Differential Network Analysis Toolbox User Manual v1.1

1. Software Description, Installation and Setup

1.1 Description

Based on the differential network analysis methods (WDNE/TDJGL/FGL/D-trace), the Differential Network Analysis Toolbox (DNAT) provides four methods to analyze gene expression data of two different states to drive the differential network between two states.

1.2 Hardware and Software Requirements

The toolbox is developed using Matlab 2018b and has been evaluated on the Windows platform.

We will adapt and evaluate it on other systems in the future.

1.3 Program Installation

The DNAT can be obtained at https://github.com/Oyl-CityU/DNAT.

After decompressing the 'DNAT.rar' file, run 'DNAT.exe' in the folder 'for_redistribution' to install DNAT. The install process needs to connect to web, if it fails to install, please do not change its default installation path and then try again, or download and install the 64-bit version of the MATLAB Runtime for R2018b from the MathWorks Web site by navigating to: http://www.mathworks.com/products/compiler/mcr/index.html and then rerun 'DNAT.exe' to install DNAT.

1.4 Referencing the software

When using this toolbox please cite: *Le Ou-Yang, Dehan Cai, Xiao-Fei Zhang, Hong Yan, Integrating differential expression analysis into differential network analysis of gene expression data with missing values.*

And specially

Dtrace: *Yuan, H., Xi, R., Chen, C., & Deng, M. (2017). Differential network analysis via lasso penalized D-trace loss. Biometrika, 104(4), 755-770.*

FGL: *Danaher, P., Wang, P., & Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(2), 373-397.*

TDJGL: *Zhang, X. F., Ou-Yang, L., Zhao, X. M., & Yan, H. (2016). Differential network analysis from cross-platform gene expression data. Scientific reports, 6, 34112.*

2. Using the DNAT

2.1 GUI layout

a. The main page

The page shown in Fig.1 is the only GUI page of DNAT. It provides several functional modules for user.

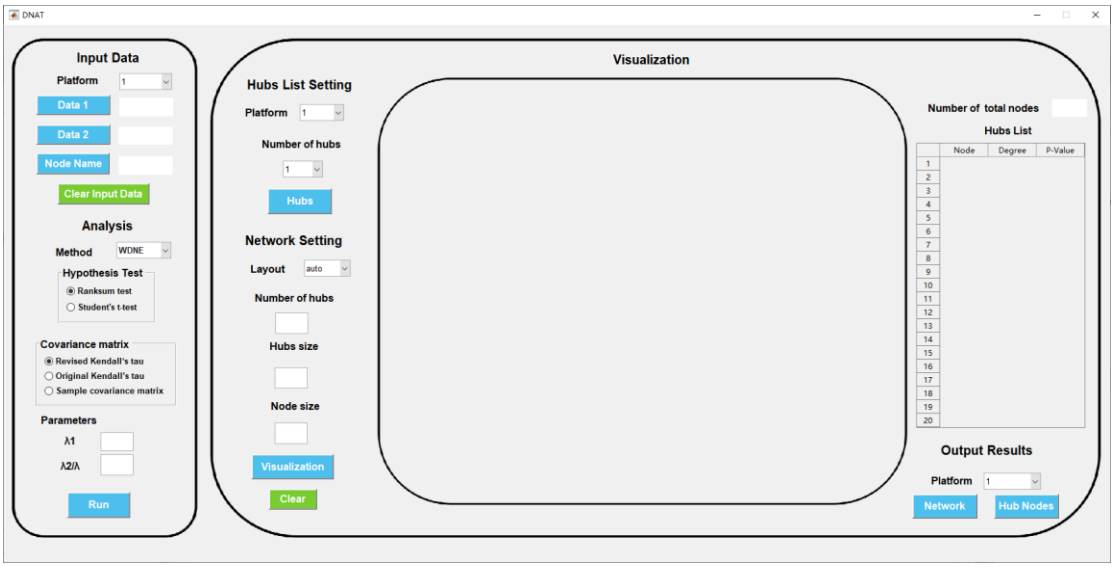

Fig.1 The main page of DNAT

In section ***2.2***, we will use an example to show how to use the methods mentioned above to analyze data and visualize the results.

