Cytoscape:un approccio sistematico all'analisi dei dati sperimentali

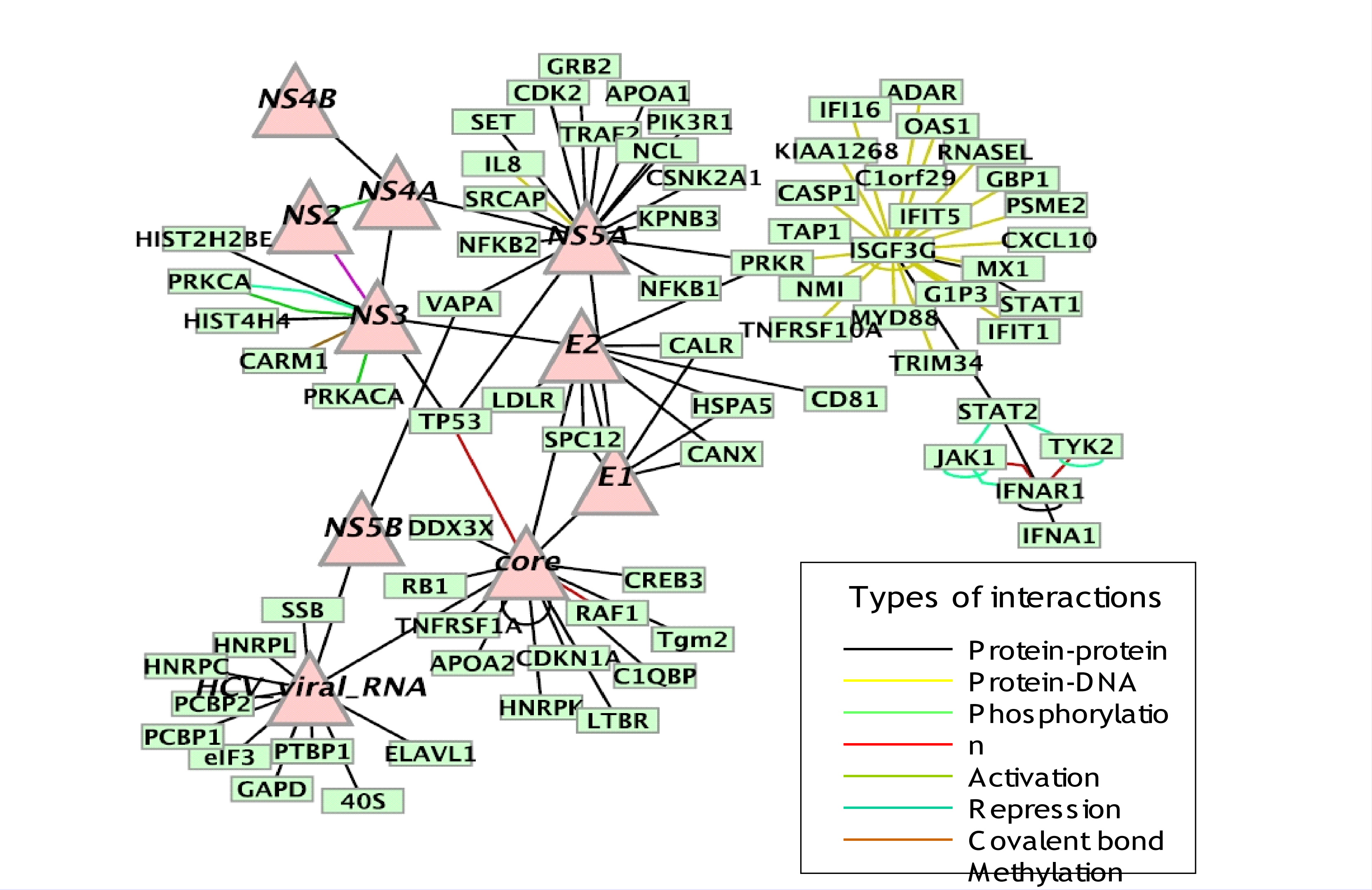
Michele Petterlini

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.

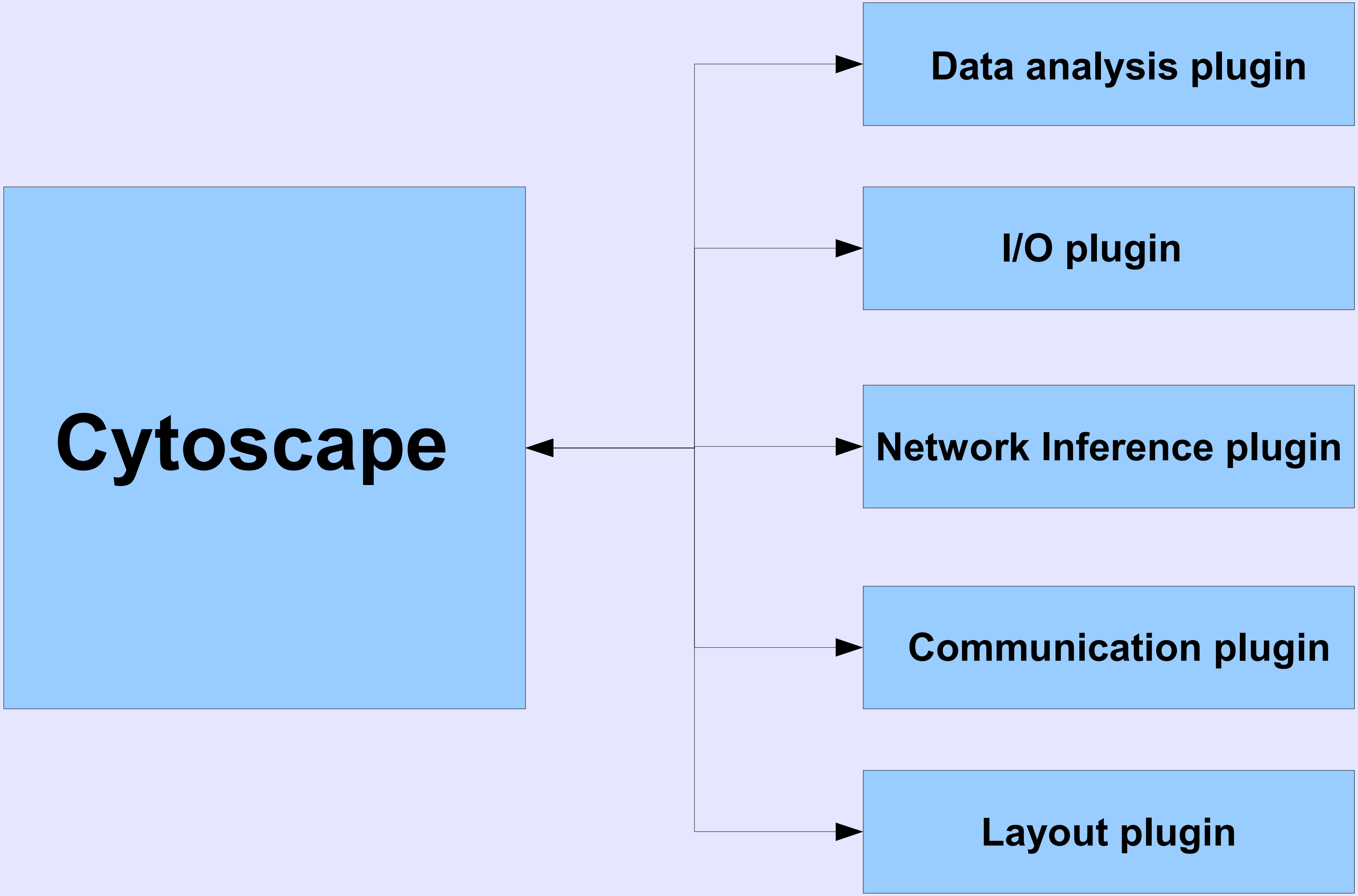
- A software “Core” provides basic functionality to layout and query the network; to **visually integrate** the network with expression profiles, phenotypes, and other molecular states; and to **link the network to databases** of functional annotations.

- 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

Data analysis plugin

I/O plugin

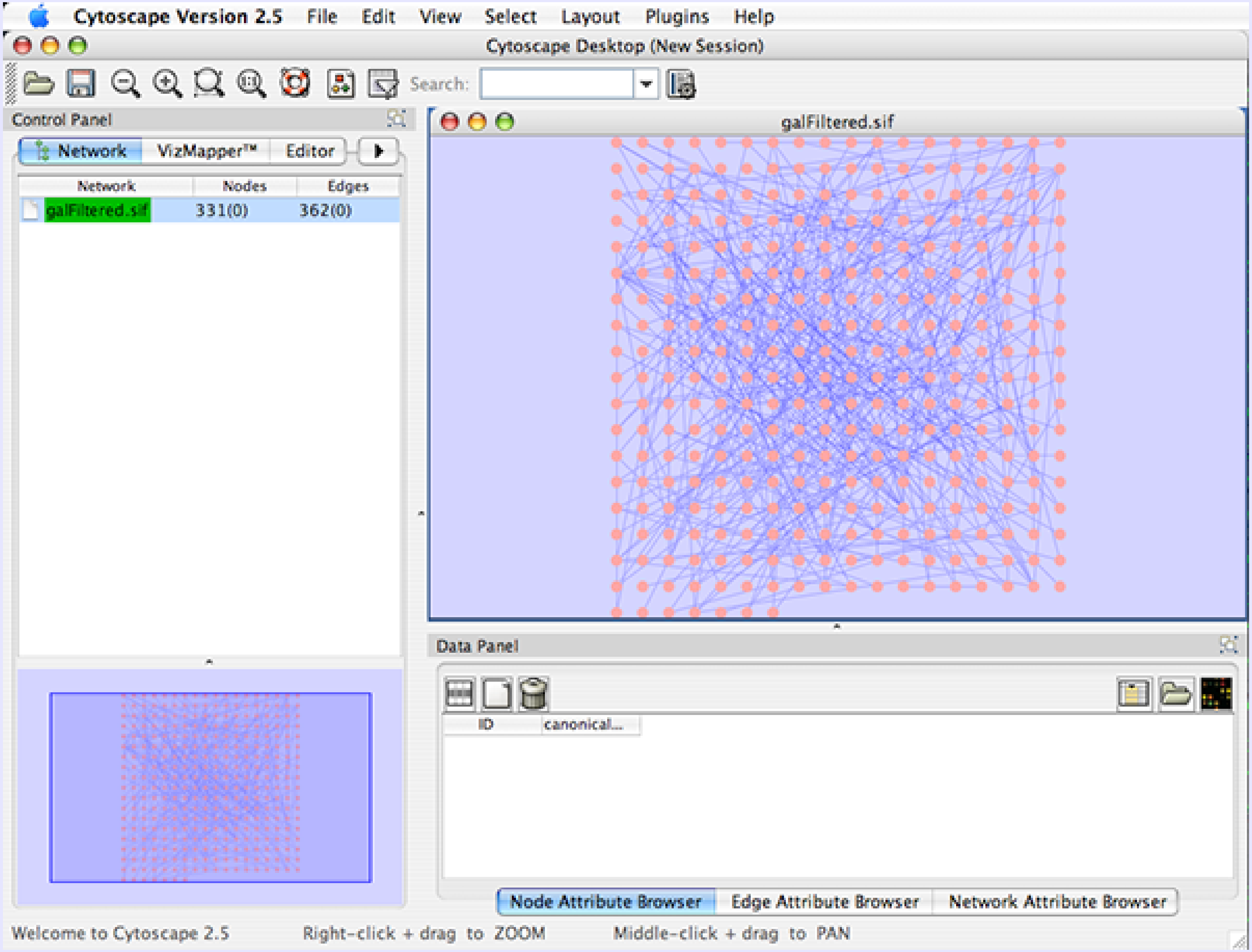
Network Inference plugin

Communication plugin

Layout plugin

Cytoscape Quick Tour
- Base features -

Cytoscape quick tour: main window

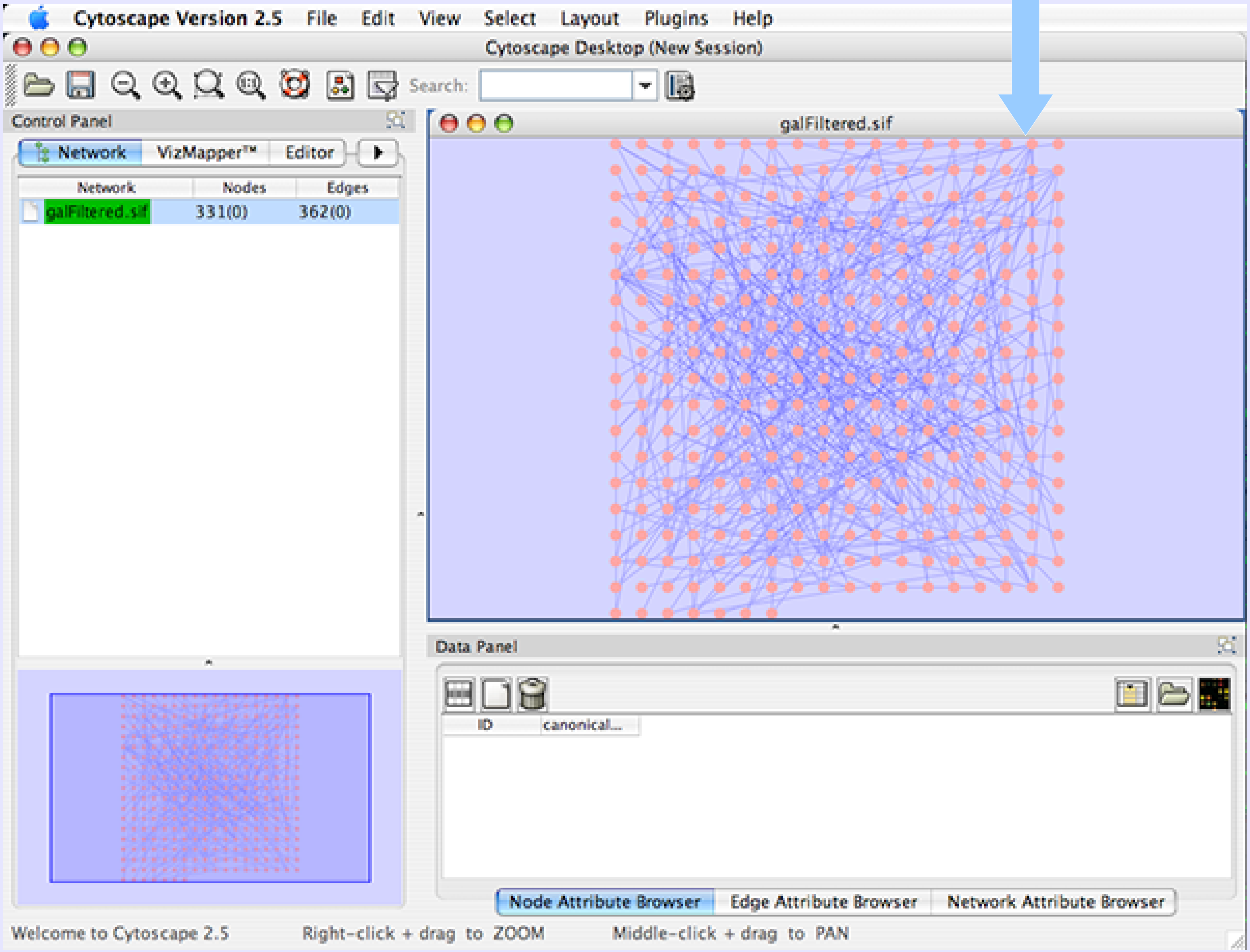


Network	Nodes	Edges
galFiltered.sif	331(0)	362(0)

