UNIVERSITÉ DU LUXEMBOURG

LSRU | LIFE SCIENCES RESEARCH UNIT

The IDARE Cytoscape Application

Thomas Pfau [a,b], Jake Lin [c], Mafalda Galhardo [a] and Thomas Sauter [a,1]

[a]Life Science Research Unit, University of Luxembourg, Campus Belval, Biotech II 423, 6 avenue du Swing, L-4367 Belvaux, Luxembourg

[b]Institute of Complex Systems and Mathematical Biology, University of Aberdeen, Meston Building, AB24 3UE, Aberdeen, United Kingdom

[c]BioMediTech, University of Tampere, Biokatu 8, 33520 Tampere, Finland

October 14, 2016

[1]thomas.sauter@uni.lu

IDARE is a two component application aiming at simplifying the navigation through complex network structures and allowing the visualisation of data in a compact manner. The two components are:

- Visualisation of multiple data on nodes in a network.
- Subnetwork generation and connection by subgroup definition.

This user manual aims at providing a comprehensive overview of the capabilities and properties of IDARE and will provide a few simple examples to get comfortable with the functionality.

Contents

Chapter 1

IDARE Data Visualisation

IDARE allows the visualisation of discrete or continuous data by automatically generating image nodes that are mapped to the appropriate nodes in a given Cytoscape network. The mapping is constructed using identifiers present in the Cytoscape network and the data file. The Data Visualisation part of IDARE offers a dataset management interface along with utilities to add and remove image nodes from visual styles in Cytoscape. We will first explain the dataset management then detail the visualisation parts of the app.

1.1 Dataset Management

In this we will introduce data formats used by IDARE and explain Visualisation Types available in the app. When initializing IDARE a new tab will be generated in the Cytoscape Control Panel (see Figure 1.1). This tab is structured to provide functionality for dataset addition and removal (top section), give an overview of the currently available datasets and their settings (middle section) and allow the generation and preview of novel nodes (bottom section). The last section also contains information how many nodes the selected sets have in common and how many image nodes could be generated from the selected sets in total. The latter number does not take into account the availability of nodes in the network, but only considers the nodes in the datasets. While the layout process is in general automatic, the order of addition of datasets can have an influence on the positioning. Therefore the order of datasets can be altered with the buttons ("move up"/"move down") provided at the bottom of the panel.

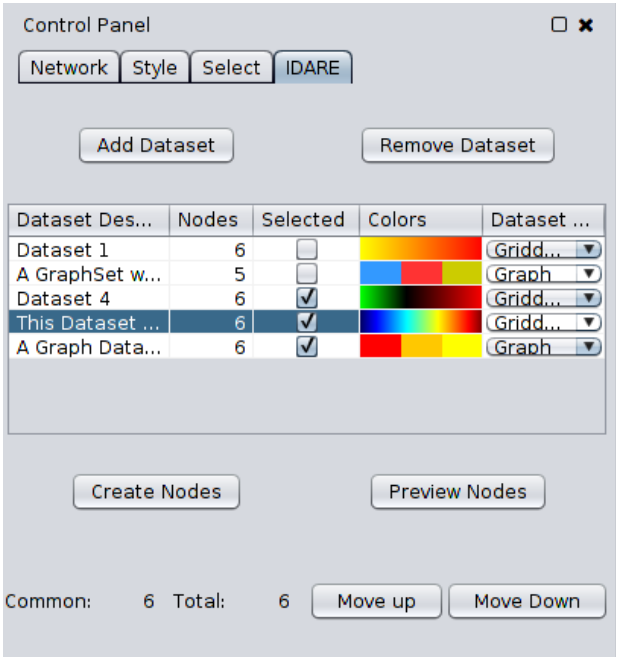


Figure 1.1: The Dataset Control Panel. The "Add Dataset" button will open a dialog to add datasets. The "Remove Dataset" button will remove the currently selected Dataset. The middle contains the current datasets and their properties and is described in more detail in Figure 1.5. The "Create Nodes" button will create ImageNodes for the current selection of datasets using the properties in the selected in the table.

1.1.1 Adding a dataset

To add a dataset simply press the "Add Dataset" button. You will be prompted to a dialog that lets you select the dataset to load (see Figure 1.2). To add a dataset you need to select a file to load the dataset from and set the properties of the dataset (The properties are explained in detail in Section 1.1.2). Once you have finished your settings click "Create Dataset". The app will potentially ask for additional information for parsing and then try to read in the dataset. If the read is successful and the data can be used with any of the available visualisation options it will be added to the IDARE Panel. If either the file format is invalid for the selected dataset type, or no visualisation can be performed for the dataset, an error dialog will be shown.

