy, xlab = "", ylab = "", main = "MINIMUM SPANNING TREE")
aux <- mst(as.matrix(dist(xy))) # Find the Euclidean minimum spanning tree
segments(xy[aux$xy0,1], xy[aux$xy0,2], xy[aux$xy1,1], xy[aux$xy1,2]) # Plot the MST
```

---

nam *Sympatry Network Analysis*

Description

Seeks groups of species cohesively sympatric (units of co-occurrence) via iterative removal of nodes with the highest betweenness score (intermediary species). It also provides geometric data to plot a cleavogram.

Usage

```
nam(input, rounding = TRUE)
```

Arguments

input
: An object of sympatry inference. Admitted classes are both 'dotinference' and 'gridinference'.

rounding
: Logical. If TRUE, betweenness scores are rounded via zapsmall set at 8 digits.

Details

The argument input contains an element $sm, i.e. a sympatry matrix **S** of order n (= number of species). Each entry Sij is 1 if there is a sympatric link between species i and j, while Sij is 0 if allopatry is suggested. The diagonal elements are 1 because sympatry is reflexive. Furthermore, sympatry matrix is symmetric due to reciprocal nature of this relationship.

Sympatry matrices are adjacency matrices, and finding the connection patterns in the associated network is the scope of nam function. In a given sympatry network, nodes represent species and edges sympatric links. nam is an iterative process of node removal to isolate subsets of nodes with within-group sympatry and between-group allopatry (units of co-occurrence). Units of co-occurrence are generally embedded into a global network due to connecting intermediary nodes.

nam identifies and removes nodes with the highest intermediacy score. Intermediacy is evaluated with the betweenness measure (Freeman, 1977). After the node removal, a sub-network is generated and the process is repeated until it arrives at a sub-network with all nodes with zero betweenness. In this way, nam produces a series of sub-networks.

Dos Santos et al. (2011) describes a new graphical tool, called a cleavogram, that depicts the structural changes of the network along the removal process. This function also retrieves the necessary information to plot a cleavogram.

The older version of NAM method focused on the structure of a single sub-network Dos Santos et al. (2008). Now, the interest relies on the cohesiveness of the various components emerging during the removal process.

The density describes the general level of linkage among the vertices (or nodes) in a network. A 'complete' graph (density = 1) is one in which all the actors are adjacent to one another. The density of a graph is the fraction of edges it actually has out of all possible vertex pairs.

Transitivity measures the probability that the adjacent vertices of a vertex are connected. This sometimes is also called the global clustering coefficient. It calculates the number of existing transitive triangles proportional to the number of connected triples.

In graph theory, the distance between two vertices in a graph is the number of edges in the shortest path connecting them (i.e., the geodesic distance). If there is no path connecting the two vertices, they belong to different connected components. The eccentricity of a vertex v is the greatest geodesic distance between v and any other vertex. It can be thought of as how far a node is from the node most distant from it in the graph. The diameter of a graph is the maximum eccentricity of any vertex in the graph. That is, it is the greatest distance between any pair of vertices. Groups of vertices tightly connected is expected to have a small diameter.

Value

An object of class `cleavogram`, which is a list with components:

mt	Input adjacency matrix with its rows and columns rearranged according to the sequence of leaves in the cleavogram.
LastNet	Integer vector indicating the last network (or sub-network) where a given node was found. Zero for the basal network and >0 for successive sub-networks.
namlast	As previously referred but adding 2 to each element.
Betweenness	Highest betweenness value recorded at the respective instance of removal process.
leaves	Character vector concerning to the labels of nodes in the network.
nsp	Integer corresponding to the size of the network.
nsub	Integer corresponding to the number of sub-networks obtained during removal process.
components	This is the core of the analysis. Data frame that indicates the structure of each component obtained along with the removal process. Additionally, there are statistical estimates for the cohesiveness of each component that have been detected (i.e. density, transitivity and eccentricities).
verticals	Data matrix that includes information about the vertical segments of the cleavogram.
kind	Character. Either 'grids' or 'points'. Describes how spatial information has been encoded.
occupancy	List of occurrences for each taxon under study (i.e. each node of the network
coords	Two-columns matrix. Spatial coordinates of the records in the data set.

Note

Betweenness measure was calculated with Newman's algorithm (Newman, 2001). Geodesic distances between vertices was calculated following Seidel (1995).

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Costa L.F., Rodrigues F.A., Travieso G., Boas V.P.R. 2005. *Characterization of Complex Networks: A Survey of Measurements.* Advances in Physics, 56(1), 167-242.

Dos Santos D.A., Fernandez H.R., Cuezzo M.G., Dominguez E. 2008. *Sympatry Inference and Network Analysis in Biogeography.* Systematic Biology 57:432-448.

Dos Santos D.A., Cuezzo M.G., Reynaga M.C., Dominguez E. 2011. *Towards a Dynamic Analysis of Weighted Networks in Biogeography.* Systematic Biology (in press).

Freeman L. C. 1977. *A Set of Measures of Centrality Based on Betweenness.* Sociometry 40:35-41.

Newman M. E. J. 2001. *Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks and Centrality.* Phys. Rev. E 64:016132.

Seidel R. 1995. *On the All-Pairs-Shortest-Path Problem in Unweighted Undirected Graphs.* In Proceedings of J. Comput. Syst. Sci.: 400-403.

See Also

The cleavogram can be explored with function cleavogram.

Examples