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

Cytoscape quick tour: main window

Network window



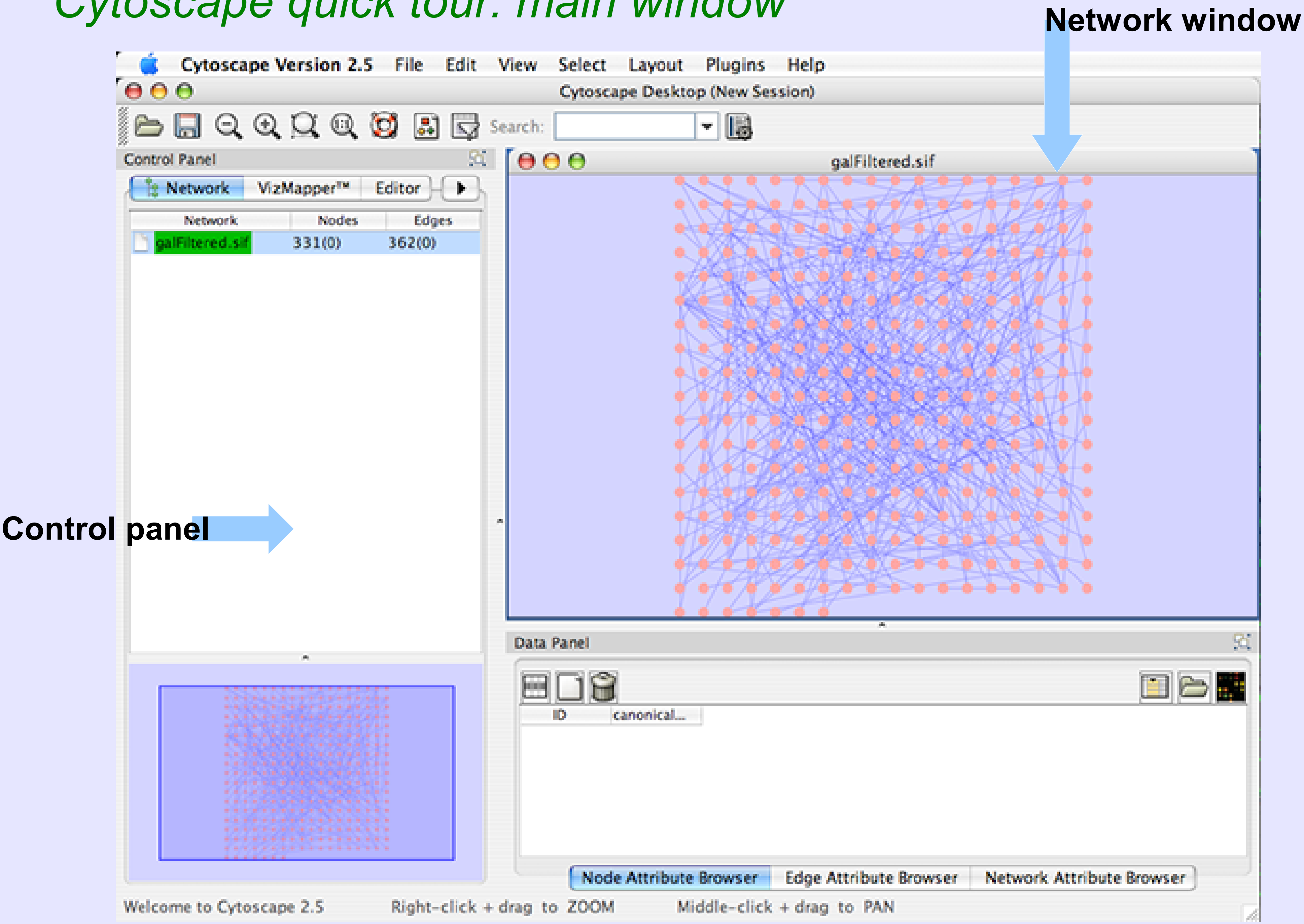
Network	Nodes	Edges
galFiltered.sif	331(0)	362(0)