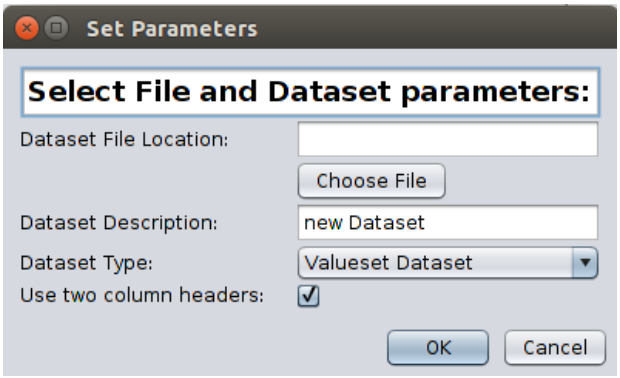


Figure 1.2: The dataset property selection dialog. You can either type the location of your file manually or use the choose file button to navigate to a file. You should fill in an alternative Dataset Description, and select which type of Dataset you are loading as well as whether to use two column headers or not. Dataset types and properties are explained in detail in Section 1.1.2

1.1.2 File Format

In the base app, IDARE uses excel sheets (xls/xlsx) and text delimited value file formats (tsv/csv). It assumes data to be presented as textual or numeric information. A plugin exists to provide support for GEO SOFT files, and allows some alternative parsing methods for those files. For Excel and delimiter separated file formats, the system expects the first row of a sheet to contain headers for the sheet's columns. These headers are used in the legend to provide information about the specific entries. After the first row, IDARE interprets each row as an entry for a node, with the first column(s) being used for node identification and labelling. Depending on your choice during dataset addition, either one column is used for both label and id of the node or, if the appropriate box is ticked, the system expects two columns, the first being the identifier that node is matched to and the second column is the label used on the images created. This allows to improve the readability, as published networks often use more computer readable identifiers that are hard to interpret. E.g. it allows the use of gene symbols as labels while the investigated network uses ENSEMBLE IDs, making visual inspection easier. The remaining file should consist of either up to five different values (e.g. 'on', 'off', 'unknown') or numeric values. These values will be used to generate the nodes. Empty entries are commonly assumed as NA values and are coloured in a light grey (which can lead to interference with some colour schemes). If there is a completely empty column (including a missing header), this column will still be included in the resulting node images (see Figure 1.3) in Array datasets but can lead to problems in Multi Array datasets

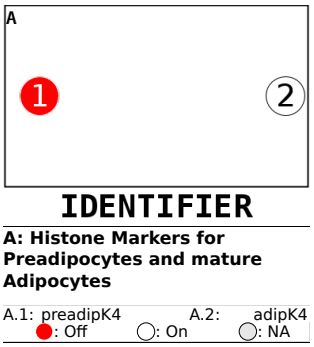
This can be used to match two datasets which have values that only partially overlap in Array data. An example would be two time series where one series has timepoint 1h,2h,4h, and 8h, while the other has 1h,4h,6h and 8h. In this instance adding an empty row behind 4h in Dataset 1 and one after 1h in dataset 2 will lead to matched positions. There is one Dataset Type that is different: GraphData. For graphs with more than one line, only excel sheets can be used. This is necessary, as the sheets in an excel file will be interpreted as independent datasets for each line on the graph and no sheets are available for csv. An example for a GraphData file is presented in Figure 1.4.

Data set types and visualisation options

IDARE currently offers the use of two different types of datasets, which are specific to different types of visualisation:

	A	B	C	D	E	F	G	H
1		preadipK4						adipK4
2	Gene1	Off						Off
3	Gene2	Off						Off
4	Gene3	Off						Off
5	Gene4	On						On
6	Gene5	Off						Off
7	Gene6	Off						Off
8	Gene7							On
9	Gene8	Unknown						On

(a) A Spreadsheet for that uses a single column for node id and label. This source file would be fitting for an Array dataset, but not for a Multiarray dataset. The empty rows indicate that there are 5 unused items between the first and the last, which allows alignment of data.)



