User Guide

Contents

1 Application Options

Application options are available on the left sidebar. It has the following options

- **Home:** Navigate to the home page

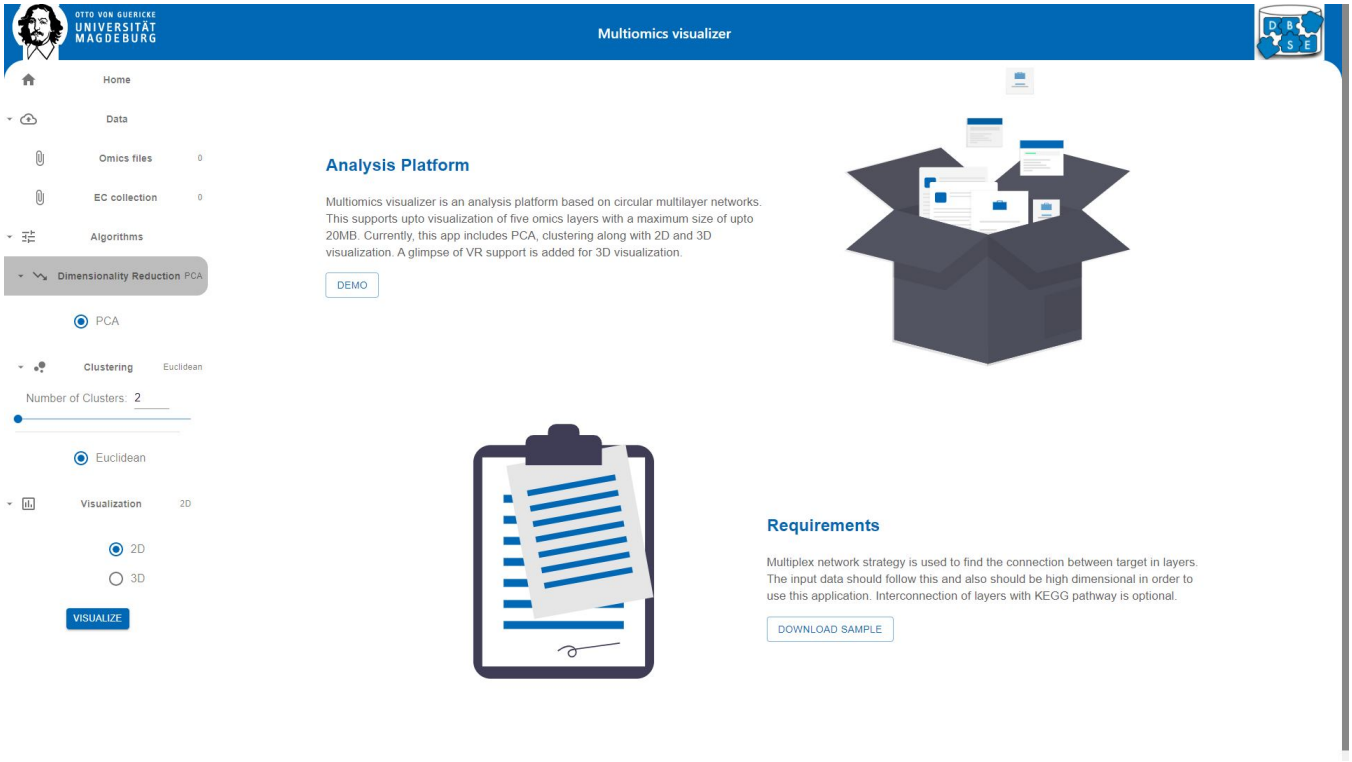

Figure 1: Application options

- **Data:** Contains an uploader where the files can be added. CSV files can only be uploaded.
  - **Omics files:** This is to upload omics files. A maximum of 5 files can be uploaded and each file can be up to only 20MB. The format of data is specified in Section 2. Samples can be downloaded from the homepage.
  - **EC collection:** In this, the EC number collection from the proteome file should be uploaded.
- **Algorithms:** Using this, algorithms used for dimensionality reduction and clustering can be changed.
  - **Dimensionality reduction:** In the current version, only PCA is included in dimensionality reduction.
  - **Clustering:** In this version, only KMeans clustering is included. Users can only change the number of clusters in clustering. The default and minimum value for the number of clusters are 2.
- **Visualization:** This is to change the type of visualization. Various options in 2D and 3D options are discussed further.

2 Data format

2.1 Omics file

In the input data, targets should be in the column and features should be defined in the row. This is applicable for both auto clustering and manual clustering. In manual clustering, the labels should be defined in the next row as shown in Table 2.

