

Cytoscape: un approccio sistematico all'analisi dei dati sperimentali

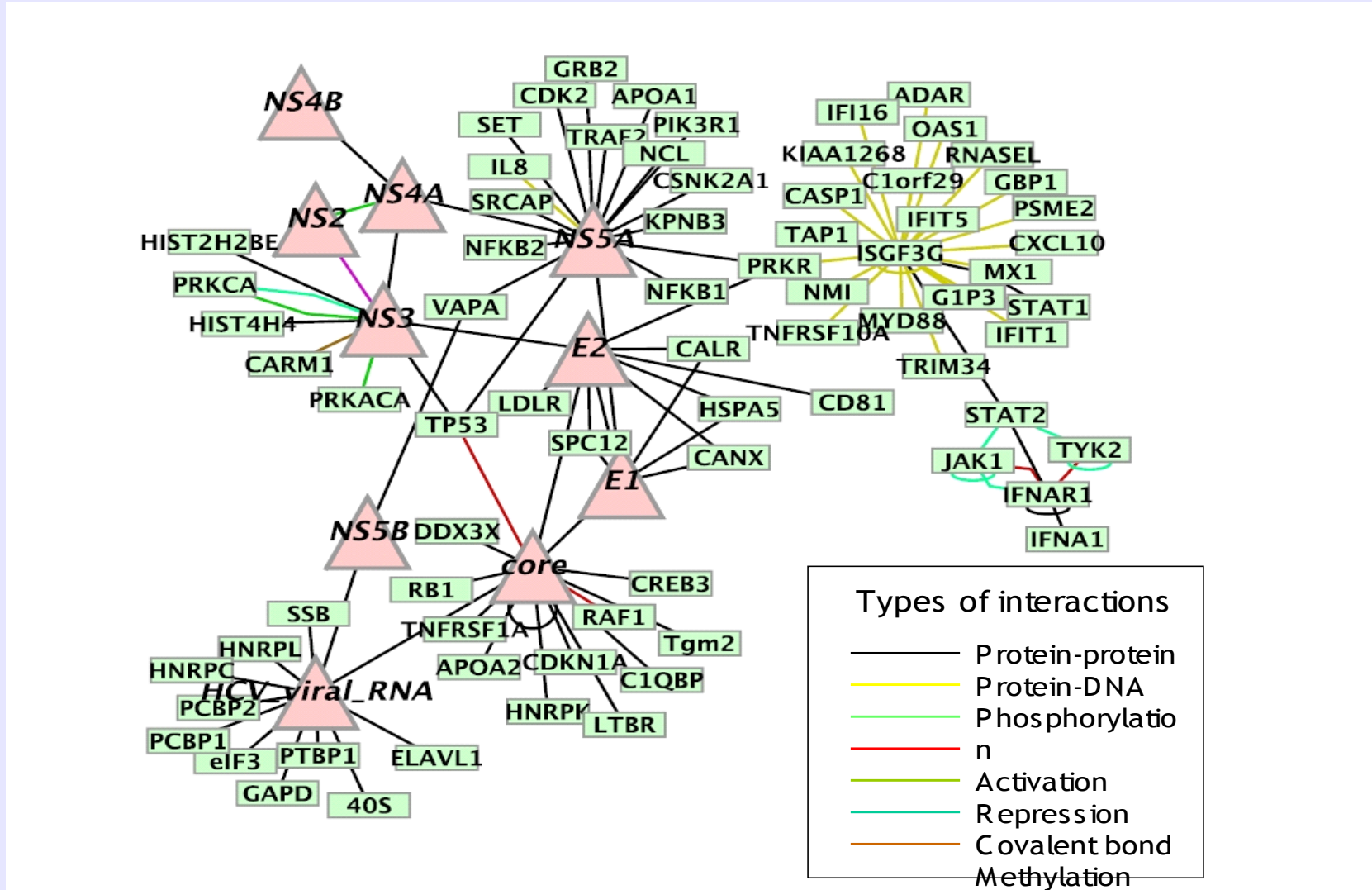
Michele Petterlini

michele@petterlini.it

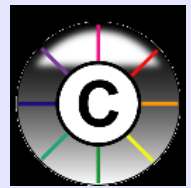
*Università di Verona
Centro di BioMedicina Computazionale*

The network

Network: a graph representation of experimental phenomena



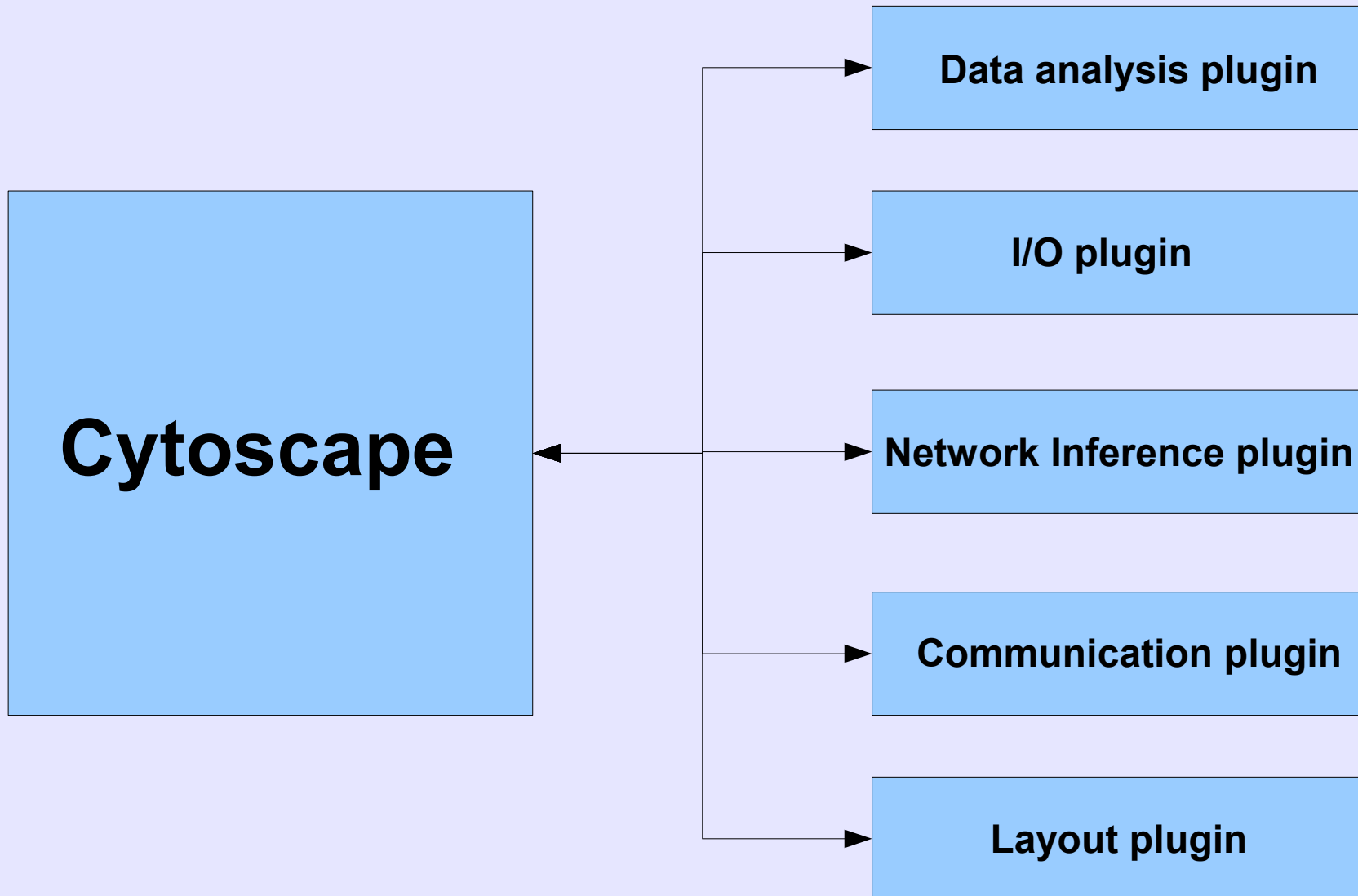
What is Cytoscape?



- Cytoscape is an open-source community software project for **integrating biomolecular interaction networks with high-throughput expression data** and other molecular states into a **unified conceptual framework**.
- Although applicable to any system of molecular components and interactions, Cytoscape is most powerful when **used in conjunction with large databases** of protein-protein, protein-DNA, and genetic interactions that are increasingly available for humans and model organisms.
- A software “Core” provides basic functionality to layout and query the network; to **visually integrate** the network with expression profiles, phenotypes, and other molecular states; and to **link the network to databases** of functional annotations.
- The Core is extensible through a straightforward **plug-in architecture**, allowing rapid development of **additional computational analyses** and features.
- The central organizing metaphor of Cytoscape is a **network graph**, with genes, proteins, and molecules represented as nodes and interactions represented as links, i.e. edges, between nodes.

Plugins

A **plugin** is an external program sharing a common workspace with a core program.



Cytoscape Quick Tour

- Base features -

Cytoscape quick tour: main window

The screenshot displays the Cytoscape Version 2.5 main window. The title bar reads "Cytoscape Version 2.5" and "Cytoscape Desktop (New Session)". The menu bar includes "File", "Edit", "View", "Select", "Layout", "Plugins", and "Help". A toolbar with various icons and a search box is located below the menu bar. The main window is divided into several panels:

- Control Panel:** Contains tabs for "Network", "VizMapper™", and "Editor". Below the tabs is a table showing the current network's statistics:

Network	Nodes	Edges
galFiltered.sif	331(0)	362(0)

- Main Visualization Area:** Displays a network graph with 331 nodes (represented by red circles) and 362 edges (represented by blue lines). The graph is titled "galFiltered.sif".
- Data Panel:** Located at the bottom right, it shows a table with one row containing the ID "canonical...".
- Node Attribute Browser:** A panel at the bottom center, currently selected.
- Edge Attribute Browser:** A panel at the bottom right.
- Network Attribute Browser:** A panel at the bottom right.

At the bottom of the window, there are three instructional text boxes:

- Welcome to Cytoscape 2.5
- Right-click + drag to ZOOM
- Middle-click + drag to PAN

Cytoscape quick tour: main window

Network window

The screenshot displays the Cytoscape Version 2.5 interface. The main window is titled "Cytoscape Desktop (New Session)" and shows a network graph with 331 nodes and 362 edges. The graph is visualized as a grid of red nodes connected by blue edges. A blue arrow points to the "Network window" title bar. The interface includes a Control Panel with tabs for Network, VizMapper™, and Editor. The Data Panel shows the ID of the selected node as "canonical...". The bottom of the window features a status bar with the text "Welcome to Cytoscape 2.5", "Right-click + drag to ZOOM", and "Middle-click + drag to PAN".

