Visualizing Networks with Cytoscape

Alex Pico

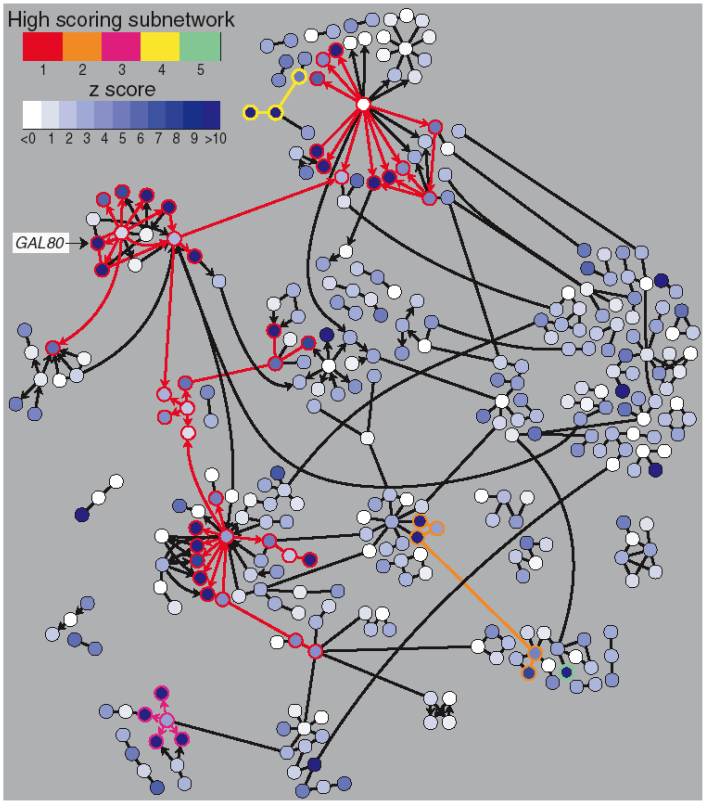
apico@gladstone.ucsf.edu

Introductions

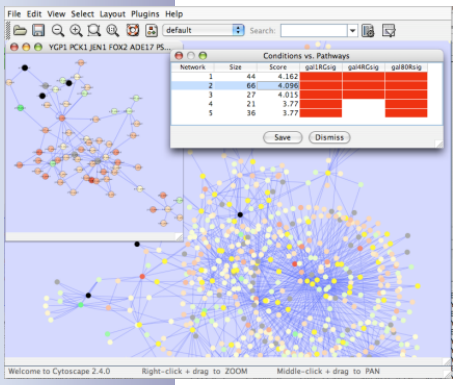
- Me
  - Executive Director, NRNB
  - Systems Biology Group, Gladstone Institutes
  - 5 Years on Cytoscape core development team
  - Author of several plugins
    - GenMAPP-CS Workspaces, criteriaMapper, GO-Elite, BubbbleRouter, Mosaic
- You?

What is a network tool?

- Networks provide an integrating context for data
  - Commonly understood diagrammatic representation for concepts and relationships
  - Provides structure that helps reduce underlying complexity of the data
  - More efficient than searching databases element-by-element
- Network tools give us functionality for studying complex processes
  - **Analysis** of global characteristics of the data, e.g. degree, clustering coefficient, shortest paths, centrality, density
    - Identify key elements (hubs) and 'interesting' subnets
    - Help elucidate mechanisms of interaction
  - **Visualization** of data superimposed upon the network
    - Help understand how a process is modulated or attenuated by a stimulus



Applications of Network Biology

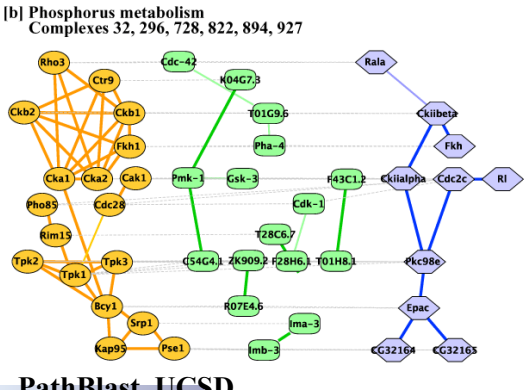


jActiveModules, UCSD

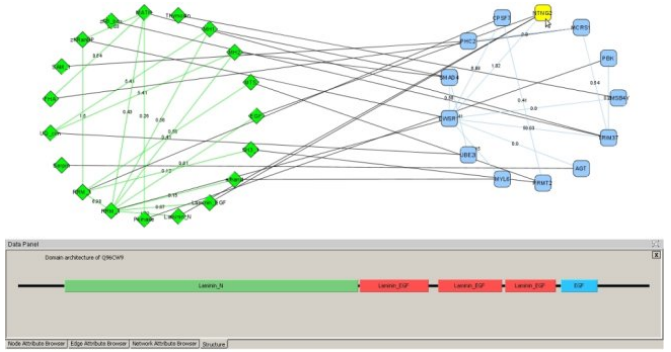
- **Gene Function Prediction** – shows connections to sets of genes/ proteins involved in same biological process
- **Detection of protein complexes/ other modular structures** – discover modularity & higher order organization (motifs, feedback loops)
- **Network evolution** – biological process(s) conservation across species
- **Prediction of new interactions and functional associations** – Statistically significant domain-domain correlations in protein interaction network to predict protein-protein or genetic interaction