(b) A legend node image (Top) and the description for Gridded circles from 1.3a

Figure 1.3: An example of the data generating a Gridded Array data node and the resulting node legend

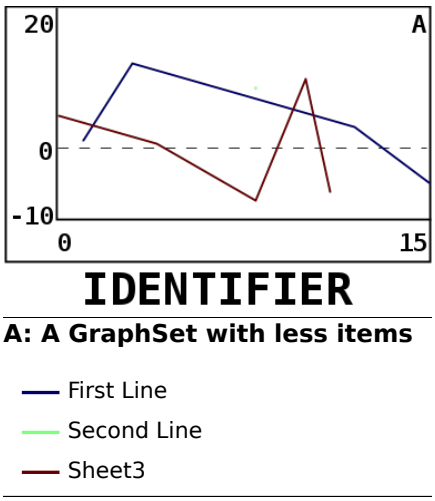
	A	B	C	D	E	F	G
1	..	..	1	3	7	12	15
2	Obj1	Objlabel1	1	2	0	2	5
3	Obj2	Objlabel2	2	3	0	4	6
4	Obj3	Objlabel3	1	12		3	-5
5	Obj4	Objlabel4	-3		0	5	7
6	Obj5	Objlabel5	1		0	4	5
7							

First Line | Second Line | Sheet3

	A	B	C	D	E	F	G
1	..	..	0	4	8	10	11
2	Obj1	Objlabel1	2.59	5.08	1.29	1.46	-0.73
3	Obj2	Objlabel2	-5.79	-4.45	-9.89	-7.93	-8.73
4	Obj3	Objlabel3			8.58		
5	Obj4	Objlabel4	-7.51	3.71	2.14	-6.44	1.60
6	Obj5	Objlabel5	-6.29	-3.25	-7.95	8.35	-4.09
7							

First Line | Second Line | Sheet3

(a) The first two sheets of a Spreadsheet containing numeric data used for a valueset dataset. The column headers indicate the different x values which will be used for the plot. The names of the sheets will be used in the legend to provide information on the different lines in the plot. The data uses a two column header, indicating that the first column is used to match ids and the second to provide labels.



(b) The legend generated for the node Obj2 as generated from the data in 1.4a

Figure 1.4: An example of the value set data and the resulting graph visualisation node legend

1. Array data and

2. Multi-array Data

The former is data that can be categorized as "one set of entries per node" (e.g. presence or absence of a histone marker at different times on a given gene or availability of a given product at a specific location). IDARE assumes, that this type of data is restricted to one spreadsheet, and will therefore ignore any sheet after the first in the data. In contrast Multi-array data can contain several sets of entries per node, assuming that each sheet represents one set of data that should be visualised simultaneously. E.g. sets of data for time series measurements of gene expression under different conditions during, with every condition being represented in one sheet in an excel file. This type of data can e.g. be used to plot graphs for different treatments in the same axis. Any data that can be used as Array data can also be used as Multiarray data and will commonly be treated as data with only one set, when viusalized (i.e. it will commonly only have one color associated and not a range of colors for the different sets). There are currently five options available for visualisation of Array datasets:

- Gridded circle visualisation
- Circle visualisation
- Time series visualisation
- Heatmap visualisation.
- Barchart Visualisation (via the BarChart plugin)

Gridded and Time series representations are inflexible representations which will be placed in the center, while items and heatmap are flexible visualizations, which can be placed at any position which has space left. Bar Charts will visualise the individual values as bars, with one bar per header in the center. For Multiarray data currently three options exist, a graph representation and a representation as a scatter plot as well as BarCharts (via the BarChartPlugin). Graphs expect, that the headers are numeric values in ascending order which will be used as x-values when drawing the graph. However, the header lines in each sheet are not required to be identical. This allows the matching of datasets with e.g. different measured timepoints. Each array (individual sheet in an excel workbook) will be plotted with the time course starting at the first non empty entry and ending at the last non empty entry (i.e. for the range of values specified in the respective sheet). If there is only one non empty point in a row, only a point will be displayed in the graph. Scatter plots require either numeric headers or at most six distinct textual headers. If textual headers are used, they have to be the same in all sheets. For numeric headers, the headers will be used as x- and the entry data as y-values for each marker. For textual headers, the data will be plotted around a center for each textual header. Bar Charts will plot sets of bars (one bar for each array - i.e. sheet in an excel workbook - of the Multiarray data) for each header. Table I offers an overview of the currently available types of visualisation.

1.1.3 The Dataset Table

The dataset table (see Figure 1.5) provides an overview of the current selection of options for each dataset. You can select which dataset to add to node you want to create, by ticking the boxes in the "selected" column. The Table also offers options to select the visualisation type of the respective dataset. You can choose both the color scale used to plot the respective dataset and the visualisation style (see Above) that shall be used. Furthermore, the Description can directly be edited (to avoid deleting and reloading due to typos).