Network	Nodes	Edges
galFiltered.sif	331(0)	362(0)

Control Panel: Network | VizMapper™ | Editor

Data Panel: ID canonical...

Node Attribute Browser | Edge Attribute Browser | Network Attribute Browser

Welcome to Cytoscape 2.5 Right-click + drag to ZOOM Middle-click + drag to PAN

Cytoscape quick tour: main window

Network window

The screenshot displays the Cytoscape Version 2.5 interface. At the top is a menu bar with options: File, Edit, View, Select, Layout, Plugins, Help. Below the menu bar is a toolbar with icons for file operations and search. The main window is titled "Cytoscape Desktop (New Session)".

On the left side, there is a "Control Panel" with tabs for "Network", "VizMapper™", and "Editor". The "Network" tab is active, showing a table with the following data:

Network	Nodes	Edges
galFiltered.sif	331(0)	362(0)

Below the table is a small thumbnail of the network graph. A blue arrow labeled "Control panel" points to this area.

The main central area is the "Network window", titled "galFiltered.sif", which displays a dense network graph with red nodes and blue edges. A blue arrow labeled "Network window" points to this area.

At the bottom of the main window is the "Data Panel", which contains a table with the following data:

ID	canonical...

Below the Data Panel are three tabs: "Node Attribute Browser", "Edge Attribute Browser", and "Network Attribute Browser".

At the very bottom of the window, there is a status bar with the following text: "Welcome to Cytoscape 2.5", "Right-click + drag to ZOOM", and "Middle-click + drag to PAN".

Cytoscape quick tour: main window

Network window

The screenshot shows the Cytoscape Version 2.5 interface. The main window displays a network graph titled 'galFiltered.sif' with a grid of red nodes and blue edges. The interface includes a menu bar (File, Edit, View, Select, Layout, Plugins, Help), a toolbar, a search bar, and a Control Panel on the left. The Control Panel shows a table with the following data:

Network	Nodes	Edges
galFiltered.sif	331(0)	362(0)

Below the Control Panel is an Overview window showing a smaller version of the network graph. The Data Panel at the bottom right shows a table with the following data:

ID	canonical...

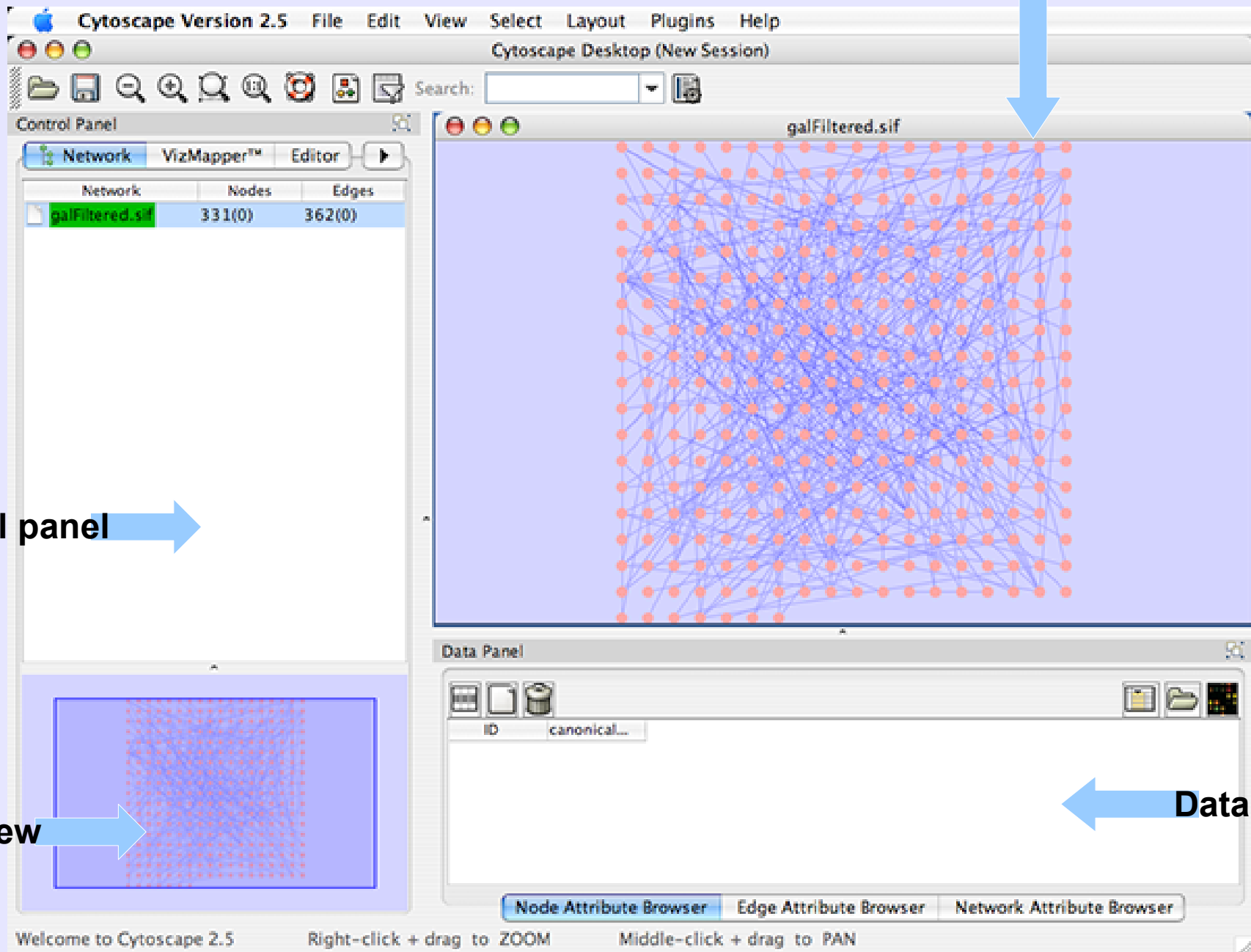
At the bottom of the window, there are three tabs: Node Attribute Browser, Edge Attribute Browser, and Network Attribute Browser. The status bar at the bottom contains the text: 'Welcome to Cytoscape 2.5 Right-click + drag to ZOOM Middle-click + drag to PAN'.

Control panel

OverView

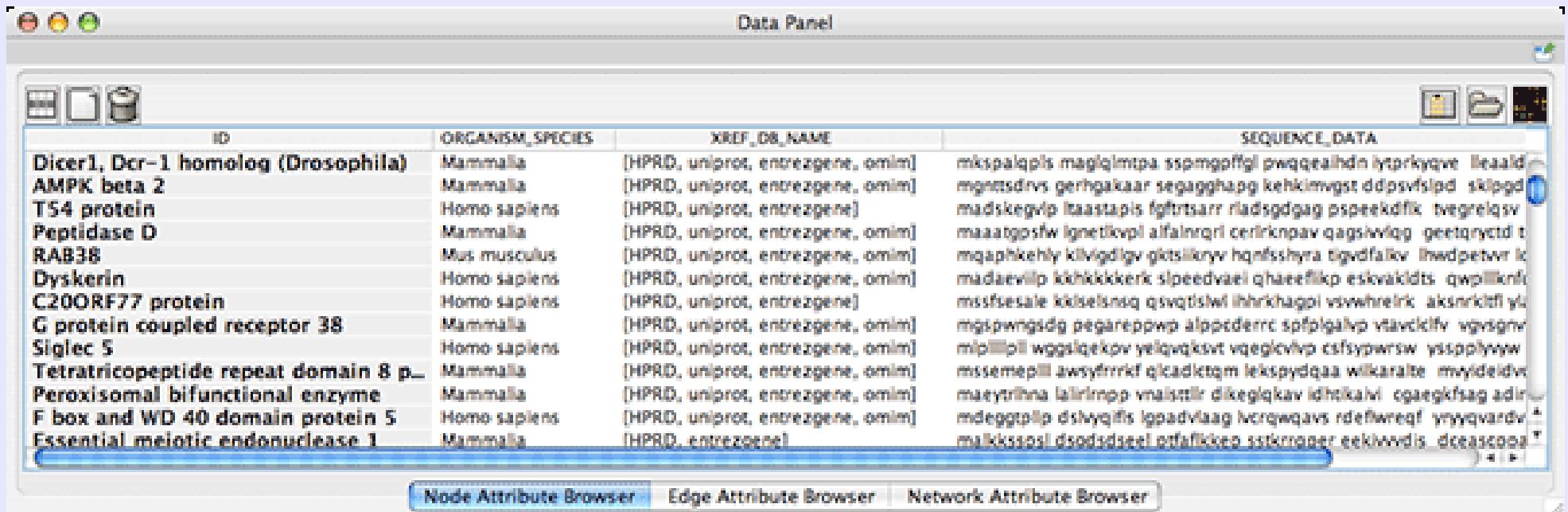
Cytoscape quick tour: main window

Network window



Cytoscape quick tour: data panel

- Node name
- Alias
- Database names
- Sequence data
- Attributes



The screenshot shows the 'Data Panel' window in Cytoscape. It contains a table with four columns: ID, ORGANISM_SPECIES, XREF_DB_NAME, and SEQUENCE_DATA. The table lists various proteins and their associated data. The 'Node Attribute Browser' tab is selected at the bottom.