```
    data(sciobius2x2)
    mtx <- gridinfer(dntable = sciobius2x2, reciprocity = FALSE) # Infers the sympatry network
    cleavosciobius <- nam(mtx) # Performs NAM analysis and create cleavogram called 'cleavosciobius'
    ## Not run:
    cleavogram()

## End(Not run)
    # Choose the previous cleavogram and dissect it!
```

---

netproperties	*Structural Properties of an Unweighted Undirected Network*

Description

Calculates i) density of the network, ii) node degrees, iii) geodesic distances between pairs of vertices. It also finds components and enumerates maximal cliques in the network.

Usage

```
    netproperties (mt, cutoff = 0, dichotomization = ">")
```

Arguments

mt	Any valued symmetric matrix.
cutoff	Numeric. It is the threshold that dichotomizes the input matrix.
dichotomization	
	Character. Dichotomization rule expressed as a comparison operator.

Details

The input matrix is transformed into a binary adjacency matrix applying the dichotomization rule over the cutoff. Then, several connecting properties of the resulting simple graph are studied.

A brief summary of graph concepts follows. The density is the proportion of dyadic connections (links) which are actually present. The degree of a node is the number of neighbors directly tied to that node. Cliques are fully connected groups of nodes, that is complete subgraphs in a graph. A clique is maximal if it cannot be extended to a larger clique. In graph theory, the distance between

two vertices is the number of edges in a shortest path connecting them. A connected component of an undirected graph is a subgraph in which any two vertices are connected to each other by paths, and which is connected to no additional vertices. If there is no path connecting the two vertices, i.e. if they belong to different connected components, then conventionally the geodesic distance can be defined as NA or infinite.

Value

An object of class `cleavogram`, which is a list with components:

Adjacency	Binary matrix resulting from dichotomization.
Components	A vector of integers indicating the component to which each vertex (or node) is allocated.
Degree	A vector of integers indicating the number of neighbors directly linked to each vertex (or node).
Geodesic	Geodesic distance matrix between vertices. If two vertices belong to different components, the respective entry is set NA.
Cliques	Data frame. Columns are maximal cliques whereas rows are vertices of the network. Membership of each node to maximal cliques are indicated through 1 (present) and 0 (absent).

Note

Geodesic distances between vertices was calculated following Seidel (1995). Maximal cliques have been enumerated following the recursive algorithm of Bron and Kerbosch (1973).

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Bron C., Kerbosch J. 1973. *Algorithm 457: Finding All Cliques of an Undirected Graph.* Commun. ACM 16(9): 575-577.

Seidel R. 1995. *On the All-Pairs-Shortest-Path Problem in Unweighted Undirected Graphs.* In Proceedings of J. Comput. Syst. Sci.: 400-403.

Examples

```
# "A" corresponds to the 0-1 adjacency matrix
# associated to a simple graph of 10 nodes
# The main diagonal has zeroes.
A <- matrix(0, 10, 10)
A[lower.tri(A)] <- ifelse(runif(5*9) < 0.5, 1, 0)
pmin(A + t(A), 1) -> A
netproperties(A)
```

---

outgearth *Generate a KML File with Marked Set of Points*

Description

Produces Google Earth placemarks arranged into folders. Each folder consists of punctual records from a given group of species (i.e. a group detected by the NAM method).

Usage

```
outgearth(partition, showlabel = FALSE)
```

Arguments

partition	Object of class 'nampartition'.
showlabel	Logical. TRUE means that labels will be added to the placemarks.

Details

Visual properties for the entire set of records in each folder can be edited through Google Earth options. For that, do no forget to click on the share style button.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

The Homepage of Google Earth is available at: <http://earth.google.com/>

See Also

Function cleavogram assists us for creating objects of class nampartition.

Examples

