
MetGem

Release 1.4.2

Nicolas Elie

May 17, 2024

CONTENTS:

1	User Manual	3
1.1	Getting Started	3
1.2	Toolbars	5
1.3	Network views	7
1.4	Parameters	10
1.5	Metadata Tables	10
1.6	Spectrum View	13
1.7	Import data	15
1.8	Databases	18
1.9	Tools	18
1.10	Settings	21
2	General Concepts	23
3	Tutorials and Howto's	25
4	MetGem's FAQ	27

Welcome to MetGem's documentation page.



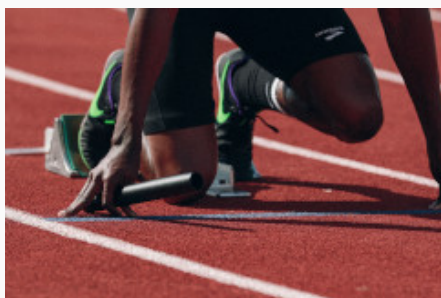
User Manual

Discover MetGem features through an online manual.



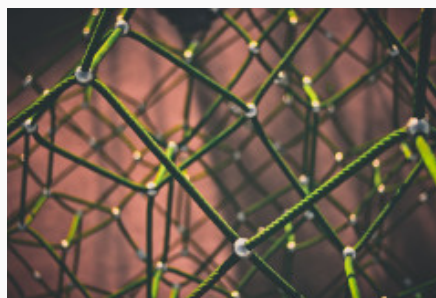
Tutorials and Howto's

Learn through tutorials to see MetGem in action.



Getting Started

New to MetGem and don't know where to start?



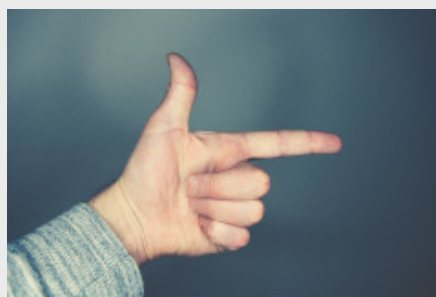
General Concepts

Learn about general concepts that are not specific to MetGem



MetGem's FAQ

Find answers to the most common questions about MetGem.



genindex

An index of the manual for searching terms by browsing.

USER MANUAL

Discover MetGem's features through an online manual.

1.1 Getting Started

Welcome to the MetGem Manual! In this section, we'll try to get you up to speed.

1.1.1 Installation

Windows

Windows users can download MetGem's installer from the [website](#).

Note: MetGem requires Windows 7 or newer.

OS X

You can download the latest binary from our [website](#).

Note: The binaries work only with Mac OSX version 10.15 and newer.

Note: Currently, MetGem lacks a signature for macOS. As a workaround, user can allow MetGem in the macOS Gatekeeper protection by running the following command in the terminal from the Applications folder.

- Download MetGem and click the .dmg installer - Drag and drop MetGem into the Applications folder.
- Open a Terminal and type the following command to tell macOS to trust the installed version of MetGem:

```
sudo xattr -cr /Applications/MetGem/MetGem.app
```

- Approve command with user password.
- Start MetGem.
- If this fails, try the following command (the app will appear in the security preferences and you will be able to choose the "Open anyway" option):

```
sudo xattr -d com.apple.quarantine /Applications/MetGem/MetGem.app
```

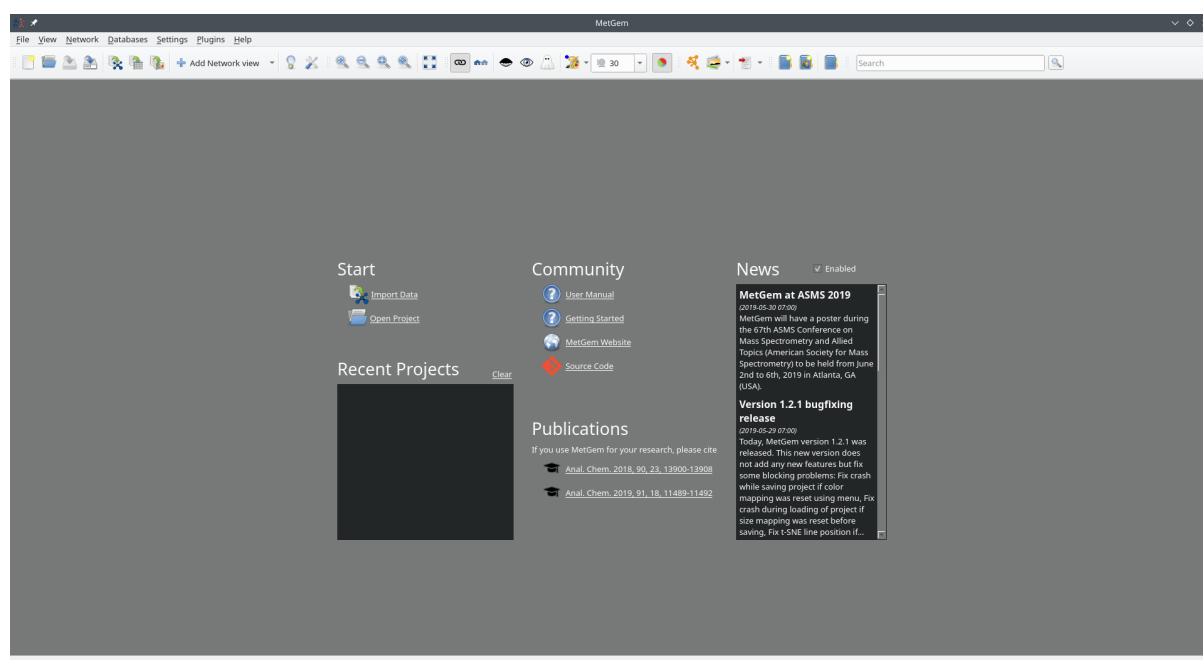
Linux

Linux binaries are available from the [website](#) as an [AppImage](#). It should be as simple as click and run in any distribution.