Welcome to Cytoscape 2.5

Right-click + drag to ZOOM

Middle-click + drag to PAN

Cytoscape quick tour: main window



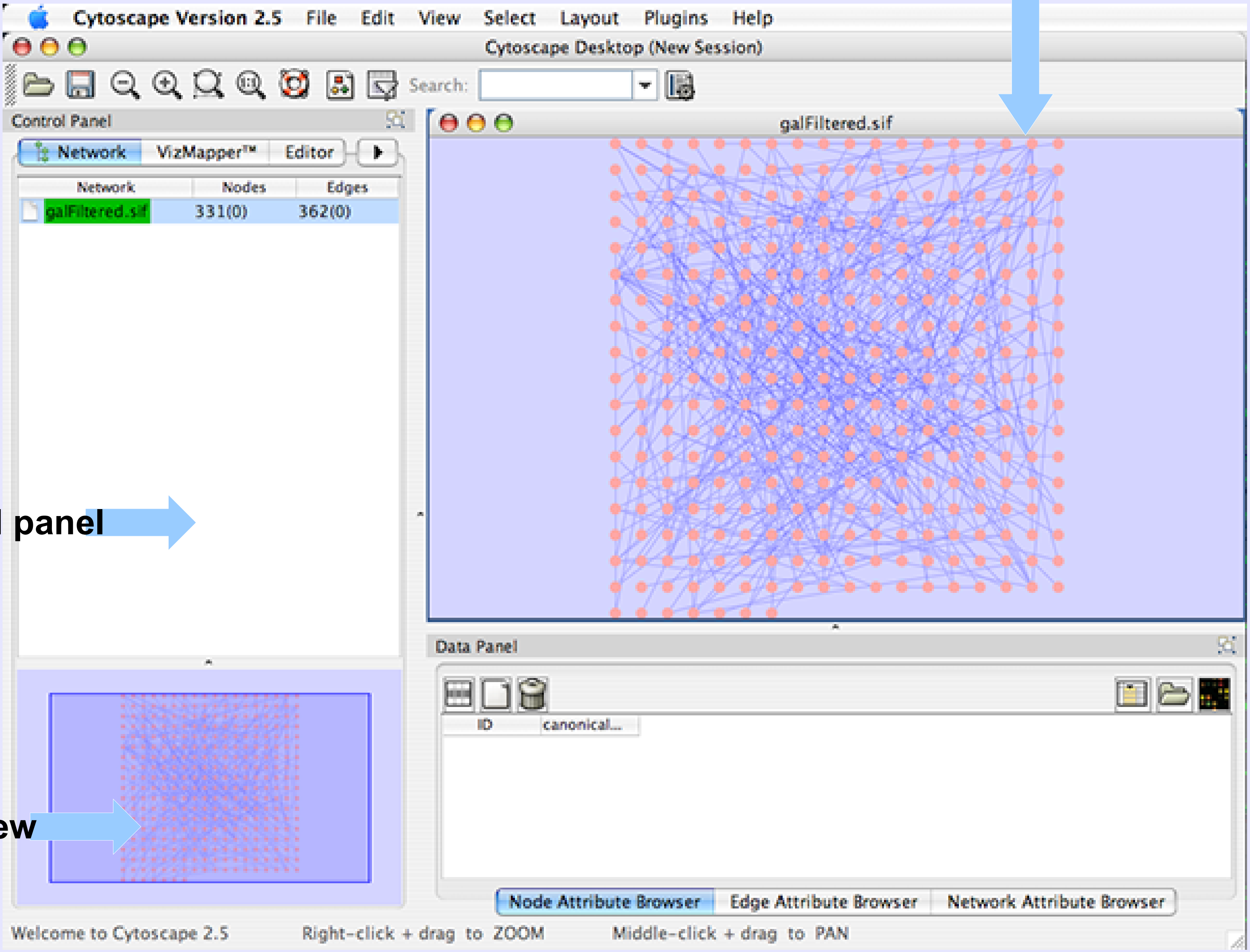
Network window

Control panel

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

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

Cytoscape quick tour: main window



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

Welcome to Cytoscape 2.5

Right-click + drag to ZOOM

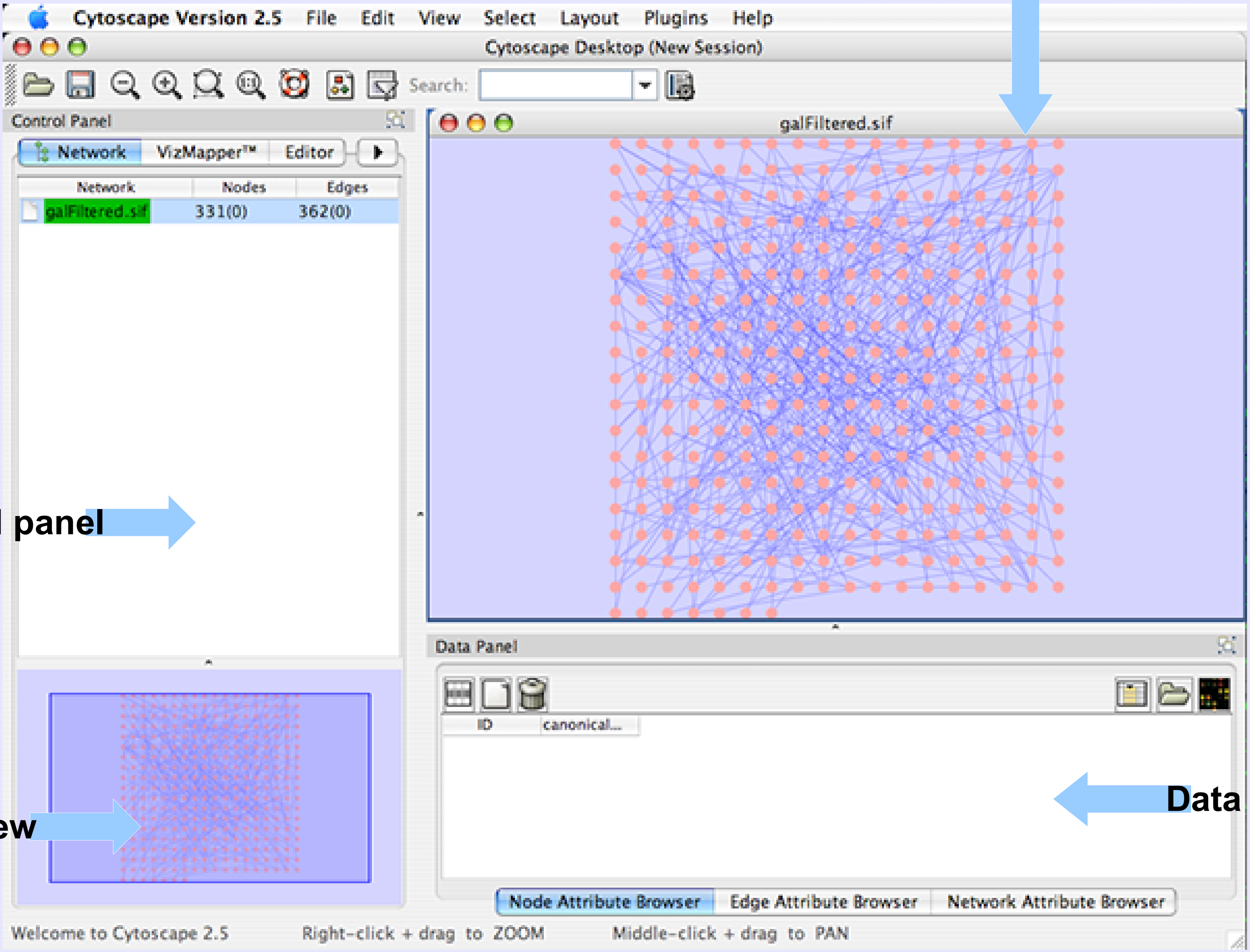
Middle-click + drag to PAN

Network window

Control panel

OverView

Cytoscape quick tour: main window



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

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

Network window

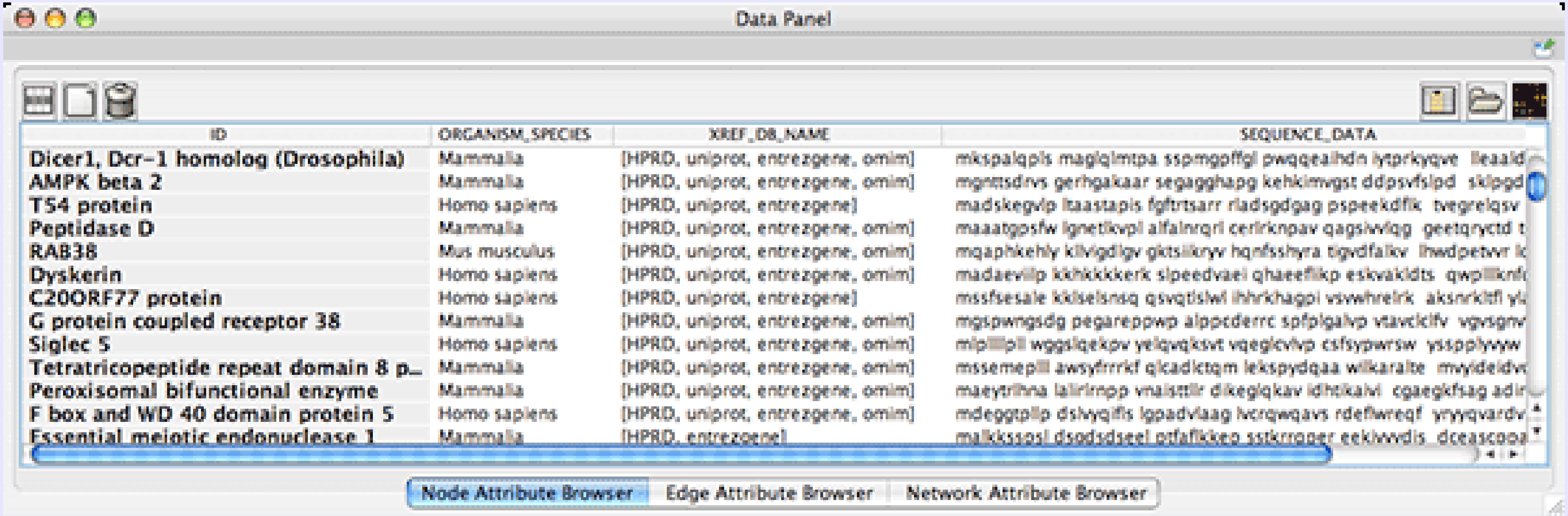
Control panel

OverView

Data panel

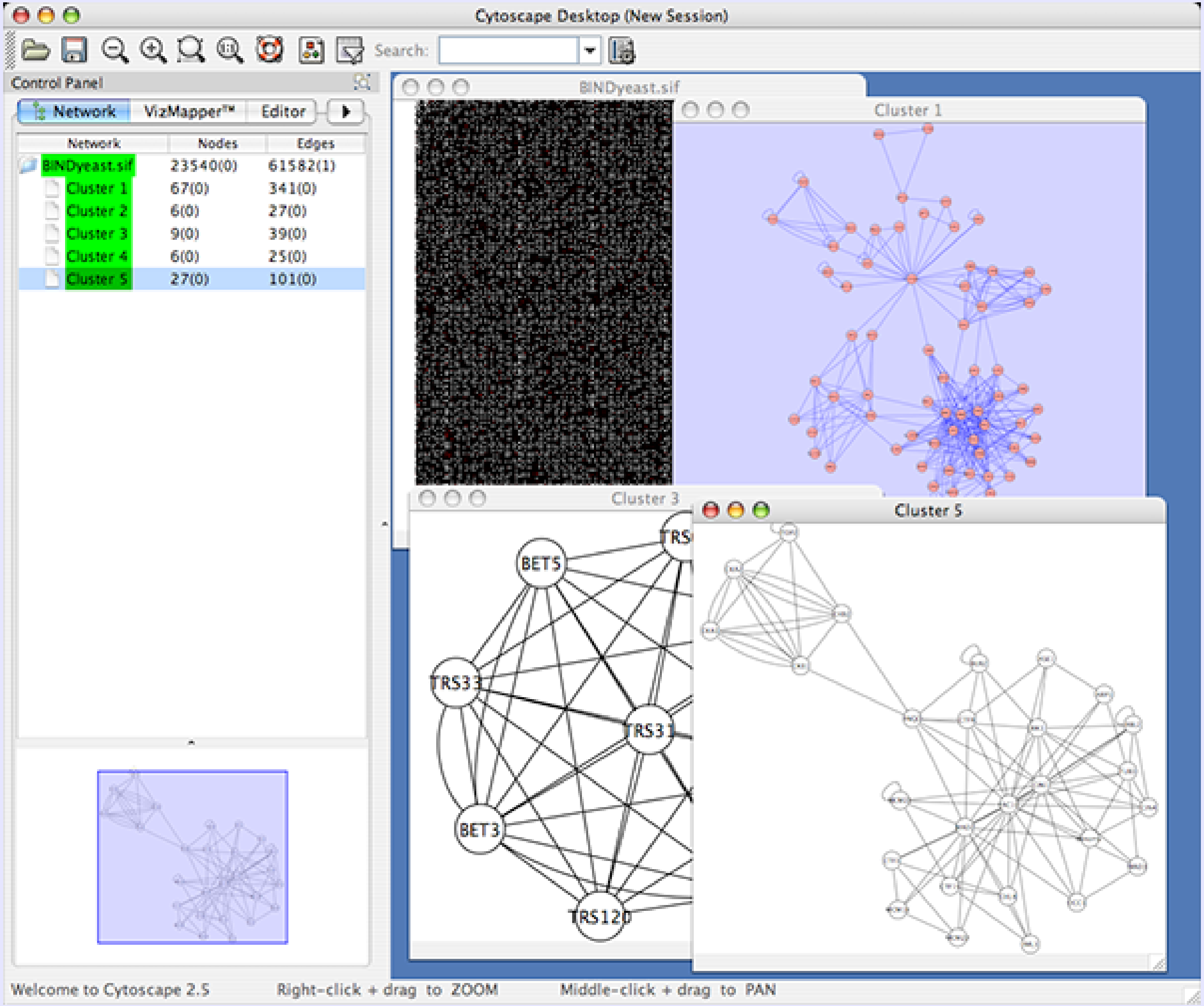
Cytoscape quick tour: data panel

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



ID	ORGANISM_SPECIES	XREF_DB_NAME	SEQUENCE_DATA
Dicer1, Dcr-1 homolog (Drosophila)	Mammalia	[HPRD, uniprot, entrezgene, omim]	mkspalqpls maglqlmtpa sspmgpffgl pwqqeaihdn iytprkyqve lleaald
AMPK beta 2	Mammalia	[HPRD, uniprot, entrezgene, omim]	mgnttsdrvs gerhgakaar segagghapg kehkimvgst ddpsvfslpd sklpgd
T54 protein	Homo sapiens	[HPRD, uniprot, entrezgene]	madskegvlp ltaastapis fgftrtsarr rladsgdgag pspeekdflk tvegrelqsv
Peptidase D	Mammalia	[HPRD, uniprot, entrezgene, omim]	maaatgpsfw lgnetlkvpl alfalnrqrl cerlrknpav qagsiwiqg geetqryctd t
RAB38	Mus musculus	[HPRD, uniprot, entrezgene, omim]	mqaphkehly klivigdlgv gktsiikryv hqnfsshyra tigvdfalkv lhwdpetvr k
Dyskerin	Homo sapiens	[HPRD, uniprot, entrezgene, omim]	madaevilp kkhkkkkerk slpeedvaei qhaeeflikp eskvakldts qwpllkn
C20ORF77 protein	Homo sapiens	[HPRD, uniprot, entrezgene]	mssfsesale kklselsnsq qsvqtlslwl lhhrkhagpi vswhrelrk aksnrkltfl y
G protein coupled receptor 38	Mammalia	[HPRD, uniprot, entrezgene, omim]	mgspwngsdg pegareppwp alppcderrc spfplgalvp vtavclclfv vgvsgnv
Siglec 5	Homo sapiens	[HPRD, uniprot, entrezgene, omim]	mipllllpll wggslqekpv yelqvqksvt vqeglcvlvp csfsypwrsw yssppiyww
Tetratricopeptide repeat domain 8 p...	Mammalia	[HPRD, uniprot, entrezgene, omim]	mssemepll awsyfrrkf qlcadlctqm lekspydqaa wilkaralte mvyideidv
Peroxisomal bifunctional enzyme	Mammalia	[HPRD, uniprot, entrezgene, omim]	maeytrlhna lalirmpp vnalstllr dlkeglqkav idhtkalvl cgaegkfsag adi
F box and WD 40 domain protein 5	Homo sapiens	[HPRD, uniprot, entrezgene, omim]	mdeggtpllp dslvyqifls lgpadvlaag lvcrqwqavs rdeflwreqf yryyqvardv
Essential meiotic endonuclease 1	Mammalia	[HPRD, entrezgene]	malkkssosl dsodsdseel otfafkkeo sstkrroper eekiwwdis dceascopa

Cytoscape quick tour: network views

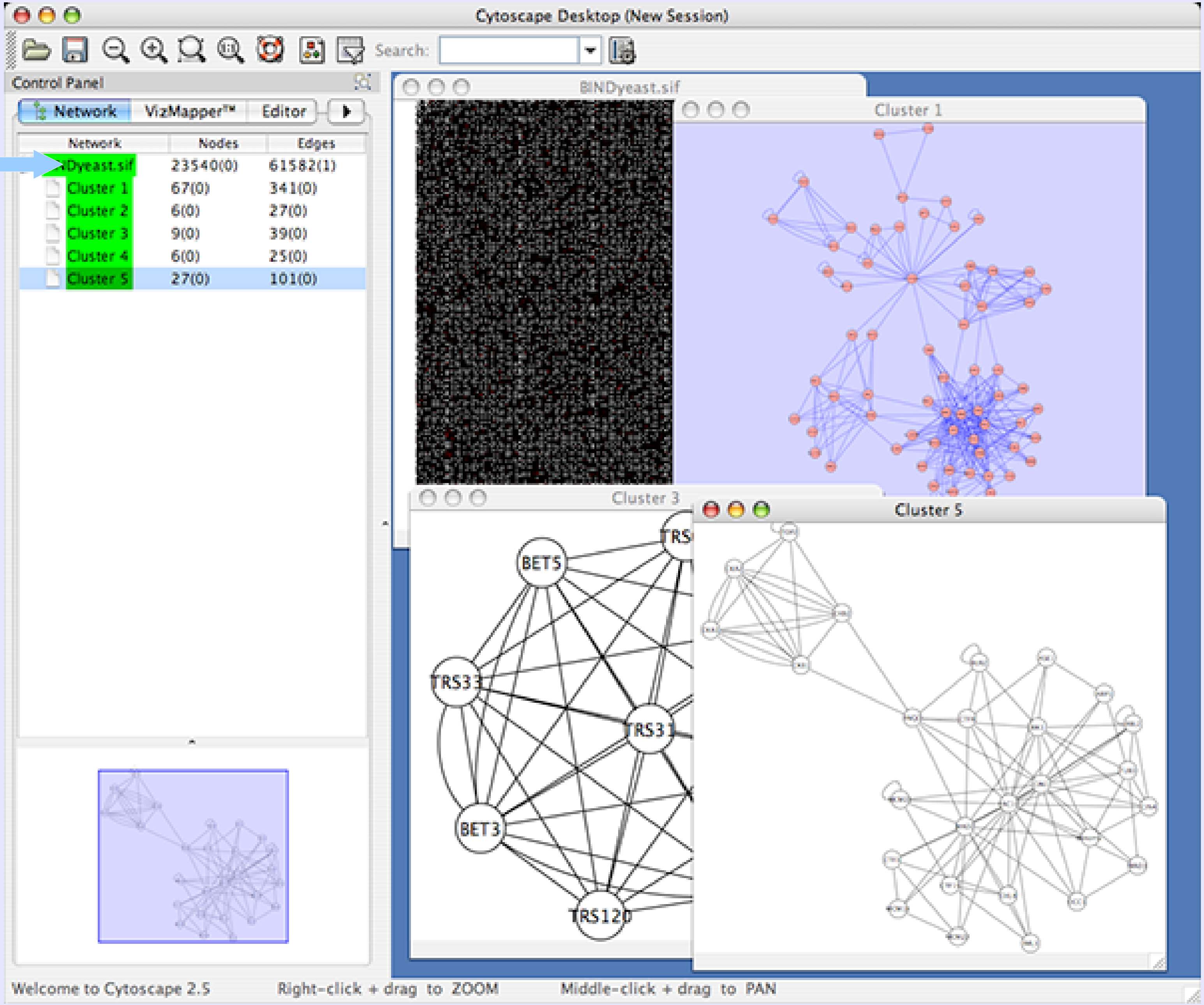


Network	Nodes	Edges
BINDyeast.sif	23540(0)	61582(1)
Cluster 1	67(0)	341(0)
Cluster 2	6(0)	27(0)
Cluster 3	9(0)	39(0)
Cluster 4	6(0)	25(0)
Cluster 5	27(0)	101(0)

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

Cytoscape quick tour: network views

Network



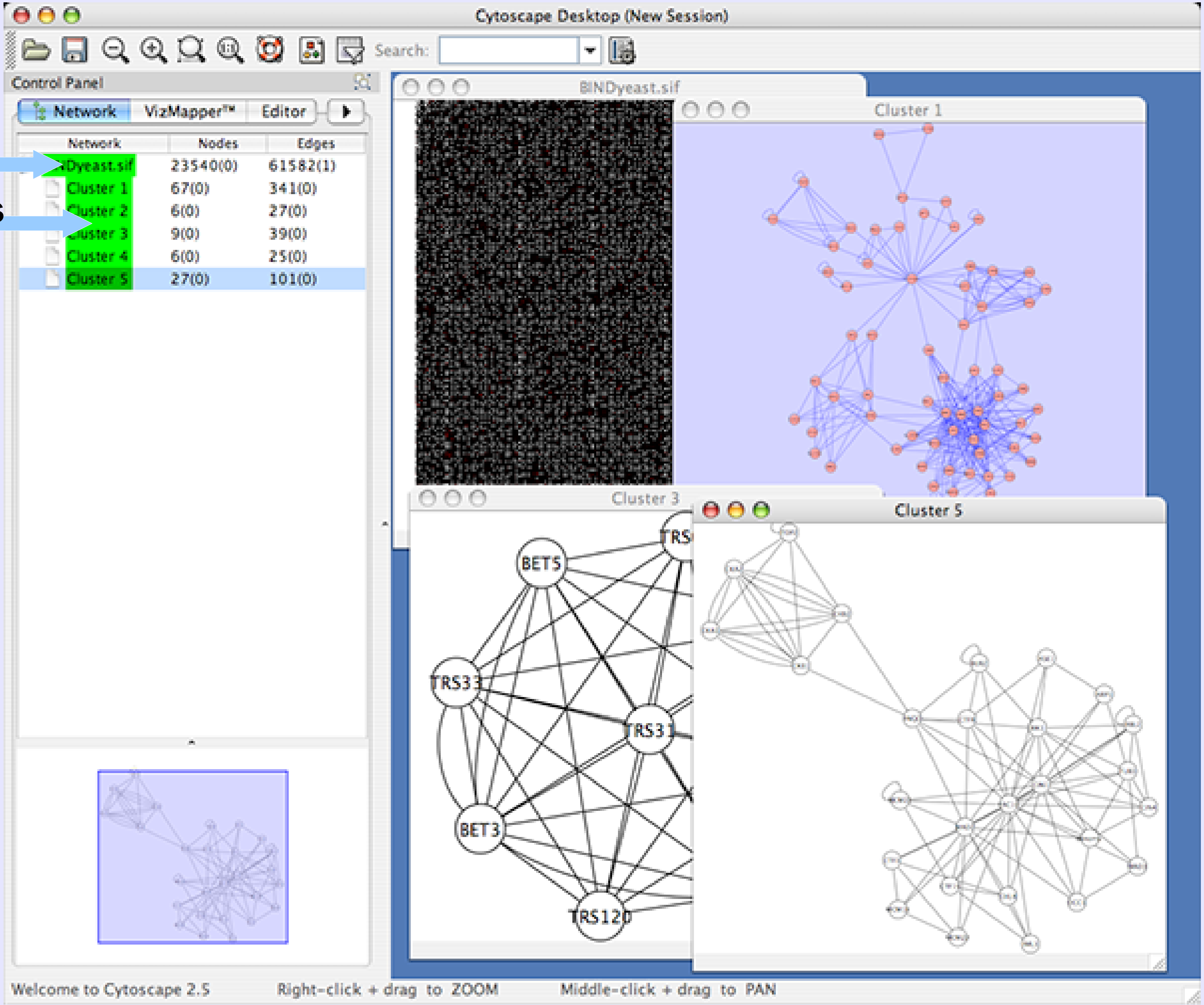
Network	Nodes	Edges
Dyeast.sif	23540(0)	61582(1)
Cluster 1	67(0)	341(0)
Cluster 2	6(0)	27(0)
Cluster 3	9(0)	39(0)
Cluster 4	6(0)	25(0)
Cluster 5	27(0)	101(0)