```
#This example is driven by a new idea of a sympatry network partitioning.
#We will implement an algorithm based on the cliques found on the network and
#we will export the final classification into a KML file.
#######
#Step 1: Infer the network of co-extensive sympatry in the Sciobius example:
data(sciobius2x2)
aux <- gridinfer(dntable = sciobius2x2)
#######
#Step 2: Obtain the cliques
cliques <- netproperties(aux$sm)$Cliques
#######
#Step 3: Perform the new alogrithm on the data frame of cliques (1/0 table of species
#by cliques). Here, the maximum clique is selected and its members removed from the
#data frame. This task is repeated until no residual group can be extracted.
inc <- apply(cliques, 1, sum) #Number of cliques where a given species occurs
flag <- sum(cliques)
i <- 1 #counter
```

```
      classes <- rep(NA, nrow(aux$sm))
      while(flag > 0){
        size <- apply(cliques, 2, sum) #Size of each clique
        clsel <- which.max(size) #Identify a single largest clique
        members <- which(cliques[,clsel]==1)
        flag <- flag - sum(inc[members])
        inc[members] <- 0
        cliques[members,] <- 0 #Indirect way for removing species already classified
        classes[members] <- paste("Group ", i)
        i <- i + 1
      }
      split(aux$Label, classes) #Print on R console the resulting partition
      #######
      #Step 3: Create an object of class nampartition by hand and .
      rslt <- c()
      rslt$kind <- "grids"
      rslt$status <- cbind(Taxa = aux$Label, Status = classes)
      rslt$occupancy <- aux$occupancy
      #Next, set coordinates in function of the geographical centre for each cell used in the
      #Sciobius' example
      rslt$coords <-  matrix(c(14, -20),  nrow = nrow(aux$coords), ncol = 2, byrow = TRUE) +
                      matrix(c(2, -2), nrow = nrow(aux$coords), ncol = 2, byrow = TRUE)*aux$coords
      class(rslt) <- "nampartition"
      #######
      #Step 4: Create the KML file
      ## Not run:
      outgearth(rslt)
      #Save and then open the file with Google Earth.

    ## End(Not run)
      #I think that this simple algorithm is worthy of refinement.
```

---

outgis   *Generate a TXT File with Marked Set of Points*

Description

Generates a .txt file that can be imported by GIS softwares (e.g. DIVA-GIS). The output corresponds
to a data set of records classified by the NAM algorithm.

Usage

```
outgis(partition)
```

Arguments

partition   Object of class nampartition.

Details

The generated file has four fields separated by commas: Label, X_Long, Y_Lat and NAMClass.
The first line of the file corresponds to the headers.

Note

Almost the totality of GIS softwares may convert our output into a shapefile of points. Particularly, DIVA-GIS users can open the generated .txt file via the menu DATA -> IMPORT POINTS TO SHAPEFILE -> FROM TEXT FILE (.TXT).

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

See Also

The homepage of DIVA-GIS is at http://www.diva-gis.org/.

Function cleavogram assists us for creating objects of class nampartition. The example developed below can be found also in the documentation for function outgearth.

Examples

```
#This example is driven by a new idea of a sympatry network partitioning.
#We will implement an algorithm based on the cliques found on the network and
#we will export the final classification into a txt file with fields separated by commas.
#######
#Step 1: Infer the network of co-extensive sympatry in the Sciobius example:
data(sciobius2x2)
aux <- gridinfer(dntable = sciobius2x2)
#######
#Step 2: Obtain the cliques
cliques <- netproperties(aux$sm)$Cliques
#######
#Step 3: Perform the new alogrithm on the data frame of cliques (1/0 table of species
#by cliques). Here, the maximum clique is selected and its members removed from the
#data frame. This task is repeated until no residual group can be extracted.
inc <- apply(cliques, 1, sum) #Number of cliques where a given species occurs
flag <- sum(cliques)
i <- 1 #counter
classes <- rep(NA, nrow(aux$sm))
while(flag > 0){
  size <- apply(cliques, 2, sum) #Size of each clique
  clsel <- which.max(size) #Identify a single largest clique
  members <- which(cliques[,clsel]==1)
  flag <- flag - sum(inc[members])
  inc[members] <- 0
  cliques[members,] <- 0 #Indirect way for removing species already classified
  classes[members] <- paste("Group_", i)
  i <- i + 1
}
split(aux$Label, classes) #Print on R console the resulting partition
#######
#Step 3: Create an object of class nampartition by hand and .
rslt <- c()
rslt$kind <- "grids"
rslt$status <- cbind(Taxa = aux$Label, Status = classes)
rslt$occupancy <- aux$occupancy
#Next, set coordinates in function of the geographical centre for each cell used in the
#Sciobius' example
rslt$coords <-  matrix(c(14, -20),  nrow = nrow(aux$coords), ncol = 2, byrow = TRUE) +
                matrix(c(2, -2), nrow = nrow(aux$coords), ncol = 2, byrow = TRUE)*aux$coords
```