ID	ORGANISM_SPECIES	XREF_DB_NAME	SEQUENCE_DATA
Dicer1, Dcr-1 homolog (Drosophila)	Mammalia	[HPRD, uniprot, entrezgene, omim]	mkspalqpls magqlmtpa sspmgoffgl pwqqealhdn lytrkyqve leaal
AMPK beta 2	Mammalia	[HPRD, uniprot, entrezgene, omim]	mgntsdnvs gerhgakaar segagghagp kehkimvgt ddpvsvslpd sklp
T54 protein	Homo sapiens	[HPRD, uniprot, entrezgene]	madskegvlp ltaastapis fgfrtsarr rladsgdgag pspeekdfk tve
Peptidase D	Mammalia	[HPRD, uniprot, entrezgene, omim]	maaatgpsfw lgnelkvpf alfalnqri cerlknqav qagslvvlgg geet
RAB38	Mus musculus	[HPRD, uniprot, entrezgene, omim]	mqaphkehly klivgdigv gktslkryv hqnfsshyra tgvdfalqv ih
Dyskerin	Homo sapiens	[HPRD, uniprot, entrezgene, omim]	madaevllp kkhkkkerk slpeedvaei qhaeeffkp eskvakids qwp
C20ORF77 protein	Homo sapiens	[HPRD, uniprot, entrezgene]	mssfsesale kkselsnq qsvqtislwl ihhrkhagol vswhrelrk aks
G protein coupled receptor 38	Mammalia	[HPRD, uniprot, entrezgene, omim]	mgspwngsdg pegareppwp alppderrc spfpalqvp vlavcklqv v
Siglec 5	Homo sapiens	[HPRD, uniprot, entrezgene, omim]	mipll wgsiqekqv yelqgkqv vqeglvlp cfsypwrsw yssplyvw
Tetratricopeptide repeat domain 8 p...	Mammalia	[HPRD, uniprot, entrezgene, omim]	mssemepll awsyfrrkf qlkadktgm lekspydqaa wilkaralte m
Peroxisomal bifunctional enzyme	Mammalia	[HPRD, uniprot, entrezgene, omim]	maeyrlhna lalrnpp vnalstlr dikeglqkav idhkalvl qeagkhsa
F box and WD 40 domain protein 5	Homo sapiens	[HPRD, uniprot, entrezgene, omim]	mdeggtplp dshyqifs lpadvlaag lvcrgwqavs rdefwregf yyyq
Essential meiotic endonuclease 1	Mammalia	[HPRD, entrezgene]	malkkssosl dsodsdgeel ofafikcep sstknaqer eeklvvdlv d

Cytoscape quick tour: network views

The screenshot displays the Cytoscape Desktop interface with a 'New Session' window. The Control Panel on the left shows a table of network statistics:

Network	Nodes	Edges
binDyeast.sif	23540(0)	61582(1)
Cluster 1	67(0)	341(0)
Cluster 2	6(0)	27(0)
Cluster 3	9(0)	39(0)
Cluster 4	6(0)	25(0)
Cluster 5	27(0)	101(0)

The main workspace contains four network views:

- binDyeast.sif**: A dense, dark network visualization.
- Cluster 1**: A network visualization with red nodes and blue edges.
- Cluster 3**: A network visualization with white nodes and black edges, featuring labels for TRS33, BET5, TRS, TRS31, BET3, and TRS120.
- Cluster 5**: A network visualization with white nodes and black edges.

At the bottom of the window, the following text is displayed: "Welcome to Cytoscape 2.5 Right-click + drag to ZOOM Middle-click + drag to PAN".

Cytoscape quick tour: network views

Network

The screenshot displays the Cytoscape Desktop interface with a 'New Session' window. On the left, the 'Control Panel' is visible, featuring a 'Network' tab and a table of network statistics. A blue arrow points to the 'Network' tab. The table lists the following data:

Network	Nodes	Edges
Dyeast.sif	23540(0)	61582(1)
Cluster 1	67(0)	341(0)
Cluster 2	6(0)	27(0)
Cluster 3	9(0)	39(0)
Cluster 4	6(0)	25(0)
Cluster 5	27(0)	101(0)

The main workspace contains several network views:

- Cluster 1:** A large, dense network graph with red nodes and blue edges.
- Cluster 3:** A smaller network graph with white nodes and black edges, featuring labels for nodes: TRS33, BET5, TRS, TRS31, BET3, and TRS120.
- Cluster 5:** A network graph with white nodes and black edges.

At the bottom of the window, there are instructions: 'Welcome to Cytoscape 2.5', 'Right-click + drag to ZOOM', and 'Middle-click + drag to PAN'.

Cytoscape quick tour: network views

Network

Views

The screenshot displays the Cytoscape Desktop interface with a 'Control Panel' on the left and several network view windows on the right. The 'Control Panel' includes a 'Network' tab with a table listing networks and clusters. The table data is as follows:

Network	Nodes	Edges
Dyeast.sif	23540(0)	61582(1)
Cluster 1	67(0)	341(0)
Cluster 2	6(0)	27(0)
Cluster 3	9(0)	39(0)
Cluster 4	6(0)	25(0)
Cluster 5	27(0)	101(0)

The main workspace contains four windows: 'BINdyeast.sif' (a dense black network), 'Cluster 1' (a network with red nodes and blue edges), 'Cluster 3' (a network with labeled nodes: TRS33, BET5, TRS31, BET3, TRS120), and 'Cluster 5' (a network with labeled nodes: TRS33, TRS31, TRS120). A small thumbnail of the main network is visible in the bottom-left corner of the control panel. At the bottom of the window, the text reads: 'Welcome to Cytoscape 2.5 Right-click + drag to ZOOM Middle-click + drag to PAN'.

Cytoscape quick tour: network views

Network

Views

The screenshot shows the Cytoscape Desktop interface with a 'Control Panel' on the left and several network view windows on the right. The 'Control Panel' has a 'Network' tab with a table listing networks and clusters. The table is as follows:

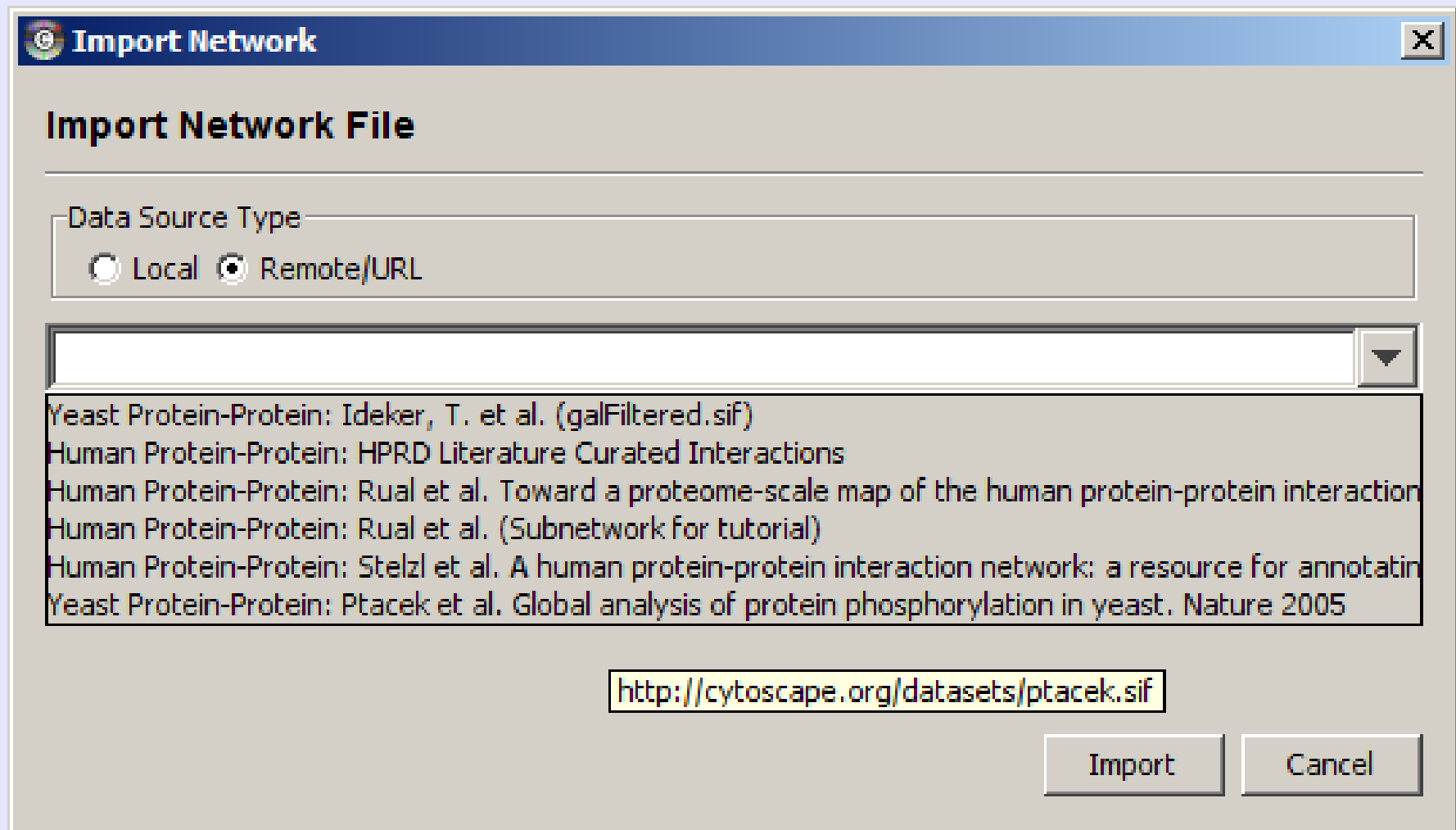
Network	Nodes	Edges
Dyeast.sif	23540(0)	61582(1)
Cluster 1	67(0)	341(0)
Cluster 2	6(0)	27(0)
Cluster 3	9(0)	39(0)
Cluster 4	6(0)	25(0)
Cluster 5	27(0)	101(0)

The main workspace contains four windows: 'BINdyeast.sif' (a dense black network), 'Cluster 1' (a network with red nodes and blue edges), 'Cluster 3' (a network with black nodes and edges, labeled with genes like TRS33, BET5, TRS31, BET3, and TRS120), and 'Cluster 5' (a network with grey nodes and edges). A blue arrow labeled 'Clusters' points to the 'Cluster 5' window. At the bottom of the window, there is a status bar with the text: 'Welcome to Cytoscape 2.5 Right-click + drag to ZOOM Middle-click + drag to PAN'.

Clusters

Cytoscape quick tour: file and URL import

- File import (from heterogeneous sources)
- URL import (from several labs)



Cytoscape quick tour: network import

- Simple interaction file (SIF or .sif format)
- Graph Markup Language (GML or .gml format)
- XGMML (extensible graph markup and modelling language).
- BioPAX (Biological PATHways eXchange)
- SBML
- PSI-MI Level 1 and 2.5
- Delimited text
- Excel Workbook (.xls)

Import Network from Table

Data Sources
Input File: file:/cellar/users/kono/workspace/cytoscape/testData/galFiltered.xls [Select File]

Interaction Definition
Source Interaction: Column 1 Interaction Type: Column 3 Target Interaction: Column 2

⚠ Columns in BLUE will be loaded as EDGE ATTRIBUTES.

Advanced
 Show Text File Import Options

Preview
Excel™ Workbook: Left Click: Edit Column Right Click: ON/OFF Column

Yeast Network Sheet 1

Column 1	Column 2	Column 3	Boolean At...	Text Attr 1	Column 6
YKR026C	YGL122C	pp	false	abcd12345	1.2344543
YGR218W	YGL097W	pp	true	abcd12346	1.2344543
YGL097W	YOR204W	pp	true	abcd12347	1.2344543
YLR249W	YPR080W	pp	true	abcd12348	1.2344543
YLR249W	YBR118W	pp	true	abcd12349	1.2344543
YLR293C	YGL097W	pp	true	abcd12350	1.2344543
YMR146C	YDR429C	pp	true	abcd12351	1.2344543
YDR429C	YFL017C	pp	true	abcd12352	1.2344543
YPR080W	YAL003W	pp	true	abcd12353	1.2344543

Import Cancel

Cytoscape quick tour: network import

Import Network and Edge Attributes from Table


Import Network from Table

Data Sources

Input File:

Interaction Definition

Source Interaction: Interaction Type: Target Interaction:

 Columns in BLUE will be loaded as EDGE ATTRIBUTES.

