



MEDICAL SCHOOL
UNIVERSITY OF MICHIGAN

MICHIGAN REGIONAL COMPREHENSIVE METABOLOMICS RESOURCE CORE



MEDICAL SCHOOL
UNIVERSITY OF MICHIGAN

COMPUTATIONAL MEDICINE AND
BIOINFORMATICS

FILIGREE

User Manual

University of Michigan

November 2020

Table of Contents

Terms of Use	3
General User Policy	3
For academic and non-profit institutions	3
For commercial and for-profit institutions	3
Citation Policy	3
Overview	4
Input File	5
Using Filigree	5
Data Tab	6
Feature Grouping	7
Calculate Feature Groups	7
Specify Feature Groups	8
Hybrid Grouping	8
Analysis Tab	10
Select Output Directory	11
Current Analysis	11
Progress	11
Run Analysis	11
Browse Networks	12
Output Files	15
Feature Grouping Output File	15
Filigree Output Files	15
Supplemental Information	15
Feature Grouping Details	15
Visualizing Complete Network	16
Using Cytoscape	16
References	17

Terms of Use

General User Policy

For academic and non-profit institutions

- Permission is granted to access, use, and/or download the Tool for internal use only.
- If the user desires to create derivative works of the Tool, source code or access to data may be made available through request by sending an email to filigree-help@umich.edu.
- Use of the Tool must be acknowledged in resulting publications (see [Citation Policy](#) below).

For commercial and for-profit institutions

- Permission is granted to access, use, and/or download the Tool for internal use only.
- To create derivative works of the Tool for commercial purposes, source code or access to data may be permitted through negotiation for a commercial license. Please send request to filigree-help@umich.edu.

Citation Policy

For the Filigree tool, please cite: Iyer GR, et al. 2020. Application of Differential Network Enrichment Analysis for deciphering metabolic alterations. Manuscript submitted.

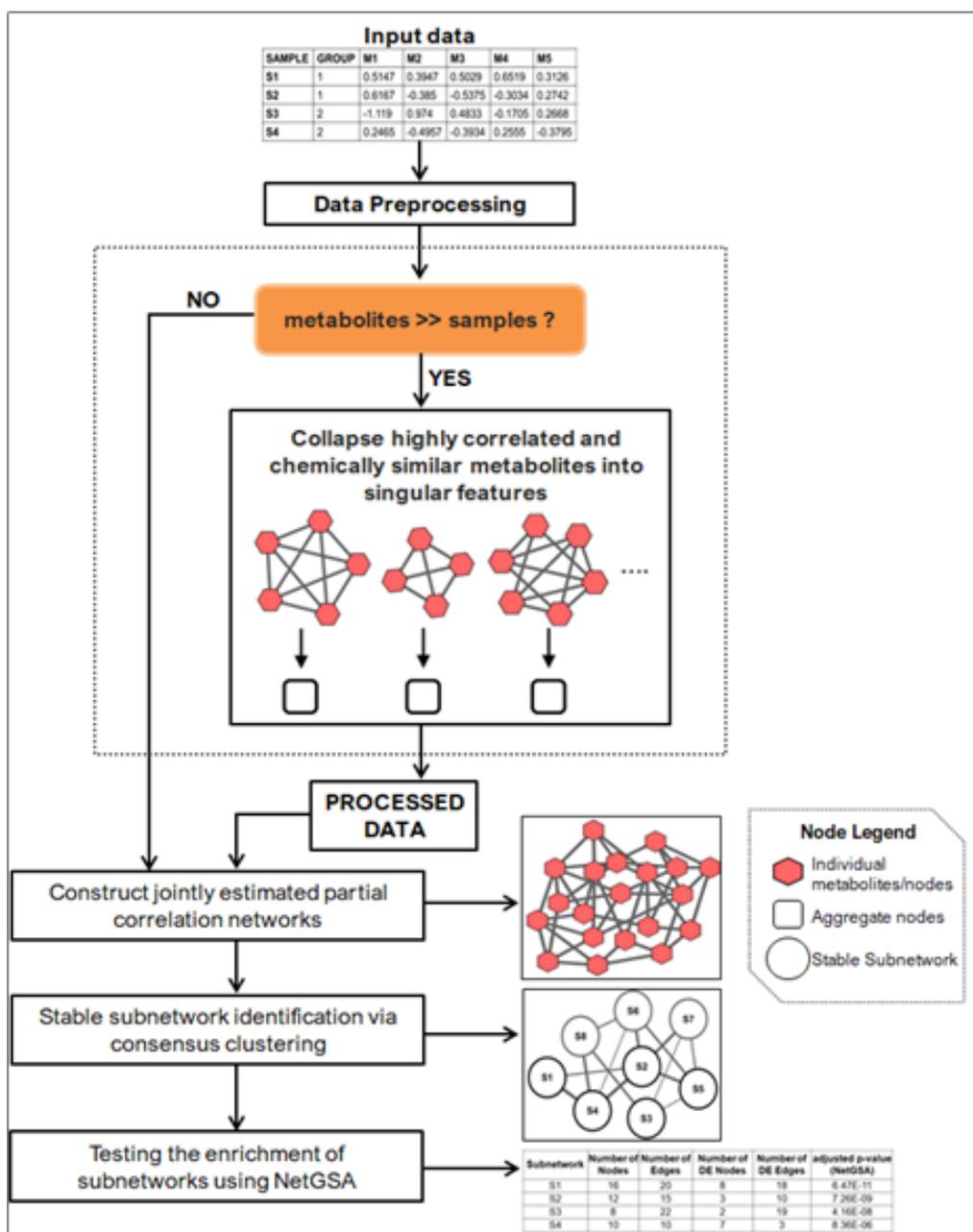
For the DNEA algorithm, please cite: Ma J., et al. 2019. Differential network enrichment analysis reveals novel lipid pathways in chronic kidney disease. *Bioinformatics* 35(18): 3441-3452.