1.1.2 Starting MetGem

When you start MetGem for the first time there will be no window open by default. You will be greeted by a welcome screen, which will have option to create a new project from MS/MS data or open an existing project.

The first step will be to import data from the **File** > **Import spectra list** menu or by clicking on **Import Data** under start section of the welcome screen. This will open the *import data file* dialog box. If you want to open an existing project, either use **File** > **Open Project...** or drag the image from your computer into MetGem's window.



Saving and opening files

Now, once you have figured out import data, you may want to save it. The save option is in the same place as it is in all other computer programs: the top-menu of **File**, and then **Save**. Select the folder you want to have your drawing, and select the file format you want to use (.mnz is MetGem's default format, and will save everything). And then hit **Save**.

Check out [Global Overview](#) for further basic information, [General Concepts](#) for an introduction on concepts MetGem is built on, or just go out and explore MetGem!

1.1.3 Global Overview

MetGem's interface is divided in four main parts:

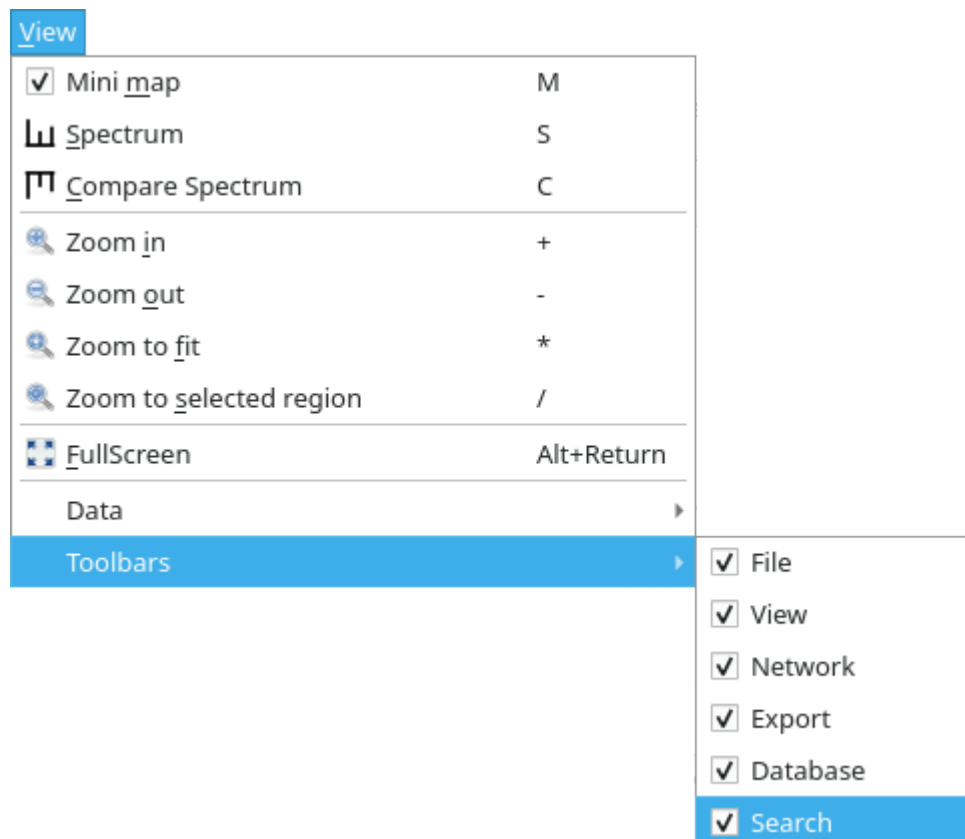
- *Toolbars*,
- *Network views*,
- *Metadata Tables*.
- *Spectrum View*.

Three dialogs are available to *import data*:

- *Import MS/MS Data* dialog,
- *Import Metadata* dialog,
- *Import Group Mappings* dialog.

1.2 Toolbars

Each toolbar can be hidden/shown using the **View >> Toolbars** menu. It can also be moved to a different position in the main window or even detached from the latter by simply dragging it to the new position with it's left handle.



1.2.1 File Toolbar



Usage:

- Create new project, save/open projects to/from file
- Compute network, load metadata table or load group-mappings
- Add Network View
- Show current parameters or change application settings

1.2.2 View Toolbar

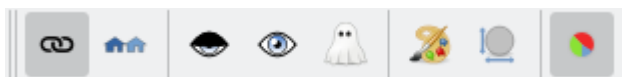


Usage:

- Adjust zoom or switch application to fullscreen

See *Navigation*

1.2.3 Network Toolbar



Usage:

- Link selection between views (See *Selection*)
- Select neighbors of selected nodes (See *Selection*)
- Hide/show nodes (See *Nodes visibility*)
- Hide/show isolated nodes (See *Nodes visibility*)
- Change selected nodes color/size (See *Mappings*)
- Hide/show pie chart on nodes (See *Mappings*)

1.2.4 Export Toolbar



Usage:

- Export networks to *Cytoscape* or as image,
- Export metadata or databases results to text files

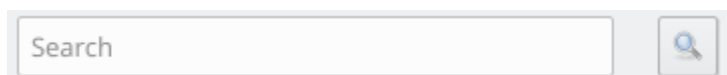
1.2.5 Databases Toolbar



Usage:

- Download, create and explore Databases

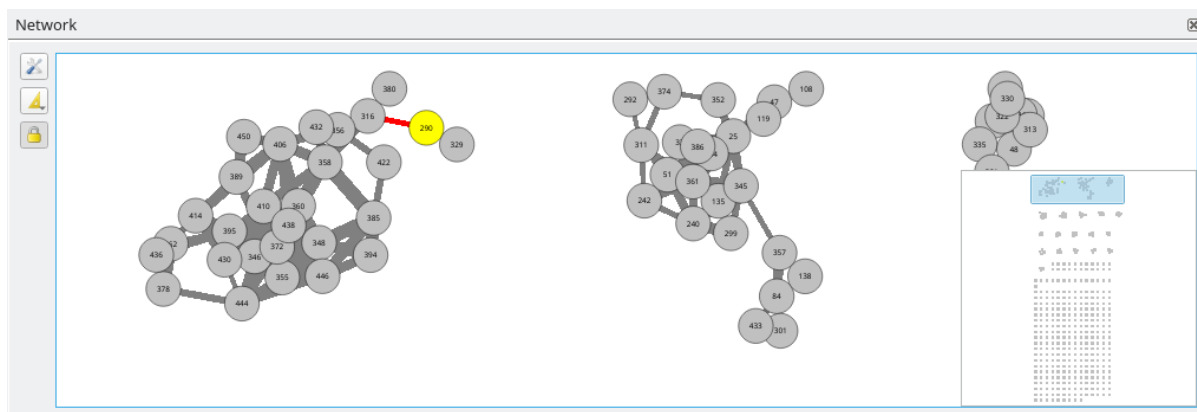
1.2.6 Search Toolbar



Usage:

- Search in nodes or edges tables

1.3 Network views



Data can be visualised using different views. MetGem offers two types of visualisations:

- **Network:** A classical Molecular Network view like what can be obtained by the [GNPS](#) platform. In this view, each node represent an MS/MS spectrum and each edge represent a the distance between two nodes (obtained via a modified [cosine-score](#) calculation). Distance between clusters is arbitrary and has no special meaning.
- **2-D Projections:** A view obtained using a dimension reduction algorithm. This is a 2-D projection of the multidimensional space, so no edge is shown but distance between clusters is informative. To simplify projection, isolated nodes are excluded from the processing and are displayed below the projection. They are arbitrarily distributed and their positions have no special meaning.

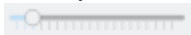
MetGem can use several algorithm to create a visualisation:

- **t-SNE:** [t-SNE](#) (t-distributed Stochastic Neighbourhood Embedding) algorithm tends to preserve local distances and distort global distances. This means that, if two clusters are close to each other in the original space, they have statistically more chance than distant clusters to be close in the [t-SNE](#) projection.
- **U-MAP:** [UMAP](#) (Uniform Manifold Approximation and Projection) is a quite new algorithm (2018) which is very similar to [t-SNE](#) but claims to preserve both local and most of the global structure in the data.
- **MDS:** [MDS](#) (MultiDimensional Scaling) is the algorithm on which [t-SNE](#) and [UMAP](#) are both based. [MDS](#) does not try to preserve local distances over global distances.
- **Isomap:** [Isomap](#) (Isometric mapping) is an extension of the [MDS](#) algorithm based on the spectral theory which tries to preserve the geodesic distances in the lower dimension.
- **PHATE:** [PHATE](#) (Potential of Heat-diffusion for Affinity-based Trajectory Embedding) is a tool for visualizing high dimensional data. [PHATE](#) uses a novel conceptual framework for learning and visualizing the manifold to preserve both local and global distances.

Changed in version 1.3: To make clear that isolated nodes positions are not meaningful in projections, an horizontal line is drawn between projected nodes and isolated nodes.

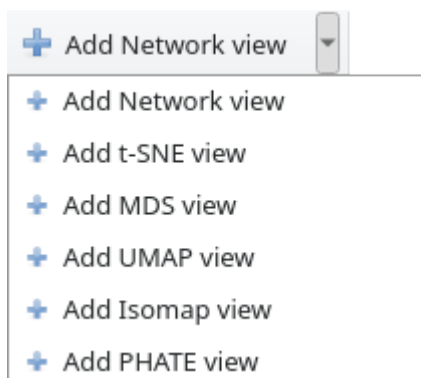
1.3.1 Toolbar



- It is possible to change parameters for each visualisation. The visualisation will automatically be re-computed and updated to match the new parameters.
- If nodes are too close to each other, you can change the scale of the visualisation using the scale that can be found in the dropdown menu 
- This option is only available for the *classical Molecular Network view*. By default, nodes can't be moved to prevent changing their positions by accident. This function let you unlock the view.

1.3.2 Adding Views

You can add view during the *import data* process or later by using the **Add View** menu in the *File Toolbar*.



1.3.3 Interaction

Navigation



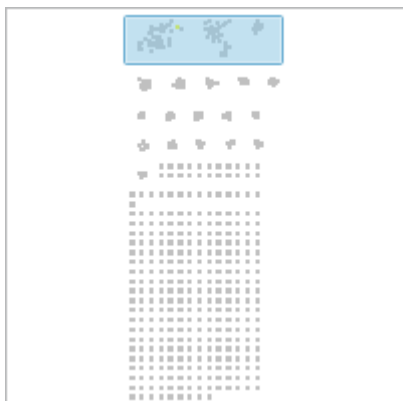
The *View Toolbar* can be used to zoom in or out the *active view*. You can also use the mouse wheel to zoom in or out.

The **Zoom to fit** button will reset the zoom to fit all the visible nodes inside the view.

If at least one node is *selected*, the **Zoom to the selected region** will adjust the zoom level to fit the selected nodes inside the view. This can be helpfull to easily locate nodes selected from the *Nodes* nodes metadata table.

Note: Aspect of nodes and edges depends on the zooming level. At lower level of zoom, nodes text, *pie charts* and edges are not rendered.

Minimap



The minimap, located on bottom-right edge of each visualisation, gives you a overview of the global representation. It can be used to navigate trough the whole visualisation in two different ways when the :

- Left click anywhere in the minimap (outside of the light blue rectangle), the global view will immediately be centered on the selected point.
- Left click anywhere in the light blue rectangle and drag it to another point to move the global view.

The minimap can be hidden by hitting the **M** key on *active view*.

Selection

Selection can be done with left click on a node/edge or by selecting a region with right mouse button. Selected nodes turn yellow while selected edges are highlighted in red. Multiple selections can also be made by holding down the **Ctrl** key while left-clicking the selection.