Advanced

Show Text File Import Options

Preview

Excel™ Workbook

Yeast Network Sheet 1

✓ Column 1	✓ Column 2	✓ Column 3	✓ Boolean At...	✓ Text Attr 1	✗ Column 6
YKR026C	YGL122C	pp	false	abcd12345	1.2344543
YGR218W	YGL097W	pp	true	abcd12346	1.2344543
YGL097W	YOR204W	pp	true	abcd12347	1.2344543
YLR249W	YPR080W	pp	true	abcd12348	1.2344543
YLR249W	YBR118W	pp	true	abcd12349	1.2344543
YLR293C	YGL097W	pp	true	abcd12350	1.2344543
YMR146C	YDR429C	pp	true	abcd12351	1.2344543
YDR429C	YFL017C	pp	true	abcd12352	1.2344543
YPR080W	YAL003W	pp	true	abcd12353	1.2344543

Cytoscape quick tour: attributes import

Attributes from file

Import Attribute from Table

Data Sources
Attributes: Node Edge Network
Input File: file:/cellar/users/kono/workspace/cytoscape/build/cytoscape-v2.4.0-b1/sampleData/galExpData.pvals [Select File]

Advanced
 Show Advanced Mapping Options Show Text File Import Options

Annotation File to Attribute Mapping
Key Column in Annotation File
Primary Key: GENE
Key Attribute for Network: canonicalName

Alias?	Column (Attribute Name)	Data Type
<input type="checkbox"/>	GENE	String
<input checked="" type="checkbox"/>	COMMON	String
<input type="checkbox"/>	gal1RG	String
<input type="checkbox"/>	gal4RG	Float
<input type="checkbox"/>	gal80R	Float

Text File Import Options
Delimiter: Tab Comma Semicolon Space Other
Preview Options: Show all entries in the file Show first 100 entries
Attribute Names: Transfer first line as attribute names Start Import Row: 1 Comment Line:
Network Import Options: Default Interaction: pp [Reload]

Preview
Text File: Left Click: Edit Column Right Click: ON/OFF Column
Legend: Key (blue) Alias (green)

GENE	COMMON	gal1RG	gal4RG	gal80R	gal1RG
YHR051W	COX6	-0.034	0.111	-0.304	3.75720e-01
YHR124W	NDT80	-0.090	0.007	-0.348	2.71460e-01
YKL181W	PRS1	-0.167	-0.233	0.112	6.27120e-03
YGR072W	UPF3	0.245	-0.471	0.787	4.10450e-04
YHL020C	OPI1	0.174	-0.015	0.151	1.40160e-04
YGR145W	YGR145W	0.387	-0.577	-0.088	5.37920e-03
YGL041C	YGL041C	0.285	-0.086	0.103	4.46050e-04
YGR218W	CRM1	-0.018	-0.001	-0.018	6.13810e-01

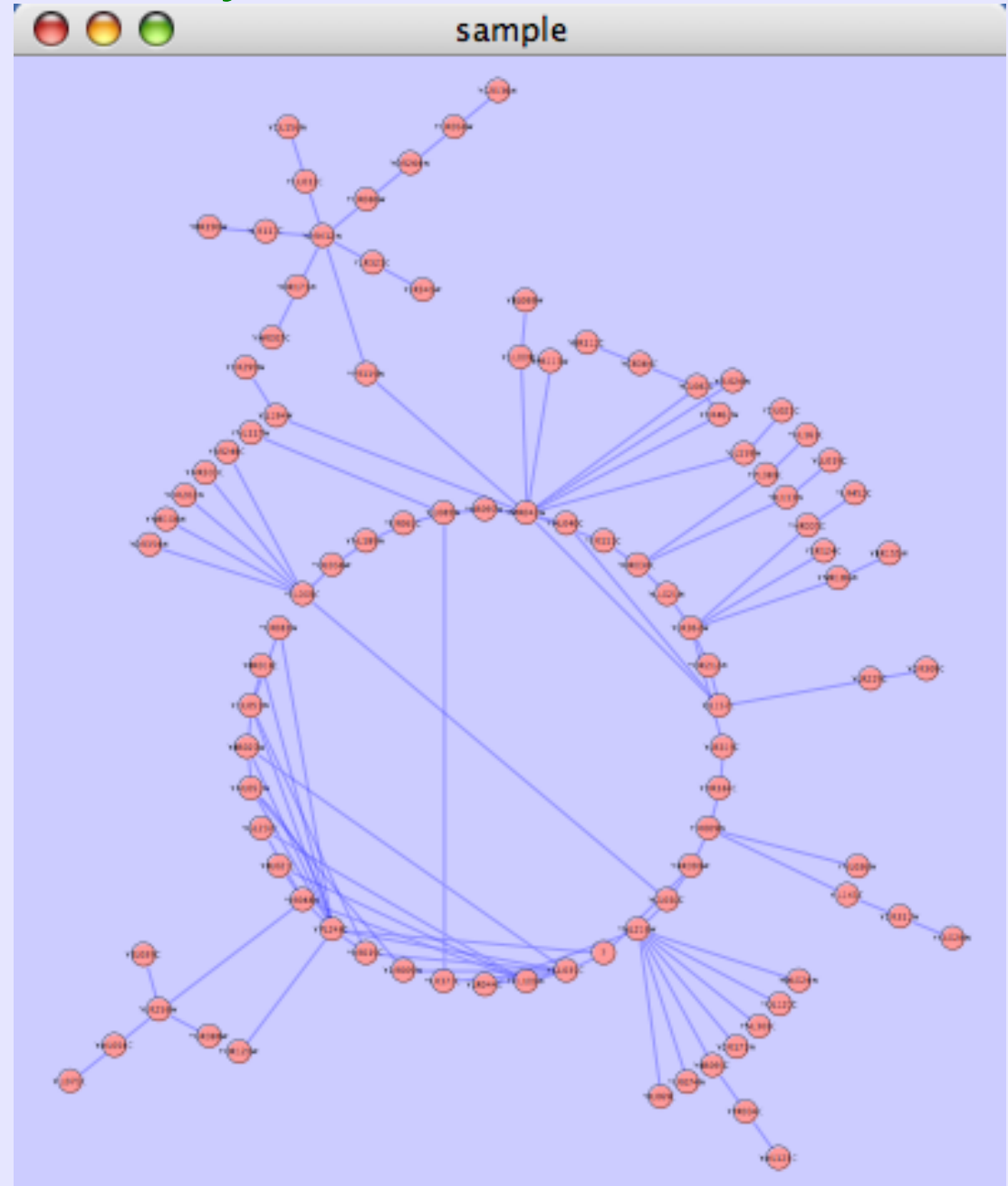
Number of matched Primary Key and Key Attribute pair: 9
First 100 entries are loaded for preview.
File Size: 415 KBytes
[Import] [Cancel]

