PIDtoSIFtoGraph Plugin: User Guide

This document explains how to install and use the Cytoscape plugin PIDtoSIFtoGraph.

# Introduction

PIDtoSIFtoGraph is used to import an XML file of a pathway downloaded from PID, convert it into SIF format and subsequently visualize it in Cytoscape. A visual style is applied, which gives the user a view of the network graph consistent with those of biological graphs.

After this, the user has the option to filter out irrelevant nodes with control and experiment microarray data of Affymetrix© barcode format, or from an Illumina© file. A new graph will be generated and shown to the user.

It is also possible to create a shortest path graph between two or more nodes. The user can select source and target nodes by input of text files containing node IDs or by direct text input of the Cytoscape node IDs. A new network graph based on the shortest path algorithm between the source and target nodes will be created and shown to the user. The source and target nodes would be highlighted to differentiate them from the nodes connecting them.

# Installing the plugin

PIDtoSIFtoGraph is based on Java technology and therefore is compatible for the most common operative systems: Linux, Windows and Mac OS.

## Cytoscape

The plugin is running with Cytoscape 2.8.2.

If you do not already have this version of Cytoscape, download the wizard of Cytoscape on the website: <http://www.cytoscape.org/download.php>, and launch it. Then follow the instructions to install it.

## Plugin

Before running the plugin, you need to put the file PIDtoSIFtoGraph.jar in the /plugins folder of Cytoscape 2.8.2.

The following files have to be copied into the same directory as cytoscape.exe, cytoscape.jar or cytoscape.sh where you installed the software Cytoscape (e.g. *C:\Program Files\Cytoscape\_v2.8.2*):

* netView.props
* AFFY-44.adf.txt : this file maps the AffymetrixID to the GeneID
* UPtoGeneIDFULL.txt : this file maps the UniprotID to the GeneID
* vizmap.props

These files are used by the plugin. They can be found in the following link:

<https://www.dropbox.com/sh/fy5x1na11fcort8/EvV-hT2zfx>

The other files in the link are also needed: PIDtoSIFtoGraph needs the Jung library (Java Universal Network Graph). Copy the 17 files of the Jung directory in the /plugins directory of Cytoscape.

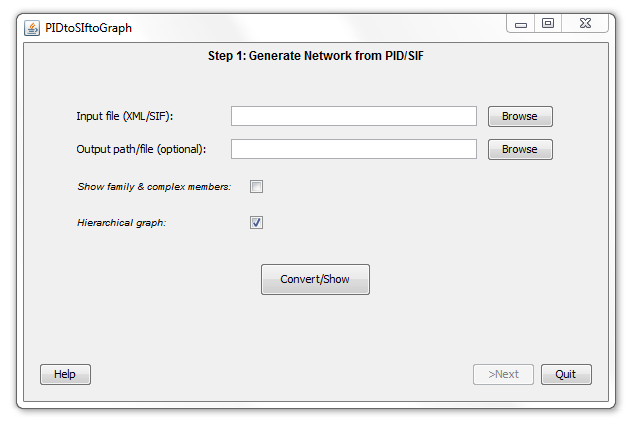
# Using PIDtoSIFtoGraph

In Cytoscape, select Plugins > PIDtoSIFtoGraph

## Step 1:

The first step consists in importing an XML file of a pathway interaction database, or a SIF file, graph of Cytoscape.

The Plugin works for well-formed XML containing PID-NCI information. It can also load the last downloadable version from PID-NCI, however the content of this file in the PID website is also related to Reactome and Biocarta databases. Also, the generated graph may contain redundancies, because of existing redundancies/inconsistencies in the xml file.



Fields/Buttons:

* **Input file (XML/SIF):**

You can use the "browse" button to find the file you want, or copy its path in the text field.

Every "browse" button opens a window similar to the following one:



* **output path/file (optional):**

Use this part to specify where you want the created files to be. You can choose a path or a file (e.g. if you want to overwrite an existing file). You can use the "browse" button or directly copy the path. If nothing is put, the output path is set to the same path of the input file.

* **Show family and complex members:**

There are different types of molecules in the graph:  
1) small molecules: ATP, GTP, which are biochemical molecules for which we don't have much information.

2) proteins: molecules with identifiers (EntrezGene and/or Uniprot), in standard data bases in molecular biology.

3) complexes: sets of proteins regrouped with ":" in the PID file. A complex depends on its member and can't exist without them.