Another way to select nodes is the **Select neighbors** button in the *Network Toolbar*.

Selecting nodes or edges will automatically filter metadata tables to show only metadata from selected nodes/edges (See *Metadata Tables*).

By default, selection is linked between view, i.e. when a node is selected in a view, the corresponding node is automatically selected in all other views. This is usefull to see correspondances between views. This behavior can be deactivated using the **Link selection between views** button from the *Network Toolbar*.

View MS/MS spectrum

When a node is *selected*, the MS/MS spectrum it represents can be loaded in the Spectrum View (See *Spectrum View*).

Nodes visibility

Selected nodes and edges can be temporarily hidden using the **Hide selected nodes and edges** button from the *Network Toolbar*. Bring them back using the **Show all button and edges** button.

Sometimes you may want to hide isolated nodes because they are not really informative and they use a lot of space in the screen. This is a job for Clyde, our cute little ghost standing on the **Show/hide isolated nodes** button!

Mappings

Nodes metadata can be used to modify how the nodes will look like (See *Mapping section*). It is also possible to bypass these mappings:

- Set the color of the selected node(s). You can use the current color (visible on the top-left corner of the button) or choose another color using dropdown window.
- Adjust size for the selected nodes(s). Select the desired size using the dropdown menu or type it in the text box. Default node size is 30.

If you added *pie charts* to nodes, you might want to temporarily disable them. The `Hide Pie Charts` button will be of great help in this task.

1.3.4 Navigate between views

You have two options to navigate between network views: either click inside a view or use the `Ctrl` + `Tab` shortcut. Current view will have a light blue outline.

1.3.5 Keyboard shortcuts

All these shortcuts apply to the *active view*.

Shortcut	Description
<code>M</code>	Show/hide the <i>Minimap</i>
<code>S</code>	Show the spectrum associated to the selected node in the <i>Spectrum View</i>
<code>C</code>	Compare the spectrum associated to the selected node to another one in the <i>Spectrum View</i>
<code>Ctrl</code> + <code>C</code>	Copy as image the visible part of the active view to the clipboard
<code>Ctrl</code> + <code>Shift</code> + <code>C</code>	Copy as image the full active view to the clipboard

1.4 Parameters

1.4.1 Cosine-score computations

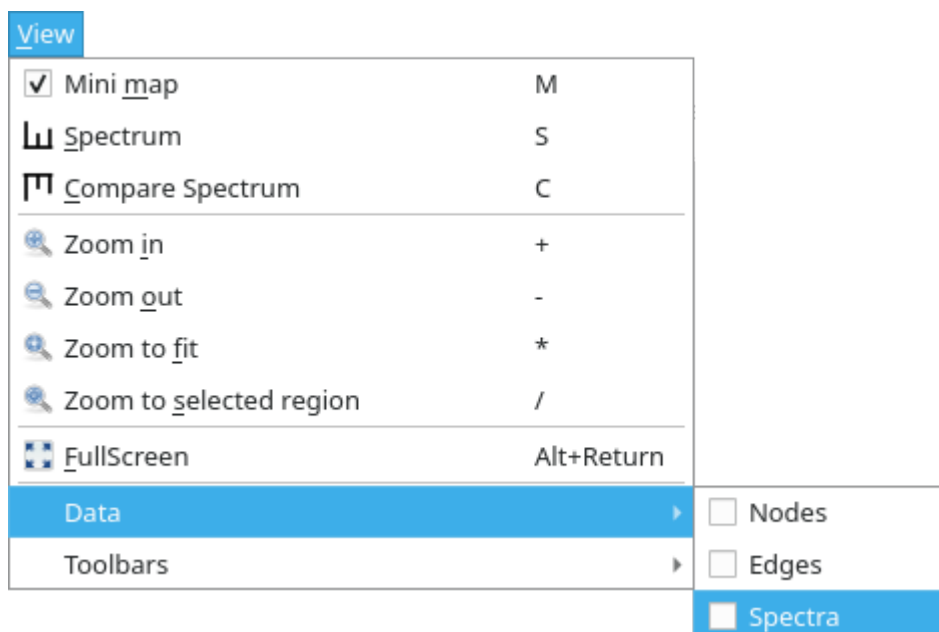
1.4.2 Molecular Networking

1.4.3 t-SNE

1.5 Metadata Tables

Nodes and edges tables will contains metadata. When nodes or edges are selected in a *Network View*, only those nodes/edges will appear in the corresponding tables. Filtering can also be performed using the *search toolbar*.

These tables can be hidden/shown using the `View >> Toolbars` menu. They can also be moved to a different position in the main window or even detached from the latter by simply dragging it to the new position or double-clicking on it's title bar.



Note: Columns can be selected via the right mouse button while the left mouse button is used to change ordering.

	m/z parent
1	474.2491149...

Default ordering can be reset by right clicking in the empty region on the upper left of the table (left side of the first column and above first line).

1.5.1 Nodes

The nodes table list all nodes and their associated metadata. There is a line per node and a column per metadata type.

The first column of this table is always **m/z parent** and contains m/z ratios of ions loaded from the *import data dialog*.

The second one is reserved to **Database search results**. This column will contain a list of standards found in the databases (See *Databases Query*). It is hidden by default and will be visible only if there is at least one results to show.

Nodes Toolbar



The toolbar located on top of the nodes table is divided in three sections.

Mapping section



The first section is dedicated to change how nodes are represented in *views*. Just select one or more metadata column(s), with a right click, then use one these buttons. The dropdown menu include a menu item to undo these changes.

- Use data from the selected column as labels for nodes in the *views*.
- Represents the data in the selected columns in the form of pie charts on the nodes.
- Adjusts each node size based on the data from the selected column.
- Defines color of the nodes from the selected column data.

