

README for Reproducing Tables and Figures in **Learning sparse directed acyclic graphs with mixed data-types nodes based on generalized linear models***

by Yongsu Lee[†] and Chunming Zhang[‡]

Information

Codes have been tested under

- macOS 12.2 on both Intel and M1 chip
- Windows 10 (Version 20H2)
- In order to compile C++ scripts, either GNU Fortran (macOS) or Rtools (Windows) should be available in your system. Check the details on <https://cloud.r-project.org>.

Pre-requisite

- Type the below link in any web browser and donwload the file:
`https://github.com/yongsu-lee/glmdag_dmkd/archive/refs/heads/main.zip`
- Or, you can visit the first author's github repository and download the zip file directly:
`https://github.com/yongsu-lee/glmdag_dmkd`
- Unzip the downloaded file and, under the 'glmdag_dmkd-main/' directory, find 'glmdag_dmkd/'. Set your working directory as './glmdag_dmkd' (where `master.R` is located).
- Open '`master.R`' file.
- In order to generate all the Tables and Figures, you have two options:
 - `use_saved_results = TRUE` (Default): Use saved results data from github (911.6 MB).
 - `use_saved_results = FALSE` : Run all the simulations directly. This is **not recommended** because our simulation/application deal with large-size network. Thus, without using parallel computing system (e.g., HTC), it takes a long time to obtain the final result. (Also, we consider 100 iterations for all cases.)
- When you use `use_saved_results = TRUE` option, all the saved results should be downloaded under './glmdag_dmkd/results_saved/'
 - Type the below link in any web browser and download the zip file (It might take a while - 911.6 MB - for downloading according to your internet speed).
`https://github.com/yongsu-lee/glmdag_results_saved/archive/refs/heads/master.zip`
 - Or, you can visit the repository directly:

*Data Mining and Knowledge Discovery

[†]yongsulee@stat.wisc.edu

[‡]cmzhang@stat.wisc.edu

```
https://github.com/yongsu-lee/glmdag\_results\_saved
```

- Unzip the file and move all the sub-directories (simu_1/, simu2_small/, simu2_large/, hcc/) to under ./glmdag_dmkd/results_saved/.

Load packages and functions

```
## install and load necessary packages
source("./scripts/00_load_r_pkgs.R")

## load required functions (need compiling tools)
subdir = "./scripts/"
source("./scripts/00_load_ftns.R")
```

Reproducing Tables

```
## Table 2 (Table, nominal-level only - 30 nodes) ####
source("./scripts/table_2.R")

## Table 3 (Table, mixed case (small) - 10 nodes) ####
source("./scripts/table_3.R")

## Table 4 (Table, mixed case (large) - 50 nodes) ####
source("./scripts/table_4.R")
```

- Note that all the scripts should be sourced via master.R.
- If you want to run each code separately, you need to run `source("./scripts/sub_master.R")` before running the body of each script.
- All the generated results will be saved under './glmdag_dmkd/tables_figures/'

Reproducing Figures

```
## Figure 2 (Graphs, nominal-level only - 30 nodes) ####
source("./scripts/figure_2.R")

## Figure 3 (Graphs, mixed case (small) - 10 nodes) ####
source("./scripts/figure_3.R")

## Figure 4 (Graphs, mixed case (large) - 50 nodes) ####
source("./scripts/figure_4.R")

## Figure 5 and Figure 6 (Graph, HCC original and trimmed graph)
source("./scripts/figure_5_and_6.R")
```

- Note that all the scripts should be sourced via master.R.
- If you want to run each code separately, you need to run `source("./scripts/sub_master.R")` before running the body of each script.
- All the generated results will be saved under './glmdag_dmkd/tables_figures/'

Miscellaneous

- Tables and Figures are already saved in './glmdag_dmkd/tables_figures_saved/' for your information. (Those files can be generated via master.R)

- './data/' contains `hcc-info.csv` file as a list of information for each variables used in HCC data (See Section 6 and Appendix F).

Possible Troubleshooting

- We directly download application data (HCC) from the repository. However, sometimes, because of status of the repository server, R is not able to download the data. So, we include `hcc-survival.zip` under './glmdag_dmkd/data/' just in case. For this case, open `./scripts/figure_5_and_6.R` and run scripts from

```
unzip(zipfile = "./data/hcc-survival.zip", exdir = "./data/.")  
# (... some codes are below ...)
```