4) family: sets of proteins regrouped by "::". A family needs only one member to exist.

When the checkbox is not checked, the graph shows the complexes and families without showing their members. When "Show family and complex members" is checked, you can see the member of these complexes and families.

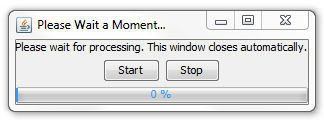
* **Hierarchical graph:**

A hierarchical graph puts the nodes without predecessors on the top of the screen, and link them to the others with interactions that are always going down. This is usually better to have a hierarchical graph, but when the graph is too big, it might take too much time to calculate it. This option is not activated by default.

* **Convert/Show:**

When you have selected a file, you can click on this button to start the conversion (if it is an XML file) and load the graph.

Actually, a small window opens. Click on "start" to start the process. You can stop the process by clicking on "Stop". A progress bar shows the progression of the process. This small window closes when the process ends.



The progress bar only gives a general indication. It is possible that the process is not finished yet but the progress bar indicates 100%. The progress should shortly end though.

* **Help:**

Show a help window.

* **>Next:**

Go to the next step.

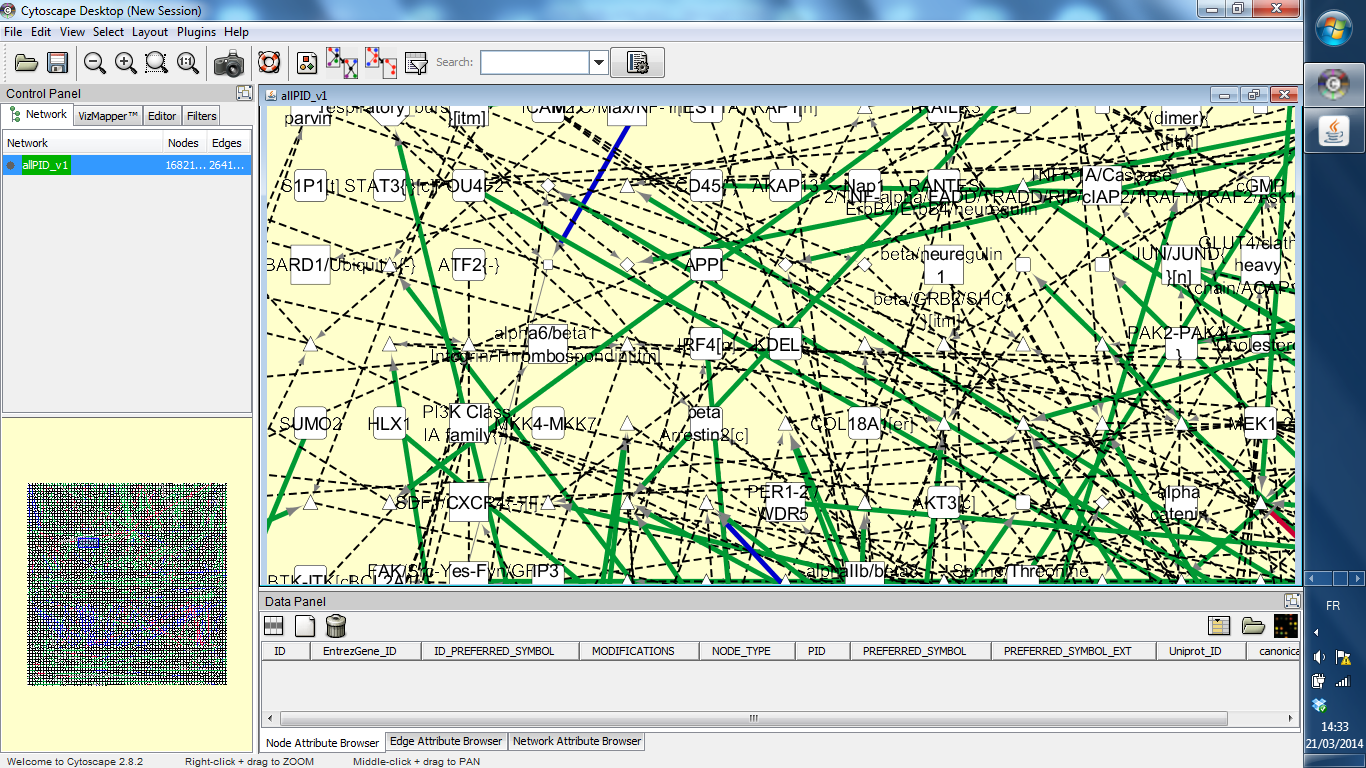
This button is disabled until a graph is created with "Convert/Show".