Nodes section



In this second section, you can interact with nodes and their associated MS/MS spectra:

- Highlight in the *views* the nodes selected in the table. You can then use the **Zoom to the selected region** function (see *View Toolbar*) to easily locate these nodes.
- Views the MS/MS spectrum associated to the selected node (shortcut **S**). You can also compare this spectrum to the one associated to a second node using the **Compare spectrum** function accessible from the dropdown menu (shortcut **C**). See *Spectrum View*.
- You can try to find similar spectra in databases by using this function. See *Databases*.

Note: Functions from this section are also accessible from a context menu that will pop up if you right click in a cell of the nodes table.

Tools section



The last section defines tools that will end up adding columns to the table:

- This tool allows you to create new data columns by combining existing columns using mathematical functions. See *Formulae*.
- This one uses a clustering algorithm (**HDBScan**) to find clusters in the visualisation and colorize nodes according to. See *Cluster*.
- Not really a tool, but you can use it to remove unwanted columns from the table.

1.5.2 Edges

The edges table list all edges and the nodes they link. This table has meaning only for the *classical network view* as the other views don't use edges.

There is a line per edge and the following columns are available:

- Source: The first node that the edge link. Source and target are arbitrarily defined and this has no special meaning since the graph is not directed.
- Target: The second node that the edge link. Same comment than the `Source` column.
- Delta MZ: Difference of m/z ratio between the parent ions associated to the two nodes. Sign is irrelevant.
- Cosine: The similarity score calculated to compare the spectra associated to the two nodes. Value lies between 0 and 1, the higher it is, the closer the spectra are to each other.
- Possible interpretation: Possible interpretation of the `Delta MZ` based on the exact mass.

Edges Toolbar

The toolbar on top the edges table offers a few functions.

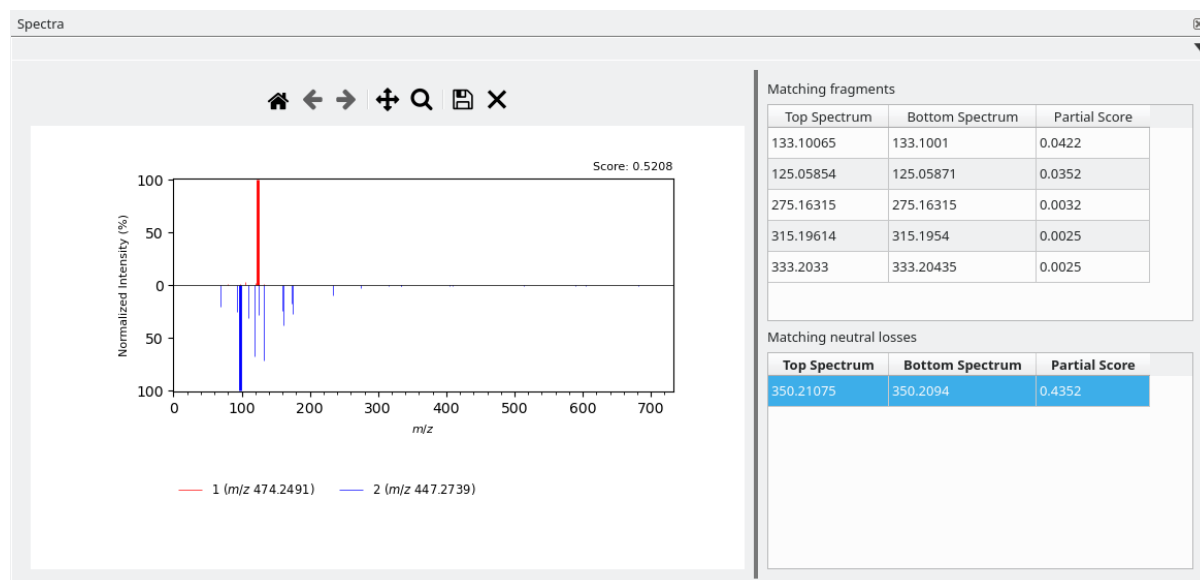


- Highlight in the *views* the edges selected in the table.
- Highlight in the *views* the nodes connected by the edges selected in the table.

Note: These functions are also accessible from a context menu that will pop up if you right click in a cell of the edges table.

1.6 Spectrum View

The spectrum view is used to visualize nodes' spectra.



The left pane shows the *loaded spectra*. You can visualize up to two spectra for comparison. The top spectrum (in red) is the first one loaded while the second one appears upside-down in the bottom part (in blue).

On the top-right corner of the spectrum, you can see the computed similarity score between the two spectra. Value lies between 0 and 1, the higher it is, the closer the spectra are to each other.

On the lowest part of the window, the legend gives information about which spectra is shown (node index and m/z of parent ion).

1.6.1 Matching Peaks List

When two spectra are loaded, the right-hand pane contains information about which fragments or neutral losses are common to both spectra:

- The top table list matching fragments,
- The lower table is for matching neutral losses.

Each table has three columns:

- *Top Spectrum*: m/z of fragment or mass of neutral loss in the first spectrum
- *Bottom Spectrum*: m/z of fragment or mass of neutral loss in the second spectrum matching the one of the first spectrum
- *Partial score*: Contribution of these fragments/losses to the overall matching score. These partial scores add up to the total score.

When a line in this table is selected, the corresponding peaks in the spectra are highlighted. Click anywhere in the table outside any cell to deselect all lines.