Data Type	Description	Properties
Circles	Data that is individual and for which the different entries are commonly unconnected.	Circles, Center or Edge
Gridded Circles	Array data that is in a specific Grid	Circles, Center
Time Series	Data that is connected and displayable in one row	Boxes, Center
Heatmap	Data that can be rearranged but is somewhat connected and displayed in the style of a heatmap	Boxes, Center or Edge
BarChart (via Plugin)	Individual or sets of data presented as bars (in groups for Multi Array Data)	Bars, Center
Graph	Data that is displayed as a line graph. A similar scale is necessary	Graph, Center
Scatterplot	Scatterplot data with x/y values or groups of y values (large and small markers are available)	random Markers, Center

Table I: A short description of the currently available types of data visualisations.

Dataset Description	Nodes	Selected	Colors	Dataset Type
Dataset 1	6	☐		Gridded Items
A GraphSet with less items	5	☐		Graph
Dataset 4	6	☑		Gridded Items
This Dataset has a very very long des...	6	☑		Gridded Items
A Graph DataSet	6	☑		Graph

Figure 1.5: The dataset table. the first column allows editing of a dataset description, the second column indicates the number of nodes in the dataset. The third column marks, whether a specific dataset will be used when generating imagenodes, the fourth column presents the colorscheme, and allows the selection of other schemes, while the last column allows the selection of visualisation type.

1.1.4 The IDARE Visual Style and adding images to other styles

IDARE offers both a specific visual style, which is aimed at working with metabolic networks, and the option to add the image nodes to the currently selected style. To use the IDARE Style the network has to be set up as he style relies on non default columns. This can be achieved, using the Apps → IDARE → "Setup Network for IDARE" menu item or clicking on an empty are in the network and select Apps → "Setup Network for IDARE". A dialog will be shown that asks you to select which network column to use to retrieve IDs to use in IDARE and which column to use to extract the node types. The former will affect the mapping between nodes and generated images, the latter is specific to the IDARE Visual style (defining the mappings used in this style).

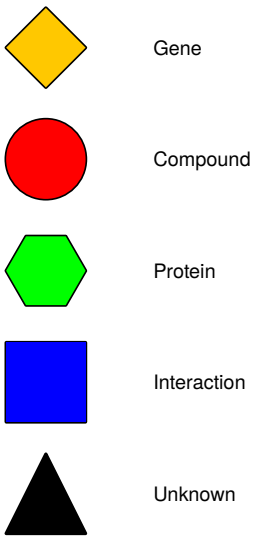


Figure 1.6: Different node types and their representation by the IDARE visual style.

The IDARE Visual Style aims at displaying bipartite networks, and in particular metabolic networks. Two types are the default assumption (compound and interaction). The different types of nodes used by IDARE are displayed as detailed in Figure 1.6. You can also add the generated image nodes to any other style (overwriting the the costum_graphics 1 property, along with node sizes). You can perform this addition via the context menu (right click → Apps → Add IDARE images). This will place the images generated from the datasets on any matched node in the network when using the currently selected style. To remove them you can use the "Remove IDARE Images" menu item. Nodes will be automatically informed of updates. Note that when saving a session and removing the nodes after reloading, only default properties (i.e. no mappings) of a style for which the nodes were added will be restored.

1.1.5 Creating user specific Visual Representations and DataSetTypes

IDARE provides interfaces to allow the addition of other visualisations that can be selected. The API for plugin development is located on the IDARE Server. Two plugins exist as samples. A Reader for GEO

SOFT files (which can be found on github at https://github.com/sysbiolux/IDAREGEOSoftReader

Chapter 2

The Subnetwork extractor tool

One issue often found in large networks is that visualisation becomes difficult due to the enormous amounts of interactions, leading to the classical image of the network-hairball. While some properties might still be visible on that level, a detailed view of a subnetwork often allows a better interpretation of the observed fluxes through a network. With biological models, there is commonly known information of pathways which form connected subnetworks. It is easy to restrict a network to the reactions found in such a pathway if only that specific pathway is of interest to the researcher. In doing so, it is however often problematic to keep track of branching pathways, which leads to a loss of overview.