mCode, University of Toronto

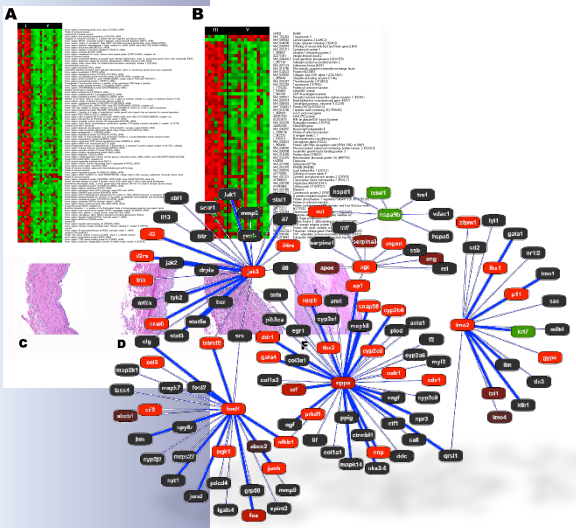


PathBlast, UCSD



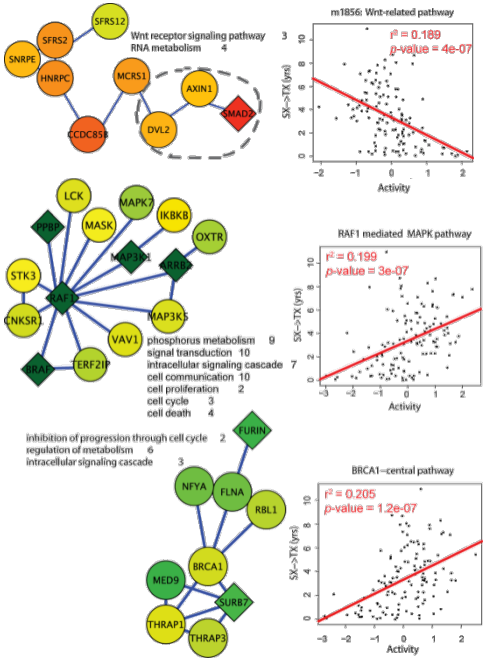
DomainGraph, Max Planck Institute

Applications of Network Informatics in Disease

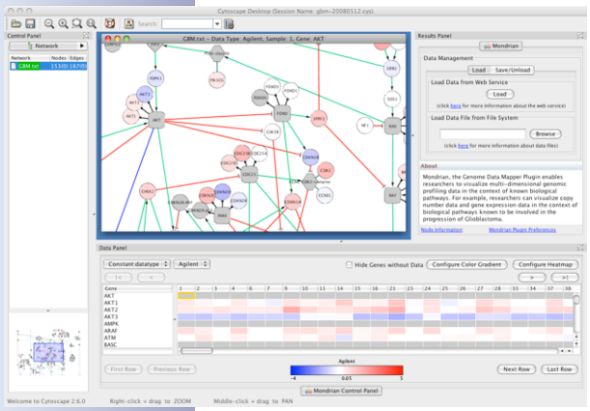


Agilent Literature Search

- **Identification of disease subnetworks** – identification of disease network subnetworks that are transcriptionally active in disease.
- **Subnetwork-based diagnosis** – source of biomarkers for disease classification, identify interconnected genes whose aggregate expression levels are predictive of disease state
- **Subnetwork-based gene association** – map common pathway mechanisms affected by collection of genotypes (SNP, CNV)