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

Cytoscape quick tour: network views

Network

Views



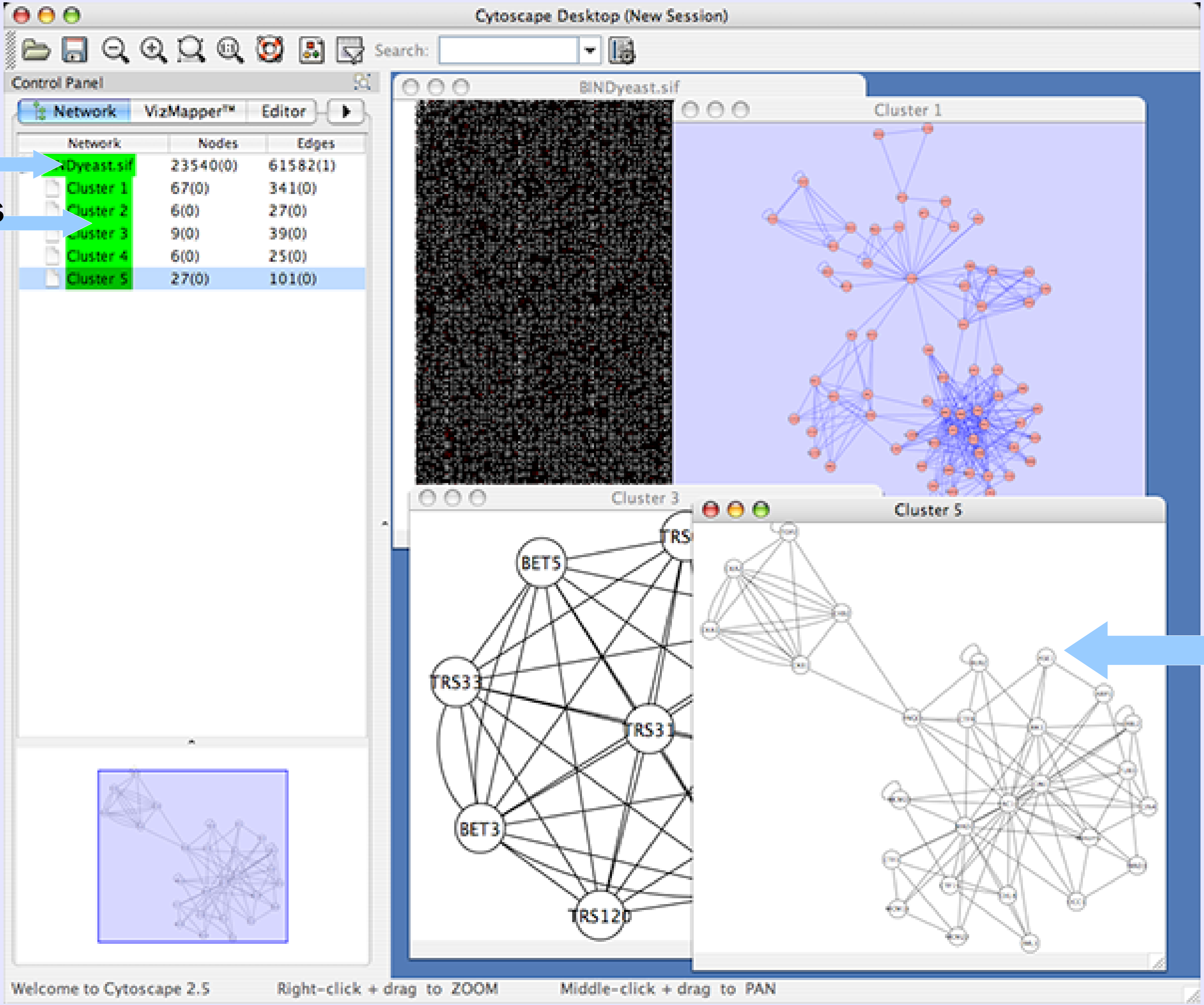
Network	Nodes	Edges
Dyeast.sif	23540(0)	61582(1)
Cluster 1	67(0)	341(0)
Cluster 2	6(0)	27(0)
uster 3	9(0)	39(0)
Cluster 4	6(0)	25(0)
Cluster 5	27(0)	101(0)

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

Cytoscape quick tour: network views

Network

Views



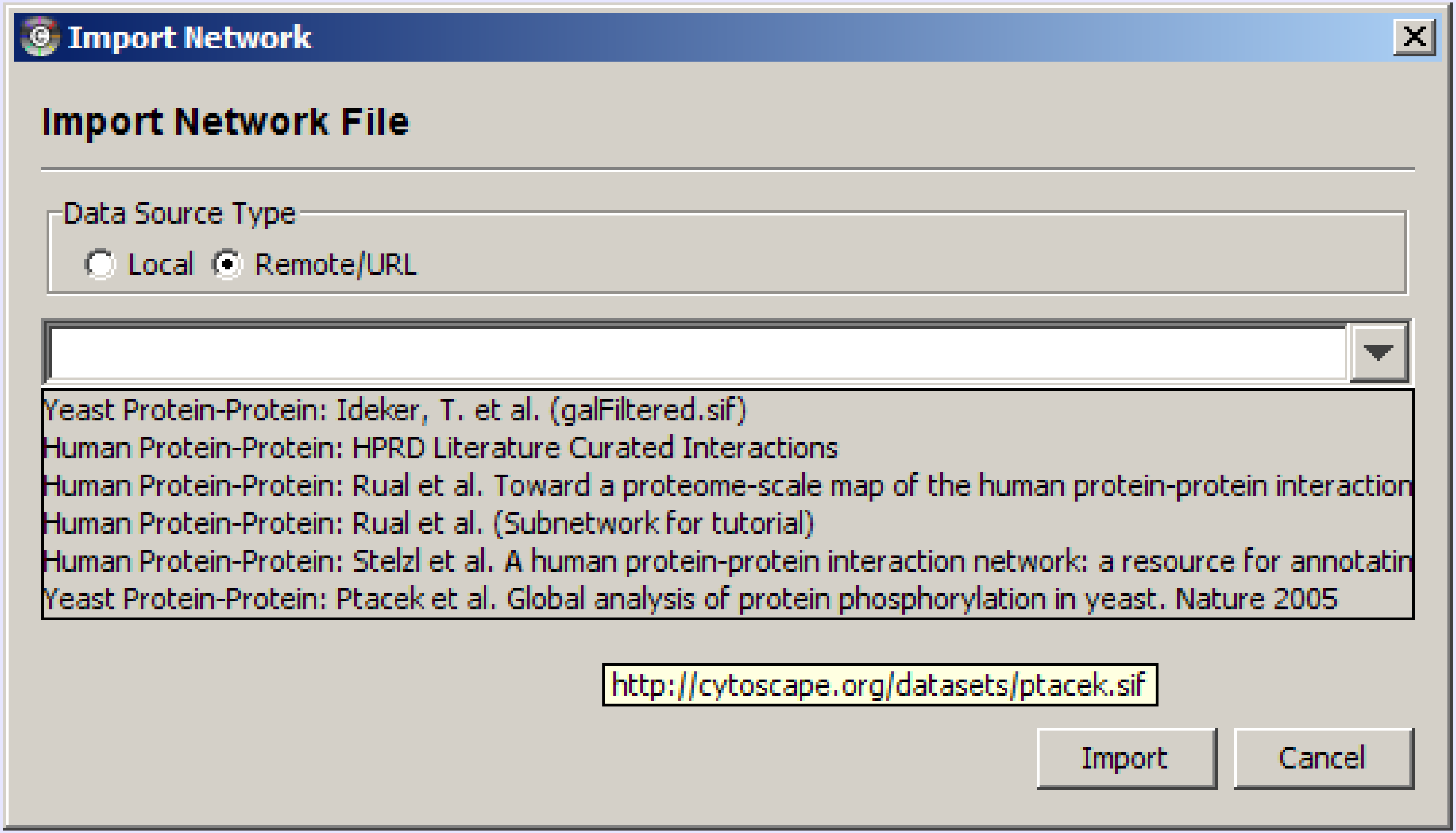
Network	Nodes	Edges
Dyeast.sif	23540(0)	61582(1)
Cluster 1	67(0)	341(0)
Cluster 2	6(0)	27(0)
Cluster 3	9(0)	39(0)
Cluster 4	6(0)	25(0)
Cluster 5	27(0)	101(0)

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)**

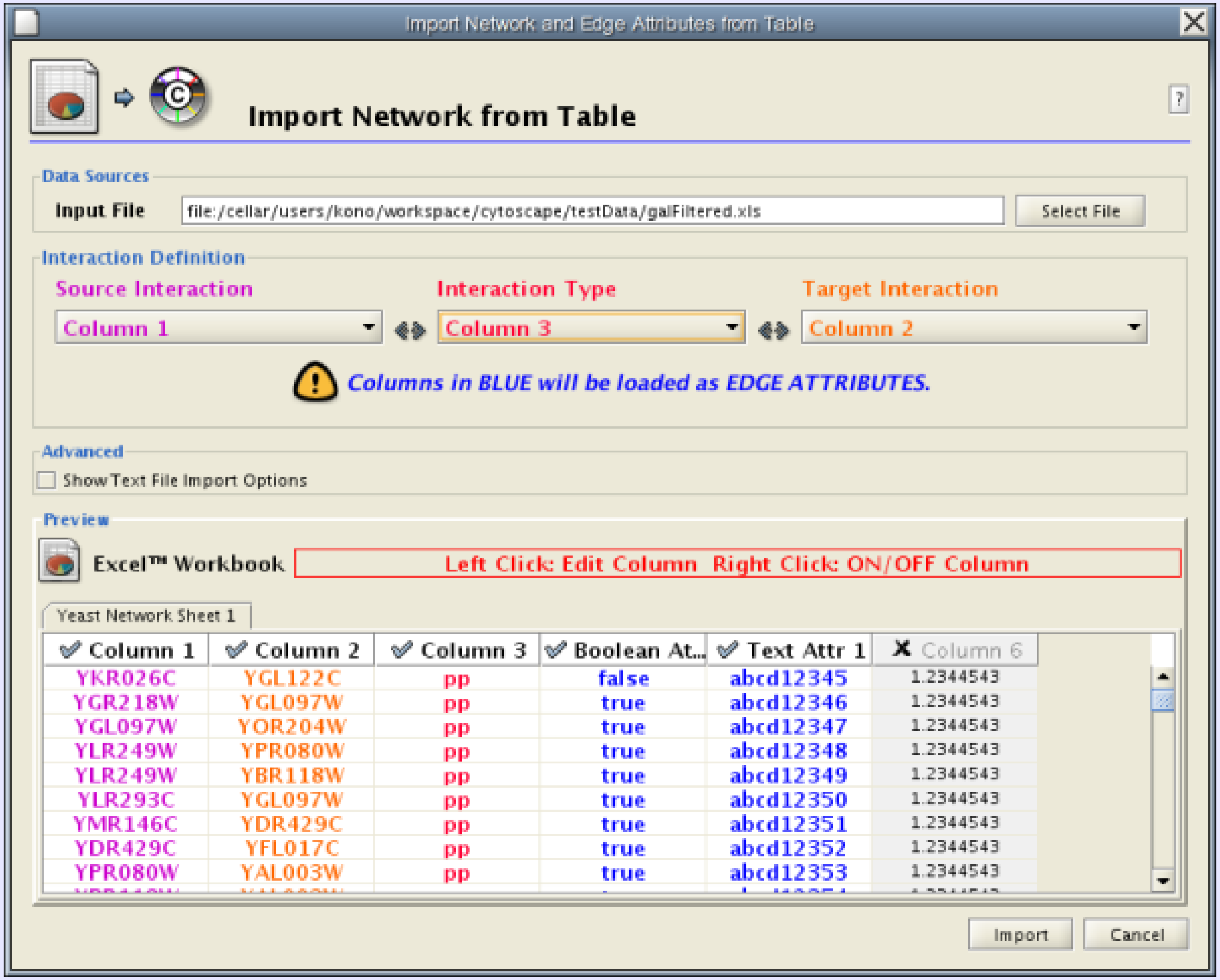


Yeast Protein-Protein: Ideker, T. et al. (galFiltered.sif)
Human Protein-Protein: HPRD Literature Curated Interactions
Human Protein-Protein: Rual et al. Toward a proteome-scale map of the human protein-protein interaction
Human Protein-Protein: Rual et al. (Subnetwork for tutorial)
Human Protein-Protein: Stelzl et al. A human protein-protein interaction network: a resource for annotatin
Yeast Protein-Protein: Ptacek et al. Global analysis of protein phosphorylation in yeast. Nature 2005