* **Quit:**

Quit the plugin.

N.B.: when the window of the plugin is closed, or when you quit the plugin, the data associated to the plugin are still present. You can use the Cytoscape menu again to open the plugin, and it is in the state you left it.

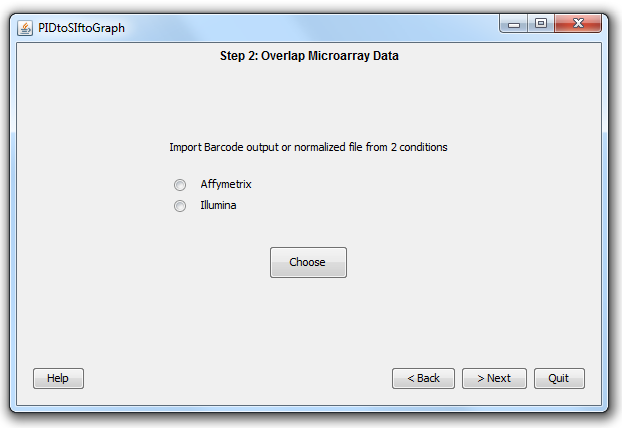
An example of graph showed by Cytoscape at the end of this step is:



## Step 2:

The step 2 is used to filter the graph loaded in step 1, based on Affymetrix barcode (<http://barcode.luhs.org>) files or Illumina files. We use these experiment files to find out if the gene associated to a molecule is expressed or not during an experiment. The corresponding nodes of the absent molecules are removed from the graph. This step is optional.

### Main window



Fields/Buttons:

* **Affymetrix/Illumina:**

Before choosing the file(s) used to filter the graph, you have to select the type of file : Affymetrix or Illumina.

* **Choose:**

Start a new window to select the 2 files needed for Affymetrix, and the file and conditions needed for Illumina.

* **Help:**

Show a help window.

* **< Back:**

Go back to step 1.

* **> Next:**

Go to step 3: Subgraph extraction.

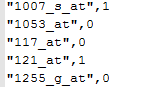
* **Quit:**

Quit the plugin.

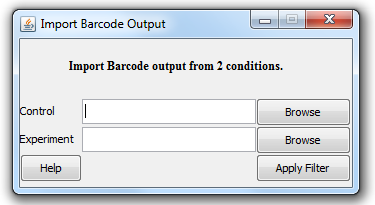
N.B. : see "quit" section of step 1.

### Affymetrix Window

When using Affymetrix files, there is an experiment done in relation to a control. 2 barcodes (<http://barcode.luhs.org>) are obtained, one for the control and the other for the experiment. These files consists in a list of lines : "Affymetrix identifier" , 0 or 1.



If both control and experiment files have "0" (not-expressed) for a given Affymetrix ID, the corresponding node is removed from the graph. Moreover, if one of the molecules of a complex is missing, the complex is removed, and if all the molecules of a family are absent, the family is removed. The files can have any filename extension.



Fields/Buttons:

* **Control:**

Put the path of to the control file, or use the browse button to find it.

* **Experiment:**

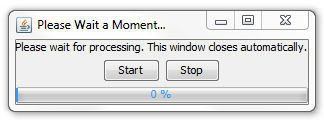
Put the path to the experiment file, or use the browse button to find it.

* **Help:**

Show a help window.

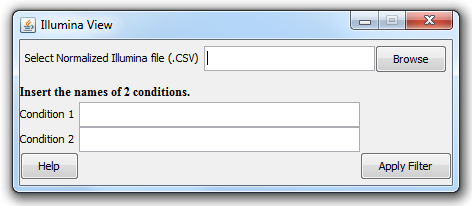
* **Apply Filter:**

When a control and an experiment, different from the control, are correctly selected, this button starts a new window :

****

This window is the same as for step 1.

### Illumina Window



Fields/Buttons:

* **Select Normalized Illumina file (.CSV):**

The Illumina file should contain a column named "EntrezGeneID"

An Illumina condition is described by many columns: experiment.mean, experiment.sd, experiment.p, experiment.nbeads

An example of a extract of a normalized Illumina file with 2 experiments can be found in /Examples/Illumina of the Dropbox folder.