```
    class(rslt) <- "nampartition"
    #######
    #Step 4: Create the txt file which is easily converted in to a shapefile
    #by any professional GIS software.
    ## Not run:
    outgis(rslt)

## End(Not run)
```

---

partitionindex   *Partition Index of a Sympatry Network*

Description

Calculates the actual partition index of a simple sympatry network and estimates its random expectancy under a Bernoulli graph model.

Usage

```
partitionindex(iptsymp, replica = 1)
```

Arguments

iptsymp	Adjacency matrix associated to the network of interest.
replica	Integer between 1 and 5000. Corresponds to the number of random adjacency matrices to be generated.

Details

The partition index (**PI**) is based on the clustering coefficient measure (**C**). For each network node, it takes into account the maximum between its own **C** value and the lowest **C** value recorded at its open neighbourhood. The selected value is known as clustering performance after Dos Santos et al. (2008). Finally, the mean of clustering performance is calculated.

(**PI**) is the statisitc to test the adequacy of the the newtork to be segregated into highly connected groups of species. The test counts the number of random simple graphs that yields a **PI** value higher or equal to the observed one.

Random expectancy for (**PI**) is based on matrices following a Bernoulli model. A random number between 0 and 1 is generated for each pair of species in the network. If this number is lower than the observed density for the the network, the respective pair of species will remain tied in the random network.

The input argument iptsymp corresponds to any adjacency matrix that reflects the incidence (1) or not (0) of a sympatric link between pairs of species. If the input matrix is valued, scores higher than zero will be automatically coded 1 (otherwise, they will be coded 0).

Value

If replica > 1, returns a list containing:

ProbTie	Density of network associated to.
PIobserved	Observed (**PI**).
PIrandomized	Statistical summary of the randomly expected indices calculated through fivenum.

ProbTie   Fraction of random scenarios where the calculated (**PI**) is higher than (or equal to) the observed one.

If `replica = 1`, the (**PI**) of the observed matrix is calculated.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Dos Santos D.A., Fernandez H.R., Cuezzo M.G., Dominguez E. 2008. *Sympatry Inference and Network Analysis in Biogeography.* Systematic Biology 57:432-448.

Examples

```
#Creates a Bernoulli graph of 100 nodes
A <- matrix(0, 100, 100)
aux <- ifelse(runif(choose(100, 2)) <= 0.3, 1, 0)
A[row(A) > col(A)] <- aux
A + t(A) -> A
#Prints the partition index on the R console
partitionindex(A)
#Produces 10 random samples and test significancy
partitionindex(A, 100)
```

---

platyrrhini   *Neotropical Primates*

Description

Distributional data of Neotropical primates arranged into an object of class `dnpoint`.

Usage

```
data(platyrrhini)
```

Format

The format is: List of 4 $ Numpoints: int 602 $ Points :'data.frame': 602 obs. of 3 variables: ..$ IDsp : num [1:602] 1 1 1 1 1 1 1 1 1 2 ... ..$ Longitud: num [1:602] 9 10 11 12 9 10 11 12 13 10 ... ..$ Latitud : num [1:602] 6 6 6 6 7 7 7 7 7 5 ... $ Label : chr [1:106] "1" "10" "100" "101" ... $ Type : chr "cartesian" - attr(*, "class")= chr "dnpoint"

Details

Goldani et al. (2006) coded the distribution of 106 Neotropical primate (Platyrrhini) species into 60 quadrats of 5 x 5 degrees as OGUs, detecting eight areas of endemism. Here, each species cell record was transformed into a punctual record so that its x and y coordinates are the respective column and row numbers of the cell in the original grid system. Numeric codes for species labels are the same as in Goldani et al. (2006).

Source

Goldani A., Carvalho G.S., Bicca-Marques J.C. 2006. *Distribution Patterns of Neotropical Primates (Platyrrhini) Based on Parsimony Analysis of Endemicity.* Braz. J. Biol. 66(1a): 61-74.

References

Dos Santos D.A., Cuezzo M.G., Reynaga M.C., Dominguez E. 2011. *Towards a Dynamic Analysis of Weighted Networks in Biogeography.* Systematic Biology (in press).

---

pointDemo   *Interactive Assessment Demo of Affinity Between Dot Clouds*

Description

Using a graphical user interface (GUI), this function demonstrates the effects of changing the topology of dot clouds on the calculation of their strength of association.

Usage