http://cytoscape.org/datasets/ptacek.sif

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

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

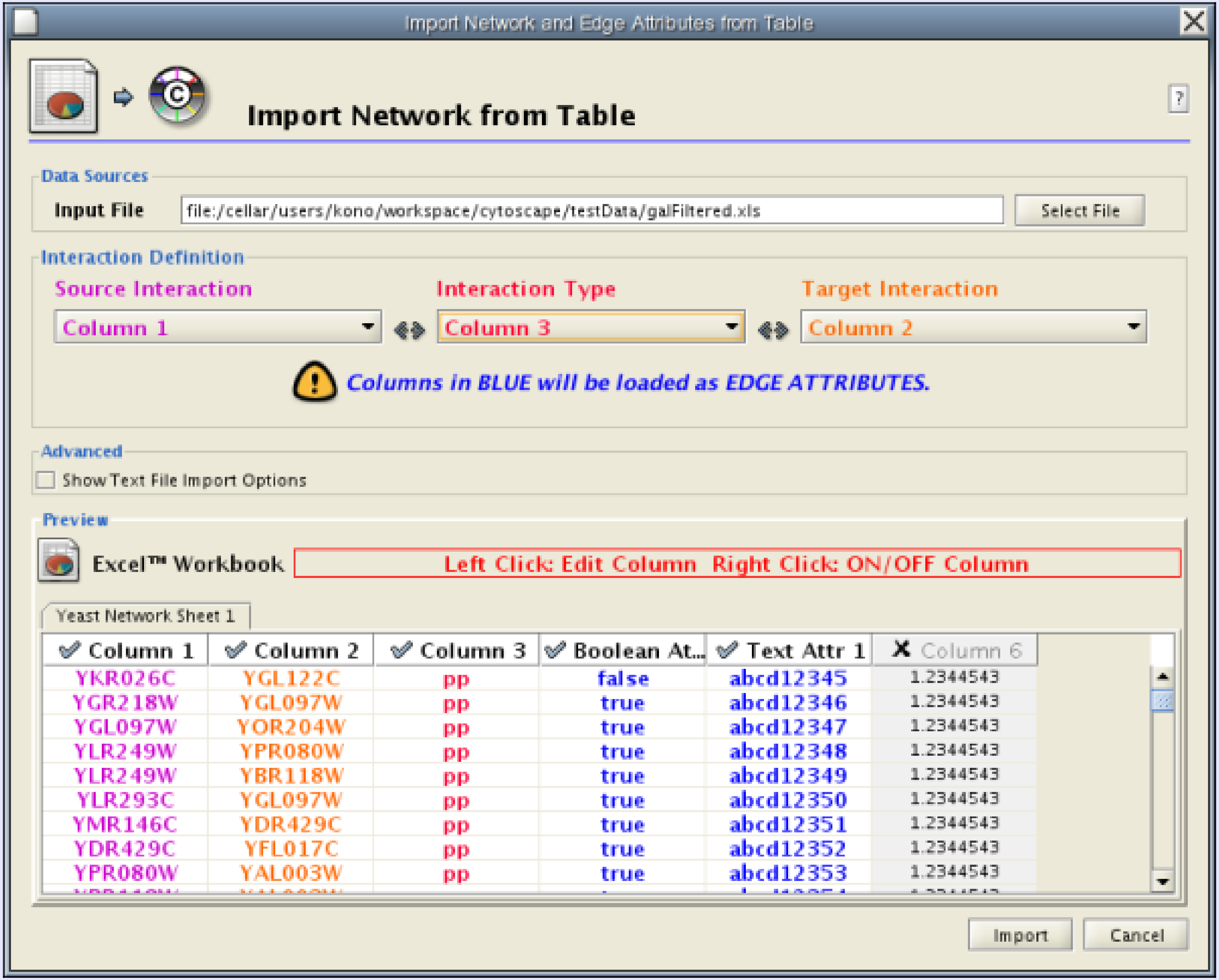
Preview

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

Cytoscape quick tour: network import



Import Network from Table

Data Sources

Input File: file:/cellar/users/kono/workspace/cytoscape/testData/galFiltered.xls

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

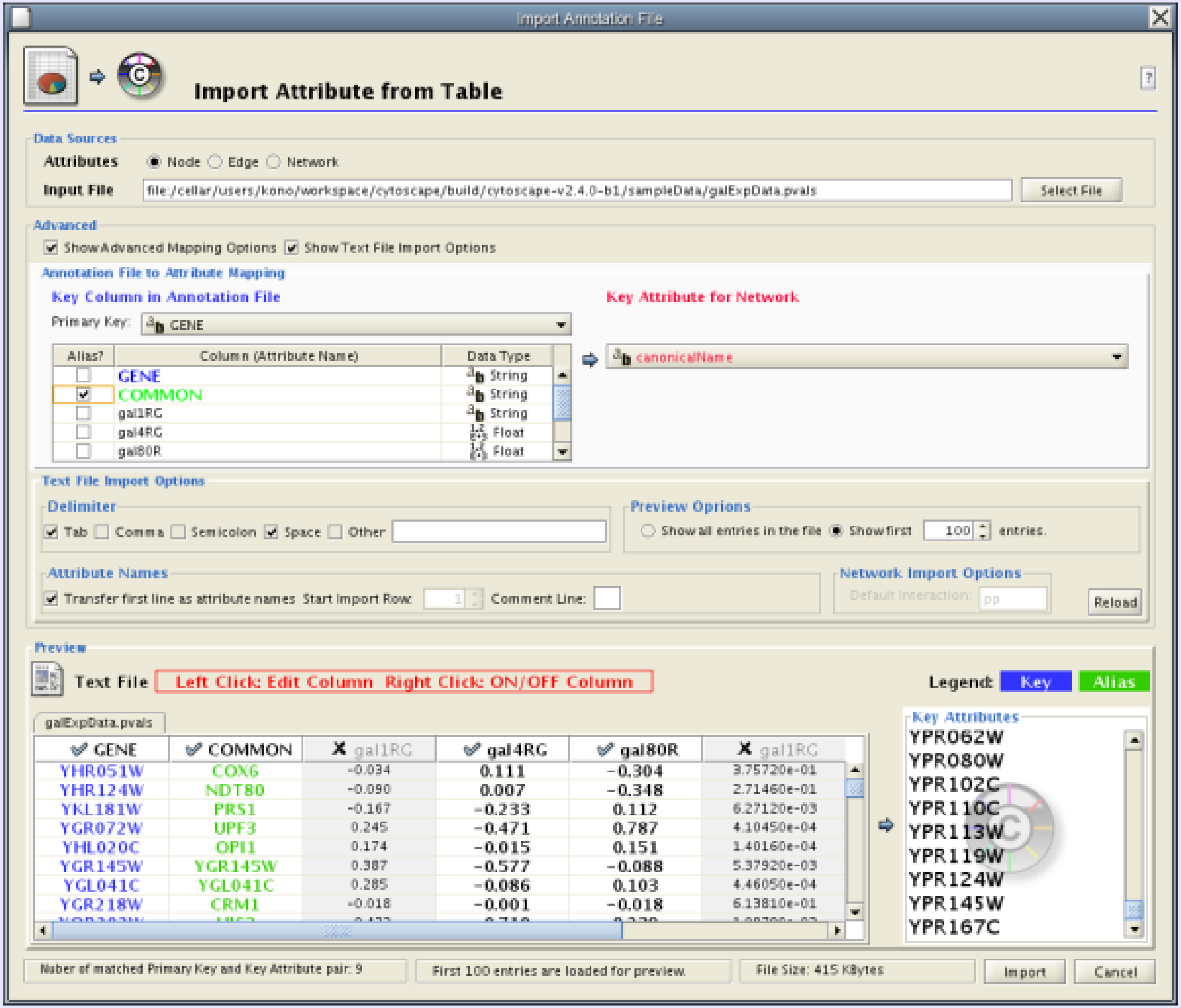
Left Click: Edit Column Right Click: ON/OFF Column

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

Cytoscape quick tour: attributes import

Attributes from file



Import Annotation 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

Advanced

☑ Show Advanced Mapping Options ☑ Show Text File Import Options

Annotation File to Attribute Mapping

Key Column in Annotation File

Primary Key: GENE

Alias?	Column (Attribute Name)	Data Type
☐	GENE	String
☑	COMMON	String
☐	gal1RG	String
☐	gal4RG	Float
☐	gal80R	Float

Key Attribute for Network

canonicalName

Text File Import Options

Delimiter

☑ Tab ☐ Comma ☐ Semicolon ☑ Space ☐ Other

Preview Oprions

○ 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

Preview

Text File Left Click: Edit Column Right Click: ON/OFF Column

Legend: Key Alias

galExpData.pvals

✓ 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

Key Attributes

YPR062W
YPR080W
YPR102C
YPR110C
YPR113W
YPR119W
YPR124W
YPR145W
YPR167C

Nuber of matched Primary Key and Key Attribute pair: 9

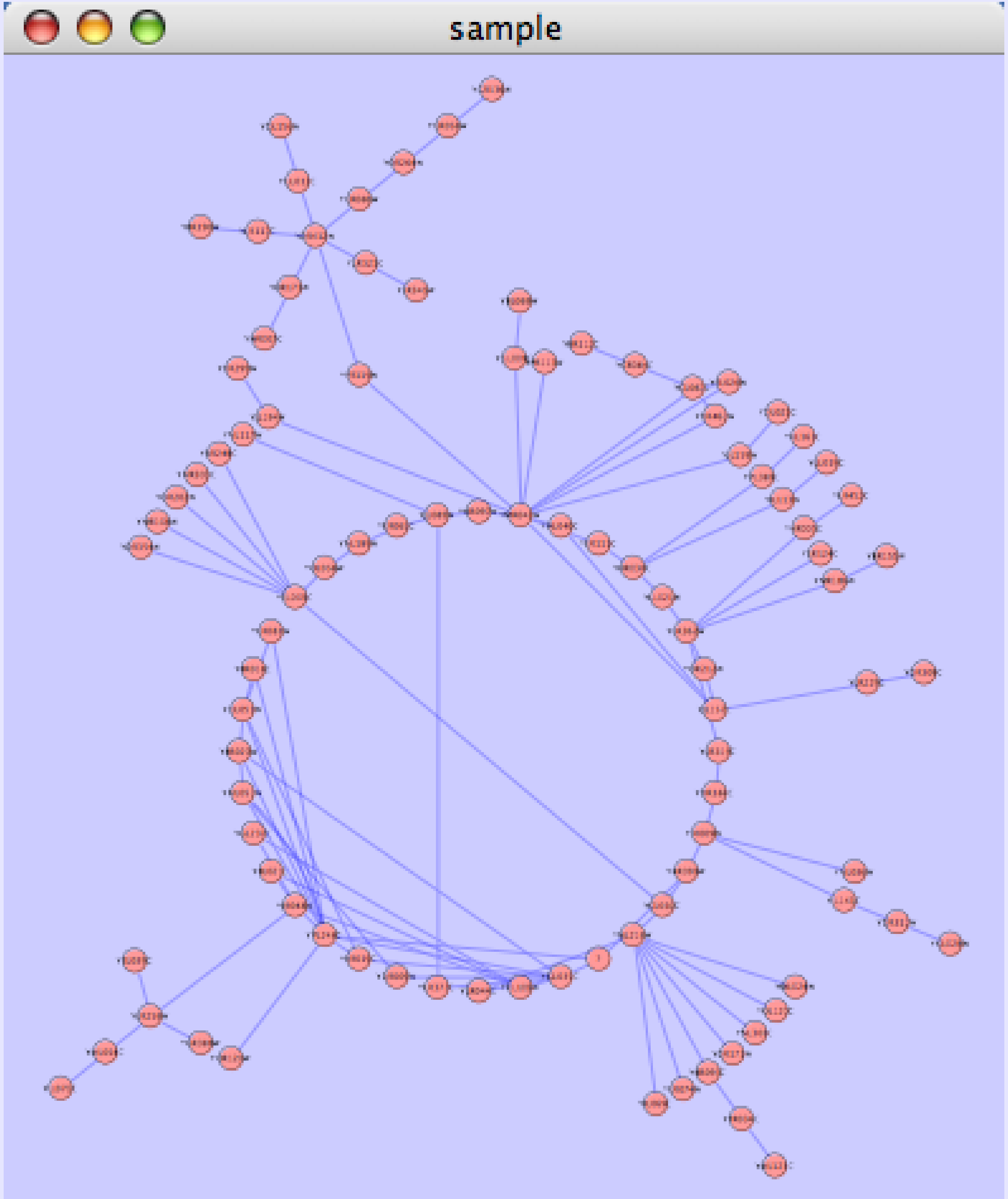
First 100 entries are loaded for preview.