The molecules are removed if for condition 1 and condition 2 (or for every condition if these 2 fields are empty), the mean value is higher than 150 and the p-value is lower than 0.01.

* **Condition 1 / Condition 2:**

You can put one or two names of experiments in these text fields. If the experiment appears as " "myExperiment.means" " in the Illumina file, put "myExperiment".

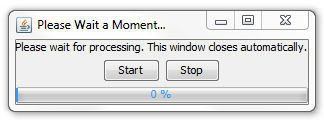
If no condition is put, every conditions of the file are taken into account.

* **Help:**

Show a help window.

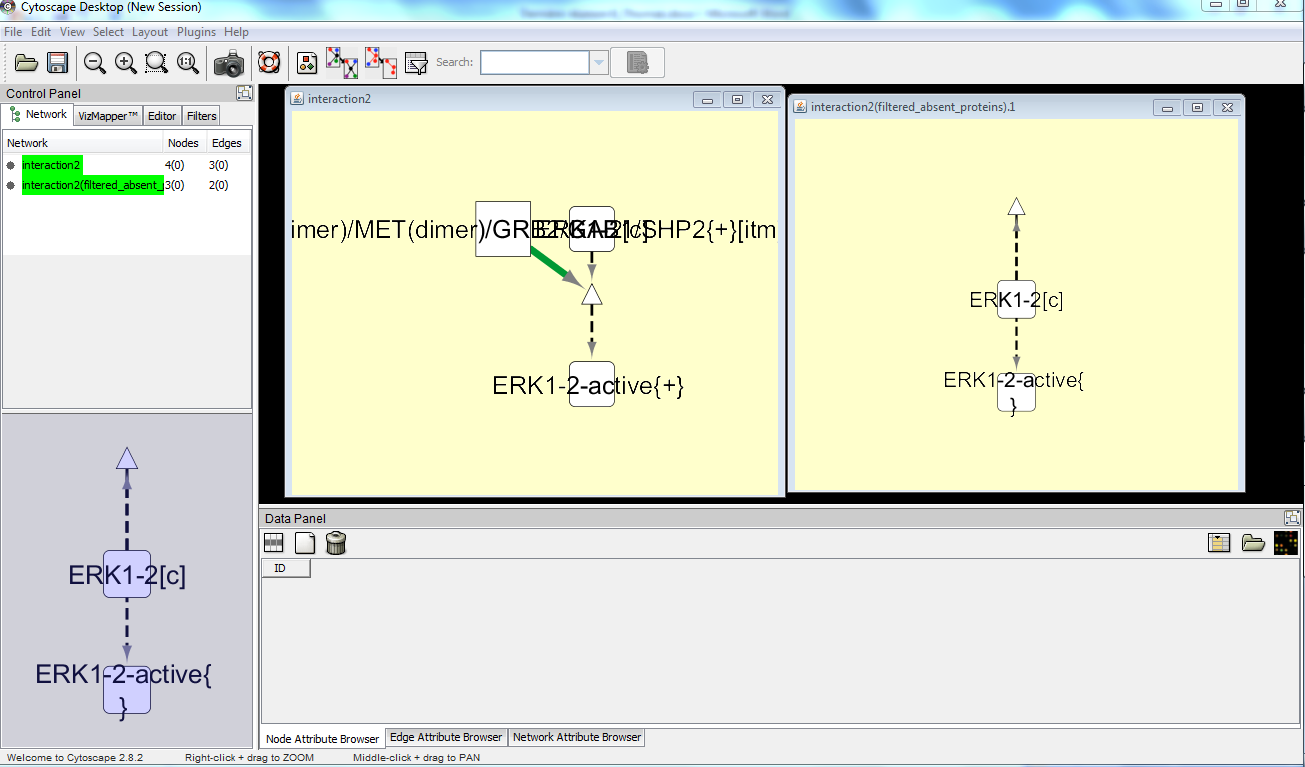
* **Apply Filter:**

When a control an Illumina file is correctly selected, this button starts a new window:

****

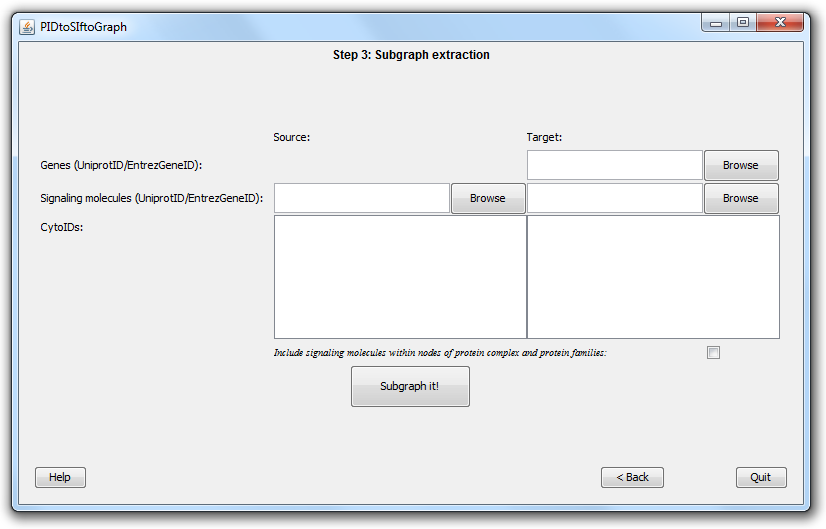
This window is the same as for step 1.

An example of graph showed by Cytoscape at the end of this step is:



## Step 3:

This step allows you to create a subgraph of the graph loaded in step 1. This subgraph is the shortest path between source molecules and target molecules. The top of the window is a table with two columns (Source and Target) and three rows (Gene, Signaling molecules and CytoIDs). This step is also optional.



Fields/Buttons

* **Source /Target:**

For the subgraph, you have to choose sources and targets. There are different ways to choose the molecules, and you can use simultaneously these ways.

* **Genes (UniprotID/EntrezGeneID):**

This area is used to import molecules IDs (Uniprot IDs or EntrezGene IDs) from a file. This fie contains a list of molecules IDs, with one ID per line.

A gene is a molecule that has a node of type “transcription” as predecessor. This entry ignores the given IDs that are not genes.

**Warning : issues can appear if there is an empty line between the IDs, especially if there is an empty line at the beginning of the file.**

* **Signaling molecules (UniprotID/EntrezGeneID):**

A signaling molecule, for the plugin, is actually any type of molecule. The type of file you can use is similar to the one of the "Genes" row.

* **CytoIDS:**

With this text field, you can copy/paste any cyto ID of the loaded graph. You have to put one ID per line.

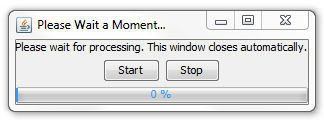
* **Include signaling molecules within nodes of protein complex and protein families:**

Check this checkbox to look for the IDs within compound molecules i.e. proteins-complexes or protein-families.

* **Subgraph it!:**

To start the subgraphing process.

It opens this window:

****

This window is the same as for step 1.

* **Help:**

Show a help window.

* **< Back:**

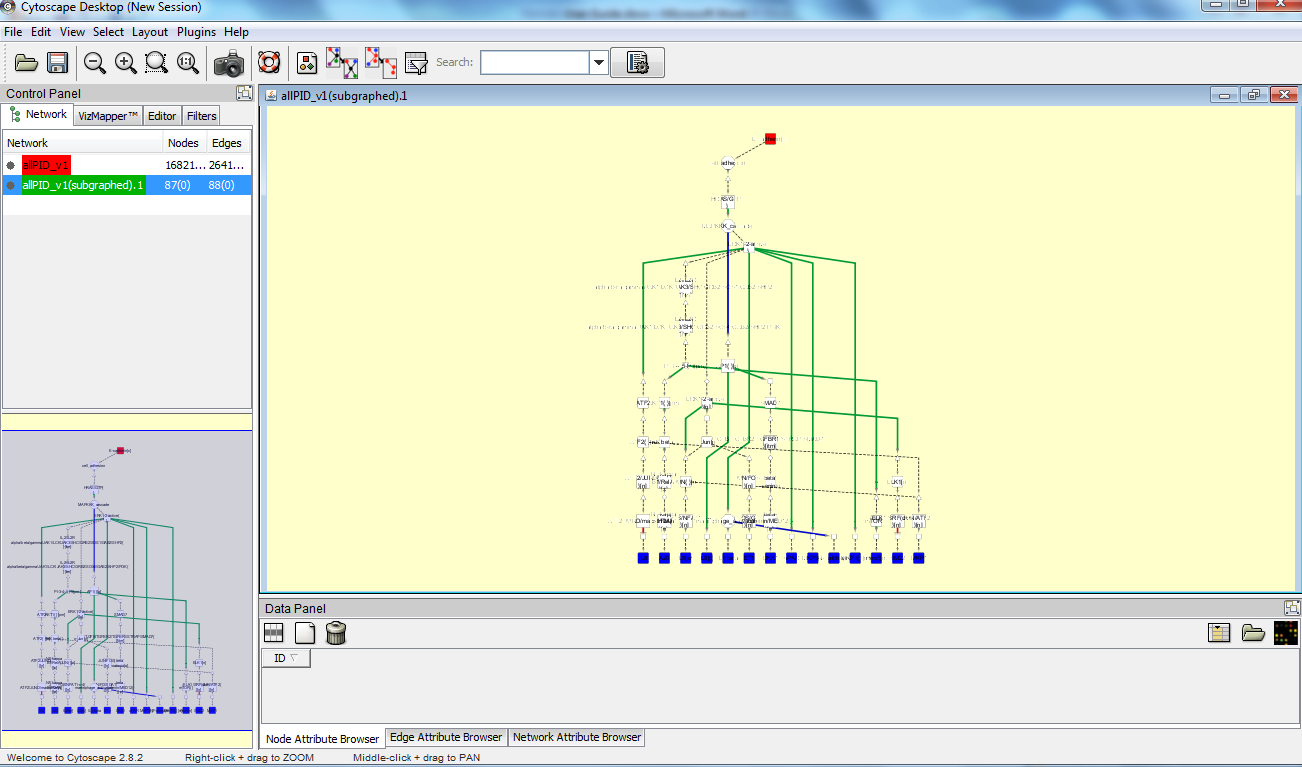
Go back to step 2.

