Title: SEMgsa: topology-based pathway enrichment analysis 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] grid   stats4   parallel stats   graphics grDevices utils
 [8] datasets methods   base

other attached packages:
 [1] huge_1.3.5   scales_1.2.0
 [3] edgeR_3.34.1   limma_3.48.3
 [5] plyr_1.8.7   data.table_1.14.2
 [7] stringr_1.4.0   ggpubr_0.4.0
 [9] gplots_3.1.3   mosaic_1.8.3
[11] ggridges_0.5.3   mosaicData_0.20.2
[13] ggformula_0.10.1   ggstance_0.3.5
[15] dplyr_1.0.9   ggplot2_3.3.6
[17] lattice_0.20-45   a4Base_1.40.0
[19] a4Core_1.40.0   a4Preproc_1.40.0
[21] clipper_1.32.0   Matrix_1.4-1
[23] PathNet_1.32.0   SPIA_2.44.0
[25] graphite_1.38.0   ROntoTools_2.20.0
[27] Rgraphviz_2.36.0   KEGGgraph_1.52.0
[29] KEGGREST_1.32.0   boot_1.3-28
[31] DEGraph_1.44.0   R.utils_2.11.0
[33] R.oo_1.25.0   R.methodsS3_1.8.2
[35] netgsa_4.0.3   qpgraph_2.26.0
[37] mirIntegrator_1.22.0   EnrichmentBrowser_2.22.2
[39] graph_1.70.0   SummarizedExperiment_1.22.0
[41] GenomicRanges_1.44.0   GenomeInfoDb_1.28.4
[43] MatrixGenerics_1.4.3   matrixStats_0.62.0
[45] org.Hs.eg.db_3.13.0   AnnotationDbi_1.54.1
[47] IRanges_2.26.0   S4Vectors_0.30.2
[49] SEMdata_0.1.2   SEMgraph_1.1.2
[51] lavaan_0.6-11   igraph_1.3.2
[53] GEOquery_2.60.0   Biobase_2.52.0
[55] BiocGenerics_0.38.0