Sample datasets can also be downloaded from the homepage.

Feature\Target	Target 1	Target 2	Target 3	Target 4	Target 5	Target 6	Target 7
Feature 1	...	...	...	...	...	...	...
Feature 2	...	...	...	...	...	...	...
Feature 3	...	...	...	...	...	...	...

Table 1: Data format for auto clustering

Feature\Target	Target 1	Target 2	Target 3	Target 4	Target 5	Target 6	Target 7
Clusters	cluster x	3	cluster x	3	3	unidentified	unidentified
Feature 1	...	...	...	...	...	...	...
Feature 2	...	...	...	...	...	...	...
Feature 3	...	...	...	...	...	...	...

Table 2: Data format for user defined clusters

2.2 EC collection

This file consists of Enzyme commission numbers. Ec number position is directly mapped to the features in protein file. This should contain **only EC numbers**, since EC-Compound pairs are obtained from the KEGG database. Sample data can be downloaded from the homepage as well.

EC-Number
EC1; EC2
;
EC1

Table 3: Data format for EC collection

For example, $EC1; EC2$ is mapped to feature 1 and it's abundance, $EC1$ is mapped to feature 3 and it's abundance.

3 Omics file upload panel

With on click of *Omics files* in the sidebar, the omics file uploader will be opened. Here the omics files or any other dataset that satisfies the required file format can be uploaded.

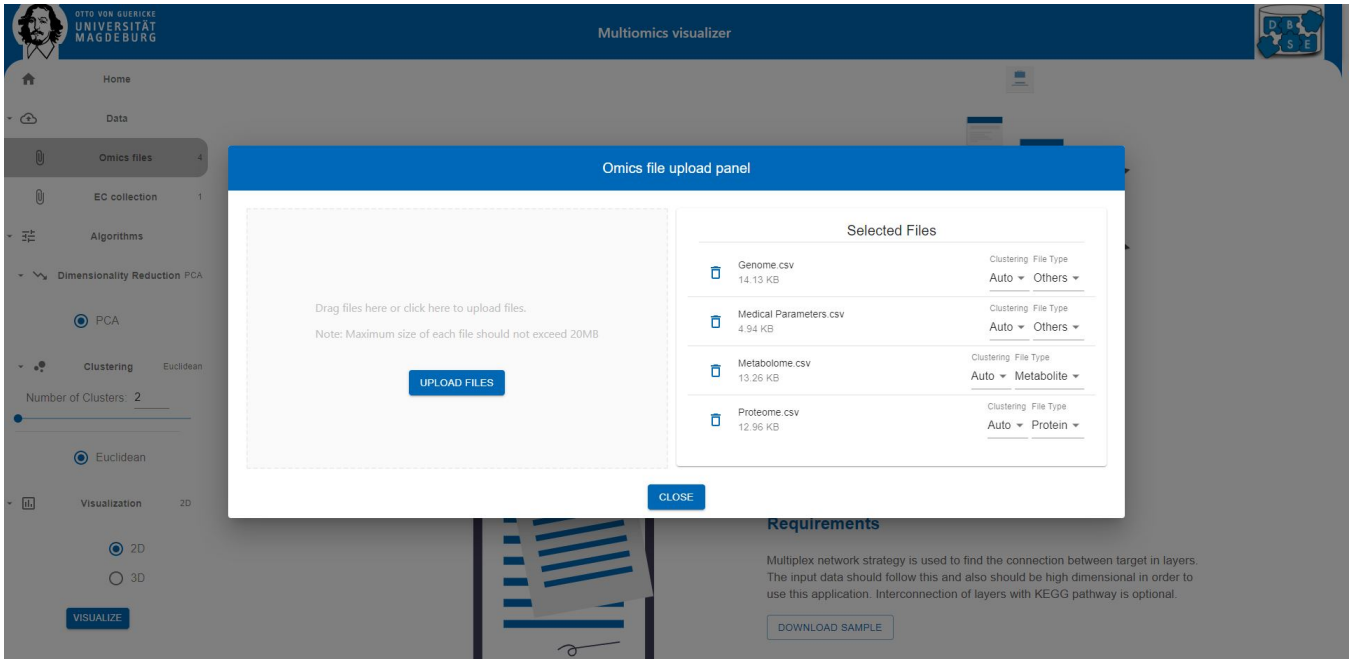

Figure 2: Omics file upload panel

A maximum of 5 files can be uploaded and each file can be up to only 20MB. File types should be mentioned properly. File types have been used to draw the connection between layers.