2.2 *An example of the use of DNAT to estimate the differential network between two different states such as platinum-resistant and platinum-sensitive ovarian tumors.*

Here is an example to illustrate how to use DNAT to estimate differential gene network between two different states.

Here we use gene expression data of platinum-resistant and platinum-sensitive ovarian tumors as an example. Users can follow the instructions below to perform the analysis.

1) After installing DNAT, open DNAT to drive the DNAT page (See Fig.1).

2) Input data (See Fig.2)

a. Choose the data platform (Suppose there are M different data platforms, select one data platform (e.g., the 3rd platform) at each time);

b. Input Data 1 and Data 2 of the selected platform (e.g., 3rd platform), each data matrix corresponds to the gene expression data of samples in one state. Each row of the data matrix corresponds to a sample, and each column of the data matrix corresponds to a gene. Users can choose a .txt file or a .mat file or an excel file, and the detailed information of data format will show in section "3. Further Information";

c. Repeat steps a-b until all the data you want to analyze have been inputted. In this study, we consider three gene expression datasets obtained from three different data platforms.

d. Input the name of the features/nodes (e.g., the name of gene). Here the names of genes are organized as a column vector.

e. If you want to reinput the data, just do the same as before. And if you want to clear all the input data, you can click the "Clear Input Data" button.

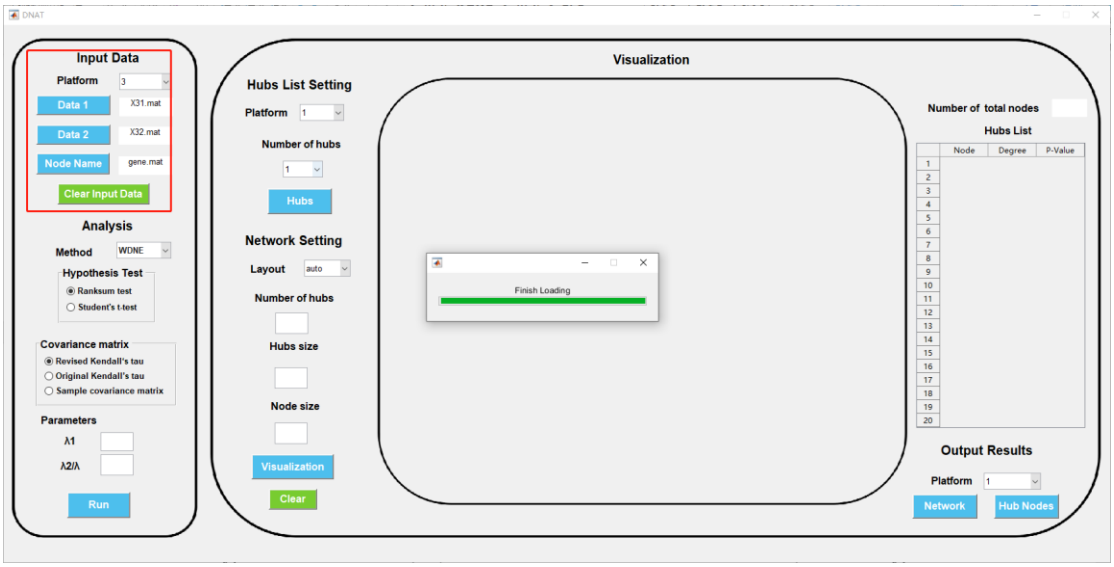

Fig.2 Input data

3) Choose the analysis methods (See Fig.3)

There are four estimation methods (i.e., WDNE/TDJGL/FGL/D-trace), two types of hypothesis test methods (i.e., student's t-test and Wilcoxon rank-sum test) and three types of computing way for empirical covariance matrix(i.e., Revised Kendall's tau matrix, Original Kendall's tau matrix and Sample covariance matrix) ready for choosing. Here we choose WDNE, Wilcoxon rank-sum test and Original Kendall's tau matrix to perform the analysis.