2.1 Creating subnetworks

The subnetwork extractor allows the user to extract subnetworks while keeping connections to the general network. This is achieved by creating links between different subnetworks that point to the position of the linked compound in the other network. Thus it is possible to follow the course of flux, or the general network structure by following the links generated. The method is applicable to any bi-partite network, that can be divided into "compound" and "interaction" nodes. The following figures provide an example of the application of the subnetwork extractor tool. The user selects one column of the Cytoscape network table which contains the class each node belongs to (Fig. 2.1).

Class 1 represents nodes connecting different subnetworks (e.g. reaction when creating compartments in metabolic networks) Class 2 represents nodes directly belonging to the subnetwork (e.g. metabolites in

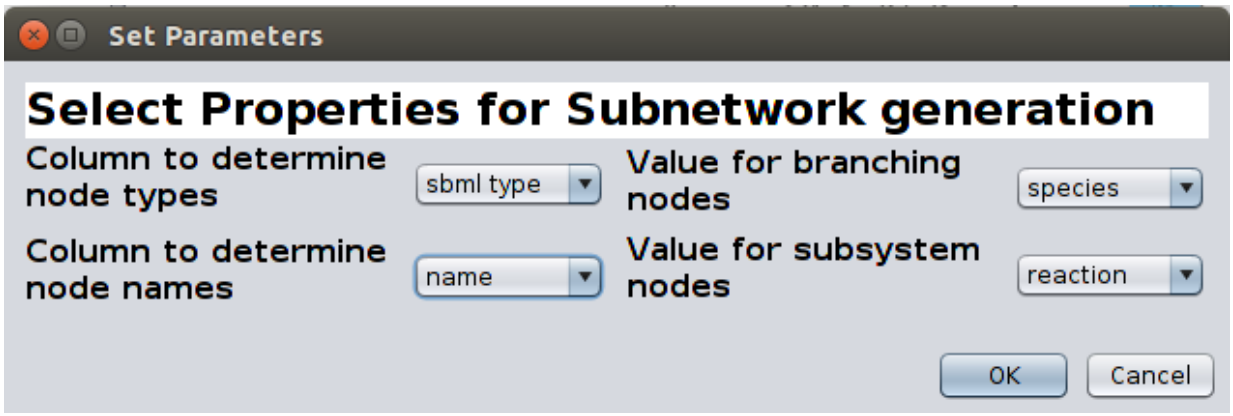


Figure 2.1: The dialog lets you select the column to determine node types and ids (left). The right side shows the options for bipartition (types for compound and interaction nodes). The right side can be missing if the selected type column does not offer appropriate options.

compartments). Any additional classes in the column are ignored for the subnetwork generation process. However, nodes will be included in the networks if they are connected to any of the subnetwork nodes (even over multiple edges). This propagation is stopped at any node of Class 2 that does not belong to the current subnetwork, or any node of Class 1 if the source node for propagation is not a noe of Class 1. Nodes belonging to Class 2 should only ever be part of the subnetwork if they are properly assigned to it. Nodes belonging to Class 1 could have indirect links (e.g. common genes) with nodes that should not belong to the same subetwork, thus they are only included if they have a direct link to a node of Class 1.

In addition to the type column a column that contains identifiers of the nodes has to be selected for the future selection of node properties.

2.1.1 Selecting the subnetworks to generate and the properties of nodes

Finally, the user is presented with a selection screen (see Figure 2.2) , containing all compounds and all subnetworks. This selection screen allows the definition of compounds, which should not become part of the final subnetworks like e.g protons or water, which are abundant and make layouting and visual inspection difficult (tick the column 'Remove'). It also allows the definition of compounds which should not be considered when creating links between subnetworks (tick the column 'Do not extend'). This is useful if considering e.g. metabolites like glyceraldehyde-3-phosphate. While removing it from the pathways would lead to gaps in the flow, branching would lead to the inclusion of an enormous amount of links. While this can be desirable the tool assumes that very common compounds are not supposed to branch and the most common are to be removed. However the final choice of compounds to be removed/declared as non branching can be done by the user in the selection screen. On the bottom the selection screen provides the not yet generated subnetworks for this network based on the selected column. In addition the system allows to remove links from the parent network.

2.1.2 Navigating between subnetworks and deleting/creating views

When subnetworks are created an additional network view for each subnetwork. It will also generate linker nodes for each combination of subnetworks sharing branching compounds. Double clicking on one of these links will automatically open the respective network view, centered on the compound in the opened view corresponding to the compound that was connected to the link in the origin network. When a network view of a subnetwork is destroyed, the corresponding linker nodes are deactivated i.e. the nodes remain but no longer open the corresponding view. When the view of a network gets restored/rebuilt the linkers are also restored and can be used to navigate again.