3.1 Clustering

Users are given a choice to either perform clustering automatically or specify their own clusters. This should be specified from the *clustering* drop-down.

- **Auto:** This is to perform clustering in the backend. In this case, the number of clusters in *clustering option* will be considered.
- **Manual:** This is to consider clustering in the input file. The number of clusters, in this case, will not be considered.

In both the options it is necessary that the input file should be in the required format. Format of the input data is defined in Section 2 and sample for both the cases can be downloaded from the home page of the application.

3.2 File type

There are three types of file types.

- Protein
- Metabolite
- Others

To visualize the interaction of the protein and metabolite file with the KEGG pathway, it is important to specify the file type. File type of the protein and metabolite files should be specified accordingly and the rest of the files will be considered by default as *Others*. **If file types of protein and metabolite are not specified, the connections with KEGG pathway will not be visualized even though EC files are uploaded**. For non-omic use cases, the file type should be *Other*.

4 2D visualization

For each uploaded file, 2D visualization has *Biplot, loading plot, score plot* and *cluster plot*. By default, principal component 1 and principal component 2 are take in x-axis and y-axis respectively.

They can be changed by *X-axis* and *Y-axis* drop-down on top of each plot. Also, the clusters will be available only for principal component 1 and principal component 2. So, whenever the components are changed clusters should be formed and can be performed using the *Form Cluster* button next to the drop-down.

In the loading plot, by default, all the vectors will be displayed. Top n loadings can be viewed based the user selection from the *Show loadings* drop-down. They are available on top of each 2D plot.

In the cluster plot when a point is clicked, the same point across all the files will only be shown. Similarly, when a centroid is clicked, the data points with the same name across all the files will only be shown. To view all points, the user should then click on the empty area in the plot.

5 3D visualization

In 3D visualization, a tooltip with details of the element will be shown on hovering over them. To rotate the 3D visualization, drag the mouse with **mouse left button** or **ctrl + mouse right button**. Options in 3D visualization offer the following.

- **Target labels:** This is to identify all the data points and clusters available in 3D space. On click of each item in the target label, the connection of the data point or all the data points in the cluster across all layers will be shown. If the same point is clicked again, the connection will be removed. It is also possible to search all the data points and clusters inside the target label.
- **EC Number and Compound Number:** These contain all the available EC and compounds in the uploaded file. On click of each item, the EC and compound that are directly connected to the item in the network will be highlighted.
  Note: The highlighted EC and compounds will be turned off if **mouse left button** is pressed on the canvas. In order to view or rotate with highlighted items, **ctrl + mouse right button** should be used.