Cytoscape Quick Tour

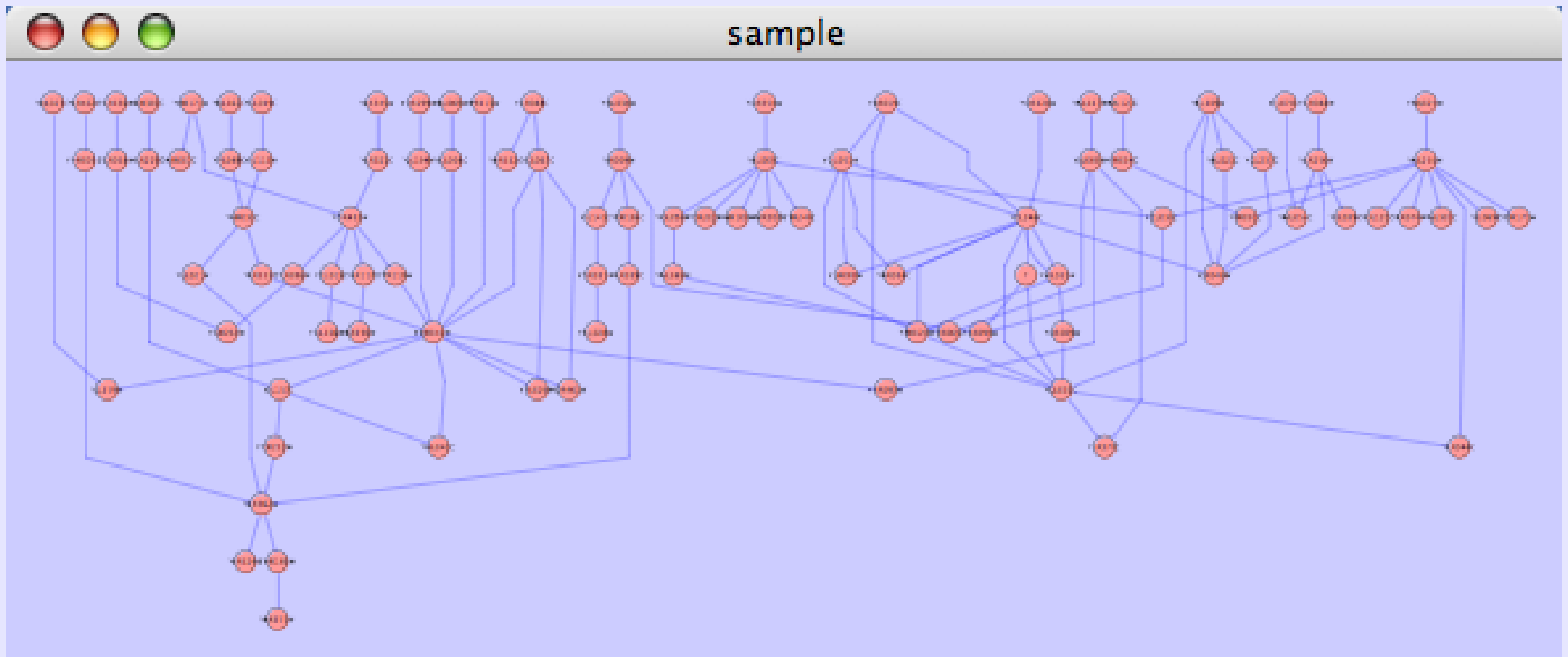
- Layouts -

Cytoscape quick tour: network layouts

Circular layout



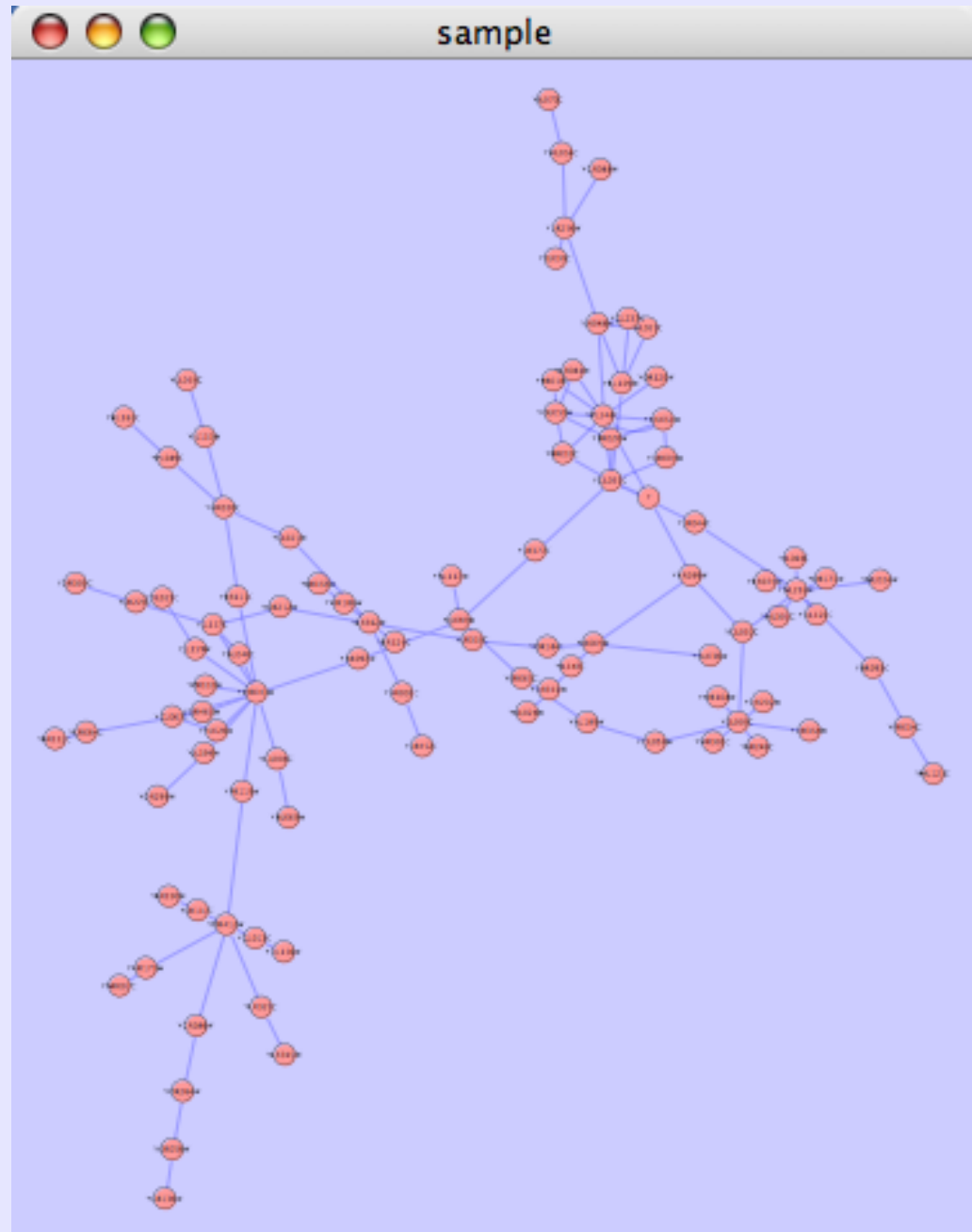
Cytoscape quick tour: network layouts



Hierarchical layout

Cytoscape quick tour: network layouts

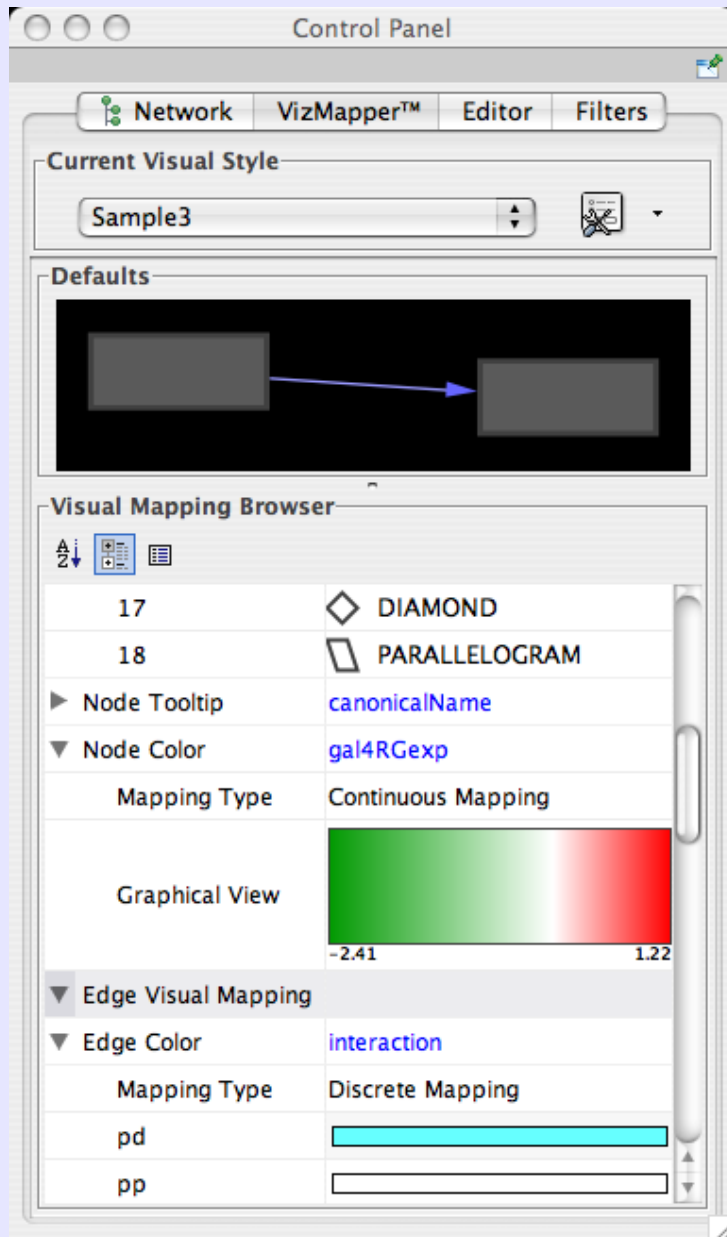
Organic layout



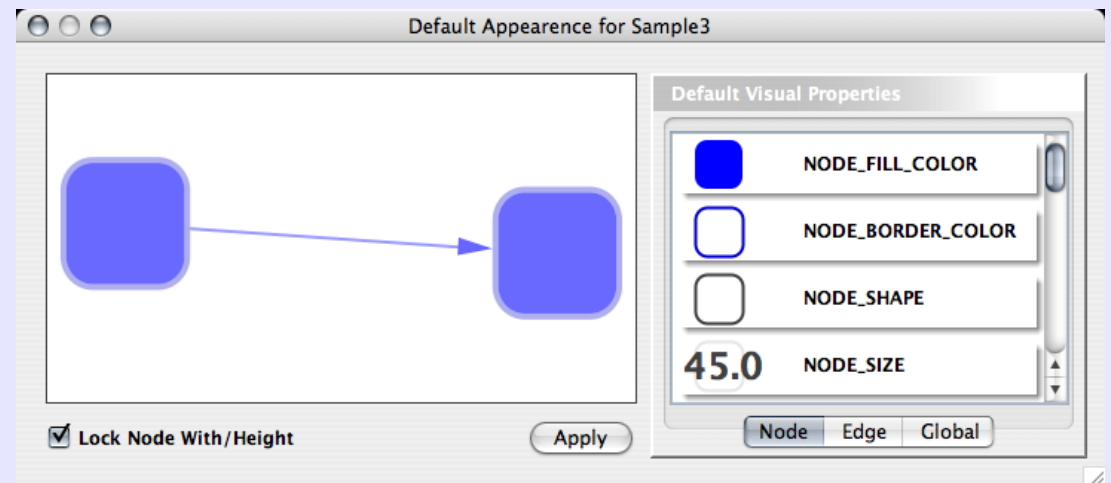
Cytoscape Quick Tour

- Visual Mapper -

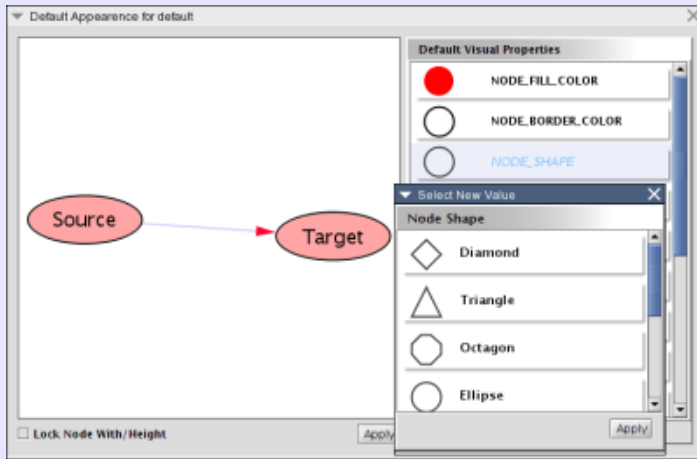
Cytoscape quick tour: Visual styles



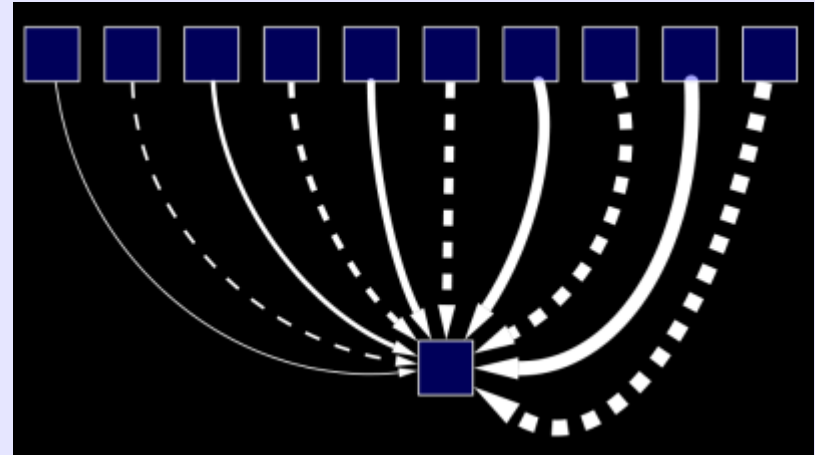
Visual Mapper



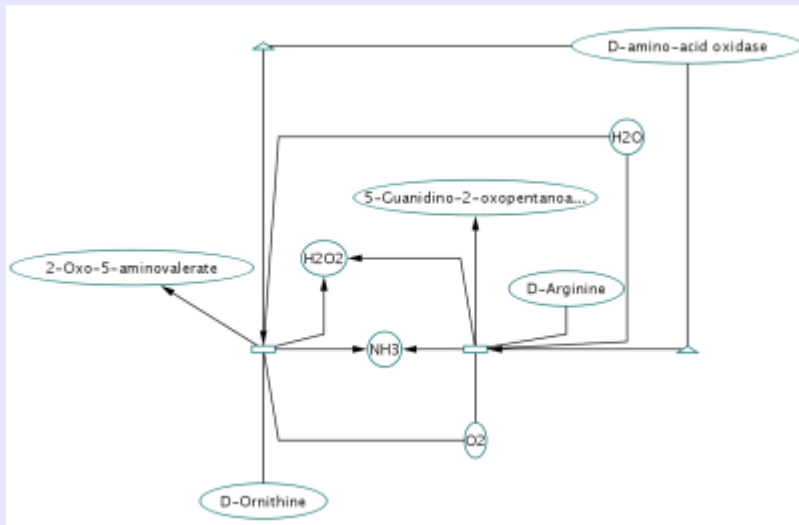
Cytoscape quick tour: Visual styles



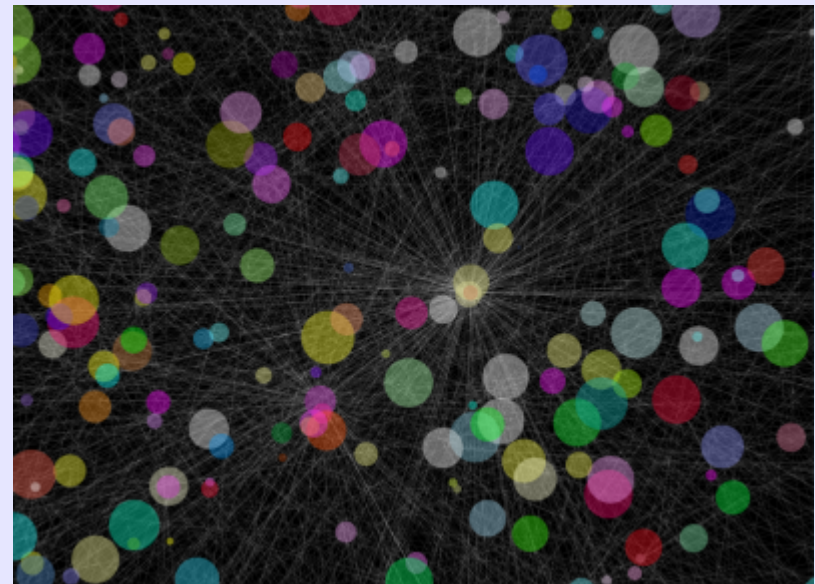
Default



Edges

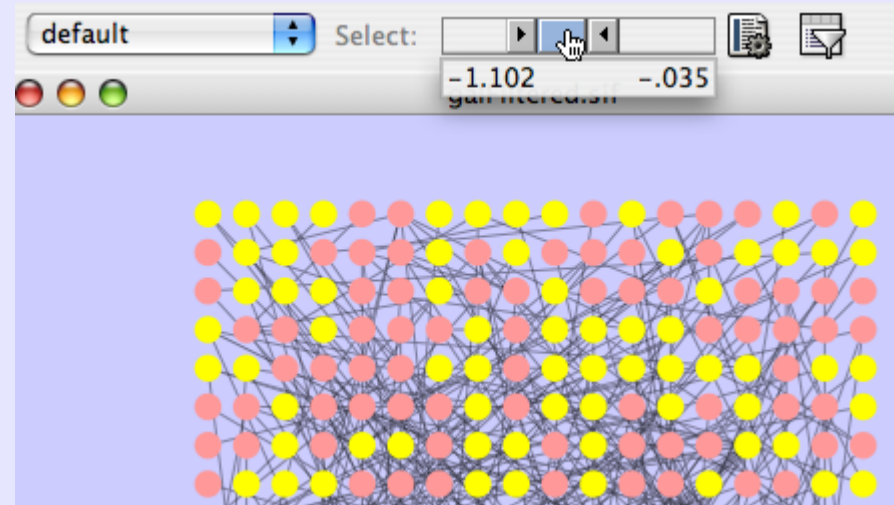