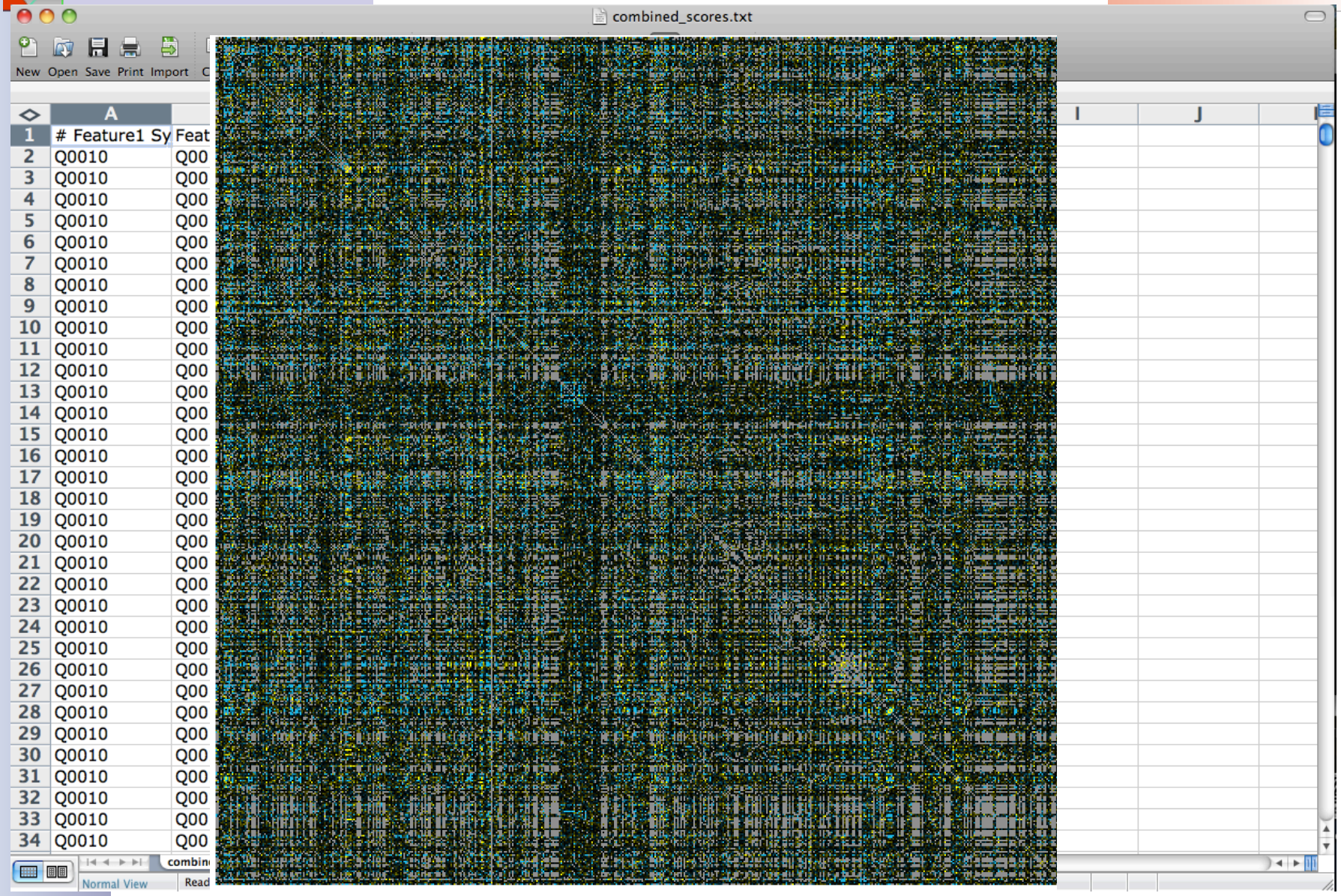


PinnacleZ, UCSD



Mondrian, MSKCC

The Challenge

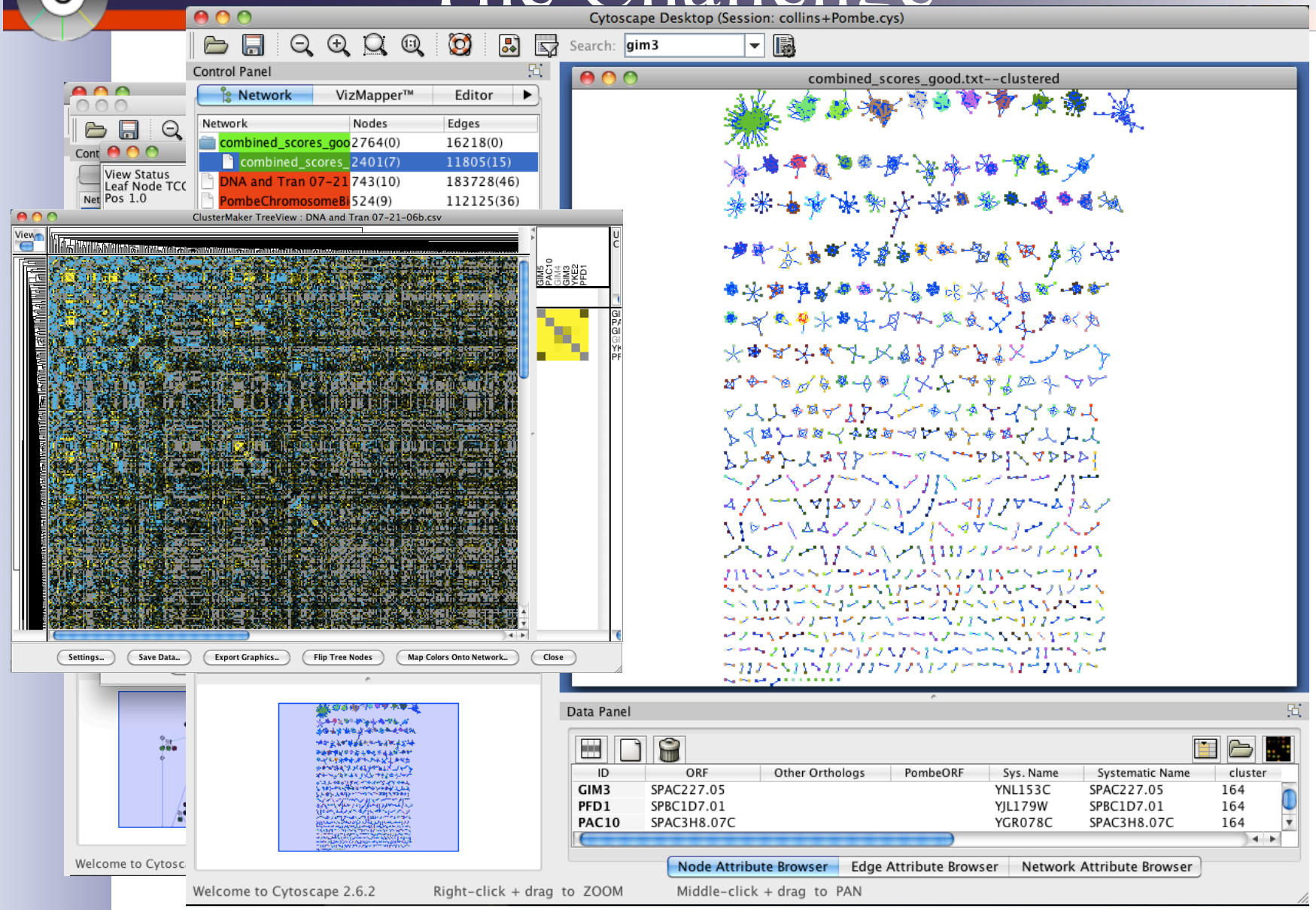




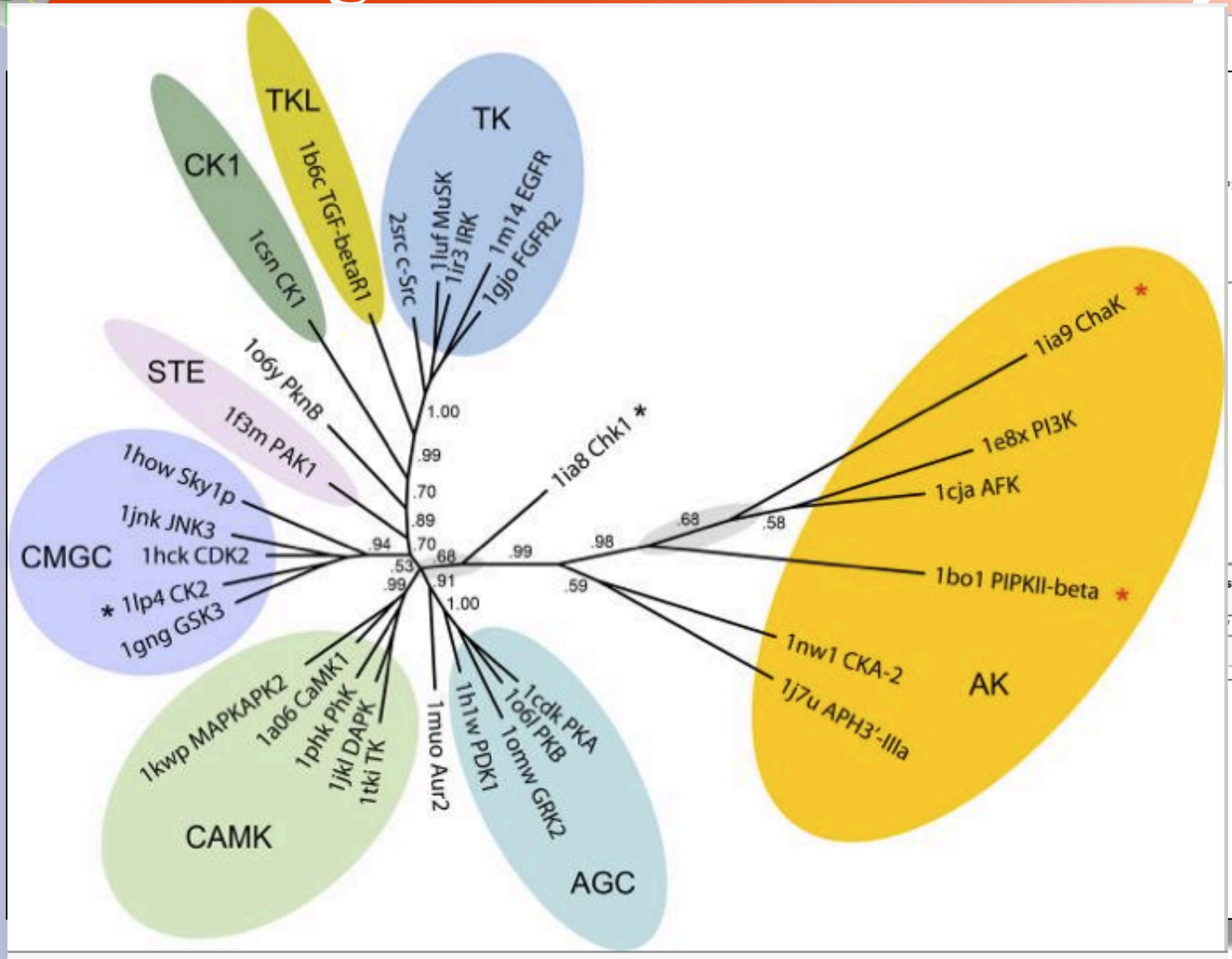
The Challenge

- Biological networks (nodes and edges)
  - Seldom tell us anything by themselves
  - Analysis involves:
    - Understanding the characteristics of the network
      - Modularity
      - Comparison with other networks (specifically random networks)
  - Visualization involves:
    - Placing nodes in a meaningful way (layouts)
    - Mapping biologically relevant data to the network
      - Node size
      - Node color
      - Edge weights