Please note that due to continuous software upgrades the images in this handout may differ from what you see on the screen.

Overview

In metabolomics and lipidomics experiments, it is common for the number of measured metabolites to far exceed the number of samples, resulting in extremely sparse partial correlation networks. The Differential Network Enrichment Analysis (DNEA) algorithm includes joint network estimation, consensus clustering, and enrichment testing.¹ The Filigree tool is a Java application that uses a data-driven approach to construct partial correlation networks, identify stable subnetworks, and perform enrichment analysis.

Below is an overview of the Filigree workflow:



Input File

Filigree accepts .csv files containing log-transformed and auto-scaled metabolite measurements, where each row is a sample. However, if performing feature-aggregation, the user should provide a .csv file with log-transformed (but not auto-scaled) metabolite measurements. The input file must include the following columns:

- Column 1 must contain the sample names. These names should not include commas, so make sure to replace all commas with dashes.
- Column 2 must contain group assignments.
- The remaining columns contain metabolites/lipids. These names should not include commas, so make sure to replace all commas with dashes.

ID	Group	xylose	xylitol	xanthosine	xanthine	valine	uridine
1	non-diabetic	0.298138	0.11356	0.29808539	0.207173	-0.15525	0.538404
2	non-diabetic	-0.14301	-0.01511	0.19139982	-0.28978	-1.15403	0.068907
3	non-diabetic	-0.46745	-0.21629	0.20816182	0.132961	-7.49916	0.208186
4	diabetic	-0.56606	-0.63871	-0.5706407	-0.22573	-0.36754	0.246013
5	non-diabetic	1.021524	0.421075	0.28726893	-0.98819	0.028634	0.324583
6	diabetic	0.386527	-0.67713	0.65082629	0.626206	-0.17298	0.555085
7	diabetic	0.622072	0.412723	-0.2319071	-0.19522	0.260781	-0.33446
8	non-diabetic	0.415355	0.462003	-0.2243708	-0.873	0.345247	-0.71879
9	non-diabetic	-0.50125	-0.7139	-0.3715864	0.290793	-0.61594	-0.08117
10	diabetic	0.139762	-1.42806	0.26205137	0.191165	0.314762	0.267119
11	non-diabetic	0.615014	-0.06666	-0.9228054	-1.05863	-0.15643	-6.47059
12	diabetic	-0.1465	0.369189	-0.0368156	0.591488	0.404657	0.982276
13	diabetic	0.83514	0.772761	0.5623044	0.445346	0.71269	0.400853
14	non-diabetic	0.511699	-0.73783	-2.531286	-1.38558	0.021345	-0.77839

Note: If the number of metabolites/lipids is much higher than the number of samples, go to the [Feature Grouping](#) section.

Note: It is recommended that the metabolite to sample ratio be approximately 1.25.

Using Filigree

After launching Filigree, the user will see two tabs. The user can step through Filigree using the tabs at the top of the window or the Next>> button at the bottom right of the window.

Data Tab

The screenshot shows the 'Data' tab of the Filigree v0.1.2 software. The interface includes the following sections:

- Select Intensity Data File (must be .csv):** A text box contains the file path 'T1D_primaryMetabolites_noIS_log_scaled_sorted.csv' and a 'Browse...' button.
- Specify Columns/Rows:** Two dropdown menus. 'Sample ID' is set to 'Sample' and 'Group' is set to 'Group'.
- Specify Sample Groups:** Two dropdown menus. 'Group 1' is set to 'diabetic' (30 samples) and 'Group 2' is set to 'non-diabetic' (41 samples).
- Feature Grouping:** Three radio buttons: 'Calculate Feature Groups' (checked), 'Specify Feature Groups', and 'Hybrid Grouping'.
- Feature reduction:** A slider set to 30.0% (1.75 feature to sample ratio) and a 'View Heatmaps' button.

Navigation buttons '<< Previous' and 'Next >>' are located at the bottom of the window.

1. Under “Select Intensity Data File (must be .csv),” use the “Browse...” button to find and load the input data file.
2. Under “Specify Columns/Rows,” use the dropdown arrows next to “Sample ID” and “Group” to map the appropriate column/row containing this information from the input file.
3. Under “Specify Sample Groups,” use the dropdown arrows next to each “Group” to define the specific groups. The dropdown menu will pull the group options from the group column in the input file (as designated in Step 2). The number of samples within a given group will appear to the right of each designated group.
4. Under “Feature Grouping,” select the desired method (see the [Feature Grouping](#) section below for details on feature grouping).
 - Calculate Feature Groups
 - Specify Feature Groups
 - Hybrid Grouping
5. Depending on the selected feature grouping method, additional steps may be required before moving to the next section. See the [Feature Grouping](#) section below for the specific details for each method.
6. To move to the next section, click the “Next>>” button or click on a different tab at the top of the window.