File Size: 415 KBytes

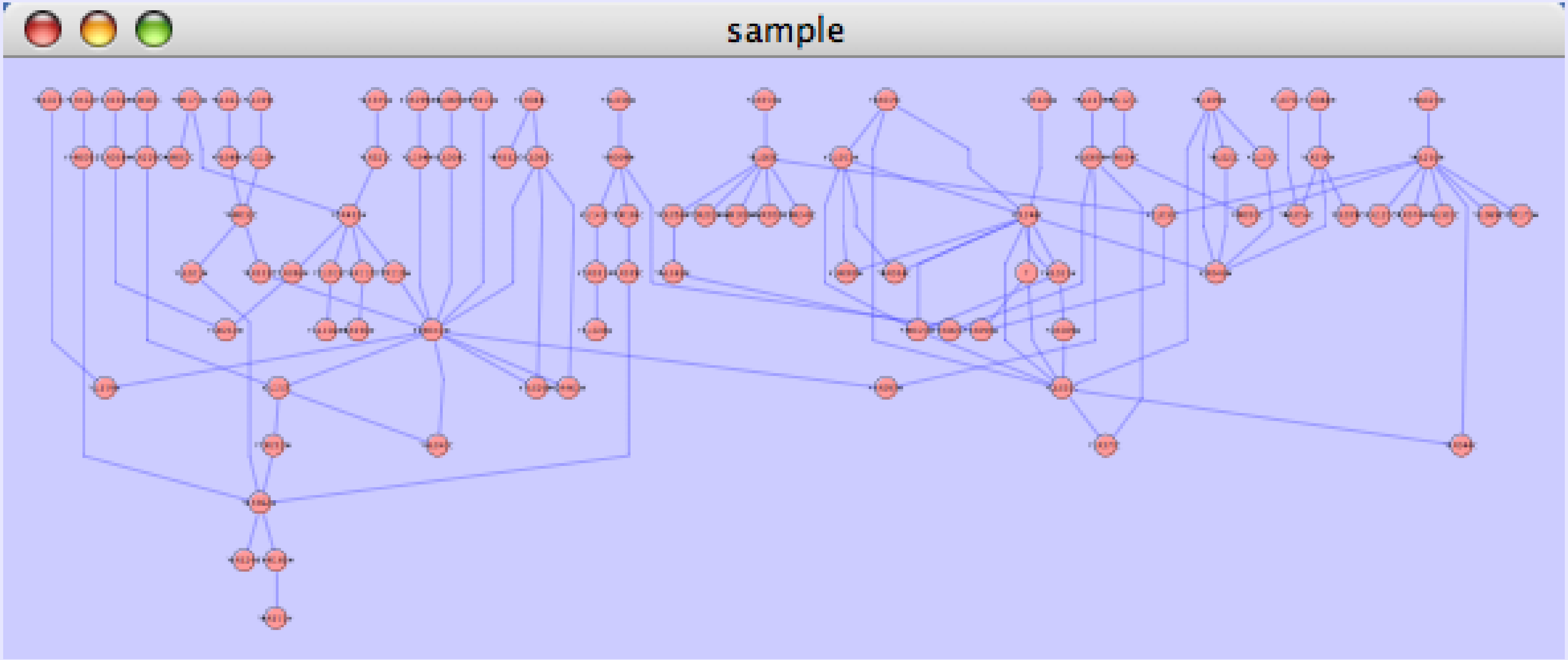
Cytoscape Quick Tour
- Layouts -

Cytoscape quick tour: network layouts

Circular layout



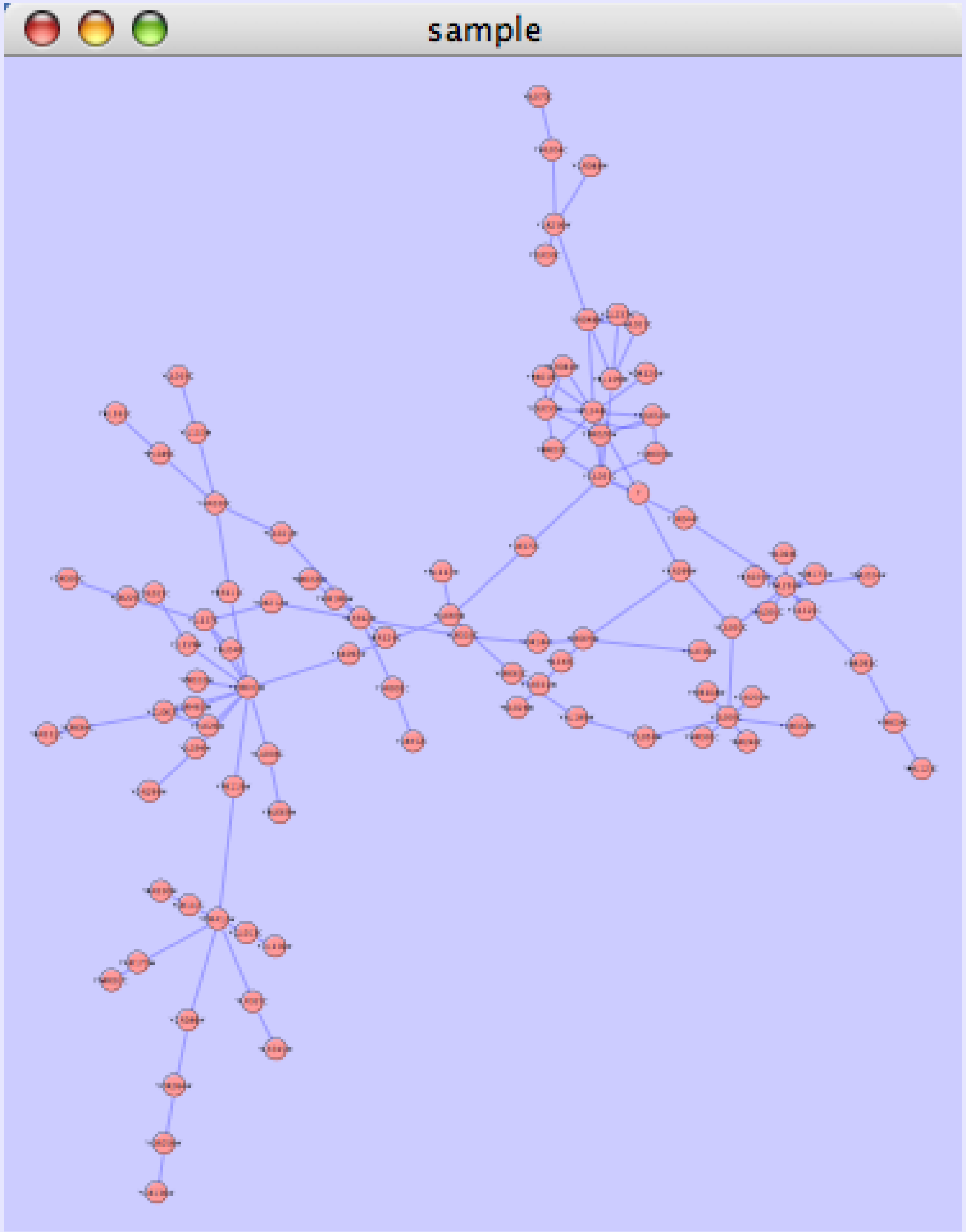
Cytoscape quick tour: network layouts



Hierarchical layout

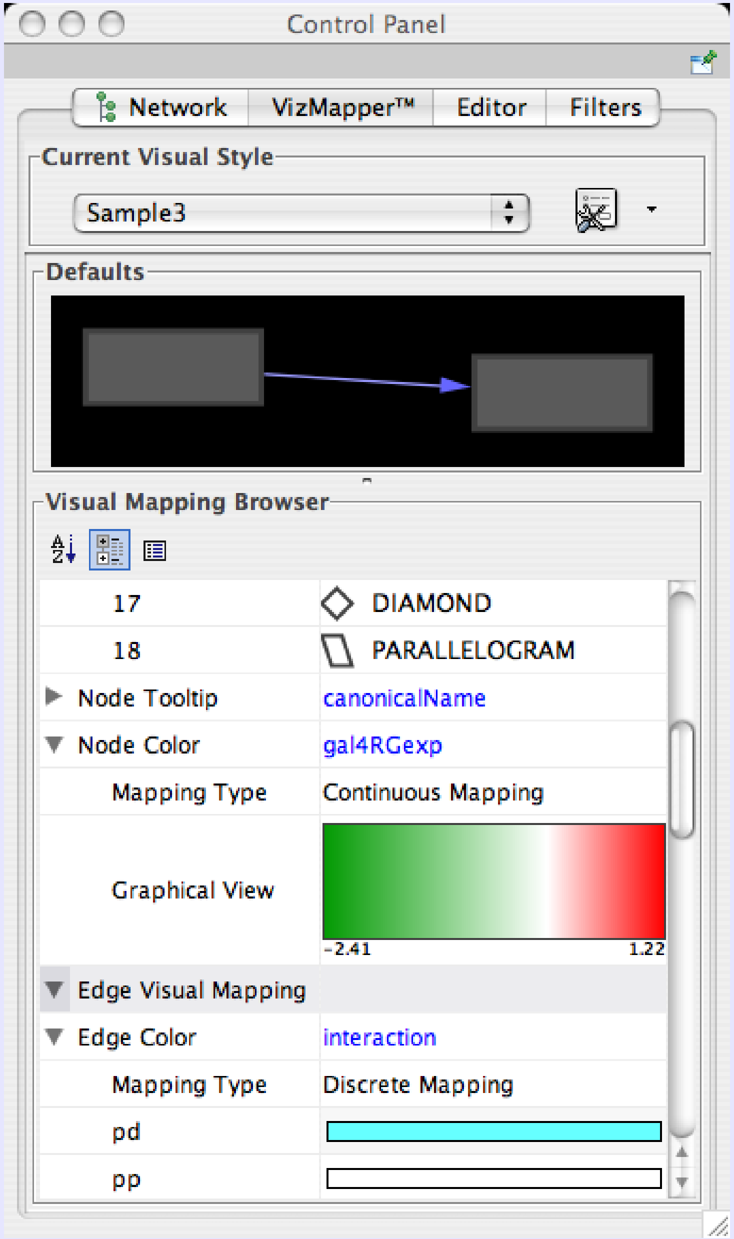
Cytoscape quick tour: network layouts

Organic layout



Cytoscape Quick Tour
- Visual Mapper -

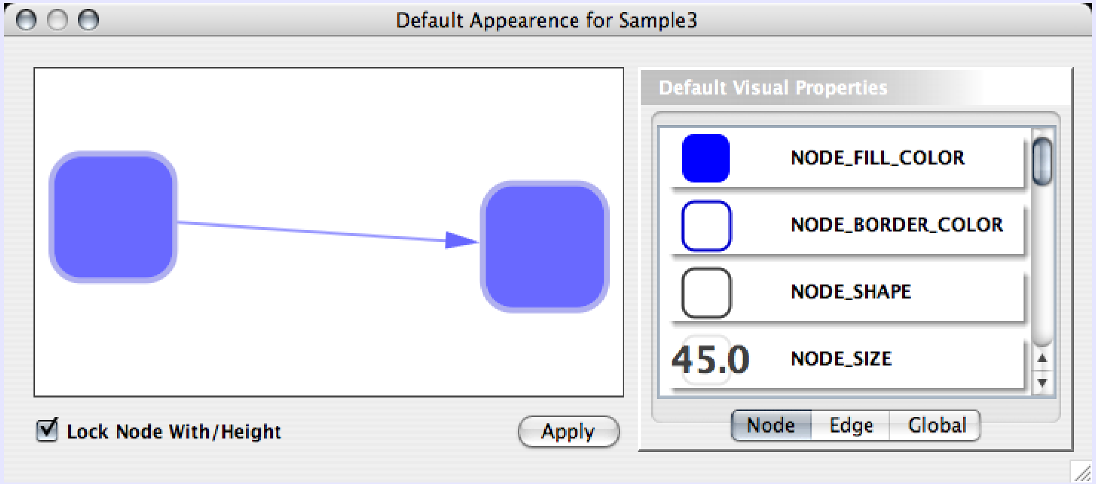
Cytoscape quick tour: Visual styles



Current Visual Style: Sample3

17	◇ DIAMOND
18	▱ PARALLELOGRAM
▶ Node Tooltip	canonicalName
▼ Node Color	gal4RGexp
Mapping Type	Continuous Mapping
Graphical View	-2.41 … 1.22
▼ Edge Visual Mapping	
▼ Edge Color	interaction
Mapping Type	Discrete Mapping
pd	🟦
pp	☐

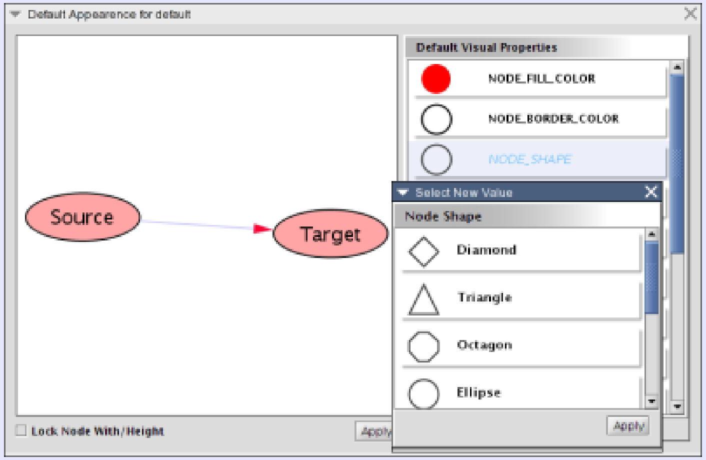
Visual Mapper



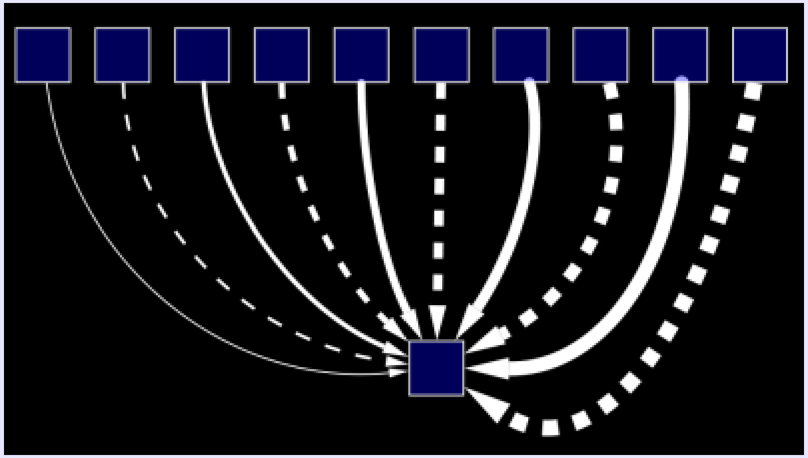
🟦	NODE_FILL_COLOR
☐	NODE_BORDER_COLOR
▢	NODE_SHAPE
45.0	NODE_SIZE

☑ Lock Node With/Height

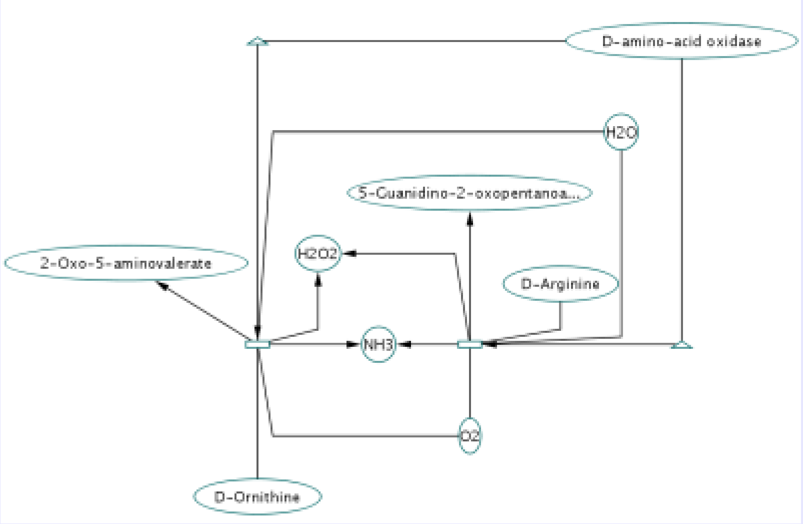
Cytoscape quick tour: Visual styles



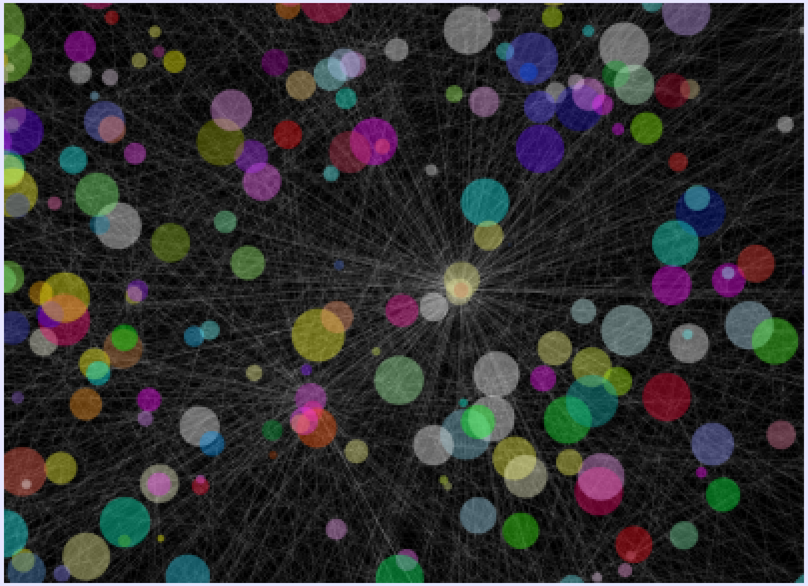
Default



Edges

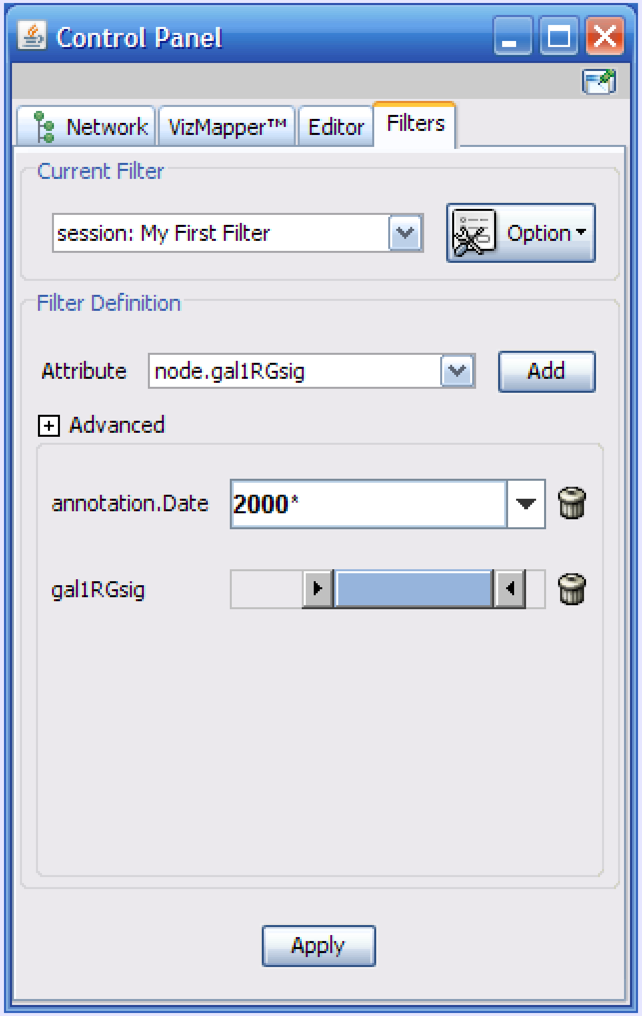


Node shapes



Nodes size and opacity

Cytoscape quick tour: Filters



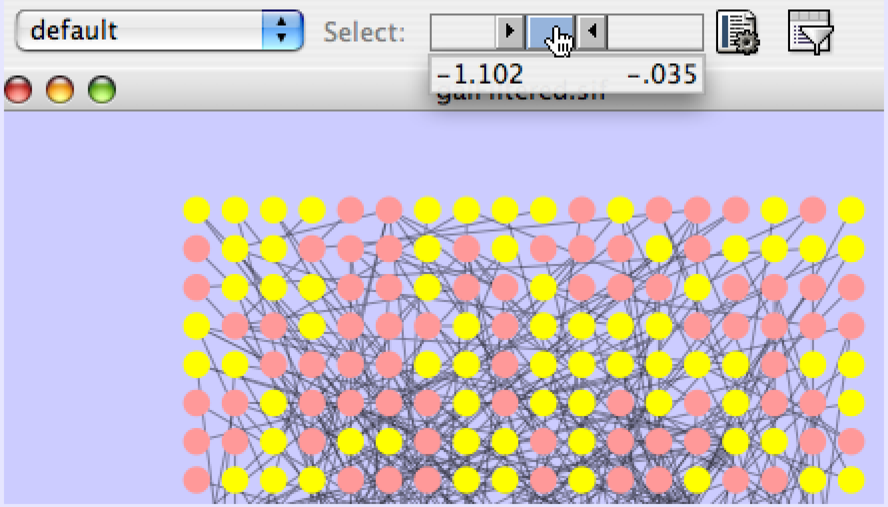
Current Filter: session: My First Filter