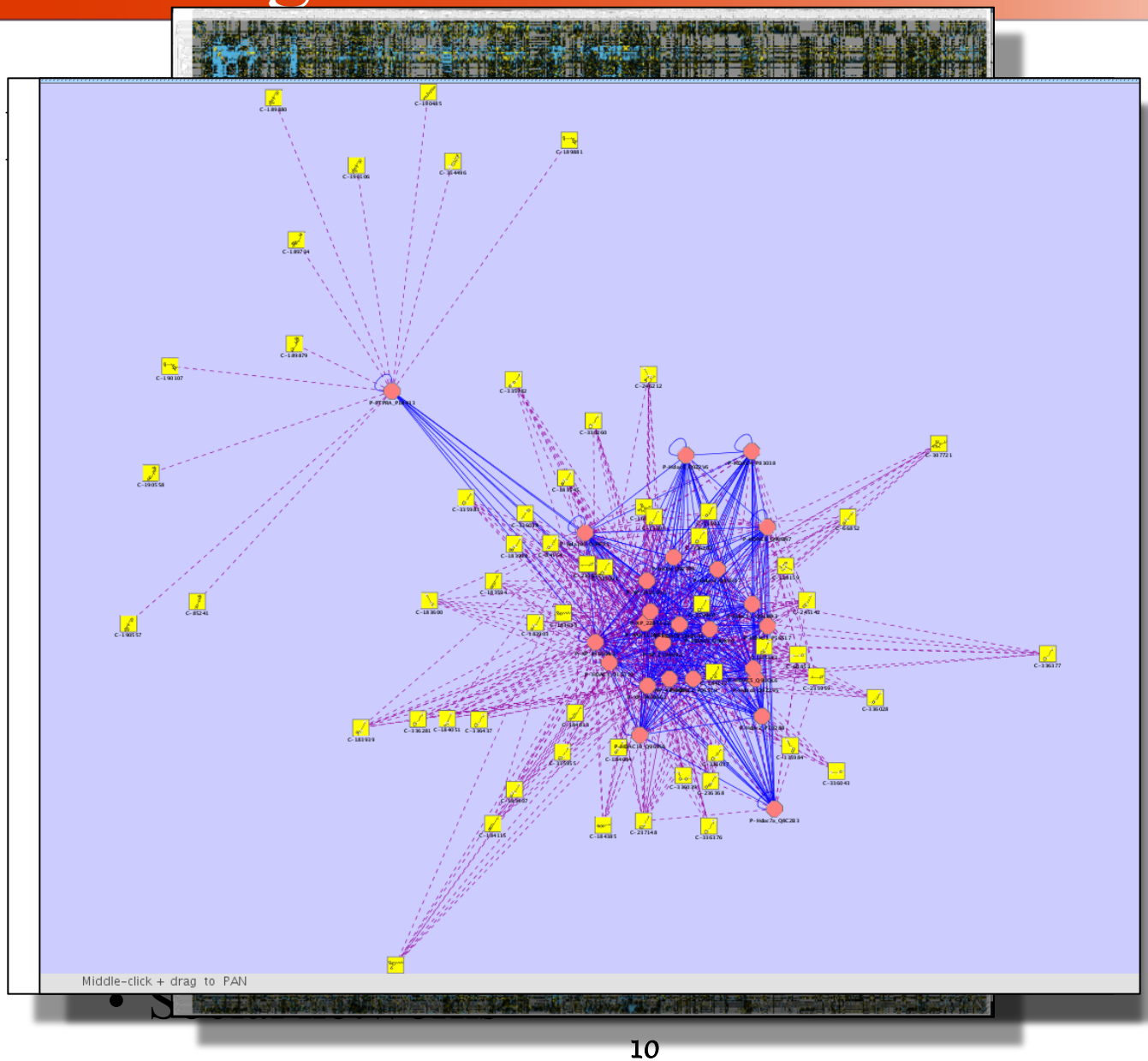
The Challenge



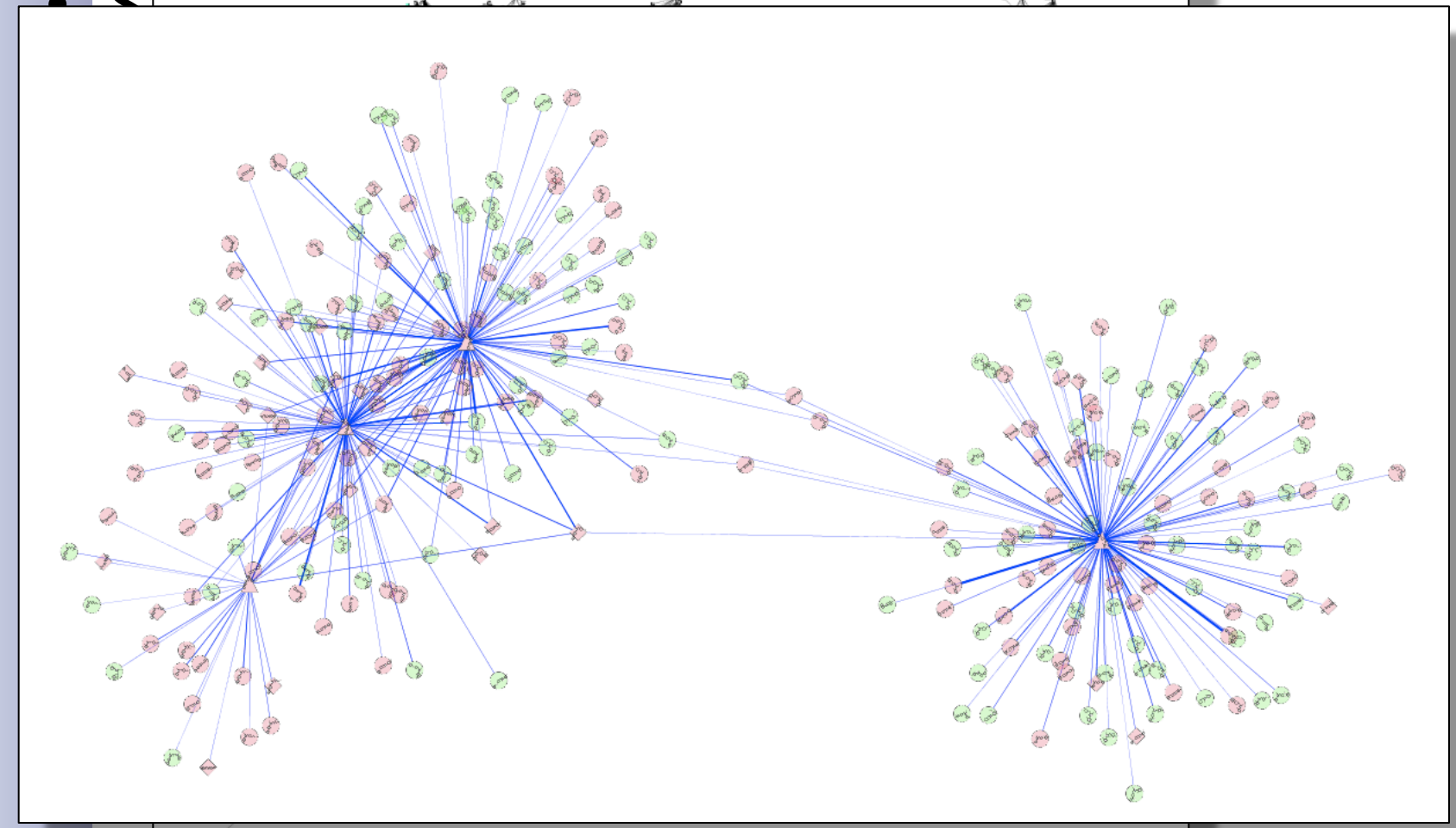
Biological Network Taxonomy