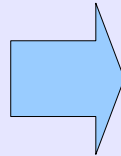
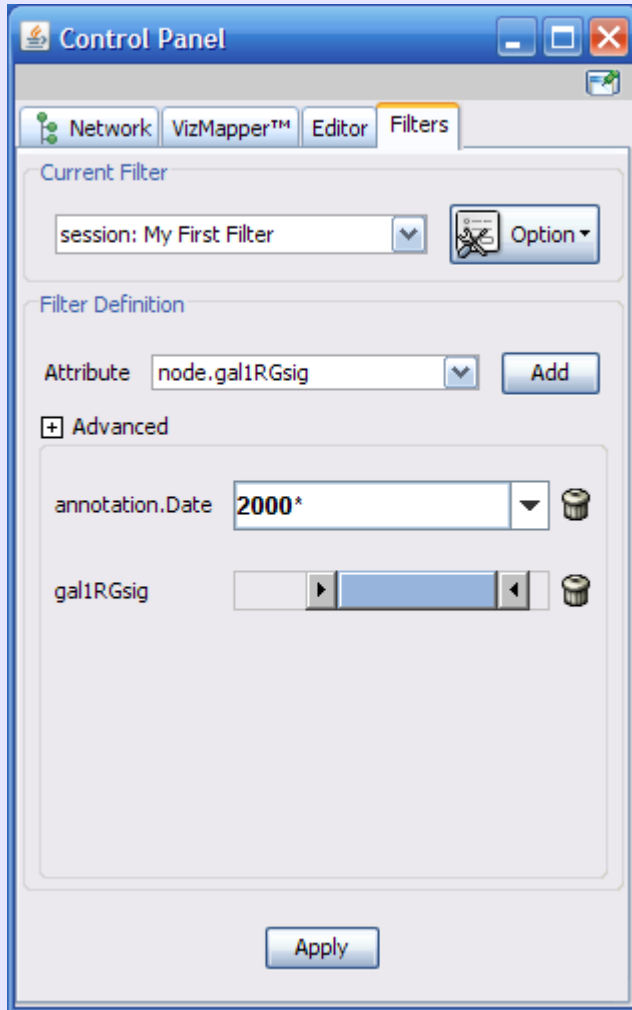


Node shapes



Nodes size and opacity

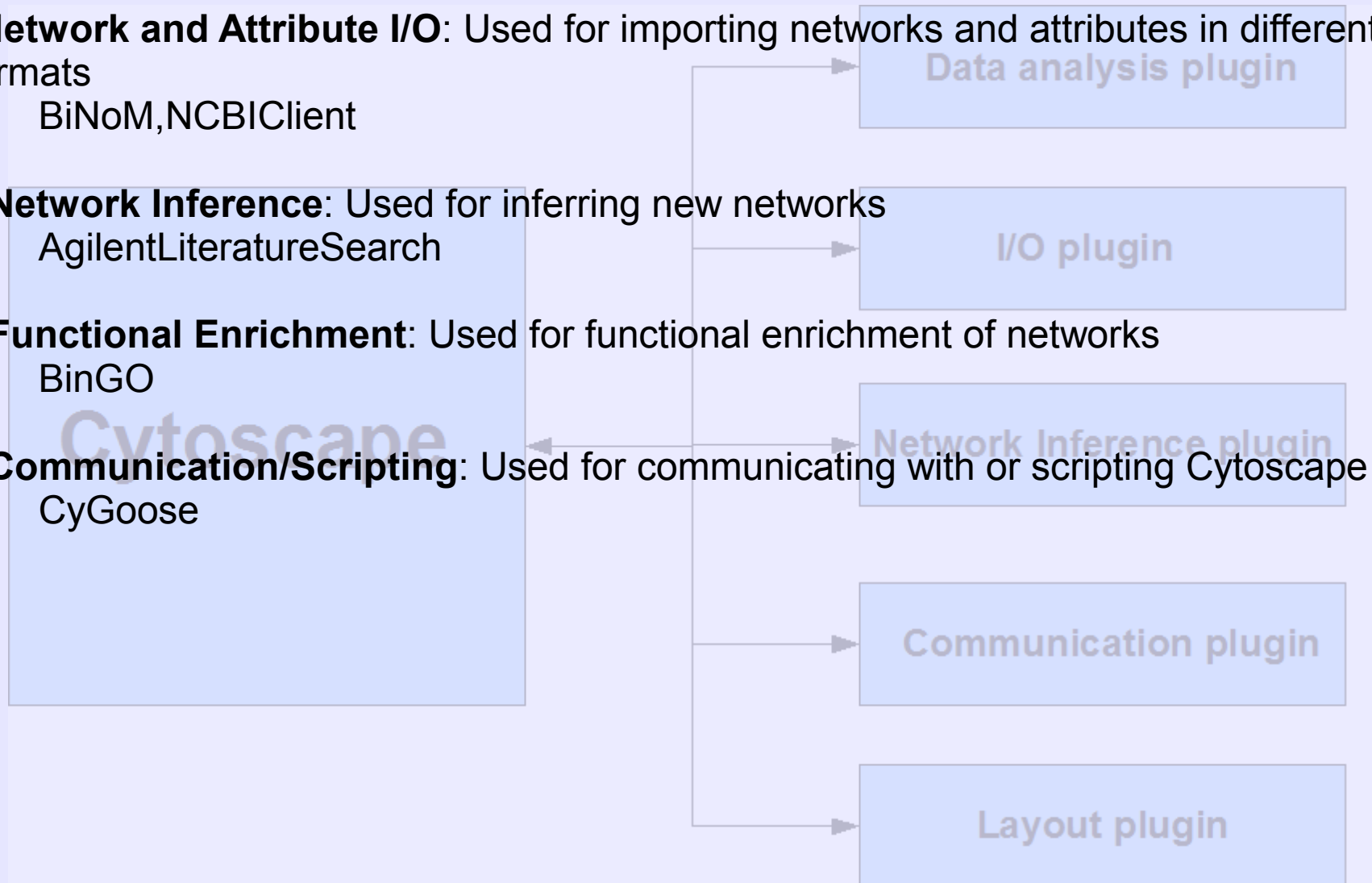
Cytoscape quick tour: Filters



- Plugins -

Cytoscape Plugins

- **Analysis:** Used for analyzing existing networks
Mcode, Metabolica, CentiScaPe (under submission)
- **Network and Attribute I/O:** Used for importing networks and attributes in different file formats
BiNoM, NCBI Client
- **Network Inference:** Used for inferring new networks
Agilent Literature Search
- **Functional Enrichment:** Used for functional enrichment of networks
BinGO
- **Communication/Scripting:** Used for communicating with or scripting Cytoscape
CyGoose