Filter Definition

Attribute: node.gal1RGsig

annotation.Date: 2000*

gal1RGsig



- Plugins -

Cytoscape Plugins

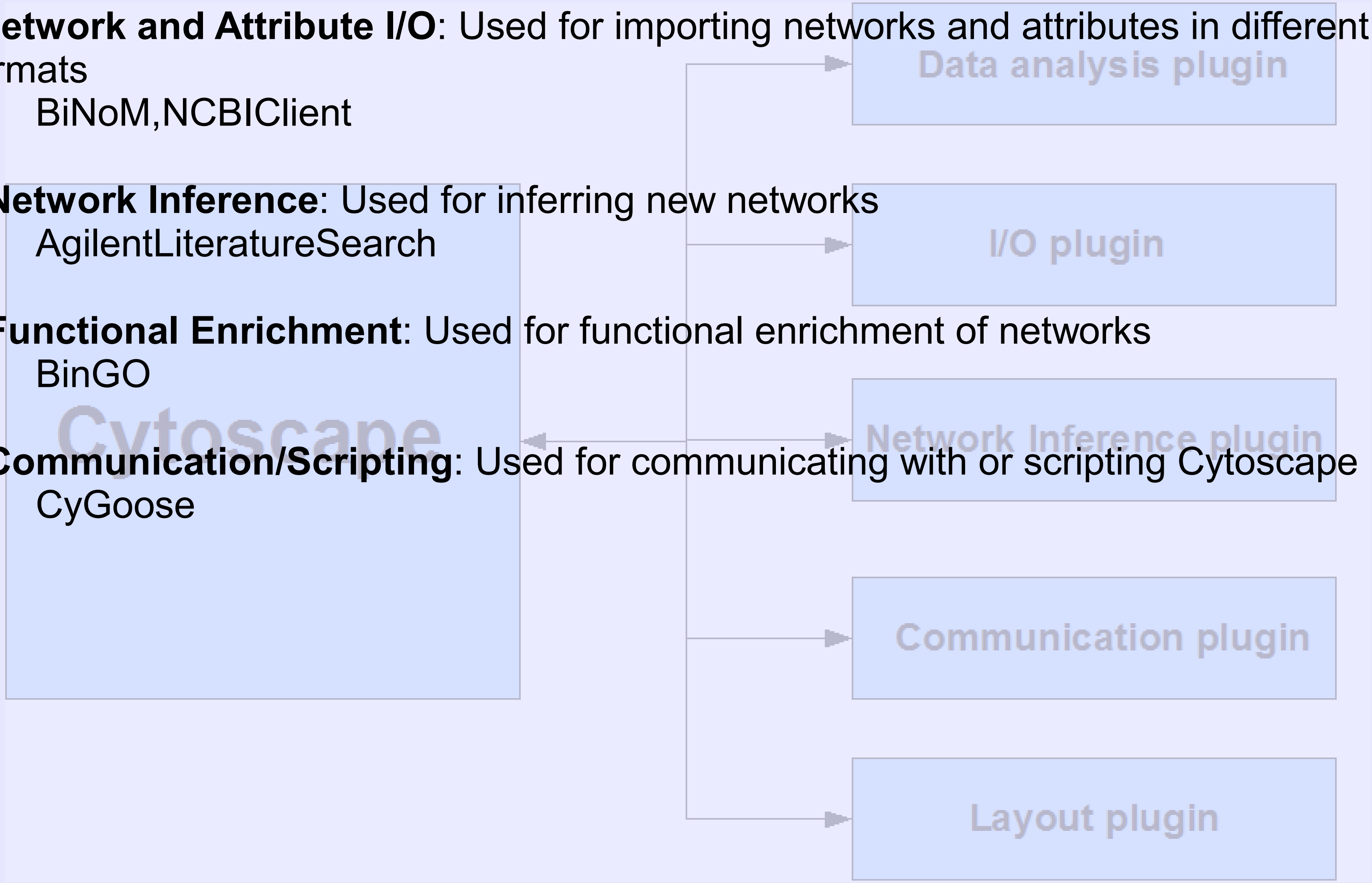
- **Analysis**: Used for analyzing existing networks
  Mcode, Metabolica, CentiScaPe (under submission)

- **Network and Attribute I/O**: Used for importing networks and attributes in different file formats
  BiNoM,NCBIClient

- **Network Inference**: Used for inferring new networks
  AgilentLiteratureSearch

- **Functional Enrichment**: Used for functional enrichment of networks
  BinGO

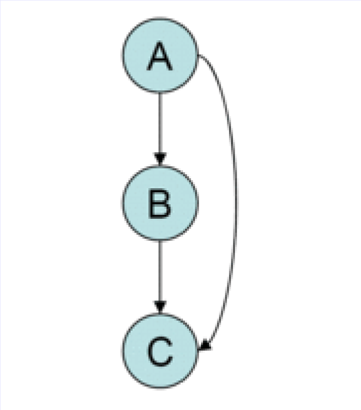
- **Communication/Scripting**: Used for communicating with or scripting Cytoscape
  CyGoose



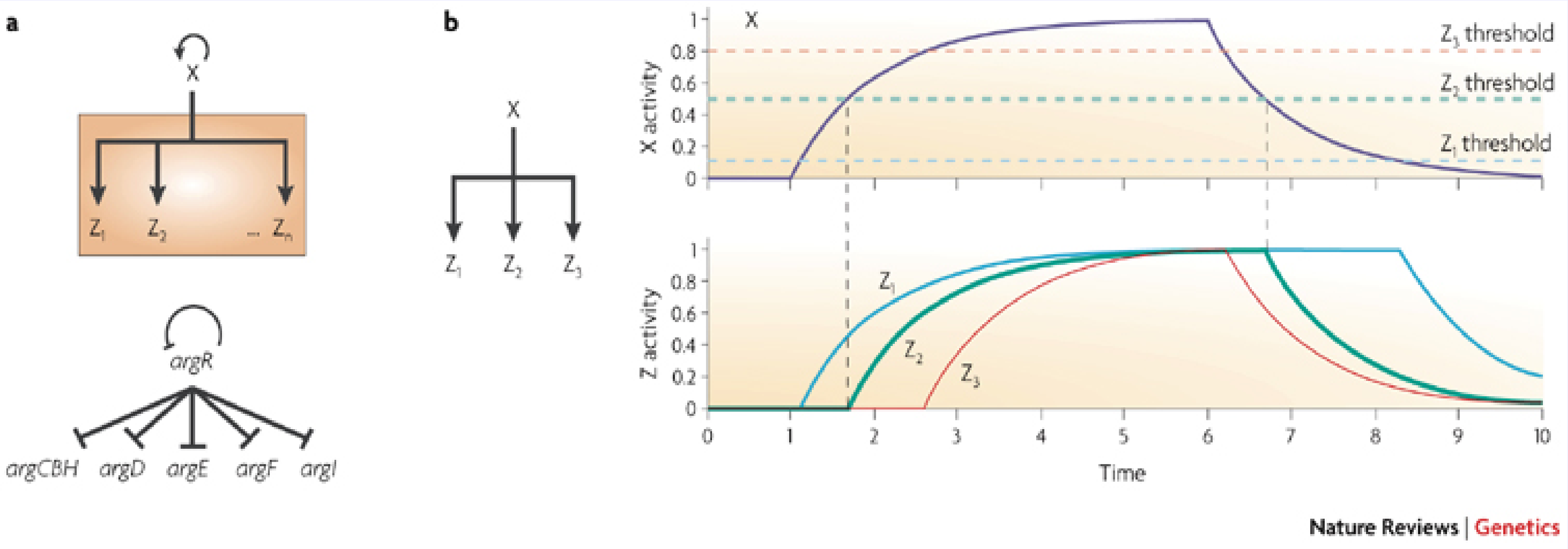
Data analysis: a small network

- Analysis Plugins - Metabolica

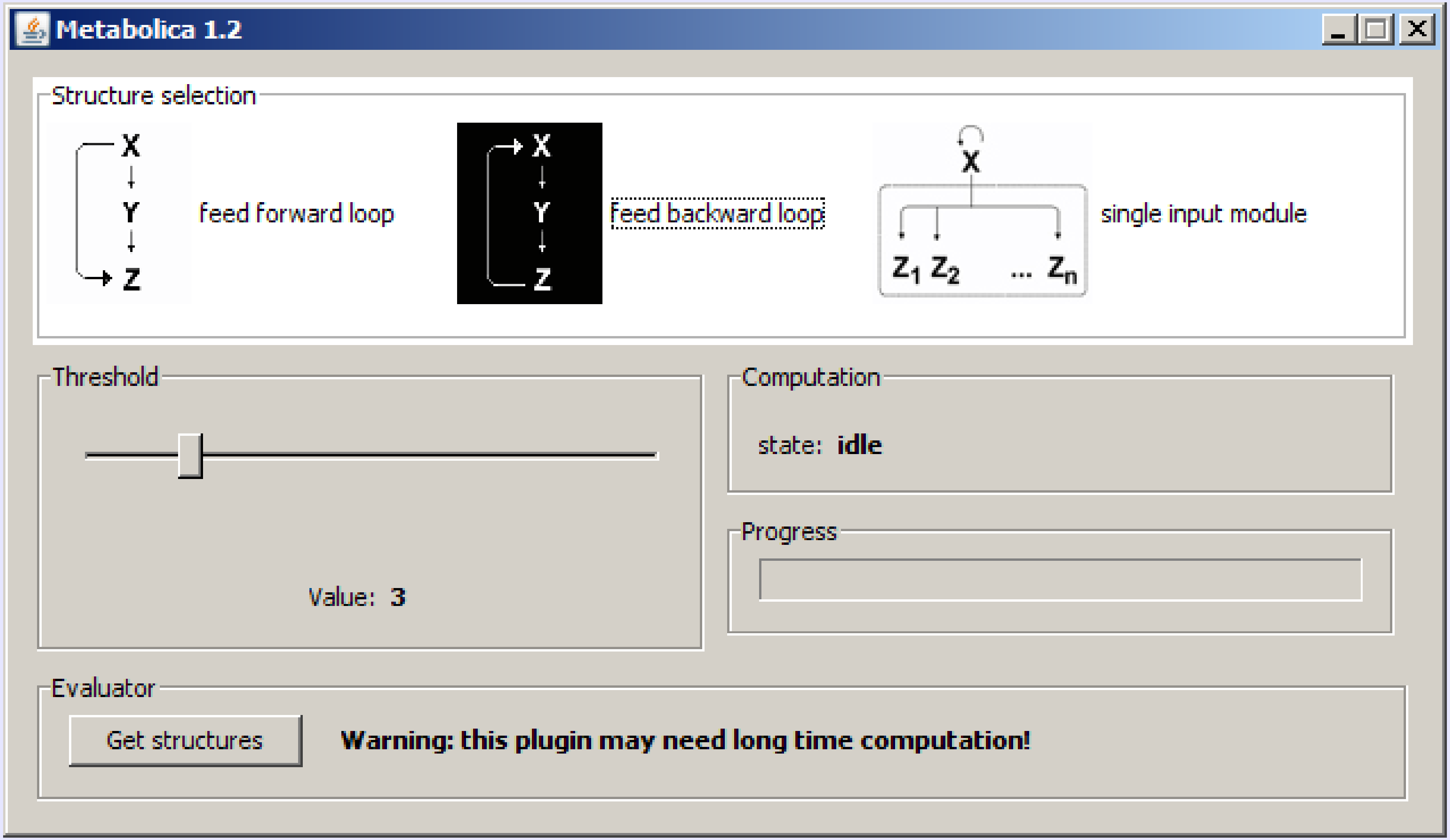
Metabolica plugin: network motifs introduction