Biological Network Taxonomy



Biological Network Taxonomy

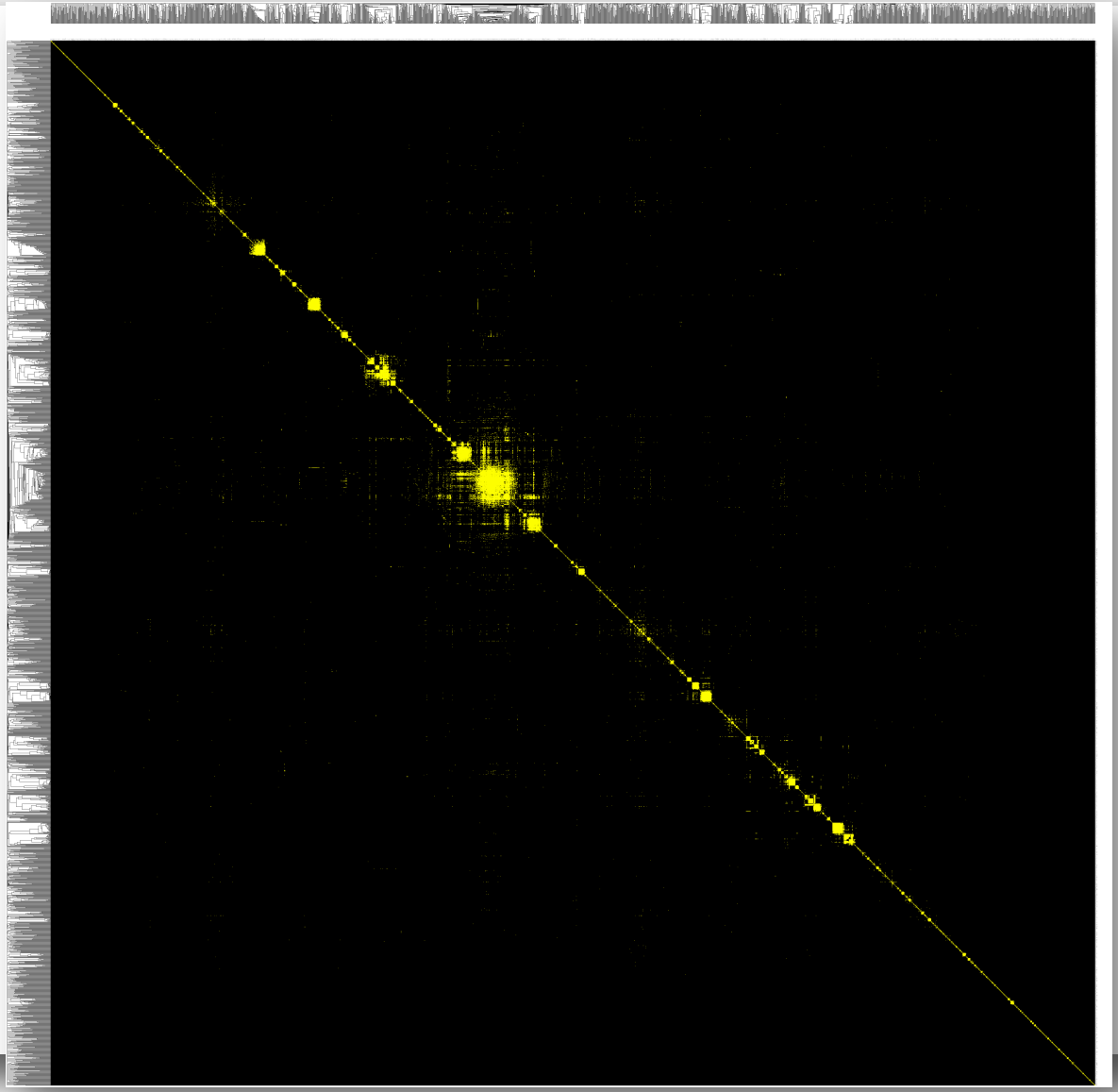




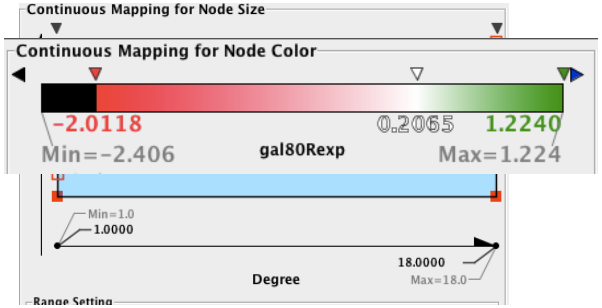
