Package 'shortr'

July 23, 2025

Title Optimal Subset Identification in Undirected Weighted Network
Models

Version 1.0.1

Description

Identifies what optimal subset of a desired number of items should be retained in a short version of a psychometric instrument to assess the "broadest" proportion of the construct-level content of the set of items included in the original version of the said psychometric instrument. Expects a symmetric adjacency matrix as input (undirected weighted network model). Supports brute force and simulated annealing combinatorial search algorithms.

Depends R (>= 4.1.0)

Imports stats, utils

URL https://doi.org/10.32614/CRAN.package.shortr

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.2

NeedsCompilation no

Author Loïs Fournier [aut, cre] (ORCID:
   <https://orcid.org/0000-0001-6743-6456>),
   Alexandre Heeren [aut] (ORCID: <https://orcid.org/0000-0003-0553-6149>),
   Stéphanie Baggio [aut] (ORCID: <https://orcid.org/0000-0002-5347-5937>),
   Luke Clark [aut] (ORCID: <https://orcid.org/0000-0003-1103-2422>),
   Antonio Verdejo-García [aut] (ORCID:
   <https://orcid.org/0000-0001-8874-9339>),
   José C. Perales [aut] (ORCID: <https://orcid.org/0000-0001-5163-8811>),
   Joël Billieux [aut] (ORCID: <https://orcid.org/0000-0002-7388-6194>)

Maintainer Loïs Fournier <lois.fournier@unil.ch>

Repository CRAN

Date/Publication 2025-04-15 15:20:02 UTC

Contents

shortr	*Optimal Subset Identification in Undirected Weighted Network Models*

Description

Identify the optimal subset such that the sum of the (absolute) values of the edge weights connecting the optimal subset with its complement is maximized.

Usage

```
shortr(
  adj.mat,
  k,
  method = base::c("brute.force", "simulated.annealing"),
  absolute = TRUE,
  start.temp = 1,
  cool.fact = 0.999,
  stop.temp = 0.001,
  max.iter = 1000,
  n.runs = 1000,
  seed = 5107,
  verbose = TRUE
)
```

Arguments

adj.mat	The adjacency matrix. Must be a symmetric adjacency matrix.
k	The size of the subset. Must be a non-null positive integer.
method	The combinatorial search algorithm. Must match either `"brute.force"` or `"simulated.annealing"`. Default is `"brute.force"`.
absolute	Whether absolute values of adj.mat, the symmetric adjacency matrix, should be computed. Must match either `FALSE` or `TRUE`. Default is `TRUE`.
start.temp	The starting temperature in the simulated annealing search. Must be a non-null positive numeric lying within the interval $(0, 1]$, and must be greater than stop.temp, the stopping temperature. Default is `1`.
cool.fact	The cooling factor in the simulated annealing search. Must be a non-null positive numeric lying within the interval $(0, 1)$. Default is `0.999`.
stop.temp	The stopping temperature in the simulated annealing search. Must be a non-null positive numeric lying within the interval $(0, 1)$, and must be less than start.temp, the starting temperature. Default is `0.001`.
max.iter	The maximal number of iterations in the simulated annealing search. Must be a non-null positive integer. Default is `1000`.
n.runs	The number of runs in the simulated annealing search. Must be a non-null positive integer. Default is `1000`.

seed
: The optional random number generator state for random number generation in the simulated annealing search. Must be a non-null positive integer. Optional. Default is 5107.

verbose
: Whether information messages should be printed to the console. Must match either FALSE or TRUE. Default is TRUE.

Details

In psychometrics, the network approach refers to a set of methods employed to model and analyze the relationships among psychological variables. Unlike traditional psychometric approaches, such as the structural equation approach focusing on latent variables, the network approach emphasizes the interconnections among observed variables. Due to the latter emphasis, modeling and analyzing network models to complement structural equation models when developing and evaluating psychometric instruments offers several advantages. Most notably, in undirected weighted network models, a subtype of network models, observed variables are represented by nodes, and associations between observed variables, each assigned a weight that represents the magnitude of associations, are represented by edges. In this perspective, undirected weighted network models provide estimates of the magnitude of associations (i.e., the shared variance) among items of psychometric instruments that structural equation models cannot, providing critical insight into the construct-level content validity of subsets of items of psychometric instruments. To illustrate, if an undirected weighted network model suggests that a subset of items of a psychometric instrument presents a high magnitude of associations with another subset of items, the shared variance of the said subsets of items is therefore high: the content they assess and the information they provide is highly similar. From the standpoint of the latter illustration, undirected weighted network modeling allows for the estimation of whether a subset of items assesses a "narrow" or a "broad" proportion of the construct-level content of a psychometric instrument. Hence, identifying an optimal subset of a desired number of items that assesses the "broadest" proportion of the construct-level content of a psychometric instrument consists of a combinatorial optimization problem.

Consider an undirected weighted network model $G = (V, E)$, where V denotes the set of nodes, and E denotes the set of edges. Each edge e_{ij} is associated with a positive or negative weight w_{ij}. Let S be a subset of nodes from V with a fixed size k, and $\bar{S}$ denote the complement of S in V. The objective is to identify the optimal subset S of size k such that the sum of the (absolute) values of the edge weights connecting S with its complement $\bar{S}$ of size $n - k$ is maximized. Formally, the combinatorial optimization problem can be expressed as:

$$\max_{S \subset V, |S| = k} \left(\sum_{i \in S, j \in \bar{S}} |w_{ij}| \right)$$

Solving the combinatorial optimization problem allows identifying what optimal subset of a desired number of items presents the highest magnitude of associations (i.e., the highest shared variance) within the set of items. In this light, combinatorial search algorithms (e.g., brute force search algorithm, simulated annealing search algorithm) allow to identify what optimal subset of a desired number of items should be retained in a short version of a psychometric instrument to assess the "broadest" proportion of the construct-level content of the set of items included in the original version of the said psychometric instrument.

Value

A list of two named objects:

optimal.S A character vector denoting the optimal subset S of size k.

optimal.function.S A numeric value denoting the sum of the (absolute) values of the edge weights connecting the optimal subset S of size k with its complement $\bar{S}$ of size $n - k$.

References

Fournier, L., Heeren, A., Baggio, S., Clark, L., Verdejo-García, A., Perales, J. C., & Billieux, J. (2025). *shortr: Optimal Subset Identification in Undirected Weighted Network Models* (Version 1.0.1) [Computer software]. doi:10.32614/CRAN.package.shortr

Examples

```
adj.mat <- (
  stats::runif(n = 25^2, min = -1, max = 1) |>
  base::matrix(nrow = 25, ncol = 25) |>
  (\(m) (m + base::t(m)) / 2)() |>
  (\(m) {base::diag(m) <- 0; m})()
)

shortr::shortr(
  adj.mat = adj.mat,
  k = 5,
  method = base::c("brute.force"),
  absolute = TRUE
)
```

Index

shortr,