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



Metabolica Plugin



feed forward loop

feed backward loop

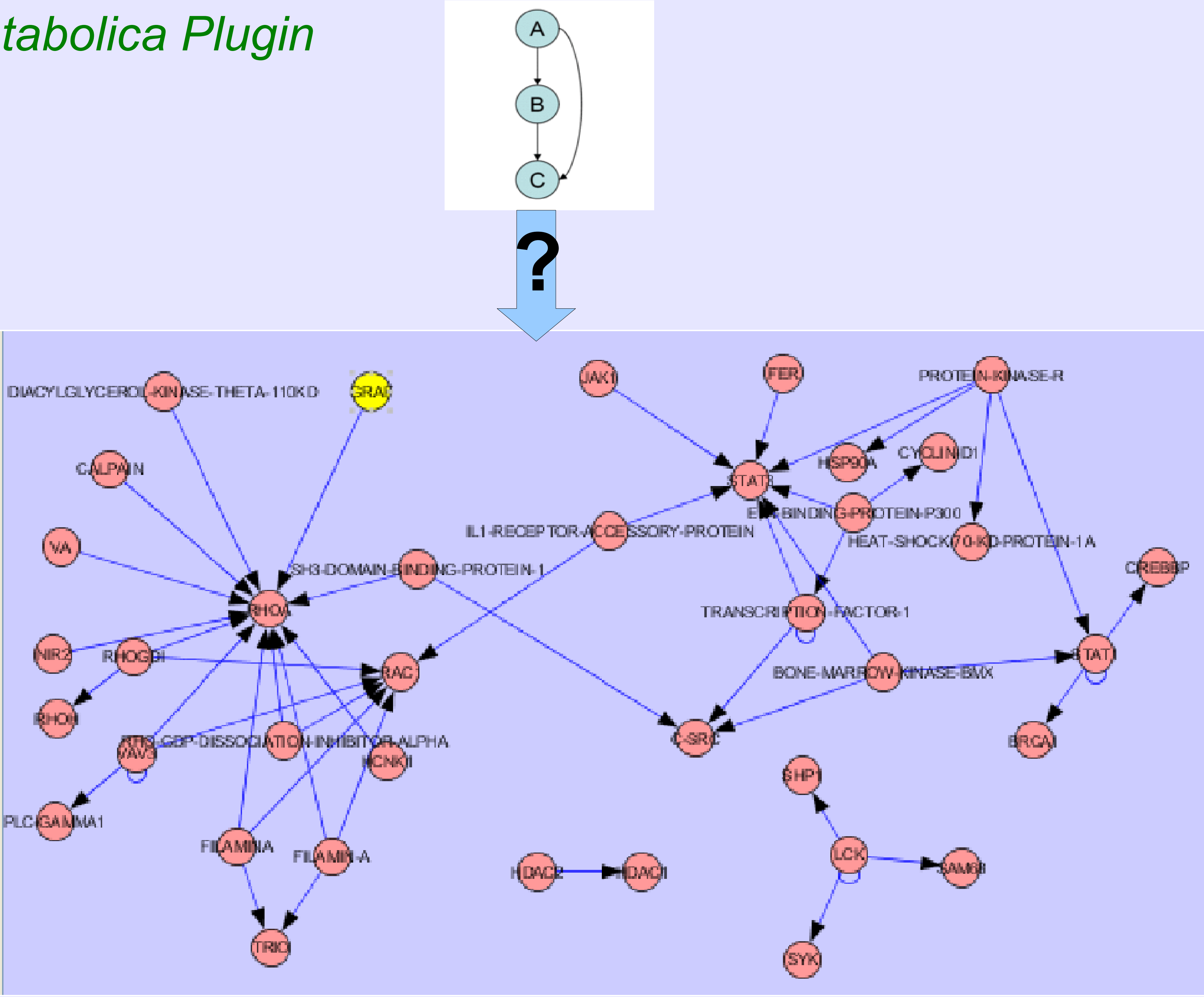
single input module

Value: **3**

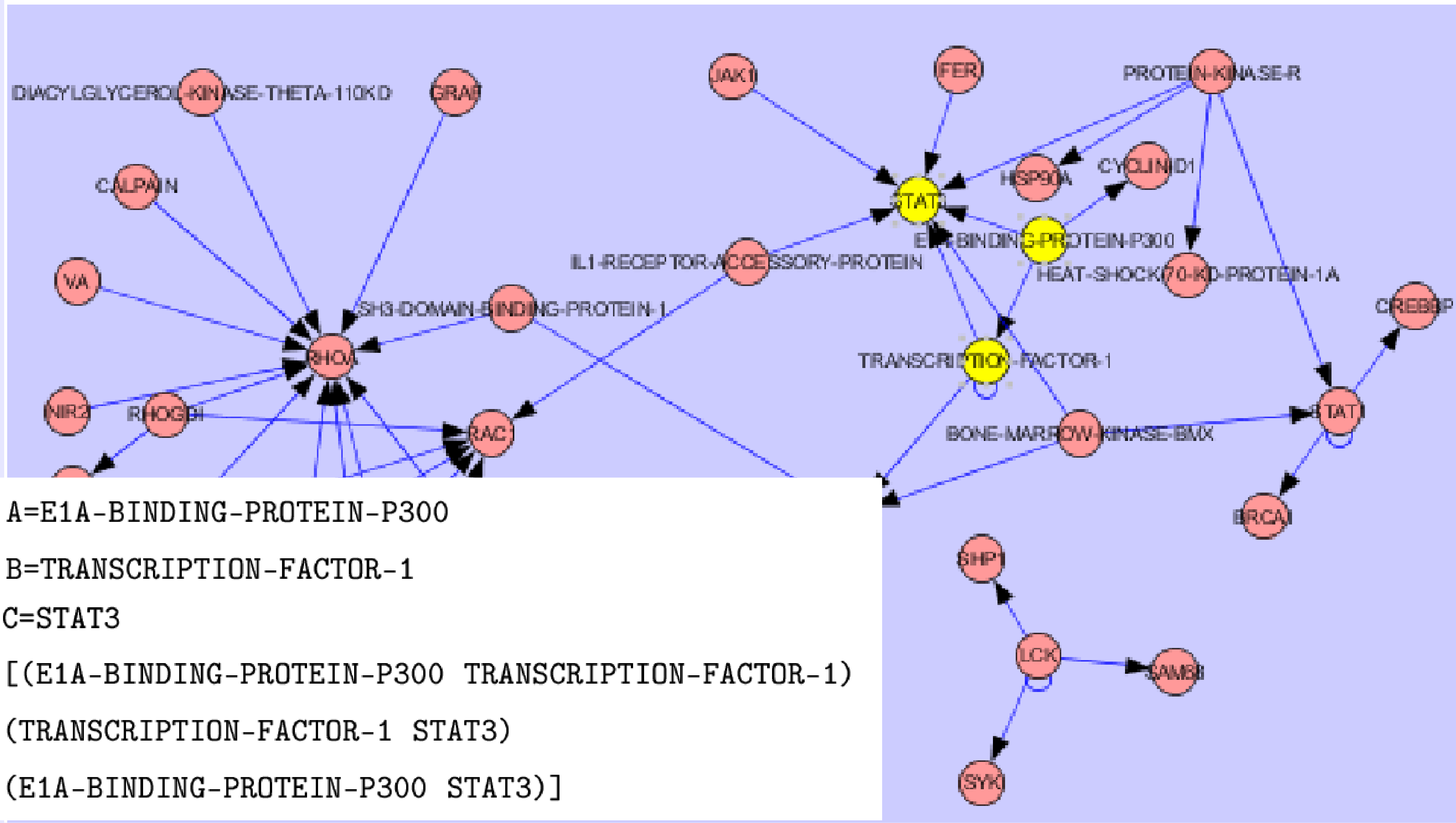
state: **idle**

Warning: this plugin may need long time computation!

Metabolica Plugin



Metabolica Plugin



A=E1A-BINDING-PROTEIN-P300

B=TRANSCRIPTION-FACTOR-1

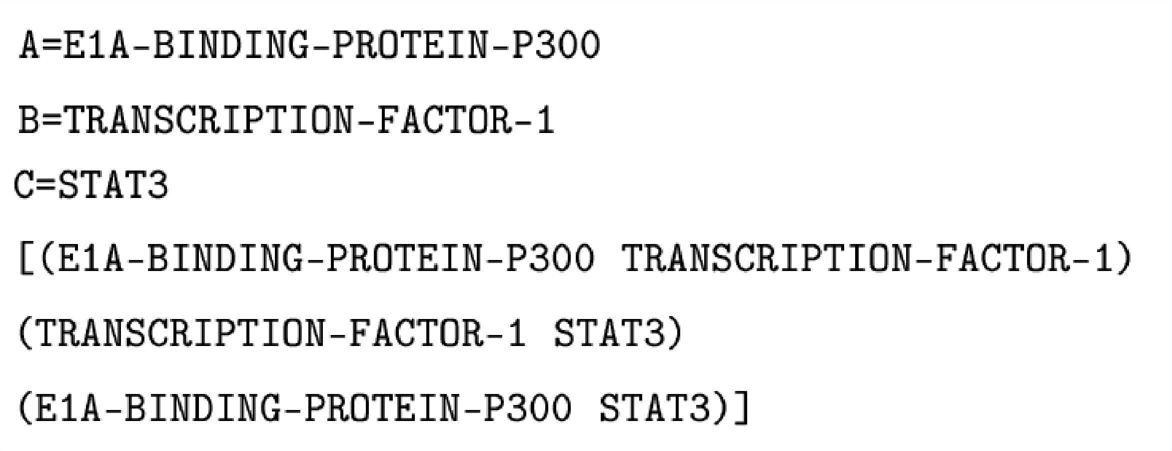
C=STAT3

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

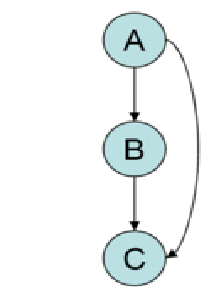
(TRANSCRIPTION-FACTOR-1 STAT3)

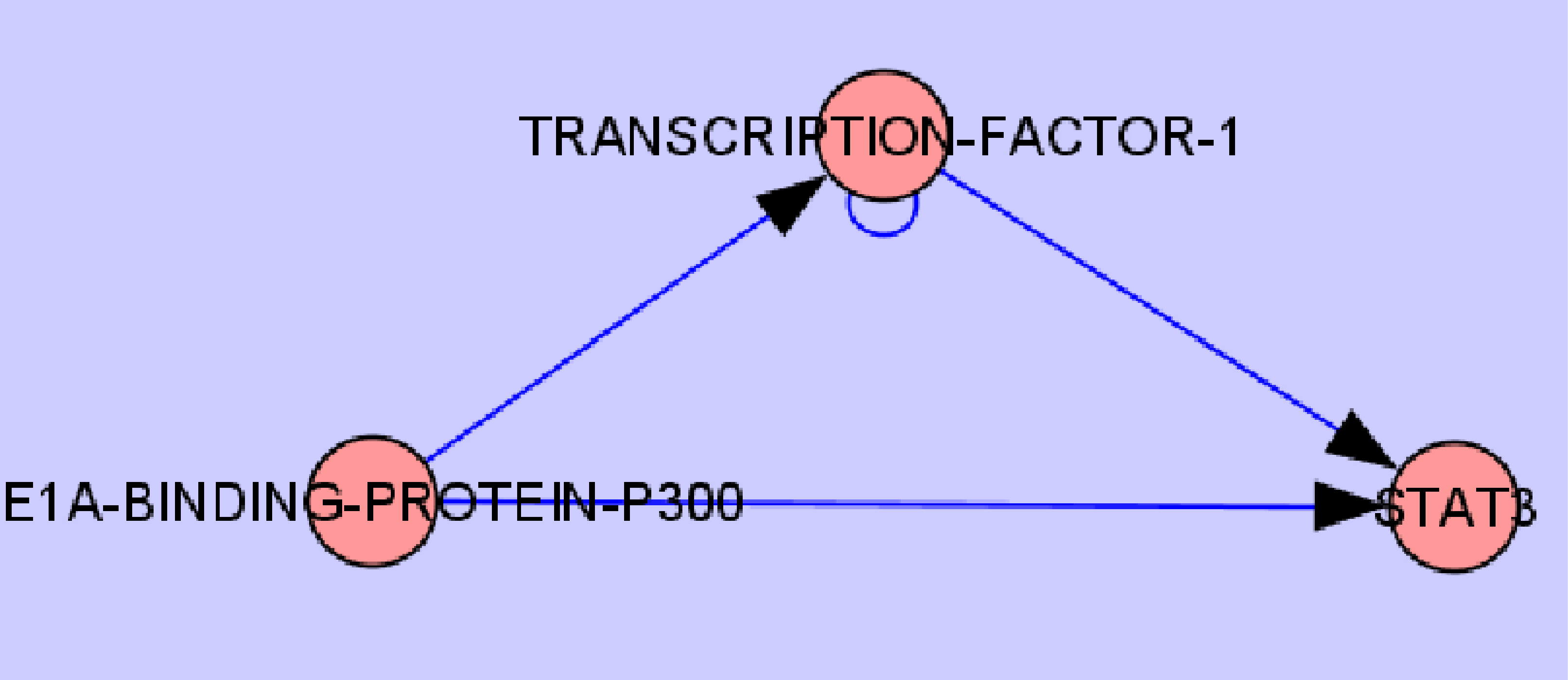
(E1A-BINDING-PROTEIN-P300 STAT3)]

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 (v1, v2) and counting the number of shortest paths linking v1 and v2 and passing through a node n. Then, the value is related to the total number of shortest paths linking v1 and v2.

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

$$\delta_{st}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}}$$

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.

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

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.

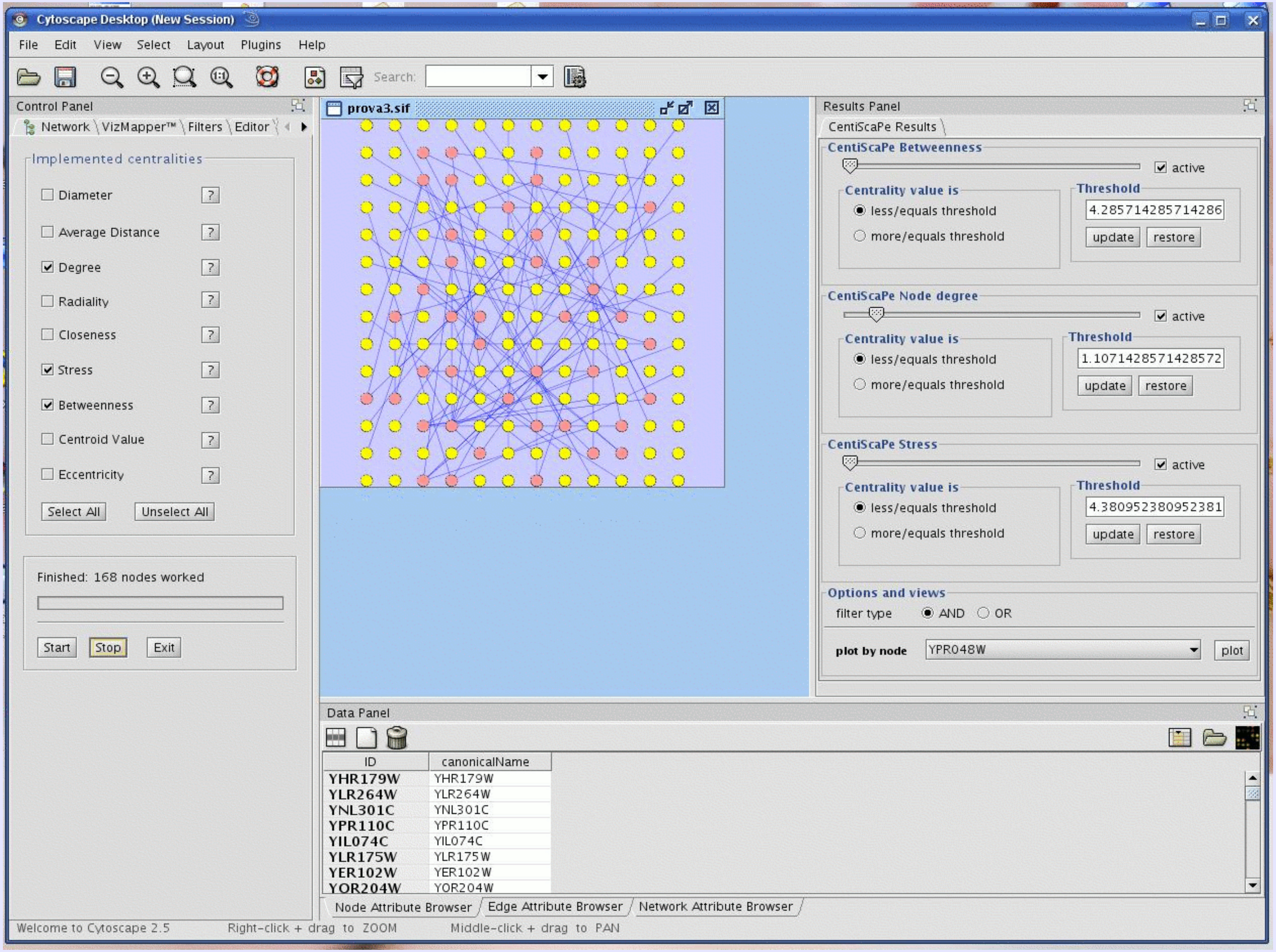
CentiScaPe plugin: centralities

- **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 form 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



Implemented centralities

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

Finished: 168 nodes worked

CentiScaPe Betweenness ☑ active

Centrality value is
● less/equals threshold
○ more/equals threshold

Threshold: 4.285714285714286

CentiScaPe Node degree ☑ active

Centrality value is
● less/equals threshold
○ more/equals threshold

Threshold: 1.1071428571428572

CentiScaPe Stress ☑ active

Centrality value is
● less/equals threshold
○ more/equals threshold

Threshold: 4.380952380952381

Options and views

filter type ● AND ○ OR

plot by node YPR048W

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

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!