2.1.3 Generating subsystems subsequently

It is possible to generate subsystems in subsequent steps and also to generate subsystems of subsystems (e.g. pathways in compartments). If additional subnetworks are created, the app checks which subnetworks, given the current column, were already created using the selected network. It will connect all networks that are based on the same column to the newly generated networks (if they have nodes in common). Thus it is possible to create subnetworks for the same classes on the network scale and on subnetwork scales, which would then be connected.

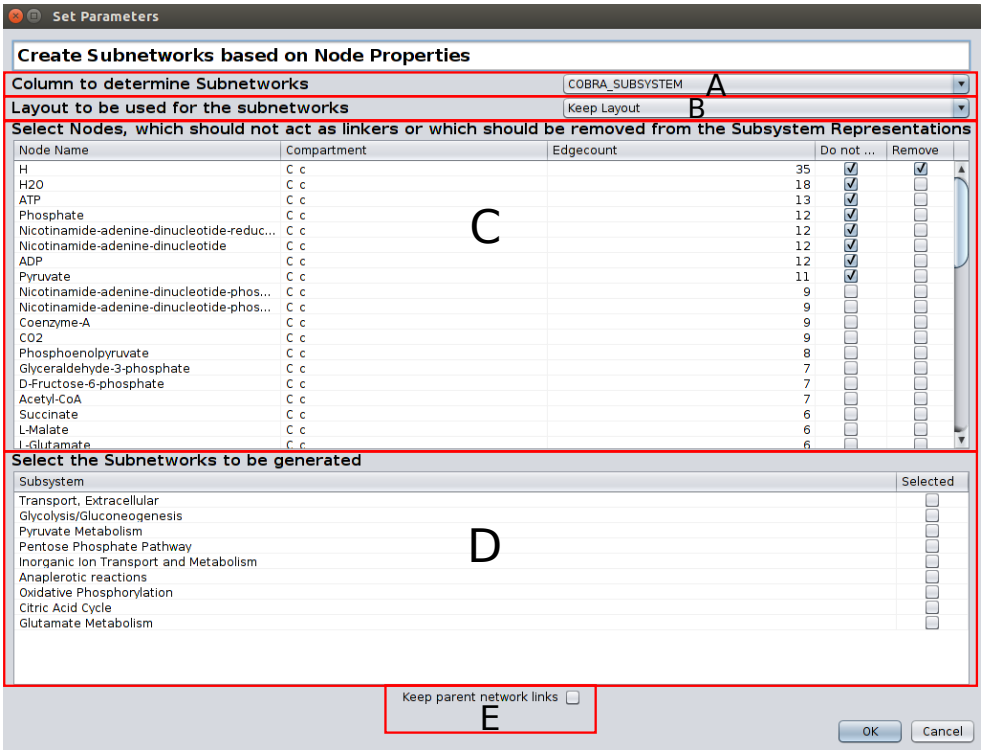


Figure 2.2: Field A: The Selection of the column used to determine the subnetworks (The default selected item is "SUBSYSTEMS" if it is present. According to this selection, Field D will adjust the available networks. Field B is the selector for the column used to select a layouting algorithm for the newly created subnetworks. To keep the Field C provides the user with the options which compounds to consider in subnetwork generation (for either linking or in general). In Field D the subsystems that should be generated are selected and Field E defines whether linkers from the parental network are kept or not.

Chapter 3

COBRA specific SBML reader

The COBRA SBML specification defines several fields in the notes section of SBML files which are specific to metabolic reconstructions. In particular those fields include information about genes associated with reactions (GENE_ASSOCIATION and GENE_LIST). To interpret transcriptomic data this information can be important as it allows to create the link between Reactions and genes, without having to compute a value for a reaction from the set of coding genes. If proper GPR rules are provided in the GENE_ASSOCIATION field, it allows to determine which combination of genes is required for a specific function (i.e. the Protein (-Complexes) that catalyse this specific reaction). GPR rules can also be incorporated by approaches using the annotation fields in SBML (as done e.g. by the Human Metabolic Reconstruction or the yeast consensus models). This approach contains additional species representing the Proteins in the SBML and annotates those proteins by bioql modifiers as follows:

1. If (protein) species has an SBO-term 0000014 (enzyme) or 0000252 (polypeptide chain), or it has a bioql modifier "is_encoded_by"
2. the "is_encoded_by" modifiers define the genes that form this specific protein.