```
pointDemo()
```

Details

Two resources are used, a tcltk control window and a screen graphics device. In the first window, the user must specify the size of the dot clouds and can manipulate the layouts of point sets on the plotting area.

Once the "Draw!" button is pressed then the graphics device is waiting for the user to click and draw points. The length of the Euclidean minimum spanning trees (MST) for each set of points is indicated. The length of the MST associated to the joint set of points is also added. It is important to note how this last measure decreases as the dot clouds get close and interpenetrated. Sympatry is suggested when the joint MST is equal or lesser than the sum of individual MSTs (i.e. counter-synergic property).

Value

This function was designed for illustrating the procedure oriented to measure the degree of association between dot clouds. Anyway, it prints a data frame into the R console specifying the coordinates for each plotted point.

Note

It seems impossible to measure sympatry between geographical entities without enclosing records into areas. However, I think that the refinement of the idea underlying this game can liberate us from the Procrustean nature of thinking distributions as bounded areas.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

See Also

toposimilar

Examples

```
## Not run: pointDemo()
```

---

procdnpoint	*Pre-processing of Raw Point Data*

Description

Transforms raw data of point distributions into formats appropriate for ulterior analysis. Please, see the details section for the different tasks that this function performs.

Usage

```
procdnpoint(dnpoint, tolerance = 1e-03)
```

Arguments

dnpoint   The object 'dnpoint' to be pre-processed.

tolerance A value specifying the maximal distance to consider any point of the data set to be 'close enough' to the known or observed records of each species.

Details

The structure of data is firstly polished in removing duplicated points. Three main tasks are then developed:

1) Distance calculations. Two matrices are produced: (i) points X points, and (ii) species X points. Great circle or Euclidean distances are calculated if the coordinate system is geographical or rectangular, respectively. The second matrix of species X points corresponds to the Hausdorff distance between any single point of the data and the set of observed points for each species.

2) Data organization. A Boolean table accounts for the real species distributions. Here, species are the rows whereas the universe of unique records is arranged into columns. Although it is redundant, a list of species occurrences is also created. This list enables us to include additional points for each species if those extra points are separated from the observed/known points by a negligible distance. The parameter tolerance dictates the criterion to consider a spatial gap between records as negligible.

The argument tolerance is assumed to be in kilometers for the geographical coordinate system. Eventually, negative values are converted to absolute ones. You can pass a numeric vector specifying the tolerance radius for each point of the data set. If the length of the vector tolerance differs from the number of points indicated by the input object, then its values are recycled through rep function until they reach the required length. Finally, if different tolerance values are set to the same point because there are different species occurring on it, then we arbitrarily consider the corresponding value in lexicographic order.

3) Minimum spanning tree report. For each species, a full report about its MST is provided (i.e. total length, endvertices of MST arcs, normalized weight for each point proportional to the mean length of its incident MST arcs).

Value

An object of class `dotdata`, which is a list with elements:

Call	All arguments passed to this function when it was called.
Label	Character vector giving the labels for each species.
dntable	Boolean distributional table of species by records.
Occupancy	List of sampled records by species. Additional points falling in the tolerance radius around observed points are also included.
Coords	Two-columns table with (Longitude, Latitude)/(x, y) coordinates for the repository of unique points found throughout the data set.
MSTsp	List with data about the minimum spanning tree for each species.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

See Also

Objects of class 'dnpoint' are created via `read.coord` from an input txt file of point coordinates. The minimum spanning tree report is provided by `mst`

Examples

```
#####
# You can recognize the format of a typical input file
# in the following created .txt:
write(c("sp", "latitude", "longitude"), file= "proof.txt", ncolumns = 3, append = TRUE, sep = ",")
# Sample 20 points from a normal distribution and segregate them into two sets equally sized.
x <- c(rnorm(10), rnorm(10, 2))
y <- rnorm(20)
for (i in 1:20)
  write(c(LETTERS[ceiling(i/10)], x[i], y[i]),file= "proof.txt", 3, TRUE, sep = ",")
# Put getwd() to identify the path where
# the file 'proof.txt' has been located
# Read the generated file.
proof <- read.coord(inputfile = "proof.txt", type = "cartesian")
procdnpoint(proof) # Show the values returned by this function
#####
unlink("proof.txt") # Delete
```

---

read.coord	*Read a List of Points from a txt File*

Description

Marked set of points are read from a given input file. Both geographical and Cartesian coordinates are accepted.

Usage

