

Title: SEMtree: tree-based structure learning methods with Structural Equation Models

Author: Mario Grassi and Barbara Tarantino

Corresponding Author for Code: Barbara Tarantino, barbara.tarantino@unipv.it

Configurations:

R version 4.1.0 (2021-05-18)
Platform: x86_64-apple-darwin17.0 (64-bit)
Running under: macOS Big Sur 10.16

Matrix products: default

LAPACK:

/Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib

locale:

[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:

[1] stats4 parallel stats graphics grDevices utils datasets
methods
[9] base

other attached packages:

[1] viridis_0.6.2	viridisLite_0.4.1	reshape_0.8.9
[4] ggplot2_3.3.6	penalizedLDA_1.1	dplyr_1.0.9
[7] readxl_1.4.1	scales_1.2.1	
clusterProfiler_4.0.5		
[10] EnrichmentBrowser_2.22.2	SummarizedExperiment_1.22.0	
GenomicRanges_1.44.0		
[13] GenomeInfoDb_1.28.4	MatrixGenerics_1.4.3	
matrixStats_0.62.0		
[16] diffusr_0.1.4	pathfindR_1.6.4	
pathfindR.data_1.1.2		
[19] BioNet_1.52.0	RBGL_1.68.0	graph_1.70.0
[22] COSINE_2.1	genalg_0.2.1	MASS_7.3-58.1
[25] org.Hs.eg.db_3.13.0	AnnotationDbi_1.54.1	IRanges_2.26.0
[28] S4Vectors_0.30.2	SEMgraph_1.1.3	lavaan_0.6-12
[31] igraph_1.3.4	GEOquery_2.60.0	Biobase_2.52.0
[34] BiocGenerics_0.38.0		

loaded via a namespace (and not attached):

[1] utf8_1.2.2	tidyselect_1.1.2	RSQLite_2.2.16
[4] grid_4.1.0	BiocParallel_1.26.2	lpSolve_5.6.15
[7] flsa_1.5.2	scatterpie_0.1.7	munsell_0.5.0
[10] codetools_0.2-18	withr_2.5.0	colorspace_2.0-3
[13] GOSemSim_2.18.1	knitr_1.40	rstudioapi_0.14
[16] dagitty_0.3-1	DOSE_3.18.3	KEGGgraph_1.52.0
[19] GenomeInfoDbData_1.2.6	flip_2.5.0	mnormt_2.1.0
[22] polyclip_1.10-0	bit64_4.0.5	farver_2.1.1
[25] downloader_0.4	coda_0.19-4	vctrs_0.4.1
[28] treeio_1.16.2	generics_0.1.3	xfun_0.32
[31] R6_2.5.1	doParallel_1.0.17	graphlayouts_0.8.1
[34] bitops_1.0-7	cachem_1.0.6	fgsea_1.18.0
[37] gridGraphics_0.5-1	DelayedArray_0.18.0	assertthat_0.2.1
[40] vroom_1.5.7	ggraph_2.0.6	enrichplot_1.12.3
[43] gtable_0.3.0	tidygraph_1.2.2	rlang_1.0.4

[46] splines_4.1.0	proclust_1.6.4	lazyeval_0.2.2
[49] yaml_2.3.5	reshape2_1.4.4	qvalue_2.24.0
[52] tools_4.1.0	ggplotify_0.1.0	ggm_2.5
[55] statnet.common_4.6.0	ellipsis_0.3.2	RColorBrewer_1.1-3
[58] hommel_1.6	Rcpp_1.0.9	plyr_1.8.7
[61] zlibbioc_1.38.0	purrr_0.3.4	RCurl_1.98-1.8
[64] pbapply_1.5-0	cowplot_1.1.1	ggrepel_0.9.1
[67] tinytex_0.41	magrittr_2.0.3	data.table_1.14.2
[70] sna_2.7	DO.db_2.9	someMTP_1.4.1.1
[73] hms_1.1.2	patchwork_1.1.2	evaluate_0.16
[76] xtable_1.8-4	XML_3.99-0.10	gridExtra_2.3
[79] compiler_4.1.0	tibble_3.1.8	V8_4.2.1
[82] GGMncv_2.1.1	crayon_1.5.1	shadowtext_0.1.2
[85] htmltools_0.5.3	ggfun_0.0.6	mgcv_1.8-40
[88] corpcor_1.6.10	tzdb_0.3.0	tidyr_1.2.0
[91] applot_0.1.6	DBI_1.1.3	tweenr_2.0.1
[94] boot_1.3-28	Matrix_1.4-1	readr_2.1.2
[97] cli_3.3.0	gdata_2.18.0.1	pkgconfig_2.0.3
[100] numDeriv_2016.8-1.1	xml2_1.3.3	foreach_1.5.2
[103] pbivnorm_0.6.0	ggtree_3.0.4	annotate_1.70.0
[106] XVector_0.32.0	yulab.utils_0.0.5	stringr_1.4.1
[109] digest_0.6.29	Biostrings_2.60.2	rmarkdown_2.16
[112] cellranger_1.1.0	fastmatch_1.1-3	tidytree_0.4.0
[115] glassoFast_1.0	GSEABase_1.54.0	curl_4.3.2
[118] gtools_3.9.3	lifecycle_1.0.1	nlme_3.1-159
[121] jsonlite_1.8.0	network_1.17.2	limma_3.48.3
[124] fansi_1.0.3	pillar_1.8.1	lattice_0.20-45
[127] KEGGREST_1.32.0	fastmap_1.1.0	httr_1.4.4
[130] GO.db_3.13.0	cherry_0.6-14	glue_1.6.2
[133] png_0.1-7	iterators_1.0.14	bit_4.0.4
[136] Rgraphviz_2.36.0	ggforce_0.3.4	stringi_1.7.8
[139] blob_1.2.3	memoise_2.0.1	ape_5.6-2

SEMtree_code_data folder structure

This folder contains the following data and files that can be used to reproduce the analysis of the manuscript. The folder structure can be summarised as follows:

./code/:

tree_sim.R
An R script to run Data simulations

tree_bench.R
An R script to run True data analysis

tree_comparison.R
An R script to run Method comparison

Help.R
An R script with functions to import for running the analysis

./data/:

C0206750_disease_gda_summaryGSE172114_rsem_gene_count_matrix_TM
M_69samples.csv
GA_result.Rdata
Gores.txt
GSE172114_rsem_gene_count_matrix_TMM_69samples.csv
GSE172114_series_matrix.txt.gz
ijmsv19p0402s2.xlsx

./graph/:

BioNet
COSINE
KEGGCovid19.graph
pathfinder
ST
STcfa
STcov
Stone
STr2z
STr2zN
STsem
tree_base
WalktrapGM
WalktrapGM1

Note:

1. Set folder SEMtree_code_data as the current working directory.
2. Specify a smaller number of iterations by argument "seed_vec" to reduce computing time for data simulations.

Source of data:

The original data used in this research can be downloaded from:
<https://www.ncbi.nlm.nih.gov/gds>.