4) Input the model parameters (See Fig.3)

For WDNE, TDJGL and FGL, you need to input the values of λ_1 and λ_2. For D-trace you need to input the value of λ. Here the values of λ_1 and λ_2 for WDNE are set to 30 and 10 respectively.

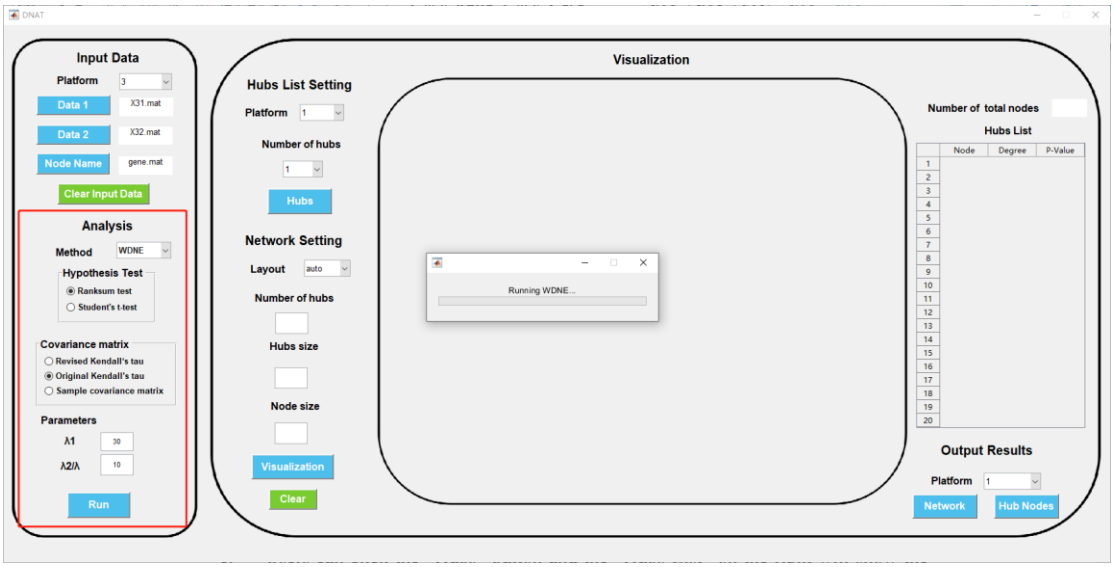

Fig.3 Choose the analysis method and input parameters to analyze the data

5) Click the "Run" button (See Fig 3)
6) Hubs List Setting
   a. When the differential network estimation is finished, users can visualize the differential network estimated form each platform by selecting the platform from the drop-down menu (each algorithm will output multiple differential networks corresponding to different platforms). Here we choose the first platform;
   b. Users can also choose the number of hub genes in the estimated differential network for visualization. Here we choose the top 20 hub genes (according to their degrees in the estimated differential network) in the differential network estimated from the first platform for visualization;
   c. Users can click the "Hubs" button and the "Hubs List" on the right will show the detailed information of the top 20 hub genes in the estimated differential network, including the degree of each hub gene and the P-value of the hypothesis test. (See Fig 4).