loaded via a namespace (and not attached):
  [1] rappdirs_0.3.3   rtracklayer_1.52.1   pbdZMQ_0.3-7
  [4] coda_0.19-4   NCIgraph_1.40.0   tidyr_1.2.0
  [7] someMTP_1.4.1.1   bit64_4.0.5   knitr_1.39

```
 [10] DelayedArray_0.18.0       RCurl_1.98-1.7            generics_0.1.2
 [13] leaflet_2.1.1             GenomicFeatures_1.44.2    RSQLite_2.2.14
 [16] bit_4.0.4                 tzdb_0.3.0                base64url_1.4
 [19] xml2_1.3.3                assertthat_0.2.1          xfun_0.31
 [22] hms_1.1.1                 ggdist_3.1.1              evaluate_0.15
 [25] DEoptimR_1.0-11           fansi_1.0.3               restfulr_0.0.15
 [28] progress_1.2.2            caTools_1.18.2            dbplyr_2.2.0
 [31] htmlwidgets_1.5.4         DBI_1.1.3                 reshape_0.8.9
 [34] purrr_0.3.4               ellipsis_0.3.2            crosstalk_1.2.0
 [37] backports_1.4.1           signal_0.7-7             GGMncv_2.1.1
 [40] V8_4.2.0                  hommel_1.6                pbivnorm_0.6.0
 [43] annotate_1.70.0           biomaRt_2.48.3           vctrs_0.4.1
 [46] abind_1.4-5               withr_2.5.0               cachem_1.0.6
 [49] RcppEigen_0.3.3.9.2       qtl_1.50                 ggforce_0.3.3
 [52] scoringRules_1.0.1        robustbase_0.95-0        checkmate_2.1.0
 [55] GenomicAlignments_1.28.0  sna_2.7                  prettyunits_1.1.1
 [58] mnormt_2.1.0              glassoFast_1.0           IRdisplay_1.1
 [61] crayon_1.5.1              uchardet_1.1.0           genefilter_1.74.1
 [64] glmnet_4.1-4              pkgconfig_2.0.3          tweenr_1.0.2
 [67] nlme_3.1-158             cate_1.1.1               ggm_2.5
 [70] rlang_1.0.2               RJSONIO_1.3-1.6          lifecycle_1.0.1
 [73] filelock_1.0.2            BiocFileCache_2.0.0      polyclip_1.10-0
 [76] distributional_0.3.0      flip_2.5.0               esaBcv_1.2.1
 [79] IRkernel_1.3              carData_3.0-5            base64enc_0.1-3
 [82] png_0.1-7                 rjson_0.2.21             bitops_1.0-7
 [85] KernSmooth_2.23-20        dplR_1.7.3               Biostrings_2.60.2
 [88] blob_1.2.3                shape_1.4.6              leapp_1.2
 [91] rstatix_0.7.0             readr_2.1.2              ggsignif_0.6.3
 [94] lpSolve_5.6.15            memoise_2.0.1            GSEABase_1.54.0
 [97] magrittr_2.0.3            gdata_2.18.0.1           zlibbioc_1.38.0
[100] compiler_4.1.0            tinytex_0.40             BiocIO_1.2.0
[103] RColorBrewer_1.1-3        rrcov_1.7-0              Rsamtools_2.8.0
[106] cli_3.3.0                 XVector_0.32.0           dagitty_0.3-1
[109] pbapply_1.5-0             MASS_7.3-57              mgcv_1.8-40
[112] tidyselect_1.1.2          forcats_0.5.1            stringi_1.7.6
[115] yaml_2.3.5                locfit_1.5-9.5           ggrepel_0.9.1
[118] tools_4.1.0               ruv_0.9.7.1
scoringutils_1.0.0
[121] cherry_0.6-14             rstudioapi_0.13          uuid_1.1-0
[124] foreach_1.5.2             gridExtra_2.3            farver_2.1.0
[127] digest_0.6.29             BiocManager_1.30.18      quadprog_1.5-8
[130] Rcpp_1.0.8.3              car_3.1-0                broom_0.8.0
[133] gRbase_1.8.7              ggdendro_0.1.23          httr_1.4.3
[136] mpm_1.0-23                colorspace_2.0-3         XML_3.99-0.10
[139] splines_4.1.0             RBGL_1.68.0              multtest_2.48.0
[142] annaffy_1.64.2            xtable_1.8-4             jsonlite_1.8.0
[145] corpcor_1.6.10            R6_2.5.1                 pillar_1.7.0
[148] htmltools_0.5.2           glue_1.6.2               fastmap_1.1.0
[151] BiocParallel_1.26.2       codetools_0.2-18         pcaPP_2.0-1
[154] mvtnorm_1.1-3             utf8_1.2.2               tibble_3.1.7
[157] sva_3.40.0                network_1.17.2           numDeriv_2016.8-
1.1
[160] curl_4.3.2                svd_0.5.1                gtools_3.9.2.2
[163] survival_3.3-1            statnet.common_4.6.0     repr_1.1.4
[166] munsell_0.5.0             GenomeInfoDbData_1.2.6   iterators_1.0.14
[169] RCy3_2.12.4               labelled_2.9.1           mosaicCore_0.9.0
[172] haven_2.5.0               gtable_0.3.0
```

SEMgsa_*code_dat*a 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:

./Data/:

    GSE172114_series_matrix.txt.gz
    GSE172114_rsem_gene_count_matrix_TMM_69samples.csv
    Raw COVID-19 source data

./R/:

    Main_sim.R
    An R script to run Data simulations

    Main_true.R
    An R script to run True data analysis

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

Note:

   1. Set folder SEMgsa_code_data as the current working directory.
   2. Specify a smaller number of iterations by argument "N" 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.