# Manual for Gmcoda that Infers Cross-domain Interactions in Microbiome Studies

Huaying Fang[1,2]

[1]Beijing Advanced Innovation Center for Imaging Theory and Technology, Capital Normal University, Beijing 100048, China.

[2]Academy of Multidisciplinary Studies, Capital Normal University, Beijing 100048, China.

This is a short introduction for gmcoda based on a paper titled "gmcoda: Graphical model for multiple compositional vectors in microbiome studies" that is under review. Gmcoda is designed to infer cross-domain interactions from multiple compositional vectors. The example data files are from the study in "Haak et al. Integrative Transkingdom Analysis of the Gut Microbiome in Antibiotic Perturbation and Critical Illness. mSystems. 2021;6(2):e01148-20. doi:10.1128/mSystems.01148-20."

0. Required: R package "glasso", "huge" and "SpiecEasi".

1. The user of gmcoda should first prepare 2 data files one corresponds to OTUs count matrix for bacteria (example file: "OTU_counts_bacteria_example_gmcoda.csv") and the other for fungi (example file: "OTU_counts_fungi_example_gmcoda.csv"). These two files should use "," as the separator and include sample IDs for rows and OTU names for columns. Some part of one data file (example file: "OTU_counts_bacteria_example_gmcoda.csv") is as follows.

```
Acidaminococcus,Actinomyces,Adlercreutzia,Agathobacter,Akkermansia
TKI_F1,0,0,0,0
TKI_F10,1,0,0,0
TKI_F11,2,4,0,1
TKI_F12,0,0,0,0
TKI_F13,0,17,1,58
TKI_F14,0,5,8,994
TKI_F15,0,0,0,11
TKI_F16,0,222,0,0
TKI_F17,0,2,0,0
```

2. The user replaces file names in the file "gmcoda_example.R" with their own 2 data files and can change "output_gmcoda_example.csv" into new file names in the file "gmcoda_example.R".

3. Open an R terminal and write "source("gmcoda_example.R");" in R terminal. There will be an output file named "output_gmcoda_example.csv" in the folder. The header of the output file is as follows.

```
type,OTU1,OTU2,gmcoda,glasso,SEGL,SEMB
1&1,Alistipes,Agathobacter,-0.0467814013289746,0,0,0
1&1,Anaerostipes,Agathobacter,0.0714541236209362,0,0,0
1&1,Bifidobacterium,Agathobacter,0.0865937605683131,0,0,0
1&1,Dorea,Agathobacter,0.120214792415115,0.00837279490182954,0.0148389879669085,0
1&1,Escherichia.Shigella,Agathobacter,-0.0417124382832421,0,-0.0412249851205499,0
1&1,Faecalibacterium,Agathobacter,0.155484914823926,0,0,1
1&1,Family_XIII_AD3011_group,Agathobacter,-0.118033045372781,0,-0.0267263464421525,1
1&1,Fusicatenibacter,Agathobacter,0.314199433117209,0.0567870854655591,0.132164766515616,1
1&1,Lactobacillus,Agathobacter,-0.00881626535076855,0,0,0
```

4. Explanation of output file. The column "type" indicates the edge is inter-domain ("1&1" for bacteria-bacteria and "2&2" for fungi-fungi) or cross-domain ("1&2" for bacteria-fungi). The columns "OTU1" and "OTU2" represent two microbes (or nodes). The columns "gmcoda", "glasso" and "SEGL" are the partial correlation coefficients from gmcoda, glasso and SEGL, respectively. The column "SEMB" indicates whether the corresponding edge is detected ("1") or not detected ("0") by SEMB.

Notes:

(i). There are 2 parameters "propFilt" and "only.gmcoda" for function "gmcoda_wrap" in Line 12 in "gmcoda_example.R." "propFilt" is the threshold for keeping OTUs and sampes if there are at least "propFilt" non-zeros for OTUs or samples. The default parameter is 0.8. "only.gmcoda" is an indicator that uses only gmcoda ("only.gmcoda = T") or also uses other methods including glasso, SEGL and SEMB (Default: "only.gmcoda = F").

(ii). The main function for gmcoda is the function "gmcoda" in the script "gmcoda.R." This function has a parameter "wgt_lamb" that controls the relative penalty weights for elements of the precision matrix. This parameter can be used if the user has some prior information about microbial interactions.