Data analysis: a small network

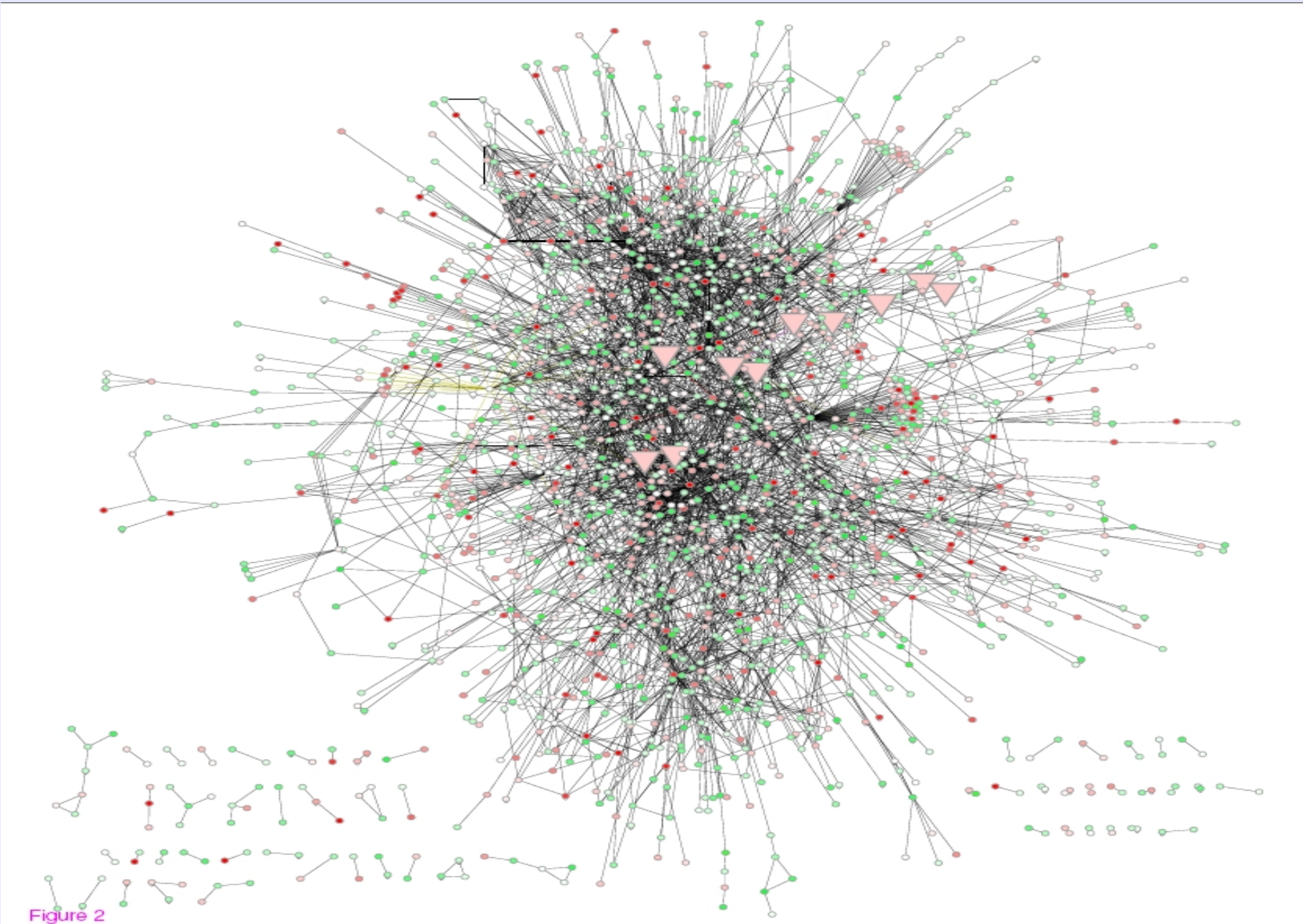
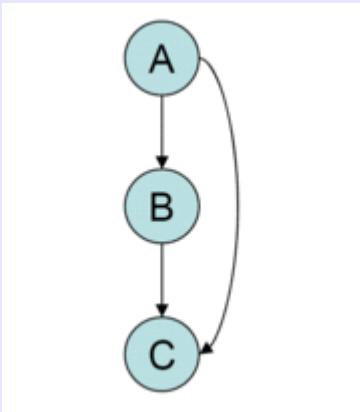


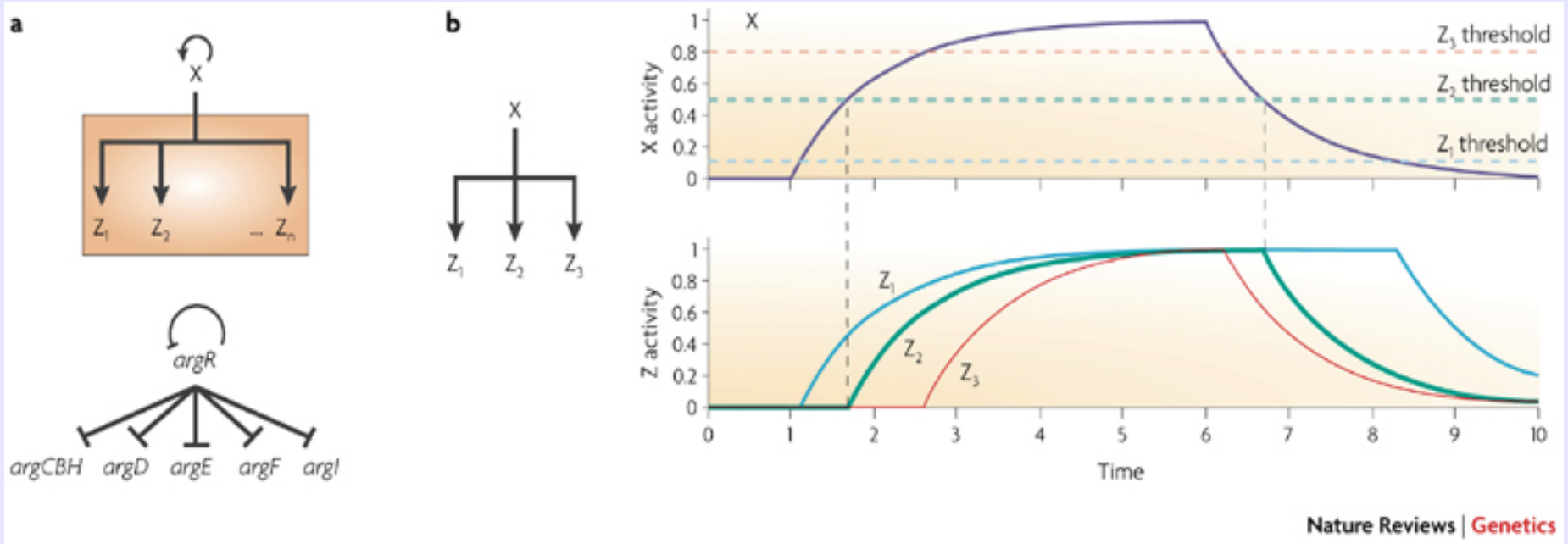
Figure 2

- Analysis Plugins - Metabolica

Metabolica plugin: network motifs introduction



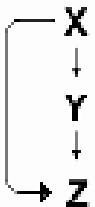
- Feed forward loops
- Feed backward loops
- Single Input Modules