Visualization of Biological Networks

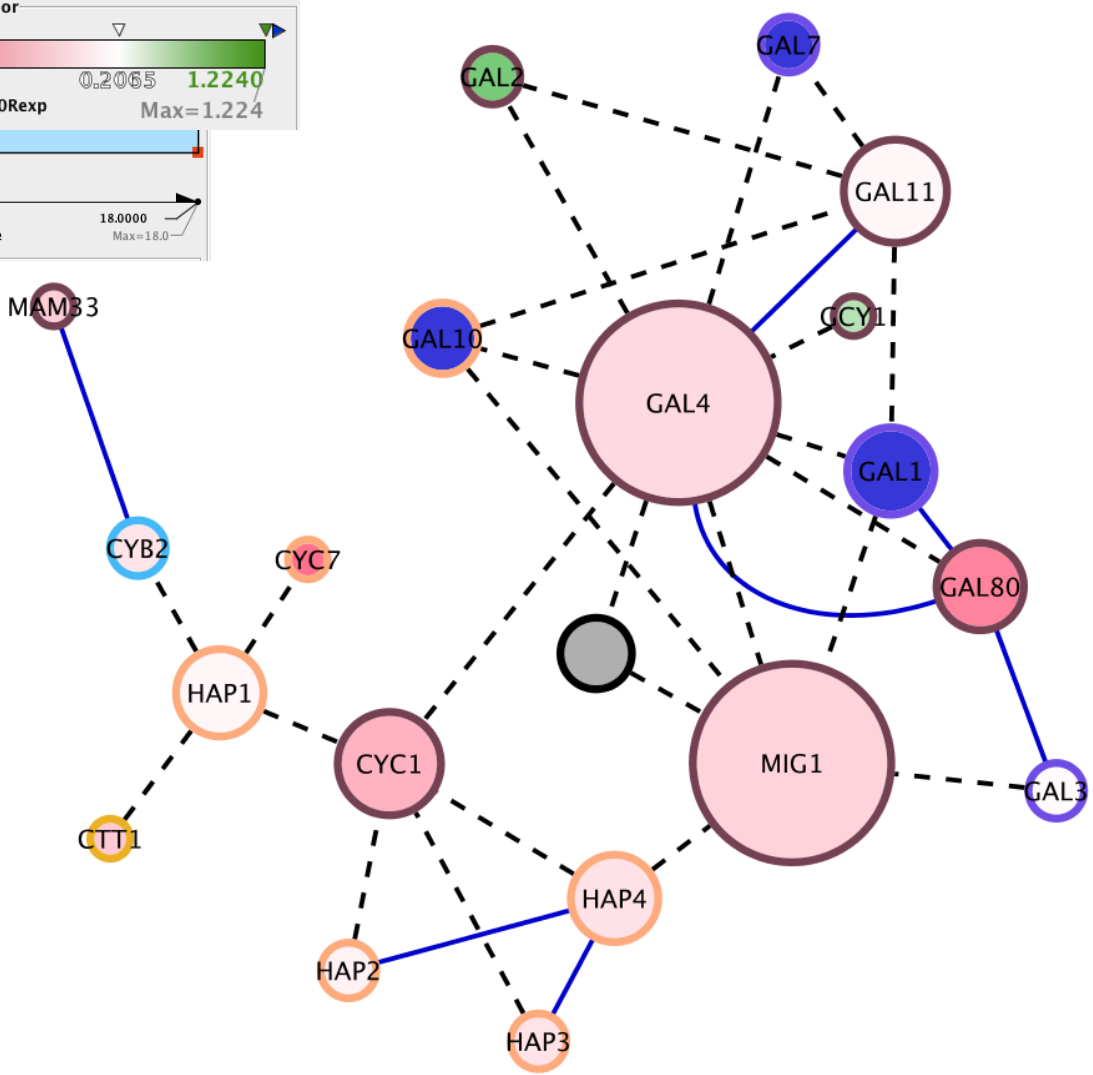
- Depiction
- Data Mapping
- Layouts
- Animation