* **Quit:**

Quit the plugin.

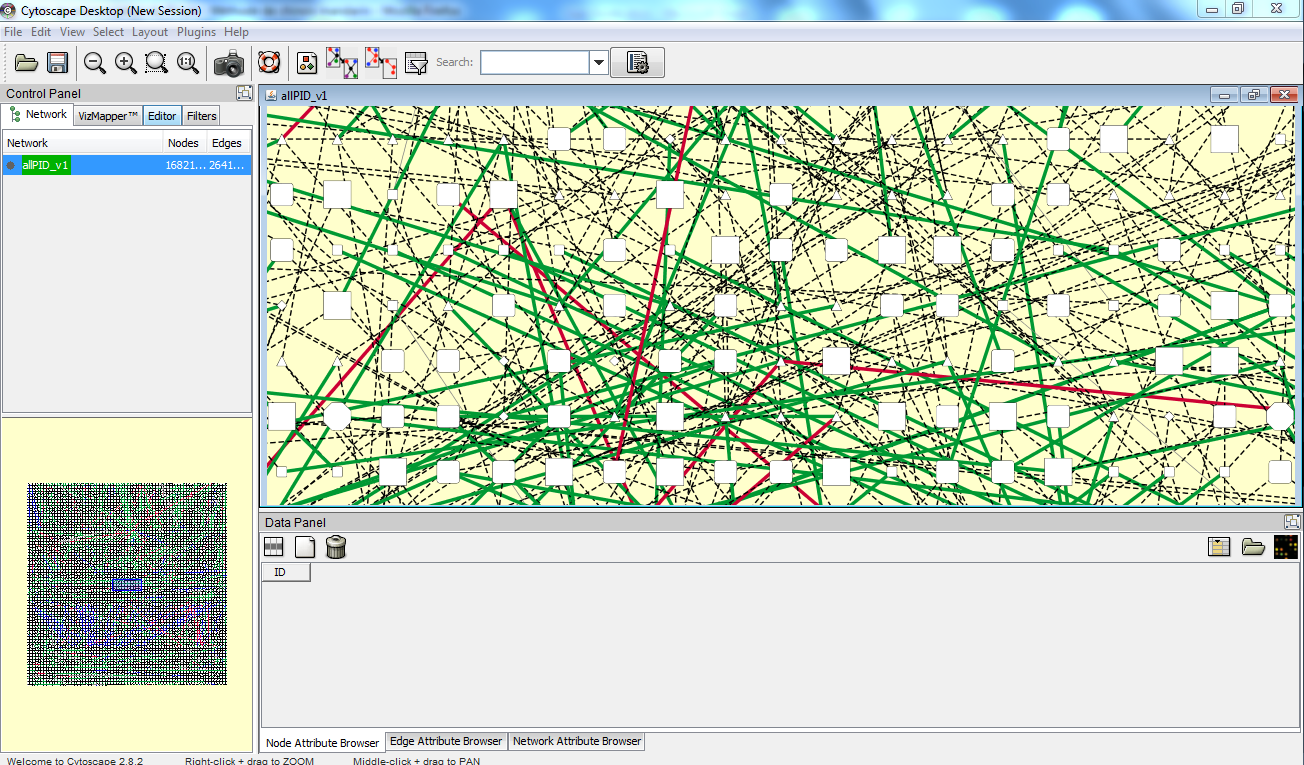
N.B. : see "Quit" section of step 1.