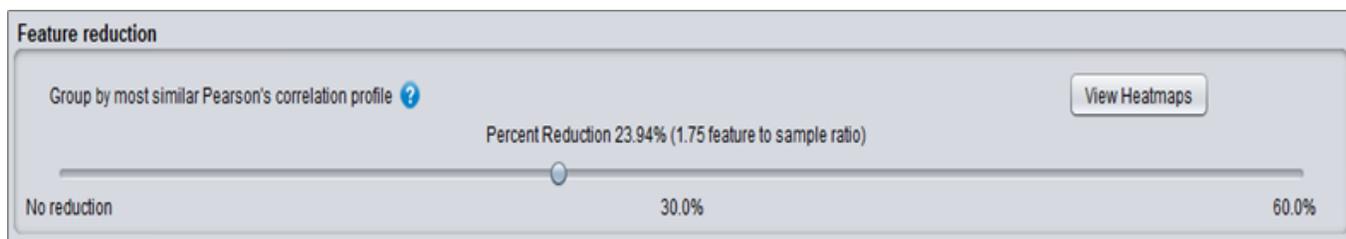
Feature Grouping

If the input data has a lot more metabolites/lipids than samples, feature grouping is recommended. Feature grouping combines metabolites/lipids that belong to the same feature group into one feature, represented by the median intensity across all samples within the group. If no feature reduction is desired, make sure the feature selection slider is positioned to the far left, at 0.

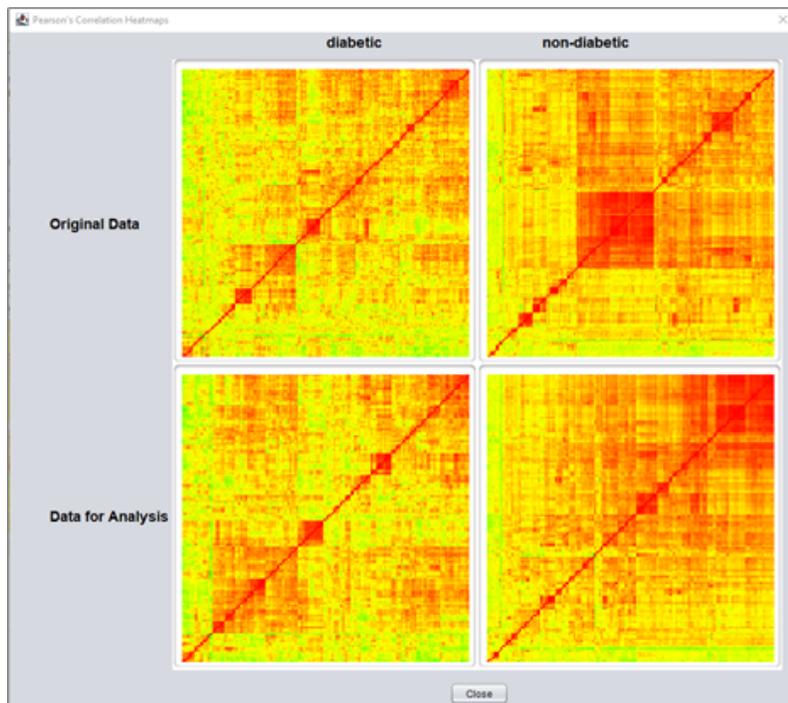
Calculate Feature Groups

This is a data-driven method that collapses nodes based on similarity, using Pearson's Correlation. Ideally, the user will want the metabolite to sample ratio to be approximately 1.25.

1. When this method is selected, the "Feature reduction" section appears. Use the slider to select the desired percent reduction of features. The feature to sample ratio for a given percent reduction will appear above the slider. See the Supplemental [Feature Grouping Details](#) section for more details on the actual feature grouping method.



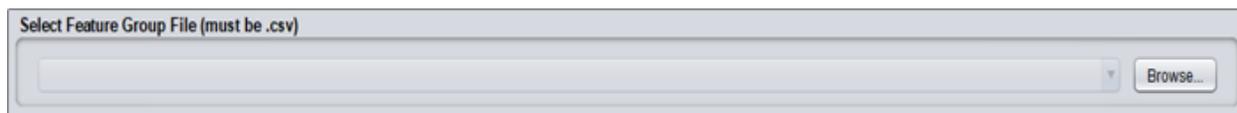
2. Click the "View Heatmaps" button to better understand the result of a specific percent reduction.
 - a. A "Pearson's Correlation Heatmaps" popup window appears showing heatmaps for the original data and the collapsed data with the chosen percent reduction.
 - b. Click the "Close" button to close the "Pearson's Correlation Heatmaps" popup window.



Specify Feature Groups

This is a knowledge-based method that collapses nodes based on the user's input. The user must upload a feature group file when this option is selected.

1. When this method is selected, the "Select Feature Group File (must be .csv)" section appears. Use the "Browse..." button to find and load the feature group file.