Unfortunately the normal SBML import of Cytoscape completely ignores (and even discards) any information that is not directly associated with species and reactions. This is improved in the cy3sbml app (which interprets the annotation fields) but even this app does not interpret the information stored in the FBA specific COBRA notes. To address this, IDARE provides an SBML Annotation task that generates gene nodes, redefines protein species as proteins or adds protein nodes for each combination of genes in the annotation and adds all COBRA specific annotations and protein annotations to the network. The COBRA specific fields added include fields like CHARGE, FORMULA, AUTHORS, EC Number or SUBSYSTEM. Especially the latter is very useful when trying to create subnetworks as it often contains the common definitions of metabolic pathways. If only COBRA annotations are available, the generated protein nodes will simply be numbered, as the COBRA annotations do not contain

Chapter 4

Example

This chapter contains the description of a small sample and can also be found online at https://github.com/sysbiolux/IDARE-QuickStart

4.1 Necessary Data

The Data-Folder available on the github repository (https://github.com/sysbiolux/IDARE-QuickStart) contains several artificial data files that are necessary for this guide. In addition, it is necessary to download the *E.coli* core network sbml file from the UCSD website,

Activity.xls - Predicted activity in four different phases of the experiment External Metabolites.xls - Data for 3 external metabolites (and the biomass amount) over a period of time Internal Metabolites.xls - Data for all internal metabolites at different textual stages. ReactionActivity.xls - experimental and predicted activity of each reaction in the network

4.2 Quickstart

The Tutorial Contains three parts. In the first part, we set up the network, which is necessary for image matching. In the second part, we add images based on the data to the network. In the third part, we will split the network into parts based on compartments and pathways.

4.2.1 Setting up the Network for IDARE in Cytoscape

In this example, we will set up the *E.Coli* core network for use with IDARE and add some additional annotations to the network.

Loading the *E.Coli* Core model

After installing the IDARE app (either from the Cytoscape app store or from here), load the *E.coli* core network by selecting File → Import → Network → File and select the *E.Coli* core file you downloaded.

Generating an initial Layout

From the Layout menu, select a layout you want to apply (for this tutorial we assume, that the y-Files organic layout was choosen).

Setting up the Network for IDARE

Right click on some empty space in the network view.
Select 'Apps' → 'Setup Network For IDARE'
A popup will appear that lets you select the properties to be used in IDARE.
On the left you can select the columns you want to use for the setup. On the right, depending on your choice of columns, you can select the values to be used for compound and interaction.
Select the sbml type column as column to determine node types and the sbml id column as column to determine the node names.
As compound node value select species, and as interaction node value select reaction.
Check the 'Overwrite existing values' checkbox.

Changing to the IDARE Visual Style

From the 'Style' tab in the 'Control Panel' select the 'IDARE Visual Style' and apply it to the network.

Adding SBML Annotations and Gene Nodes

Right click on some empty space in the network view.
Select 'Apps' → 'Add SBML Annotations'
Select the *E.Coli* core network file again. If you have cy3sbml this step is not necessary, as Cytoscape will automatically use the SBML structure provided by cy3sbml for the network.
You will be asked, whether you want to add Gene Nodes. Click Yes
You will notice, that additional nodes have been created, which represent the gene nodes.
If the SBML contains enzyme species (annotated by the "isEncodedBy" bio qualifier), you would be asked if you also want to create protein nodes and which labeling pattern (i.e. database) to use for the genes and proteins.

4.2.2 Automated image generation and loading images to the Network.

In this example you will generate a few images based on artificial data and map it to the previously created and initialized network. This example assumes that you have at loaded the E.coli core network and set up the network for IDARE (Steps 1 and 3 of the previous example).

Load Data into IDARE

In the 'Control Panel', select the IDARE tab.
Click on 'Add Dataset'
In the dialog, click on 'Choose File'
Select "Activity.xls" and click 'open'
In the 'DataSet Description' field write the text: "Indication of reaction activity at different simulation phases"
As 'DataSet Type', select 'Array Dataset' (since this dataset contains only one sheet).
Uncheck the 'Use two column headers' option, as we only have one header.
Click ok.
Repeat the process with "External Metabolites.xls" and "ReactionActivity.xls" which should both be loaded as 'Multi Array Datasets'. Even though ReactionActivity.xls contains only one sheet, the number of datapoints is unsuitable for itemized representation. Thus importing it as a 'Valueset Dataset' with one

set allows it to be displayed e.g. as a graph.
Finally repeat the process with "Internal Metabolites.xls" which should be loaded as 'Array Dataset' and is a two column header file, so the checkbox has to be ticked. Example descriptions would be "External Metabolite amount information" and "Reaction Activity during the experiment" and "Internal Metabolite Amounts at specific stages".