An example of graph showed by Cytoscape at the end of this step is:



## Visualisation of the graph:

When a network is loaded, Cytoscape will look like this :



To see the attributes of a node, click on the icon "Select attributes" on the top of the data panel: .

For more information about how to use Cytoscape, you can refer to the User Manual, especially the "Quick Tour of Cytoscape" paragraph:

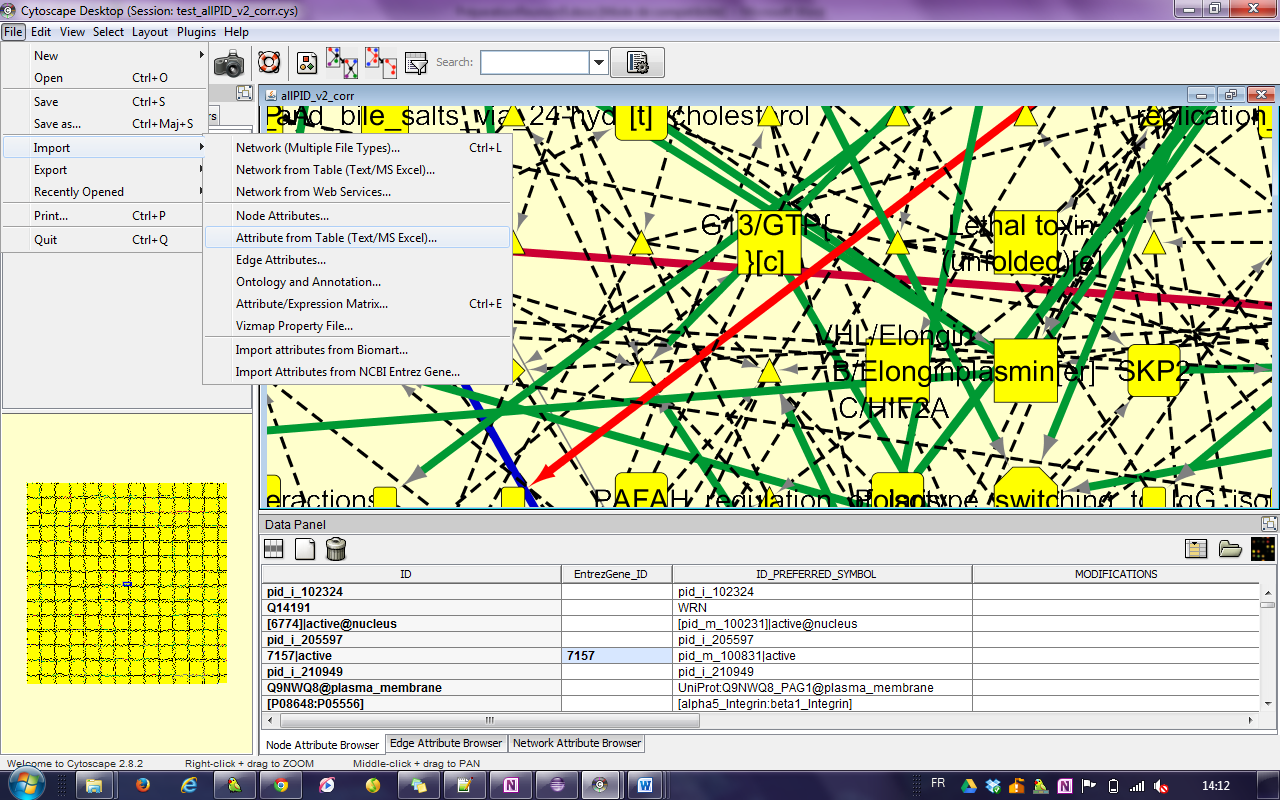
<http://cytoscape.org/manual/Cytoscape2_8Manual.html>