The feature group file must include the following columns:

- A column containing the metabolite name
- A column containing the group name

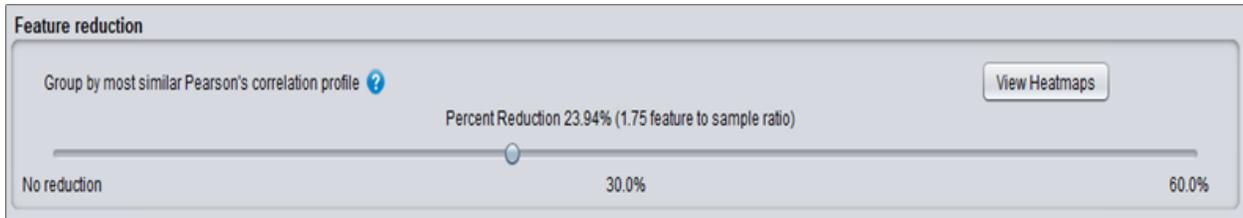
Feature	Group
Metabolite 1	Group 1
Metabolite 2	Group 1
Metabolite 3	Group 1
Metabolite 4	Group 1
Metabolite 5	Group 2
Metabolite 6	Group 3
Metabolite 7	Group 3
Metabolite 8	Group 3

Lipid	Lipid class
PC 39:5	PC
PC 40:9	PC
PC 44:4	PC
TG 39:0	TG
TG 46:1	TG
TG 50:4	TG
TG 52:6	TG
LysoPC 17:1	LPC
LysoPC 26:0	LPC
LysoPE 22:5	LPC

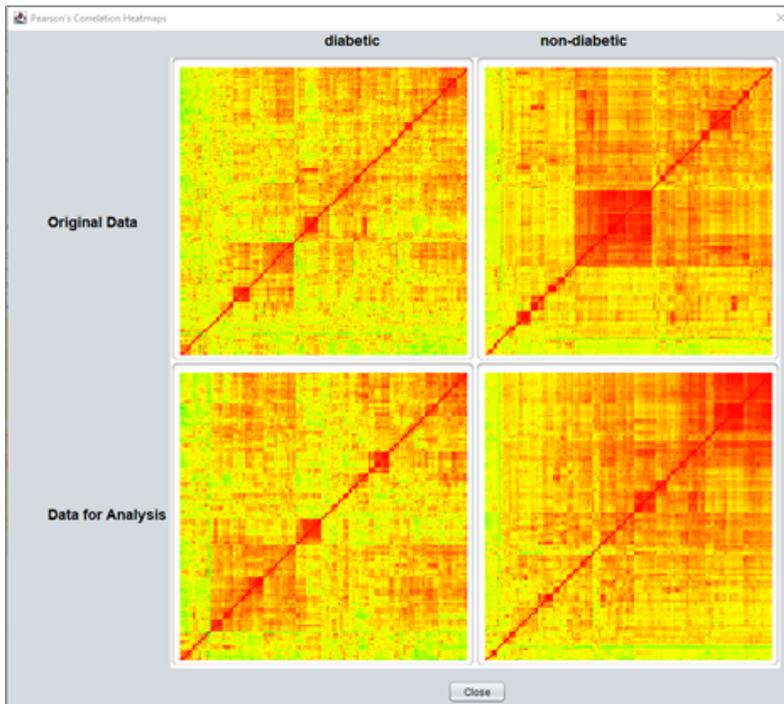
Hybrid Grouping

This method combines the data-driven Calculate Feature Groups method and the knowledge-based Specify Feature Groups method. The tool collapses chemically similar nodes, based on the user's input, that are also highly similar, based on Pearson's Correlation.

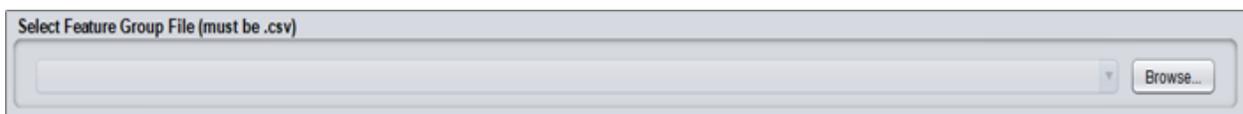
1. When this method is selected, the "Feature reduction" section appears. Use the slider to select the desired percent reduction of features. The feature to sample ratio for a given percent reduction will appear above the slider. See the Supplemental [Feature Grouping Details](#) section for more details on the actual feature grouping method.



2. Click the “View Heatmaps” button to better understand the result of a specific percent reduction.
 - a. A “Pearson’s Correlation Heatmaps” popup window will appear showing heatmaps for the original data and the newly analyzed data with the chosen percent reduction.
 - b. Click the “Close” button to close the “Pearson’s Correlation Heatmaps” popup window.



3. In addition, the “Select Feature Group File (must be .csv)” section appears. Use the “Browse...” button to find and load the feature group file.



The feature group file must include the following columns:

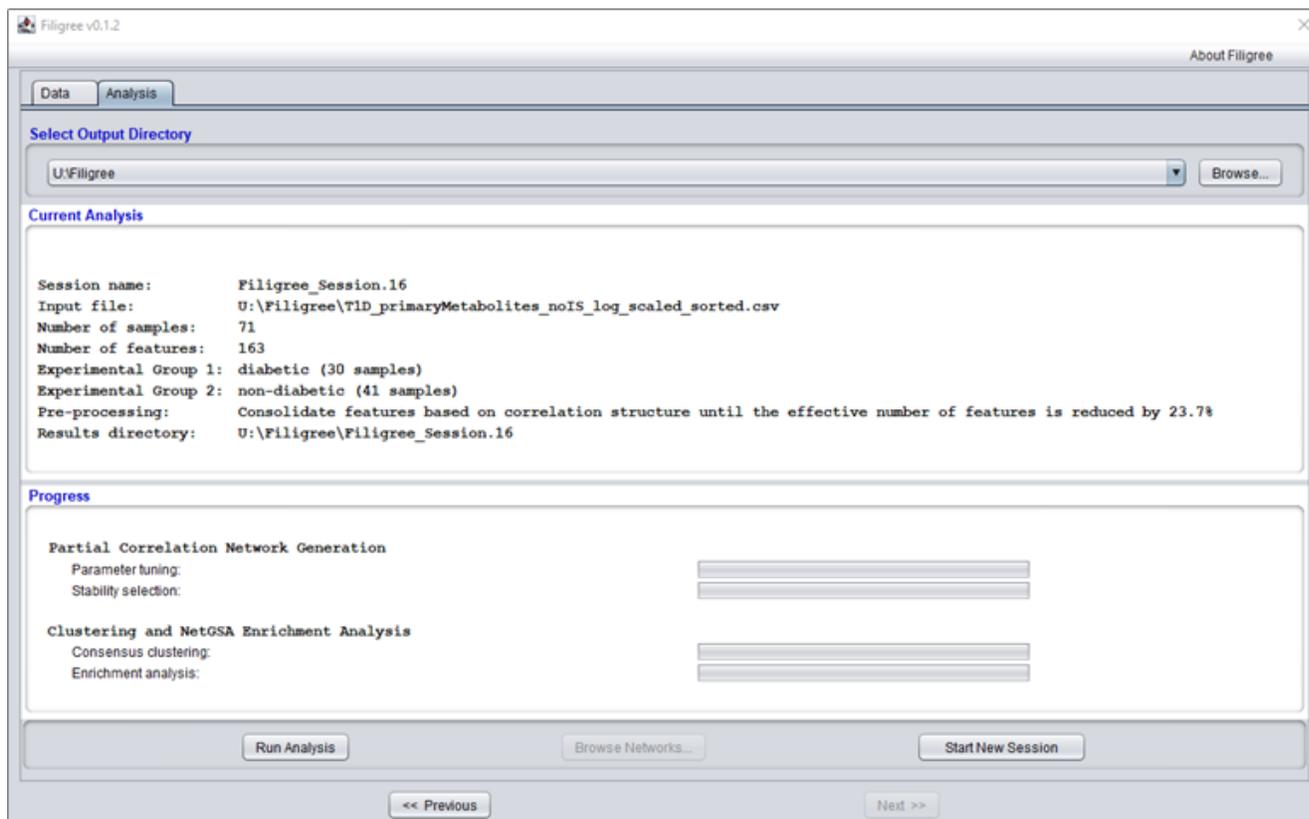
- A column containing the feature name
- A column containing the group name

Feature Names	Group Names
Feature	Group
Metabolite 1	Group 1
Metabolite 2	Group 1
Metabolite 3	Group 1
Metabolite 4	Group 1
Metabolite 5	Group 2
Metabolite 6	Group 3
Metabolite 7	Group 3
Metabolite 8	Group 3

Lipid	Lipid class
PC 39:5	PC
PC 40:9	PC
PC 44:4	PC
TG 39:0	TG
TG 46:1	TG
TG 50:4	TG
TG 52:6	TG
LysoPC 17:1	LPC
LysoPC 26:0	LPC
LysoPE 22:5	LPC

Analysis Tab

This tab includes a dashboard containing details about the user's current analysis and a progress section containing details about the state and progress of the analysis. In addition, the button for running the analysis is located on this tab.



Select Output Directory

1. Under “Select Output Directory,” use the “Browse...” button to select the directory location for storing the output files generated from the analyses.

Current Analysis

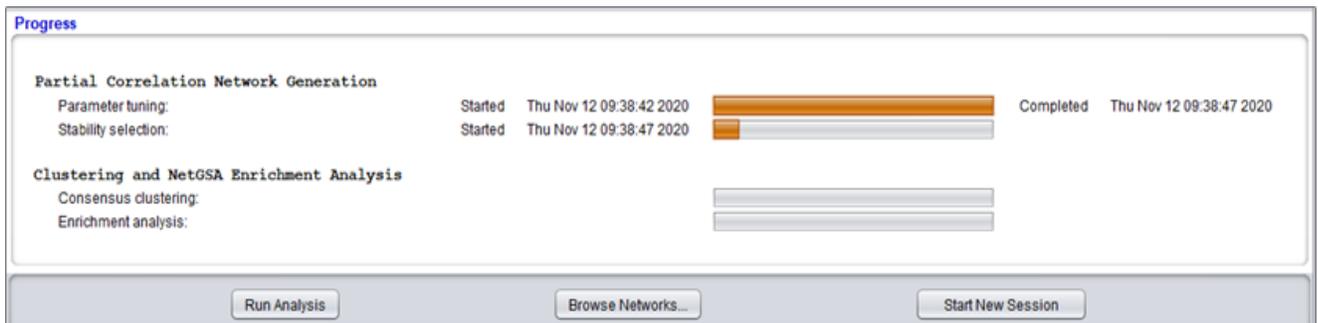
1. Under “Current Analysis,” review the information about your planned analysis. This section includes characteristics of the data set, such as the number of samples and number of features. It also lists the computer directory where the results will be saved.

Progress

This section is where the user can view the progress of their current analysis. The tool runs the analysis in 2 main components:

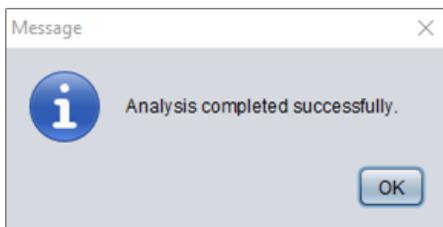
- “Partial Correlation Network Generation” which has 2 sub-components: “Parameter tuning” and “Stability selection.”
- “Clustering and NetGSA Enrichment Analysis” which has 2 sub-components: “Consensus clustering” and “Enrichment analysis.”