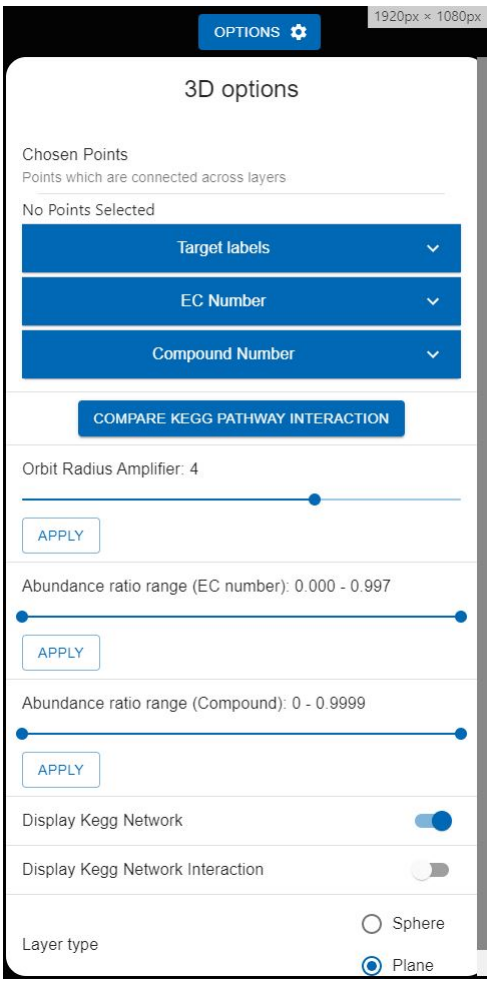

Figure 3: 3D options

- **Compare KEGG pathway interaction:** This is to compare the chosen or selected points in 3D visualization with the KEGG pathway. A pop-up with 2D visualization will be shown. In this, two points from the selected points in 3D can be compared. Also, the abundance ratio of each of them can be varied and compared as shown in Figure 4.
- **Orbit Radius Amplifier:** This is to increase or decrease the distance between the KEGG pathway and orbit. Users can vary the range between 2 to 6 and by default, it's set to 4.
- **Abundance ratio range:** The interaction of protein and metabolite files with the KEGG pathway can be modified by varying the abundance ratio of EC and compound respectively. By default, the abundance ratio range includes the entire values in protein and metabolite files.
- **Display KEGG network:** This is to toggle visibility of the KEGG network. This will be disabled if the EC collection file is not uploaded. By default, the KEGG network is enabled.
- **Display KEGG network interaction:** This is to toggle visibility of protein and metabolite files with the KEGG pathway. By default, this is disabled.
- **Layer type:** The layers in 3D can be visualized either as a plane or sphere. Using this, the layer type can be toggled.
  Note: While toggling between layers, the connection displayed will be removed.

6 VR

3D supports experimental VR support. For now, only limited support is provided. Users can only navigate through the 3D space in the current version. The browser extension, *WebXR API Emulator* can be used to emulate VR devices.

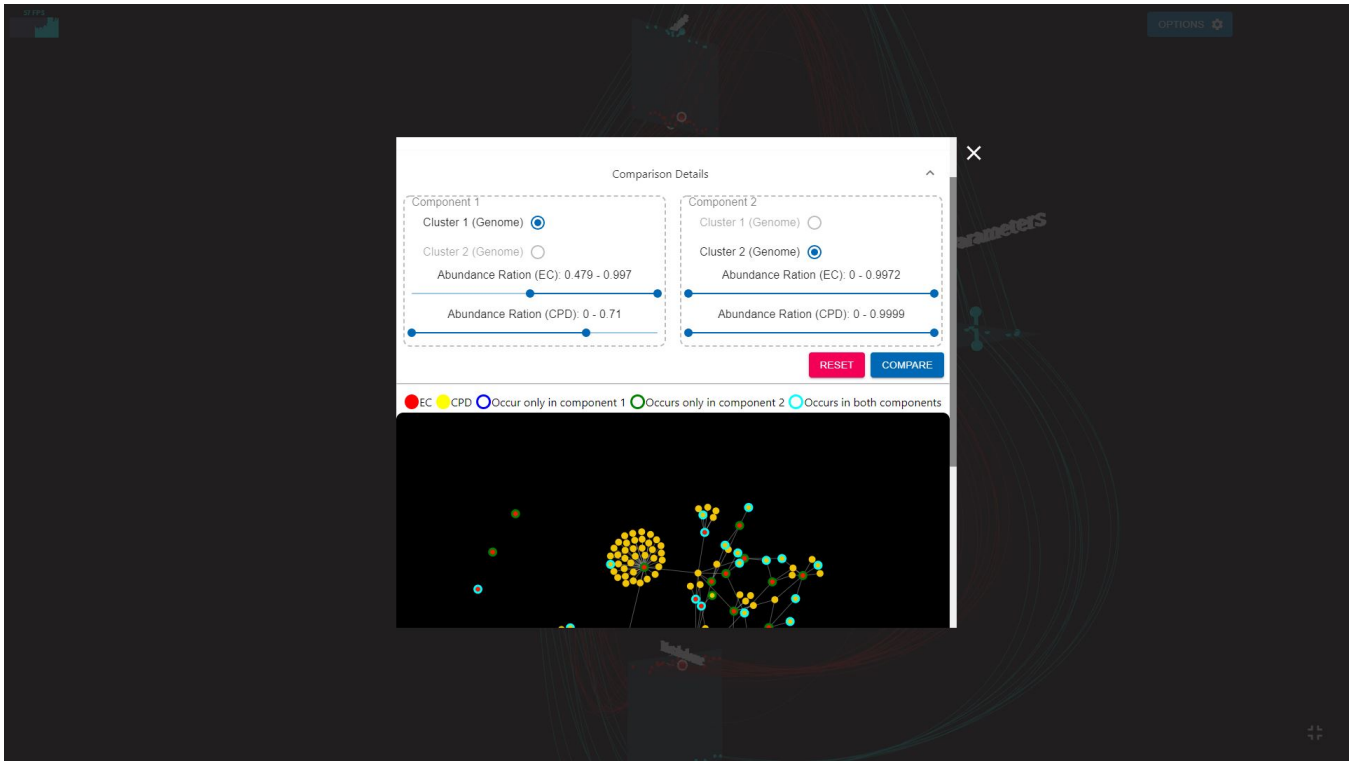

Figure 4: KEGG pathway comparison