1.6.2 Toolbar



Spectrum View come with a navigation toolbar on top of the window, which can be used to navigate through the spectrum. Here is a description of each of the buttons:

- *Home*, *Forward* and *Back* buttons are akin to a web browser's home, forward and back controls. **Forward** and **Back** are used to navigate back and forth between previously defined views. They have no meaning unless you have already navigated somewhere else using the pan and zoom buttons. This is analogous to trying to click **Back** on your web browser before visiting a new page or **Forward** before you have gone back to a page – nothing happens. Home always takes you to the first, default view of your data. Again, all of these buttons should feel very familiar to any user of a web browser.
- The *Pan/Zoom* button has two modes: pan and zoom. Click the toolbar button to activate panning and zooming, then put your mouse somewhere over an axes. Press the left mouse button and hold it to pan the figure, dragging it to a new position. When you release it, the data under the point where you pressed will be moved to the point where you released. Press the right mouse button to zoom, dragging it to a new position. The x axis will be zoomed in proportionately to the rightward movement and zoomed out proportionately to the leftward movement. The point under your mouse when you begin the zoom remains stationary, allowing you to zoom in or out around that point as much as you wish.
- Click the *Zoom-to-rectangle* button to activate this mode. Put your mouse somewhere over an axes and press the left mouse button. Drag the mouse while holding the button to a new location and release. The axes view limits will be zoomed to the rectangle you have defined. There is also an experimental 'zoom out to rectangle' in this mode with the right button, which will place your entire axes in the region defined by the zoom out rectangle. You can also zoom in and out using the mouse wheel.
- Click the *Save* button to launch a file save dialog. You can save as images with the following extensions: jpg, png, ps, eps, svg, pdf, pgf, tif, raw, and rgba. You can also save spectra as text in the following formats: **mgf** and **msh**.
- The *Reset* button will simply unload any previously loaded spectrum.

1.6.3 Load spectra

To load a spectrum, simply select a node in *view* and use the **View >> Spectrum** menu or the **S** shortcut. You can also compare two spectra: select a different node and use the **View >> Compare Spectrum** menu or the **C** shortcut. Second spectrum will appear upside-down. Spectra can also be loaded from the *nodes metadata table*.

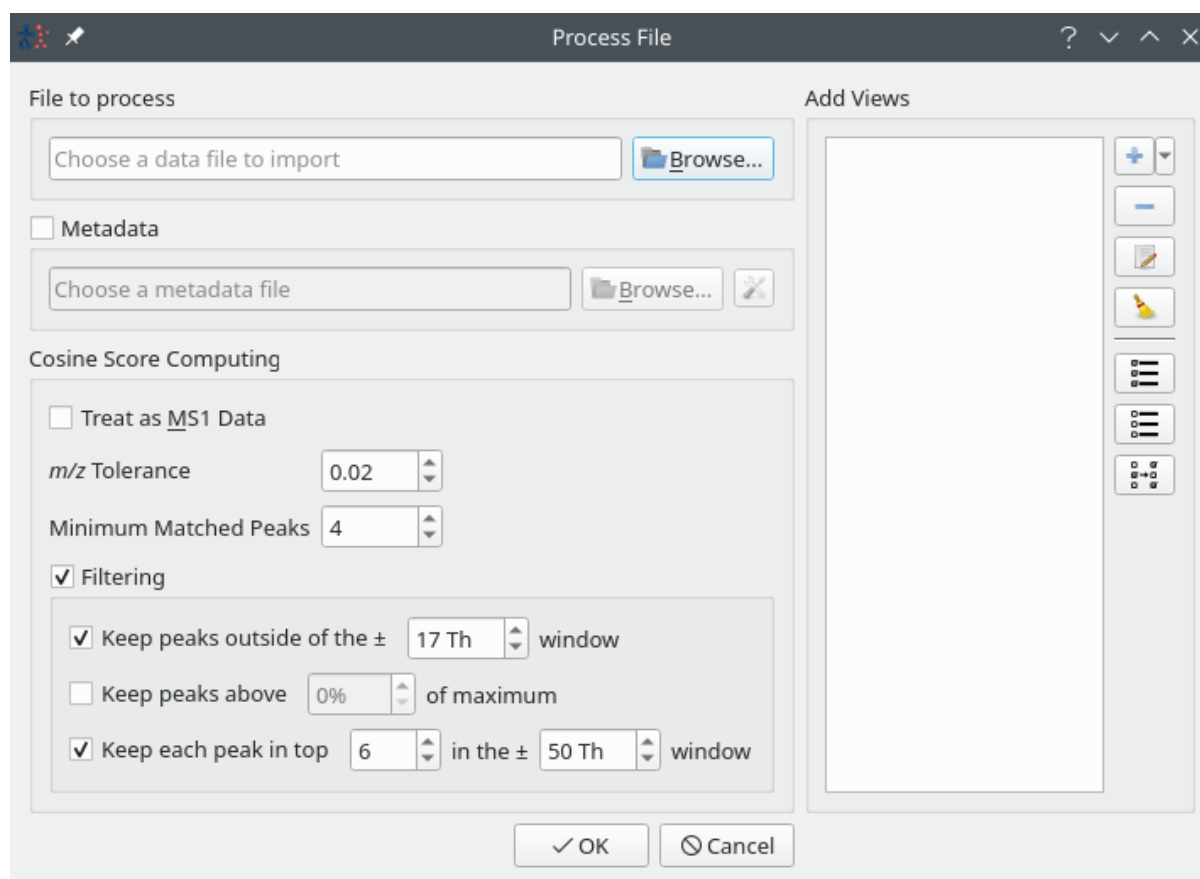
1.6.4 Shortcuts

Shortcut	Description
H , R , Home	<i>Home</i>
Left , C , Backspace	<i>Back</i>
Right , V	<i>Forward</i>
P	<i>Pan/Zoom</i>
Shift	Hold to temporarily activate <i>Pan/Zoom</i>
O	<i>Zoom To Rect</i>
Ctrl	Hold to temporarily activate <i>Zoom To Rect</i>
Ctrl + S	<i>Save</i>
g when mouse is over an axis	Toogle major grids
G when mouse is over an axis	Toogle minor grids

1.7 Import data

1.7.1 Data

Networks can be created from a data file (in *mgf* or *msp* format) using the **Import Data** icon. This will open the following dialog:



This dialog lets you open a data file and, optionally, a metadata table (*csv* file) using the **Browse...** buttons. Separator for the *csv* file should be auto-detected but you can change this parameter directly in this dialog and more parameters are available via the **Options...** button. See [Metadata](#).

Parameters used by MetGem for the cosine computations step can be tuned in the **Cosine Score Computing** section. See [Cosine-score computations](#).