```
read.coord(inputfile = "", type = "geographical")
```

Arguments

inputfile	Filename with complete path. By default, a Windows file dialog is open to choose the input file.
type	A character string naming the coordinate system in which points have been referenced. A partial matching is performed against these arguments: "geographical" and "cartesian".

Details

The input file should have three fields separated by commas. A header (i.e. first line) is required to differentiate latitude from longitude accurately. The first field must hold the species IDs. The second and third fields corresponds to the longitude and latitude indistinctly. If points are derived from a Cartesian system of coordinates, longitude and latitude fields will be interpreted as the abscissa and ordinate respectively.

Value

An object of class dnpoint, which is a list with elements:

Numpoints	Number of read points.
Points	Data frame with columns (1) indices for species, (2) Longitude and (3) Latitude. Coordinates are in decimal format.
Label	Character vector of species labels.
Type	String specifying the type of coordinate system passed as argument.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

Examples

```
#####
# You can recognize the format of a typical input file
# in the following created .txt:
write(c("sp", "latitude", "longitude"), file= "proof.txt", 3, TRUE, sep = ",")
for (i in 1:10)
  write(c(LETTERS[i], i*1.5, i*-8.6),file= "proof.txt", 3, TRUE, sep = ",")
proof <- read.coord("proof.txt") # Put getwd() to identify the path where
                                 # the file 'proof.txt' has been located

#####
unlink("proof.txt") # Delete
```

---

reweight	*Reweighting of Links from an Adjacency Weighted Matrix*

Description

Transforms a weighted matrix into another one constrained into a common scale of relative strength. Values are reweighted by a combination of rewards and punishments. For a given link, its relative strength increases with the number of triads where it is the strongest and decreases with the number of triads where it is the weakest.

Usage

```
reweight(wm, similarity = TRUE, t1t2 = if(similarity) c(0, 1) else c(Inf, 0),
  normalized = TRUE)
```

Arguments

wm	Square and symmetric matrix filled with valued score of spatial affinity.
similarity	Logical, depending on the existence of direct or inverse relationship between matrix scores and the spatial affinity between species (i.e. TRUE: high scores - high affinity; FALSE: low scores - high affinity).
t1t2	Non-negative numeric vector, conventionally of length 2.
normalized	Logical, if TRUE output values are linearly transformed into the scale [0, 1].

Details

The vector t1t2 holds the thresholds to delimit intervals for meaningful association. If similarity is TRUE, any value of the original matrix lower than min(t1t2) is discarded for further consideration because it represents the lower bound for meaningful association. On the other hand, any value of the input matrix equal or higher than the upper bound (i.e. max(t1t2)) is deemed to be maximally similar and coded 1 in the output. If similarity is FALSE, any link lower than min(t1t2) represent close affinity between species, and conversely any link equal or higher than max(t1t2) denote no affinity at all between its endvertices.

The dichotomization is a common procedure to transform weighted matrices into binary ones. Values are set 1 or 0 depending on the location of some cutoff value. This function also enables us to perform a hard thresholding when the parameter t1t2 has been fed with a single value. Under this setting, the final reweighted matrix will be binary. If similarity is TRUE (FALSE), entries are 1 if original values are equal or higher (lower) than t1t2, otherwise entries are 0.

Value

A similarity matrix composed of reweighted values. Values are constrained into the interval [0, 1] under the option of normalization. Here, null association is coded with a zero (0), whereas the unity (1) corresponds to the optimal meaningful association.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Dos Santos D.A., Cuezzo M.G., Reynaga M.C., Dominguez E. 2011. *Towards a Dynamic Analysis of Weighted Networks in Biogeography*. Systematic Biology (in press).

Examples

```
#Example to show an interesting numerical property of the reweighting formula, i.e.
#its convergence to the same rescaled matrix after several self-iterations
nitems <- 25
ej <- matrix(0, nitems, nitems)
ej[row(ej) > col(ej)] <- runif(nitems*(nitems - 1)/2)
t(ej) + ej -> ej
diag(ej) <- 1
### Display three graphics
```

```
  op <- par(mar = rep(3, 4), mfrow = c(3, 1))
  plot(unlist(ej),unlist(reweight(ej)), xlab = "Input matrix", ylab = "Reweighted matrix", main = "First re
  #Iterative reweighting
  histcor <- c()
  #Perform 100 iterations
  for(i in 1:100) {
    reweight(ej) -> rej
    b <- c()
    for(k in 1:(nrow(ej) - 1))
      for(j in (k + 1):nrow(ej)) {
        a <- (ej[k, j] >= ej[k, ]) + (rej[k, j] >= rej[k, ])
        b <- c(b, sum(a==2)/sum(a > 0))
    }
    histcor <- c(histcor, mean(b))
    ej <- rej
  }
  plot(unlist(ej),unlist(reweight(ej)), xlab = "Input matrix", ylab = "Reweighted matrix",
        main = "Reweighting after several iterations")
  plot(histcor, xlab = "Iteration", ylab = "Cross adjustment",
        main = "Evolution of the resemblance between input and reweigthed network")
  ## At end of plotting, reset to previous settings:
  par(op)
```

---

sciobius2x2   *Sciobius Grid Occupancy Data*

Description

Presence/absence data of 47 South African species of weevils over a squared grid of 2 x 2 degrees.

Usage