Create the Visualisation

You can now select individual datasets you want to visualize. To do so, choose a selection of datasets from the table by ticking the "selected" boxes. In the lower left corner, (below the 'Create' and 'Preview' buttons) are indicators, how many nodes the datasets represent in total and how many shared nodes the selection contains. When creating visualisations for multiple datasets at once, there should be an overlap of the sets, otherwise parts of the generated image will always stay blank since no data is available. However, you can of course create multiple different nodes by selecting different sets of data. Images will only be created for nodes which are part of a selected dataset and old images will be kept unless a new image is created.

Adding IDARE images to other Styles

If you did not select the IDARE visual style, but want to use a different style for your network, you can do so by right-clicking anywhere in a network view using the style you want to add your nodes to. Select 'Apps' → 'Add IDARE Images'. Now the images will be shown for your choosen style.

4.2.3 Subnetwork Generation

The first step of this this part is independent on the first two examples. For the second step, it is necessary to have the SBML annotations added to the network (step 1.5), as otherwise the corresponding table column does not exist. You can add the sbml annotation to the Cytosol network (C_c) after step 1 of this example, but this would lead to gene nodes not being added in the external compartment subnetwork.

Create Networks for the External compartment and the cytosol

We first want to split the network at the transporters which translocate metabolites from the external compartment to the cytosol, and vice versa.
To do so:
Select 'Apps' → 'IDARE' → 'Create Subnetworks'
or
Right Click in an empty space in the network you want to create subnetworks in → 'Apps' → 'Create Subnetworks'
In the resulting dialog, you can select the column to determine the node types as well as the node names. If the network is set up for IDARE (see Part 1), the corresponding columns choosen there are automatically selected.
You further have to choose a value for nodes which form the branching points between subnetworks, and a value for nodes that form the "base" of the subnetwork. Since the aim is to branch at the transporters between compartments reactions will be used as branching nodes and species as subnetwork nodes.
Click accept, when you finished your selection.
In the next dialog, you can select the column that contains the subnetwork identifiers as well as the type of layout that should be used for the subnetwork (or the "keep layout" option, if the current layout shoudl be kept) The table allows you to select nodes, which should not branch and those which should be removed entirely (e.g. nodes with too many connections). Depending on the network size some nodes are already

suggested by the tool.
Since we want to create the networks for the cytosol and external compartment, we select "sbml compartment" as column to determine sub-networks. We leave the 'Keep Layout' option as it is and select remove for the Biomass reaction (as this reaction mainly leads to a hairball structure).
We also select C_c, the cytosol, and C_e, the external medium, in the sub-networks to be generated table.
If we now click on the "network" tab in the 'Control Panel' we will see two new networks that were generated.

Create pathway networks in the Cytosol

There is not much to see in the external compartment, as it only contains the transporters and the cytosol is still a hairball. To get a better overview, we would like to generate pathway networks in the cytosol.
To do so, again, select 'Apps' → 'SubNetworkGenerator'
However, this time, since reactions belong to pathways and metabolites can be shared between different pathways, we select species as values for branching nodes and reaction as value for subnetwork nodes.
By default, the "COBRA_SUBSYSTEM" column generated by the SBMLAnnotator is selected, if available, so this selection is already fine.
From the metabolites we would like to exclude currency metabolites and most cofactors.
For larger networks, the pre-selection is better, but on small networks (like the e.coli core), the relative amounts of presence tend to be higher, so we have to select additional nodes to remove during the generation process.
Select M_h_c, M_h2o_c, M_atp_c, M_pi_c, M_pi_c, M_nadh_c, M_nad_c, M_adp_c, M_nadph_c, M_co2_c, M_nadp_c and M_coa_c, as metabolites to remove.
As pathways, we would like to create 'Glycolysis', the 'Citric Acid Cycle', 'Anaplerotic reactions' and the 'Pentose phosphate pathway'.
Click ok.
You can now navigate through the different pathways. If you wish to create additional pathways, select the source network again and repeat the steps. New links will be added to the existing networks.
You will notice, that links from parental networks will be included in the newly created networks, but no links are generated in these parental networks. In general, links will be generated for all networks built on the same column, that share linking nodes. If you create additional subnetworks at a later stage, links to and from old networks will be created for all newly generated networks.