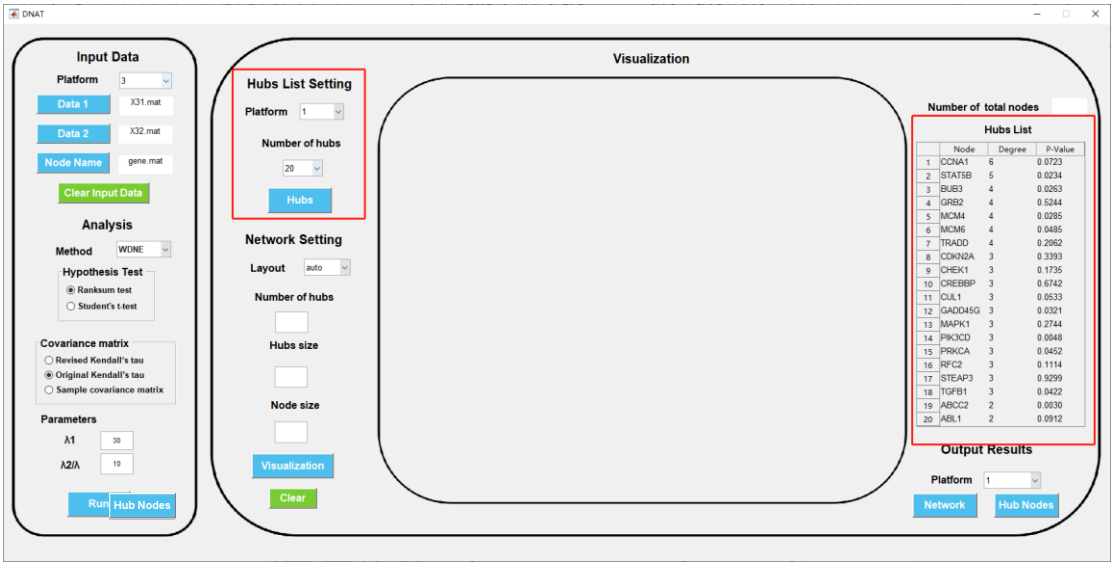

Fig.4 Hubs List Setting

7) Network setting
   a. Choose the "Layout" of network, here we choose "force";
   b. Set the "Number of hubs" in the network (the default number is 10, which means we would like to see the connection patterns of the top-10 hub nodes in the estimated differential network). If the total number of nodes whose degrees are greater than 1 is less than the setting number, it will shows the whole network;
   c. Set the sizes of hubs and the sizes of other connected nodes. The default sizes of hubs and other nodes are 8 and 4;
   d. Click the "Visualization" button to show the network. (See Fig 5)
   e. There will be a float window also showing the network, you can save figure of the network according to this float window. (See Fig 5)
   f. If you want to clear the "Hubs List" or the network visualization, click the "clear"

button.