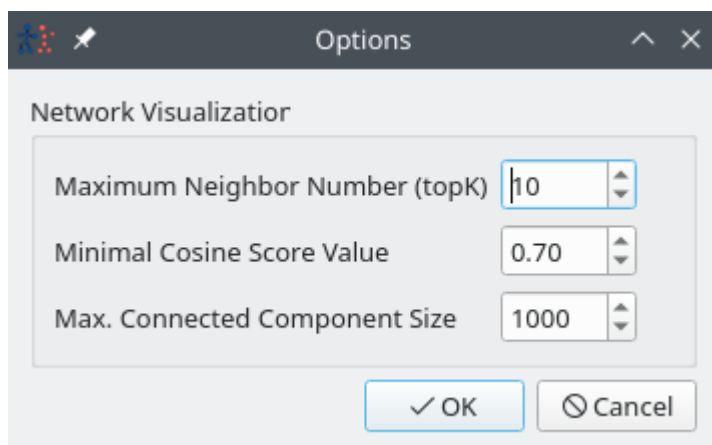
In the **Add Views** section, you can optionally add visualisations like *Molecular Network* or *2-D projections*. See [Add visualisations](#).

When you have loaded an data file and you are satisfied with the parameters, you can click **OK** to start the process.

Add visualisations

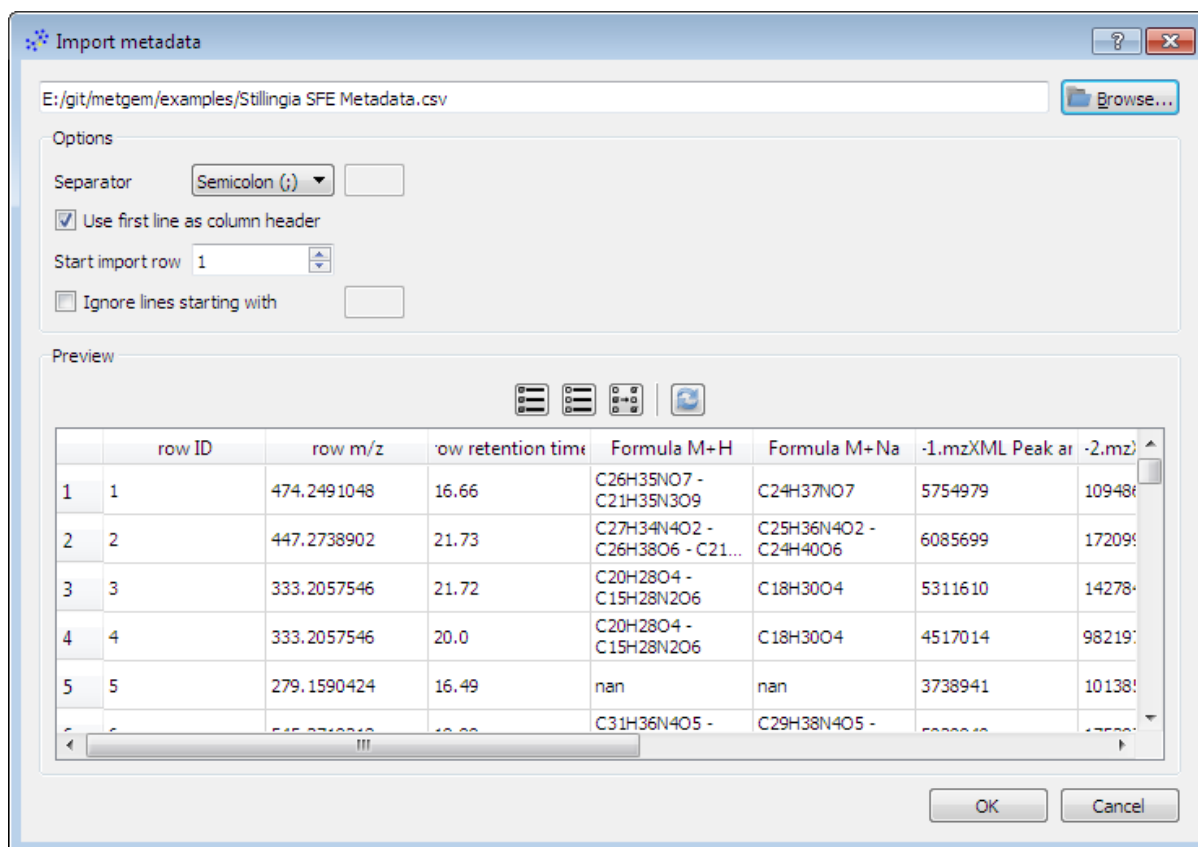
MetGem offers two types of views that you can add to your project. See [Network views](#).

To add a view, use the dropdown menu next to the button. Choose the desired visualisation and a dialog will open to let you set a few parameters. See [Parameters](#) for more informations about the parameters.



1.7.2 Metadata

You can associate metadata to the spectra loaded during *Data Import data* step. You can load these metadata from a *csv* or from a *spreadsheet* file (like *LibreOffice Calc* or *Microsoft Excel*) or want to load new metadata, you can do so using the tool button from the *File Toolbar*. The following dialog will pop-up:



Metadata file (*csv*) can be selected using the **Browse...** button. Separator for *csv* file should be auto-detected but can this be changed in the *Options* section. More parameters like whether the file contains headers or not can be also be tuned in this section.

You can see the first 100 lines that will be imported in the *Preview* section. You can select which column to import by clicking on the corresponding headers or via the upper toolbar. If no column is selected, all columns will be imported.

The button can be used to reload file from disk using the parameters defined in the *Options* section.

1.7.3 Group Mappings

Group mappings file can be used to group columns and sum values they contains. You can load such a file via the tool button.

Group mapping files are simple text files that should follow the following scheme:

```
GROUP_group1=filename1.mzXML
GROUP_group2=filename2.mzXML;filename3.mzXML
```

The example below can be translated as

Create a group named *group1* containing columns *filename1.mzXML* and a group *group1* containing columns *filename2.mzXML* and *filename3.mzXML*

If a column does not exists, it is simply ignored. Groups can be empty. Group columns are identified with the icon.