As the analysis progresses, the user will see progress bars update for each analysis component. Once an analysis component is completed, the progress bar will be completely filled-in, the start time will be shown to the left of the progress bar, and the completion time will be shown to the right of the progress bar. Once a step is completed, the next step will begin.



Run Analysis

1. Click the “Run Analysis” button at the bottom left.
 - a. Once the appropriate resources are available, the analysis will run.
 - b. When the analysis is completed, a popup window will appear with the message “Analysis completed successfully.” Click the “OK” button to close this window.



2. To visualize the results, click the “Browse Networks...” button. Learn more about viewing networks in the below [Browse Networks](#) section.
3. To start a completely new Filigree session, click the “Start New Session” button. This will reset all parameters to their default and begin a completely new session.

Browse Networks

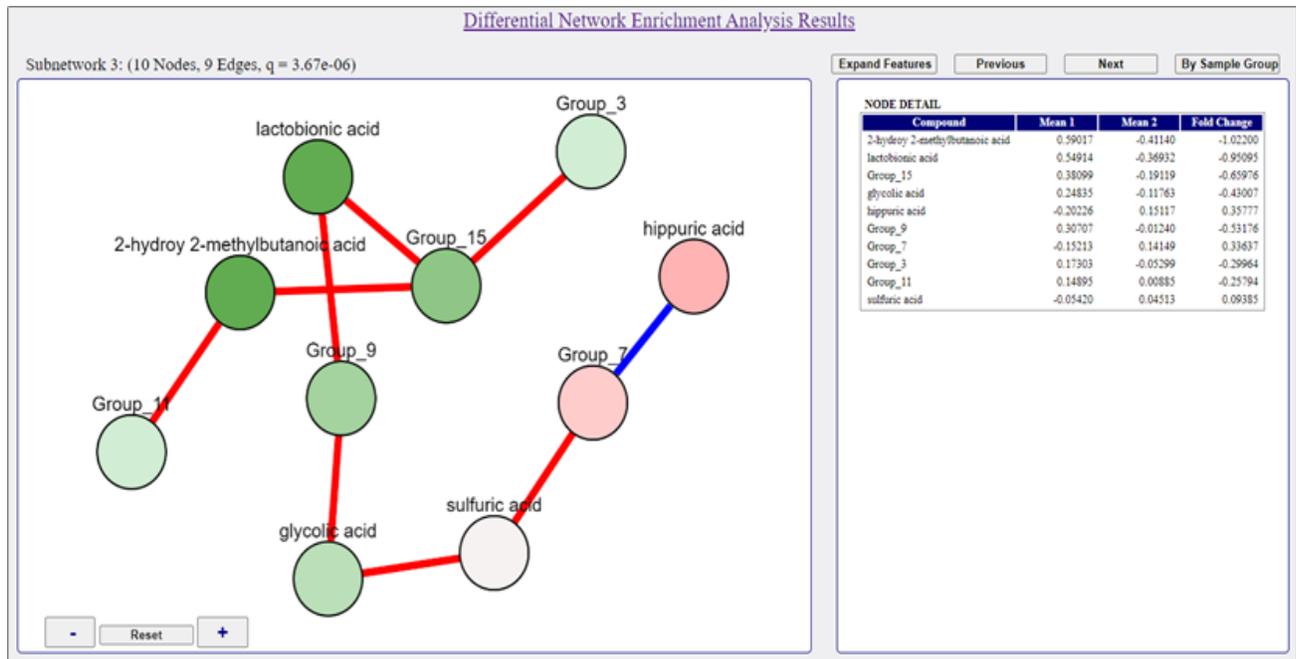
Filigree provides HTML files that allow the user to interactively visualize results as subnetworks. To visualize the complete network, see the Supplemental [Visualizing Complete Network](#) section.

1. Click the “Browse Networks...” button at the bottom of the Analysis tab.
2. A menu may appear asking which browser to use to open the HTML file. Select the preferred browser and click “OK.” (Note: the networks should be viewable in any browser, but Chrome is recommended.)
3. The interactive Filigree subnetworks open in the selected browser tab, initially showing the Differential Network Enrichment Analysis Results summary table. For each subnetwork, this table includes the number of nodes and edges, the number of significant nodes and differential edges, the p-value, and the q-value.