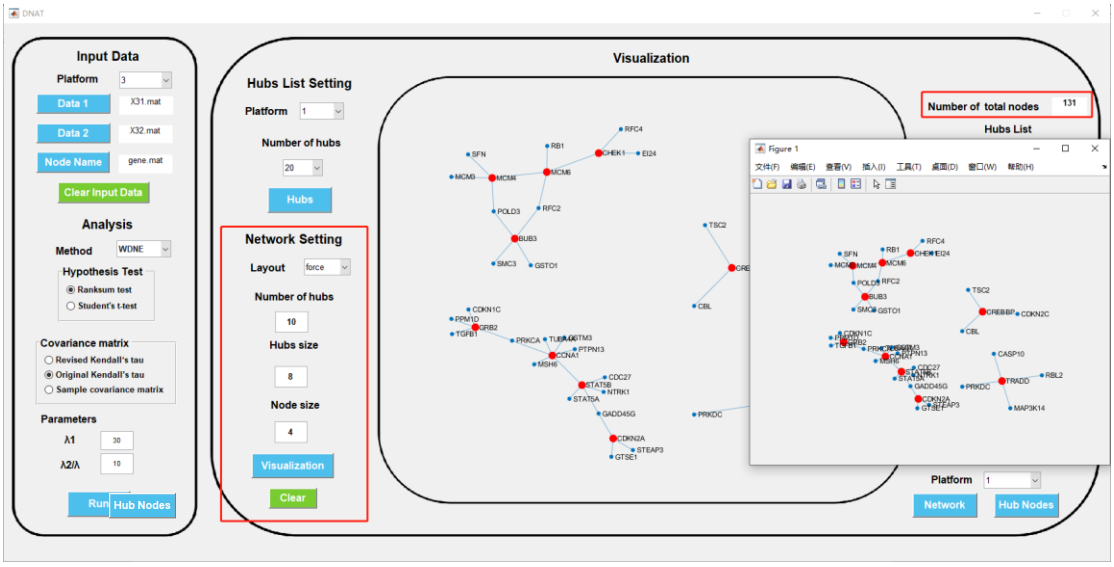

Fig.5 Network Setting

9） Output result

If you want to see the whole network and the information of all nodes for each platform, you can choose the platform and then click the "Network" button or the "Hub Nodes" button to save the result in a .txt file. (See Fig 6)

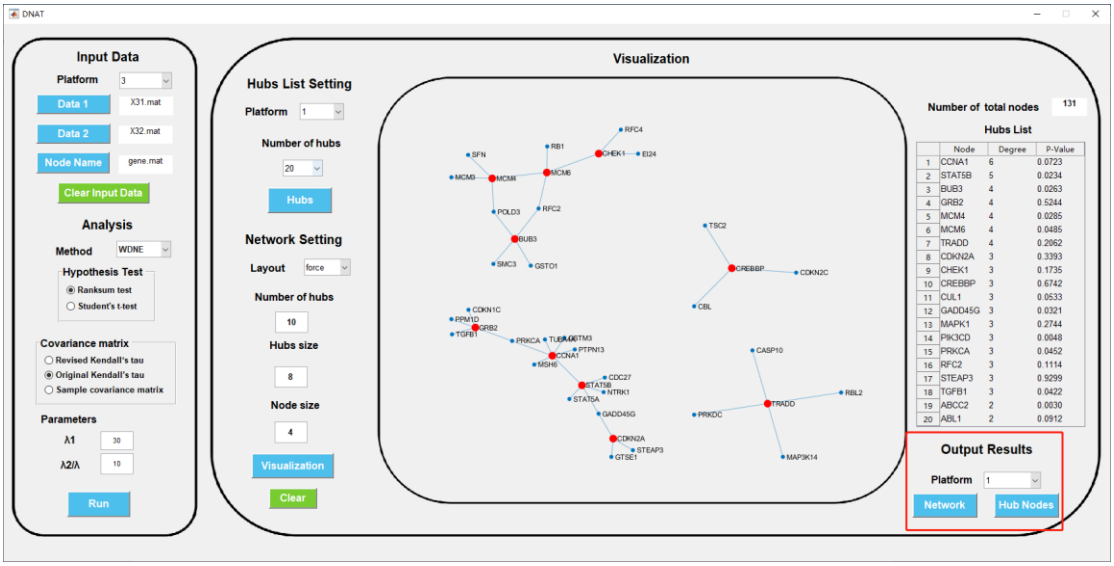

Fig.6 Output Result

3. Further information.

3.1 Each data matrix denotes the gene expression collected from one platform over subjects in one state. Each row in the data matrix denotes a sample, and each column in the data matrix denotes a feature.

3.2 You should make sure that only one data matrix in one file, (we support the follow files: *.mat / *.xls / *.xlsx / *.txt). And there should be only one data matrix in one file. Besides, for *.xls / *.xlsx file , do not include any empty sheet in the excel file.

3.3 If you want to gain a more sparse network, please input bigger values for λ (for D-trace) or λ_1 , λ_2 (for others) and analyze the data. If there is a hint "No edge in the network", please input smaller values for λ or λ_1 , λ_2 and analyze the data.

3.4 Model parameters: WDNE(λ_1 , λ_2) ,TDJGL (λ_1 , λ_2), FGL (λ_1 , λ_2),D-trace (λ).

4. Help:

If the program fails to run, make sure that the data format is right or you have closed windows from a previous run.

If a window hangs, use the task manager to stop any DNAT-related process that is using excessive CPU time.

5. Contact information:

Email: caidehan2016@email.szu.edu.cn; leouyang@szu.edu.cn.