1.8 Databases

1.8.1 Download

1.8.2 User databases

1.8.3 Explore

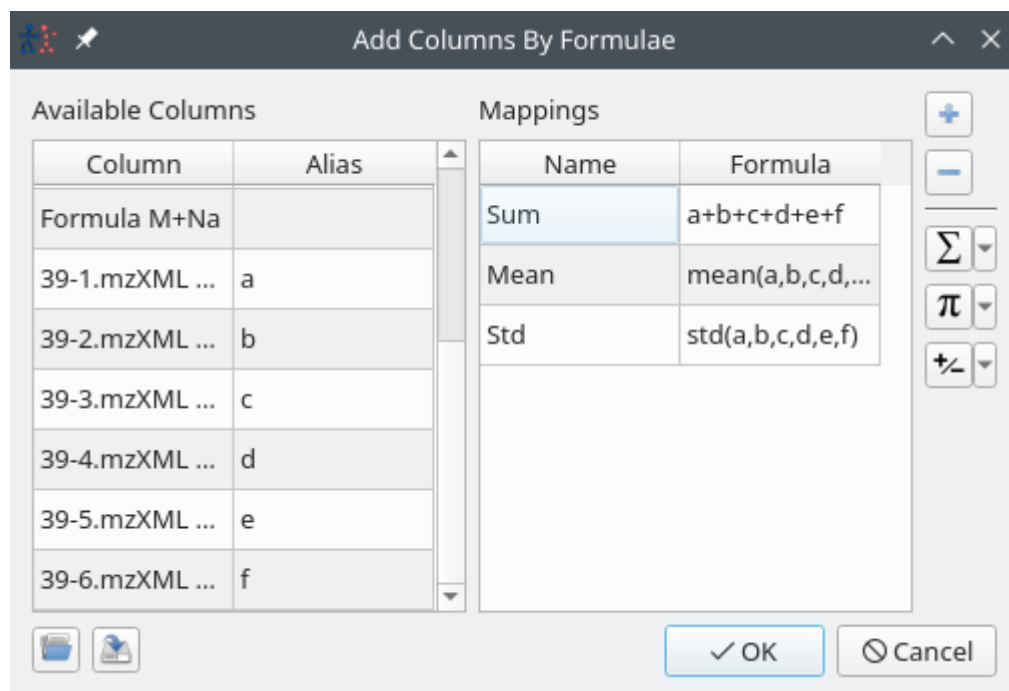
1.8.4 Query

1.9 Tools

MetGem includes a few tools to manipulate data.

1.9.1 Cluster

1.9.2 Formulae




This tool is designed to allow you to combine multiple columns using simple or complex formulae. See [Usage](#) for practical information and [Syntax](#) to get a description of the formulae's syntax.

Usage

The window is divided in two parts:





- *Available columns*: on the left lies a table of the metadata columns used to create aliases for metadata column titles,
- *Mappings*: on the right, you can create new columns by combining existing ones.

Creating a new formula will then need two steps:

1. To be used in a formula, a column needs an alias (that needs to be a valid [Python identifier](#)). You just need to double-click inside a cell of the *Alias* column to define an alias for the corresponding column. Validate by hitting **Enter**.
2. Last step is to add a new column by clicking on the , set its name (a column with this name will be overwritten if it already exists) and define a formula that includes constants and/or alias set in step before.

Toolbar

The toolbar located on top right of the dialog includes a few buttons:

-  *Add new formula*: Add a new empty line in the *Mappings* table. It's up to you to fill *Name* and *Formula* cells,
-  *Remove selected formulae*: Remove the selected rows on the *Mappings* table.
-  *Add Function*: A drop-down sectional list of available functions. A function is used by adding a comma-separated list of arguments in parentheses after it's name, e.g. `mean(a,b)`.
-  *Add Constant*: A list of available constants. A constant is just a replacement for

Syntax

The syntax used to describe formulae is a subset of [Python](#) programming language. As MetGem use the [Pandas](#) library internally, the following operations are supported:

- Arithmetic operations except for the left shift (`<<`) and right shift (`>>`) operators, e.g., `x + 2 * pi / y ** 4 % 42 - pi`
- Comparison operations, including chained comparisons, e.g., `2 < x < y`
- Boolean operations, e.g., `x < y` and `x < y or not column1`
- List and tuple literals, e.g., `[1, 2]` or `(1, 2)`
- Math functions: `sum`, `mean`, `median`, `prod`, `std`, `var`, `quantile`, `min`, `max`, `sin`, `cos`, `exp`, `log`, `expm1`, `log1p`, `sqrt`, `sinh`, `cosh`, `tanh`, `arcsin`, `arccos`, `arctan`, `arccosh`, `arcsinh`, `arctanh`, `abs`, `arctan2` and `log10`
- Constants: `pi`, `e`

This Python syntax is not allowed:

- **Expressions:**
 - Function calls other than math functions.
 - `is/is not` operations
 - `if` expressions
 - `lambda` expressions
 - `list/set/dict` comprehensions
 - Literal dict and set expressions
 - `yield` expressions
 - Generator expressions
 - Boolean expressions consisting of only scalar values
 - Attribute access, e.g., `df.a`
 - Subscript expressions, e.g., `df[0]`
- **Statements**
 - Neither simple nor compound statements are allowed. This includes things like `for`, `while`, and `if`.

Note: A Python identifier is a name used to identify a variable, function, class, module or other object. An identifier starts with a letter A to Z or a to z or an underscore (_) followed by zero or more letters, underscores and digits (0 to 9).

Python does not allow punctuation characters such as @, \$, and % within identifiers. Python is a case sensitive programming language. Thus, *Manpower* and *manpower* are two different identifiers in Python.

1.10 Settings

1.10.1 Theme

1.10.2 Edges

GENERAL CONCEPTS

TUTORIALS AND HOWTO'S

METGEM'S FAQ