```
data(sciobius2x2)
```

Format

A data frame with 49 observations on 25 columns. Columns are the grid cells in the study area (South Africa). The first pair of rows correspond to the coordinates of cells in the grid system (i.e. numeric indices of the cells in the rectangular layout of the grid).

The rest of rows are the species. Presence is indicated with 1 and absence with 0.

Details

Morrone (1994) provides the original matrix for this classical example, but ulterior corrections are introduced by Mast and Nyffeler (2003). The corrected version of the data are here represented. Species names are coded through two letters following the nomclature of Mast and Nyffeler (2003).

Source

Schoeman, A. S. 1983. *Revision of the Genus Sciobius Schonherr (Coleoptera: Curculionidae).* Entomol. Mem. Dep. Agric. Repub. S. Afr. 59:1-50.

References

Morrone, J. J. 1994. *On the Identification of Areas of Endemism.* Syst. Biol. 43:438-441.

Mast, A. R., and R. Nyffeler. 2003. *Using a Null Model to Recognize Significant Co- occurrence Prior to Identifying Candidate Areas of Endemism.* Syst. Biol. 52:271-280.

---

stablecouple	*Stable Matching Derived from Spatial Affinity Matrices*

Description

Produces a partition into stable couples from a set of species arranged into a symmetric matrix of spatial affinity. Self-matched elements are allowable.

Usage

```
stablecouple(mt, selfprefmethod = "customized", similarity = TRUE,
initselfpref = if(similarity) 1e-07 else median(mt[lower.tri(mt)]),
prob = 0.5)
```

Arguments

mt	Square and symmetric matrix filled with valued score of spatial affinity.
selfprefmethod	String specifying the method to set the main diagonal of selfpreferences. Options are "two-means", "mean", "quantile" and "cutomized".
similarity	Logical, depending on the existence of direct or inverse relationship between matrix scores and the spatial affinity between species (i.e. TRUE: high scores - high affinity; FALSE: low scores - high affinity).
initselfpref	Numeric vector, it is forced to have the same length as the number of matrix rows. If NULL, it is directly replaced by the main diagonal of the input matrix itself. This vector bounds the domain of selfpreferences for each species. Thus, for a given species with initial selfpreference set at k, associations with other species are taken into account if their values belong to the interval [k, 1] (or [0, k]) in the similarity (or dissimilarity) matrix.
prob	Single numeric value between 0 and 1 used as probability for the method **"quantile"**. The default value of 0.5 corresponds to the median.

Details

The stable marriage problem is that of matching *n* men and *n* women, each of whom has ranked the members of the opposite sex in order of preference, so that no unmatched couple both prefer each other to their partners under the matching. The stable marriage assignment problem was introduced by Gale and Shapley (1962) in the context of assigning applicants to colleges, taking into account the preferences of both the applicants and the colleges.

The roommates problem is essentially a version of the stable marriage problem involving just one set. Each item in the set ranks the *n* - 1 others in order of preference. The object is to find a stable matching, which is a partition of the set into pairs of roommates such that no two items which are not roommates both prefer each other to their actual couples.

In biogeography, pairs of co-distributed species are considered pointers to some underlying factor driving that particular pattern (either historical or ecological one). It seems then reasonable to

recognize stable couples of species in the matrix of spatial affinity in order to capture its main structure of co-distribution. The bulk of strong links between species is expected to be included in the pool of stable pairs. So, the measures of affinity associated to that pool can be used to extract a cutoff value to differentiate strong from weak links relating to the data under analysis.

The profile of `preferences` for each species are dictated by its respective row in the input matrix. Self-matchings are possible. Selfpreferences are coded along the main diagonal and can be set by the user through a numeric vector. Alternatively, different methods can be applied on the set of association scores bounded by `initselfpref`, namely 1) "two-means": splits the scores into two classes of magnitude through the standard k-means algorihtm. The selfpreference is obtained from the midpoint between both classes; 2) "mean": takes the mean from the sampled values; 3) "quantile": produces the sample quantile corresponding to the given probability `prob`; 4) "customized": the intial reference provided by `initselfpref` becomes the final setting.

In dealing with symmetric matrices of affinity, there is a simple algorithm to find the stable matching (Rodrigues-Neto 2007). Let *n* be a strictly positive integer. Let A = {1, 2, ..., *n*} be the set of items in a population. Find the pair of elements (i, j) from A X A with the maximum (or minimum) score in the similarity (or dissimilarity) matrix. Match i with j and remove them from the population. Then repeat the procedure with the remaining population until no more items could be matched.

Value

This function returns a list with three components:

stpairs	Integer indices corresponding to the stable couple for each element. Remember that self-matchings may be also reported.
valref	Numeric vector. Spatial affinities associated to the selected pairs.
valdiag	Numeric vector. Values of used selfpreferences.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Gale D., Shapley L.S. 1962. *College Admissions and the Stability of Marriage.* Amer. Math. Monthly 69: 9-15.

Rodrigues-Neto J.A. 2007. *Representing Roommates' Preferences with Symmetric Utilities.* Journal of Economic Theory 135(1): 545-550.

See Also

Quantiles are estimated by the quantile function provided by the `stats` package. The method "two-means" uses the function kmeans with its argument `centers` set to 2.

Examples