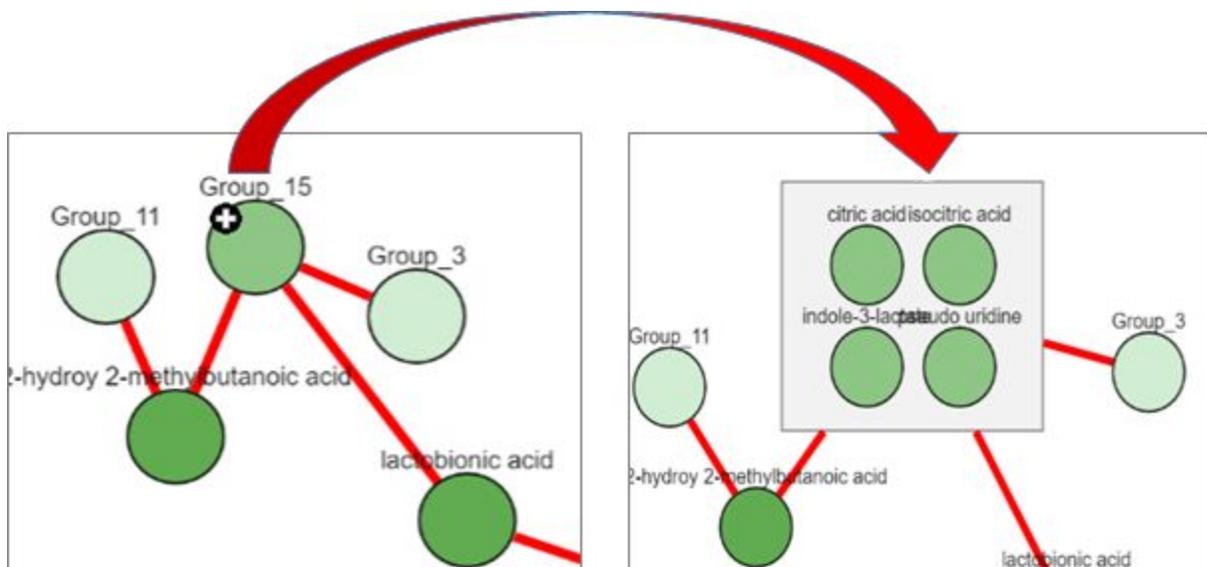
Differential Network Enrichment Analysis Results

Subnetwork Name	# Nodes	# Edges	# Sig. Nodes	# Diff. Edges	P-Value	Q-Value FDR
Subnetwork 1	7	7	3	7	1.687e-10	1.518e-09
Subnetwork 2	8	9	4	9	5.288e-07	2.380e-06
Subnetwork 3	10	9	4	9	1.223e-06	3.669e-06
Subnetwork 4	7	8	2	8	2.882e-06	6.485e-06
Subnetwork 5	7	9	1	9	6.932e-05	0.0001248
Subnetwork 6	7	8	2	8	0.0009670	0.001450
Subnetwork 7	16	22	2	22	0.001157	0.001487
Subnetwork 8	12	15	2	15	0.3591	0.4040
Subnetwork 9	8	9	0	9	0.9952	0.9952

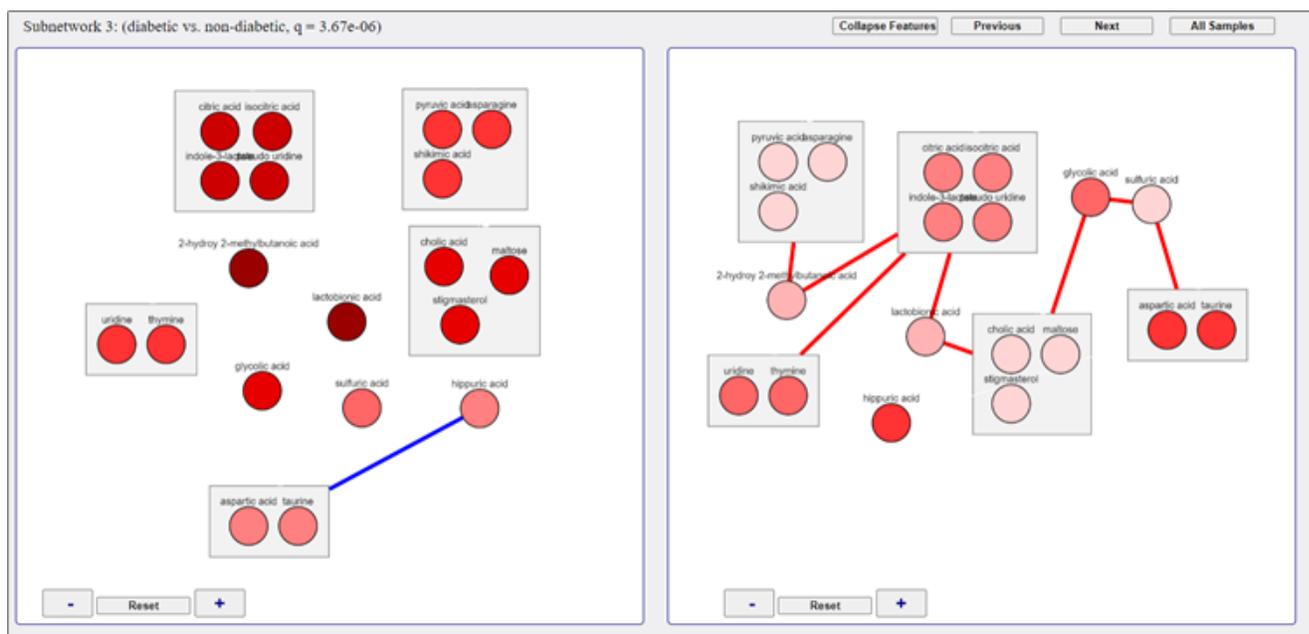
1. To view a specific subnetwork, click the associated link in the “Subnetwork Name” column in the table. (Note: the subnetwork may open in the same window as opposed to opening a new tab.)
2. For each subnetwork, there is a network visualization and a table providing node details.
 - Nodes represent metabolites or groups of metabolites.
 - Edges represent partial correlations between two metabolites.
3. To get back to the Differential Network Enrichment Analysis Results summary table, click the “Differential Network Enrichment Analysis Results” link, located at the top of each subnetwork page.



