This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

Part 1: Data

☐ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Abstract

This paper analyzes both simulated and real data.

The simulated data are generated with random number seed.

The first real dataset is the OTU abundance data downloadable from the *Tara* Oceans Project data repository (http://taraoceans.sb-roscoff.fr/EukDiv/), including:

- Database_W5_OTU_occurences.tsv: OTU abundance data (zipped due to storage limit on GitHub)
- taxa_mapping.csv: taxonomic hierarchy for each OTU
- TARA_truth.csv: literature validated genus interactions

The second real dataset is the OTU abundance data from the OSU Zebrafish Project, including:

- asv.tab: OTU abundance data
- tax.tab: taxonomic hierarchy for each OTU
- metadata.tab: meta data for samples (zebrafish)

Availability

☒ Data **are** publicly available.
☐ Data **cannot be made** publicly available.

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available data* section, below.

Publicly available data

☒ Data are available online at: https://github.com/yuanjiang-osu/Comp-gLASSO-JASA

☐ Data are available as part of the paper's supplementary material.

☐ Data are publicly available by request, following the process described here:

☐ Data are or will be made available through some other mechanism, described here:

Non-publicly available data

Description

File format(s)

☒ CSV or other plain text.
☐ Software-specific binary format (.Rda, Python pickle, etc.): .rds
☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
☐ Other (please specify):

Data dictionary

☒ Provided by authors in the following file(s): README.md
☐ Data file(s) is(are) self-describing (e.g., netCDF files)
☐ Available at the following URL:

Additional Information (optional)

Part 2: Code

Abstract

The code is available online at https://github.com/yuanjiang-osu/Comp-gLASSO-JASA. The code is organized in different folders for different sections of the paper:

- Simulation: code to produce the ROC curves and the bar graphs in simulation
- TARA: code to produce the results for data analysis of the TARA Oceans Project data
- Zebrafish: code to produce the results for data analysis of the OSU Zebrafish Project data

The folders may also include necessary files such as original and/or intermediate data files for convenience of reproducing the results in the paper. See the workflow for details about how to run the code to reproduce the numerical results in the paper.

Description

Code format(s)

☒ Script files
  ☒ R
  ☐ Python
  ☐ Matlab
  ☐ Other:
☒ Package
  ☒ R
  ☐ Python
  ☐ MATLAB toolbox
  ☐ Other:
☐ Reproducible report
  ☐ R Markdown
  ☐ Jupyter notebook

    - ☐ Other:
- ☐ Shell script
- ☐ Other (please specify):

Supporting software requirements

Version of primary software used

- R 4.0.4

Libraries and dependencies used by the code

- R packages
  - MASS 7.3-53.1
  - glasso 1.11
  - huge 1.3.4.1
  - propagate 1.0-6
  - foreach 1.5.1
  - doParallel 1.0.16
  - space 0.1-1.1
  - tidyverse 1.3.0
  - igraph 1.2.6
  - reshape2 1.4.4
  - cowplot 1.1.1
  - fields 11.6
  - xtable 1.8-4
  - ggplot2 3.3.3
  - snow 0.4-3

Supporting system/hardware requirements (optional)

Parallelization used

- ☐ No parallel code used
- ☒ Multi-core parallelization on a single machine/node
  - Number of cores used: 10
- ☐ Multi-machine/multi-node parallelization
  - Number of nodes and cores used: 3 nodes, 243 cores

License

- ☒ MIT License (default)
- ☐ BSD
- ☐ GPL v3.0
- ☐ Creative Commons
- ☐ Other: (please specify below)

Additional information (optional)

Scope

The provided workflow reproduces:

☐ Any numbers provided in text in the paper
☒ All tables and figures in the paper
☐ Selected tables and figures in the paper, as explained and justified below:

Workflow

Format(s)

☐ Single master code file
☐ Wrapper (shell) script(s)
☐ Self-contained R Markdown file, Jupyter notebook, or other literate programming approach
☐ Text file (e.g., a readme-style file) that documents workflow
☐ Makefile
☒ Other (more detail in *Instructions* below)

Instructions

Set the R working directory to the main folder and run the R wrapper file "Wrapper.R" to reproduce all figures and tables in the manuscript. Below are additional details.

SIMULATION

To reproduce Figures 1-2 in Section 3.2, run the following R scripts:

```
setwd("./Simulation")
source("Simulation_Dense.R")
```

To reproduce Figures S1-S2 in the supplementary materials, run the following R scripts:

```
setwd("./Simulation")
source("Simulation_Sparse.R")
```

REAL DATA/TARA

Run the following R scripts:

```
setwd("./TARA")
source("data_preprocessing.R")
source("TARA_path.R")
```

To produce Figure 3(a) in Section 4.1, run the following R scripts:

```
source("TARA_table_plot_path.R")
```

To produce Figure 3(b) in Section 4.1 and Figure S3 in the supplementary materials, as well as Table 1 in Section 4.1 and Table S1 in the supplementary materials, run the following R scripts:

```
source("TARA_StARS.R")
source("TARA_table_plot_StARS.R")
```

REAL DATA/Zebrafish

Run the following R scripts:

```
setwd("./Zebrafish")
source("data_preprocessing.R")
source("fish_StARS.R")
```

To produce Figures 4-5 in Section 4.2 and Table S2 in the supplementary materials, run the following R scripts:

```
source("fish_table_plot_StARS.R")
```

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

- ☐ < 1 minute
- ☐ 1-10 minutes
- ☐ 10-60 minutes
- ☐ 1-8 hours
- ☒ > 8 hours
- ☐ Not feasible to run on a desktop machine, as described here:

Additional information (optional)

Notes (optional)