# Troubleshooting

## The '+' problem:

If you have attributes with '+' signs and they are replaced with white spaces, you can manually import these attributes:

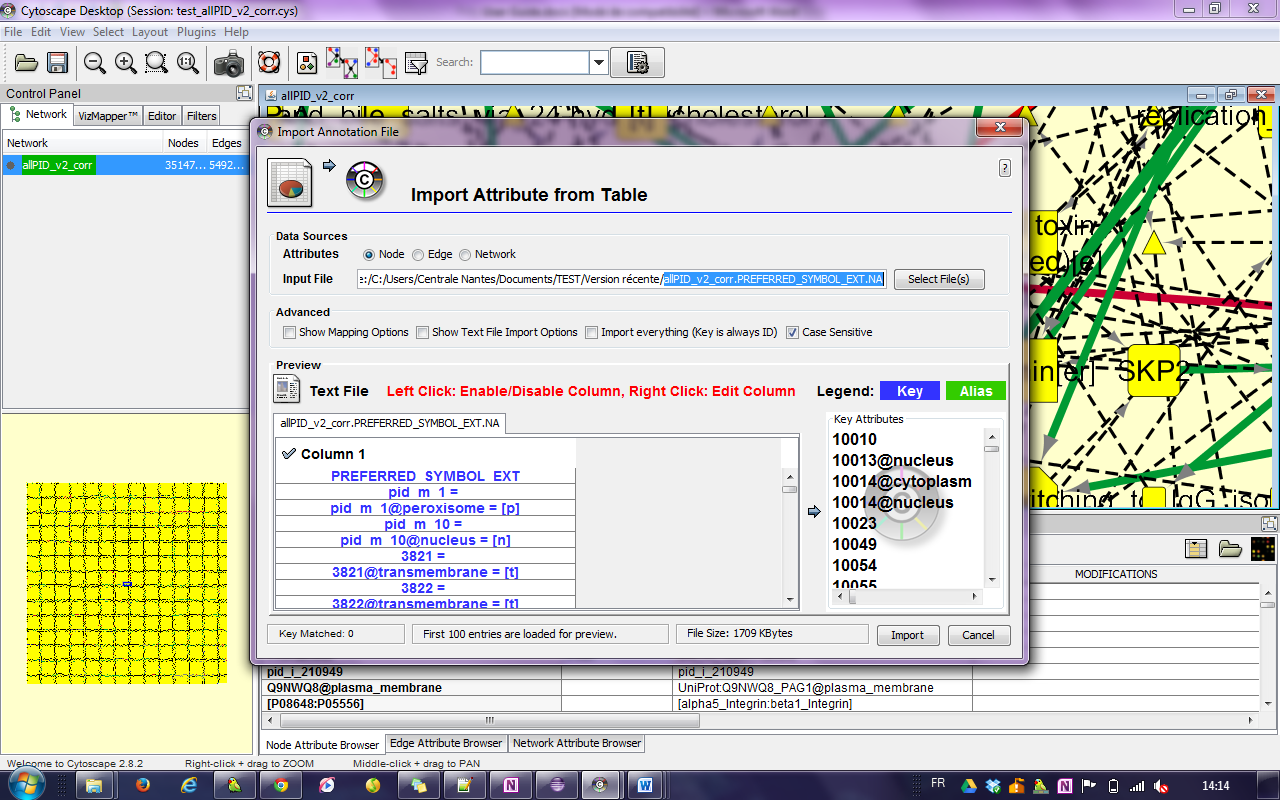
Select File->Import->Attribute from Table (Text/MS Excel).



We call "myAttribute" the name of the attribute with the '+' problem, and "myPIDfile" the name of the imported file in step 1.

Find the directory where the .sif file was created and select the file that ends with "myPIDfile.MYATTRIBUTE.NA" and open it.

For example, to see the names of the molecules with '+' symbols, choose "myPIDfile. PREFERRED\_SYMBOL\_EXT.NA".



You should now see the '+' symbols.

## XML file loading problem:

For some XML files, the step 1 fails and returns the following warning message :



The reason of this problem is that the XML file is not correct. For example, when the XML file defines an interaction between nodes that have not been defined previously, this error appears.

## The small window does not closes

Sometimes, the windows used to start/stop the process does not immediately closes because the process is actually not entirely finished. It should close if you wait a bit more.

## Using the same session to modify several graphs

It is possible to load several graphs during the same session. However, executing step 2 or step 3 on several loaded graphs is not recommended, as errors might appear.