- Node color represents fold change of experimental/sample group 2 relative to experimental/sample group 1 (as designated in the [Data Tab](#)).
 - Pink nodes represent up-regulated nodes.
 - Green nodes represent down-regulated nodes.
 - The darker the node, the higher the fold change; the lighter the node, the lower the fold change
- Edge color represents differential status between experimental/sample groups (as designated in the [Data Tab](#))
 - As an example, the above screenshot is based on the input file shown earlier in this user manual with Group 1 defined as diabetic and Group 2 defined as non-diabetic. The red edges are present in diabetic individuals, while the blue edges are present in non-diabetics. If there were black edges, those would be present in both diabetic and non-diabetic individuals.
- Use the “+” and “-” buttons at the bottom left of the page to zoom in and out of the network.
- Click on and drag node(s) to reposition them.
- The entire network can be repositioned by clicking in white space and dragging.
- Use the “Previous” and “Next” buttons at the top right of the page to scroll through the various subnetworks generated by the specific Filigree analysis.
- If the network has group nodes, use the “Expand Features” button at the top right of the page to expand all the groups, showing the individual nodes that make up a feature group.
- To expand an individual group node, mouse over the node and click the black and white + sign () that appears.



- Use the “By Sample Group” button at the top right of the page to view a single subnetwork as multiple subnetworks further divided by group (as designated in the [Data Tab](#)). The user can return to the original subnetwork view by clicking the “All Samples” button from this new view (located at the top right).



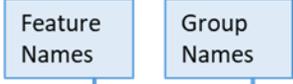
- In the grouped side-by-side networks, node color represents expression level, proportional to group mean.

Note: These side-by-side networks are designed with parallel node layouts to more easily compare them. If the node layout differs between the two networks, resize the window and reload the network. (Layouts were manually altered in above screenshot to better fit on the page).

Output Files

Feature Grouping Output File

Feature aggregation combines metabolites/lipids that belong to the same feature group into one feature, represented by the median intensity across all samples within the group. The feature grouping output file (filigree_feature_groups.csv) is a .csv file with the identical structure as the original feature grouping input file. The below example contains a column listing the features and a column with the associated group names.



Feature	Group
MG_18_4	MG_18_4
MG_18_0	MG_18_0
MG_18_2	MG_18_2
MG_18_1	MG_18_1
DG_36_2	Group_23
PC_38_7	PC_38_7
TG_60_0	TG_60_0

Filigree Output Files

The final output files are .csv files. The files are:

- edgelist.csv: An edgelist containing metabolites with their partial correlation values.
- nodelist.csv: A complete nodelist containing metabolites, fold change, direction, statistics such as p-value and q-value, and subnetwork membership.
- netgsa_results.csv: A table containing the NetGSA² results that include number of nodes/edges, number of differential nodes/edges, p-value, and FDR for each subnetwork.
- consensus_clusters.pdf: Static Images of the subnetworks.

See the Supplemental [Visualizing Complete Network](#) section for guidance on how to generate network visualizations from these files.

Supplemental Information

Feature Grouping Details

- DNEA determines (data-driven) feature groups by performing hierarchical clustering with complete linkage.
- The targeted # of features determines the clustering stop condition. It establishes a bound on how loose clusters/grouped features are allowed to be.
- At each clustering step, sample groups are clustered independently. Specifically:

For sample group l ($l = 1, 2$), each feature is represented by a vector of correlations $v_i = (\sigma_{i1}, \sigma_{i2}, \dots, \sigma_{iN})$ where

- σ_{ik} is the Pearson correlation between intensities for feature i and the intensities for feature k in sample group l .
- N is the number of features in the dataset.

The distance between feature i and feature j is equal to the Manhattan distance between v_i and v_j :

$$\sum_k (|\sigma_{ik} - \sigma_{jk}|)$$

At each step, the feature closest to an existing cluster is merged.

- After clustering each sample group independently, the feature groupings determined for each are recombined by calculating their intersection.
- After combining, the achieved # of features is compared to the targeted number of features.
- Clustering stops when the achieved # of features is less than or equal to the targeted number.

Visualizing Complete Network

Filigree output files can be used in various software tools, such as Cytoscape, to create useful network visualizations.

Using Cytoscape

To create a network in [Cytoscape](#), the input file must have at least 3 columns: source interaction, target interaction, and interaction type. Additional columns, containing edge attribute(s), are optional. The Filigree edgelist output file contains the necessary information and appropriate format for importing into Cytoscape.

1. Make sure Cytoscape is installed on the computer.
2. Launch Cytoscape.
3. Go to File -> Import -> Network from file...
4. The Network File to Load window appears.
5. Navigate to the data file and click "Open."
6. The Import Network From Table window should appear. Notice that the data from the input file shows up in this window.
7. Use the dropdown arrows next to each column header to designate which column is the source () , target () , interaction type () , and attributes (). For example, the Metabolite.A column could be used as the source, the Metabolite.B column could be used as the target, and the edge column could be used as the interaction type. The rest of the columns could be edge attributes.
8. Click "OK."
9. After the data are imported, a network graph appears in the Cytoscape main view.

For more information on using Cytoscape, please see the [Cytoscape Documentation](#).

References

1. Ma J., et al. 2019. Differential network enrichment analysis reveals novel lipid pathways in chronic kidney disease. *Bioinformatics* 35(18): 3441-3452.
2. Ma J., Shojaie A, and Michailidis G. 2016. Network-based pathway enrichment analysis with incomplete network information. *Bioinformatics* 32(20): 3165–3174.