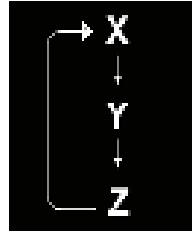


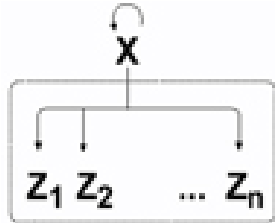
Metabolica Plugin

Metabolica 1.2


Structure selection

 feed forward loop

 feed backward loop

 single input module

Threshold

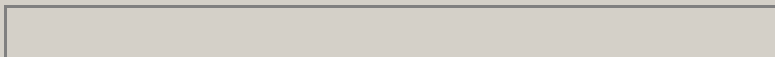


Value: 3

Computation

state: **idle**

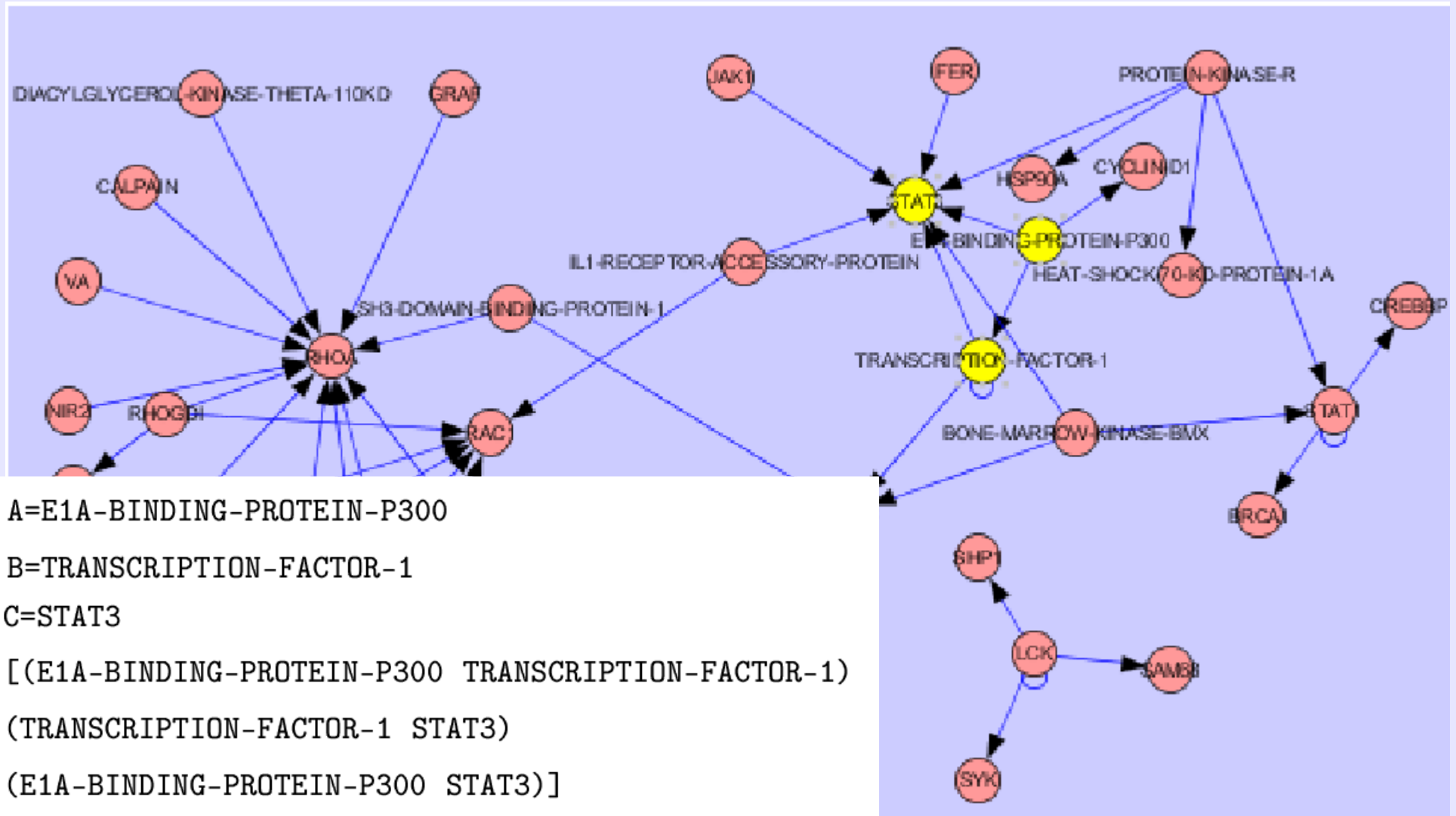
Progress



Evaluator

Warning: this plugin may need long time computation!

Metabolica Plugin



Metabolica Plugin

A=E1A-BINDING-PROTEIN-P300

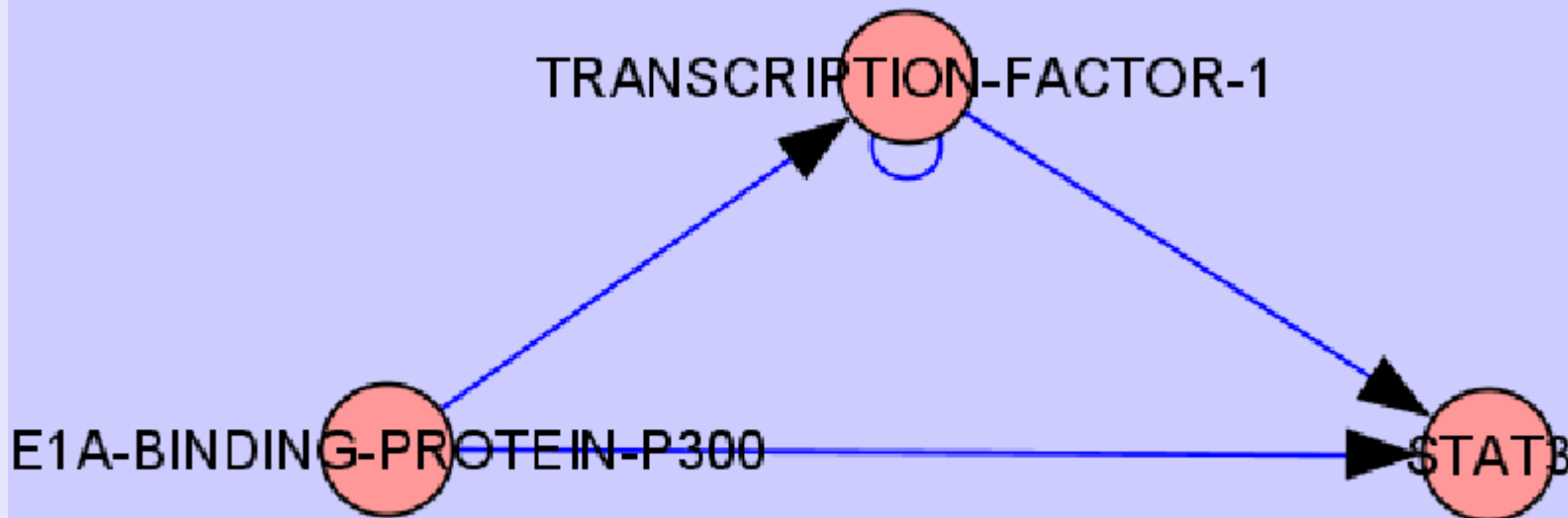
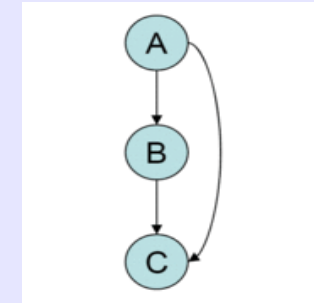
B=TRANSCRIPTION-FACTOR-1

C=STAT3

[(E1A-BINDING-PROTEIN-P300 TRANSCRIPTION-FACTOR-1)

(TRANSCRIPTION-FACTOR-1 STAT3)

(E1A-BINDING-PROTEIN-P300 STAT3)]



- Analysis Plugins - CentiScaPe

CentiScaPe plugin: centralities

- CentiScaPe can find **centralities** for a specified network. A centrality is a **topological property** for each node of a graph.

Computed centralities are:

- **Average Distance:** biologically, a big protein network with low average distance may suggest that the proteins within the network have the tendency to generate functional complexes and/or modules

- **Diameter:** biologically, gives a measure of the easiness of the proteins to communicate and/or influence their reciprocal function

- **Degree:** Biologically, gives immediate evaluation of the regulatory relevance of the node

- **Stress:** measures the number of shortest paths passing through a node. A “stressed” node is a node traversed by a high number of shortest paths .

$$C_{str}(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \sigma_{st}$$

In biological terms, the stress of a node in a biological network can indicate the relevance of a protein as functionally capable of holding together communicating nodes.

CentiScaPe plugin: centralities

- **Betweenness:** It is similar to the stress. It is calculated considering couples of nodes (v_1 , v_2) and counting the number of shortest paths linking v_1 and v_2 and passing through a node n . Then, the value is related to the total number of shortest paths linking v_1 and v_2 .

$$C_{spb}(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \delta_{st}(v) \qquad \delta_{st}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}}$$

biologically, can indicate the relevance of a protein as functionally capable of holding together communicating proteins.

- **Radiality:** if the radiality is high this means that, with respect to the diameter, the node is generally closer to the other nodes, whereas, if the radiality is low, this means that the node is peripheral.

$$C_{rad}(v) = \frac{\sum_{w \in V} (\Delta_G + 1 - dist(v, w))}{n - 1}$$

Biologically can be interpreted as the “probability” of a protein to be functionally **relevant for several other proteins**, but with the possibility to be **irrelevant for few other proteins**.

CentiScaPe plugin: centralities

- **Closeness:** high values of closeness should indicate that all other nodes are in **proximity** to node v . In contrast, low values of closeness should indicate that all other nodes are distant from node v .

$$C_{clo}(v) = \frac{1}{\sum_{w \in V} dist(v, w)}$$

Biologically, the closeness of a node indicate the “probability” of a node to be part of a **cluster**.

- **Centroid Value:** The centroid value suggests that a specific node has a central position within a graph region characterized by a high density of interacting nodes.

$$C_{cen}(v) = \min \{ f(v, w) : w \in V / \{v\} \}$$

where $f(v, w) := \gamma_v(w) - \gamma_w(v)$, and $\gamma_v(w)$ is the number of vertex closer to v than to w .

Biologically, centroid can be interpreted as the “probability” of a protein to be **functionally capable** of organizing discrete protein clusters or modules.

CentiScaPe plugin: centralities

- **Eccentricity:** if the eccentricity of the node v is high, this means that **all other nodes are in proximity**. In contrast, if the eccentricity is low, this means that there is at least one node (and all its neighbors) that is far from node v .

$$C_{ecc}(v) = \frac{1}{\max\{dist(v, w) : w \in V\}}$$

Biologically, can be interpreted as the easiness of a protein to be functionally reached by all other proteins in the network

CentiScaPe plugin

Control Panel

Network \ VizMapper™ \ Filters \ Editor

Implemented centralities

- Diameter
- Average Distance
- Degree
- Radiality
- Closeness
- Stress
- Betweenness
- Centroid Value
- Eccentricity

Select All Unselect All

Finished: 168 nodes worked

Start Stop Exit

Results Panel

CentiScaPe Results

CentiScaPe Betweenness

Centrality value is: less/equals threshold more/equals threshold

Threshold: 4.285714285714286

update restore

CentiScaPe Node degree

Centrality value is: less/equals threshold more/equals threshold

Threshold: 1.1071428571428572

update restore

CentiScaPe Stress

Centrality value is: less/equals threshold more/equals threshold

Threshold: 4.380952380952381

update restore

Options and views

filter type: AND OR

plot by node: YPR048W plot

Data Panel

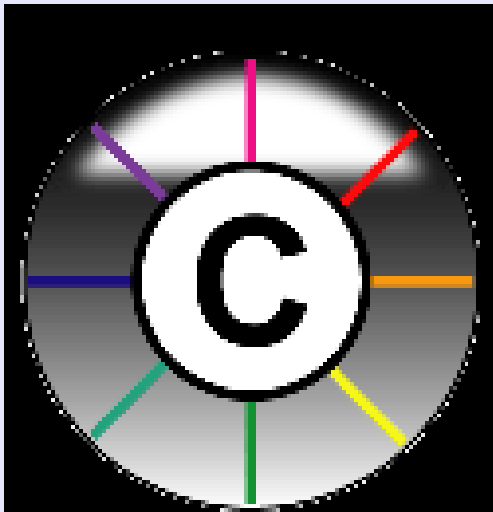
ID	canonicalName
YHR179W	YHR179W
YLR264W	YLR264W
YNL301C	YNL301C
YPR110C	YPR110C
YIL074C	YIL074C
YLR175W	YLR175W
YER102W	YER102W
YOR204W	YOR204W

Node Attribute Browser Edge Attribute Browser Network Attribute Browser

Welcome to Cytoscape 2.5 Right-click + drag to ZOOM Middle-click + drag to PAN

Conclusions

- Cytoscape is a “**must have**” tool for biological and biotechnology research
- **Homogeneous** platform
- **Open source** (nearly free)
- **Hardly updated and many contributions**



<http://www.cytoscape.org>

**THANK
YOU!**