```
data(mayflynz)
aux <- procdnpoint(mayflynz)
mtx1 <- acsh(aux) # Construct the dissimilarity matrix between species sets of points.
mtx2 <- toposimilar(aux) # Now, construct the similarity matrix
st1 <- stablecouple(mtx1, similarity = FALSE) #Stable matchings under mtx1
st2 <- stablecouple(mtx2) #Stable matchings under mtx2
#Following, plots the same set of species along three parallel axes.
plot(rep(1:3, each = 40), rep(1:40, 3), main = "STABLE MATCHINGS", axes = FALSE,
```

```
        ylab = "", xlab = "", pch = 19)
    mtext("Couplings induced by mtx1", side = 1, line = 1, at = 1.5)
    mtext("(dissimilarity matrix)", side = 1, line = 2, at = 1.5)
    mtext("Couplings induced by mtx2", side = 1, line = 1, at = 2.5)
    mtext("(similarity matrix)", side = 1, line = 2, at = 2.5)
    #Each matching is represented by a segment. Self-matchings are horizontal
    #line segments. Note that symmetry means identical behavior of the involved
    #pair of taxa along the profiles of stable couplings.
    segments(rep(1, 40), 1:40, rep(2, 40), st1$stpairs, col = 2, lwd = 2)
    segments(rep(2, 40), 1:40, rep(3, 40), st2$stpairs, col = 2, lwd = 2)
```

---

toposimilar   *Resemblance between Process Point Patterns*

Description

Produces a weighted matrix of sympatric association between species from their points of occurrence.

Usage

```
toposimilar(dotdata)
```

Arguments

dotdata   Object of class 'dotdata'

Details

This function accounts for the strength of sympatric relationships between species. In considering traditional approaches to measure sympatry, distributions are coded into some contour maps and the percentage of overlap between distributions is then calculated. Here, there is independence from explicit species areas, so arbitrary enclosing areas for the raw points of occurrence are avoided. On the contrary, this function focuses on the resemblance between patterns of spatial occupancy followed by the records themselves.

The minimum spanning tree (MST) of species dots are used as the operational tool to capture the skeleton of their spatial pattern. If two sets of points associated to a pair of species are both close and interpenetrated (i.e. exhibit a similar pattern and likely belong to an underlying area of common membership), the MST of the joint set of points is expected to be formed by several arcs connecting interspecific pair of points. For a pair of species, sympatry is considered when their joint set of points gives rises to a MST not longer than the sum of the individual MST lengths associated to each species. If two species have a single record each, they are considered maximally sympatric if they co-occur at the same single point.

Value

Returns a matrix that expresses the spatial similarity between species point sets. Non-negative values suggest sympatry and denote the relative strength of association in the interval [0, 1]. Negative values suggest allopatry and correspond (their absolute image) to the shortest geographical gap between dot sets.

Author(s)

Daniel A. Dos Santos <dadossantos@csnat.unt.edu.ar>

References

Dos Santos D.A., Cuezzo M.G., Reynaga M.C., Dominguez E. 2011. *Towards a Dynamic Analysis of Weighted Networks in Biogeography.* Systematic Biology (in press).

Dos Santos D.A., Deutsch R. 2010. *The Positive Matching Index: a New Similarity Measure with Optimal Characteristics.* Pattern Recognition Letters 31: 1570-1576.

See Also

Objects of class 'dotdata' are created via `procdnpoint`.

Examples

```
data(mayflynz)
aux <- procdnpoint(mayflynz)
# Infer the weighted sympatry matrix and display
# a portion of its entries.
head(toposimilar(aux))
```

Index