Depiction



Data mapping



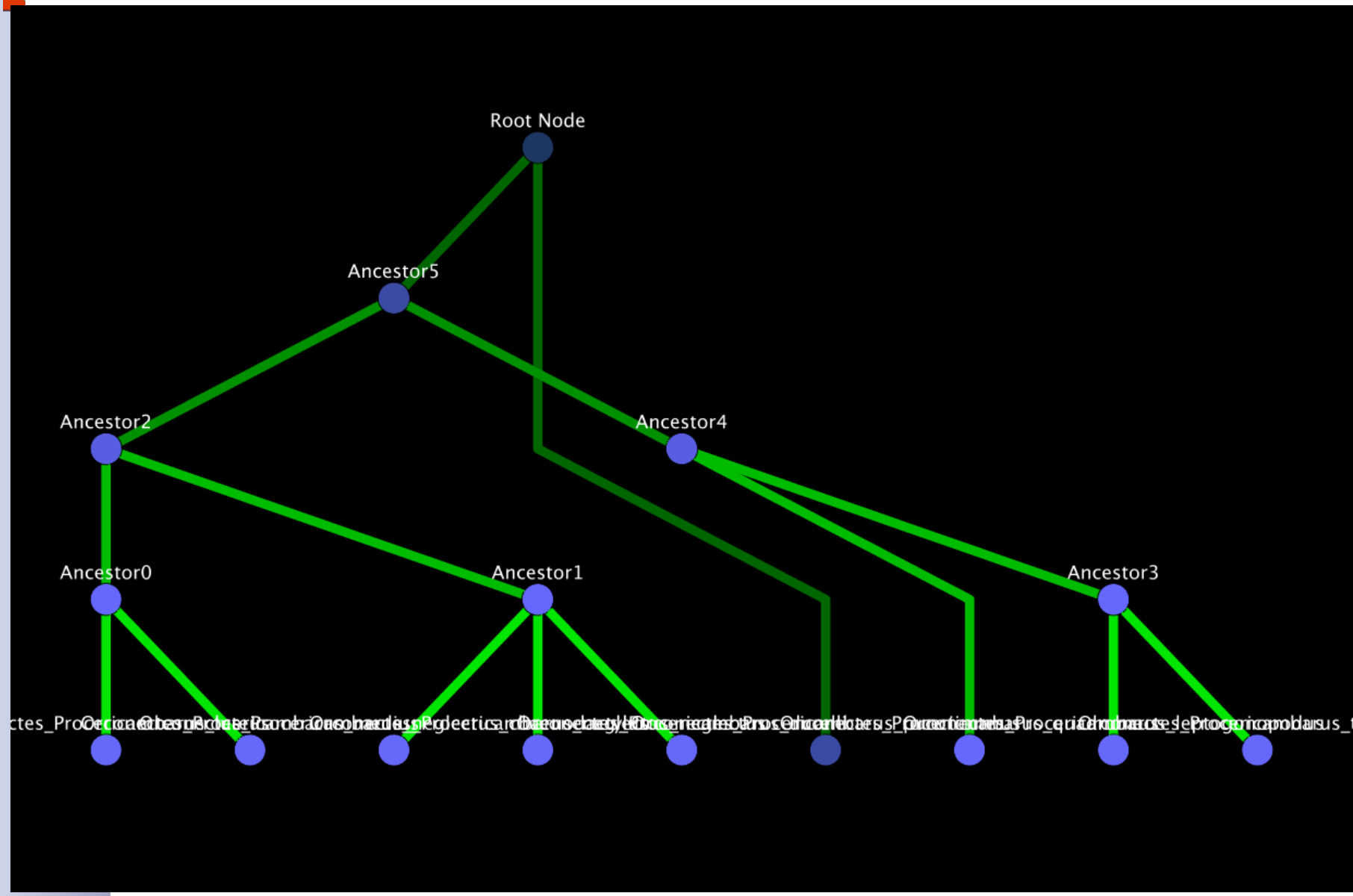




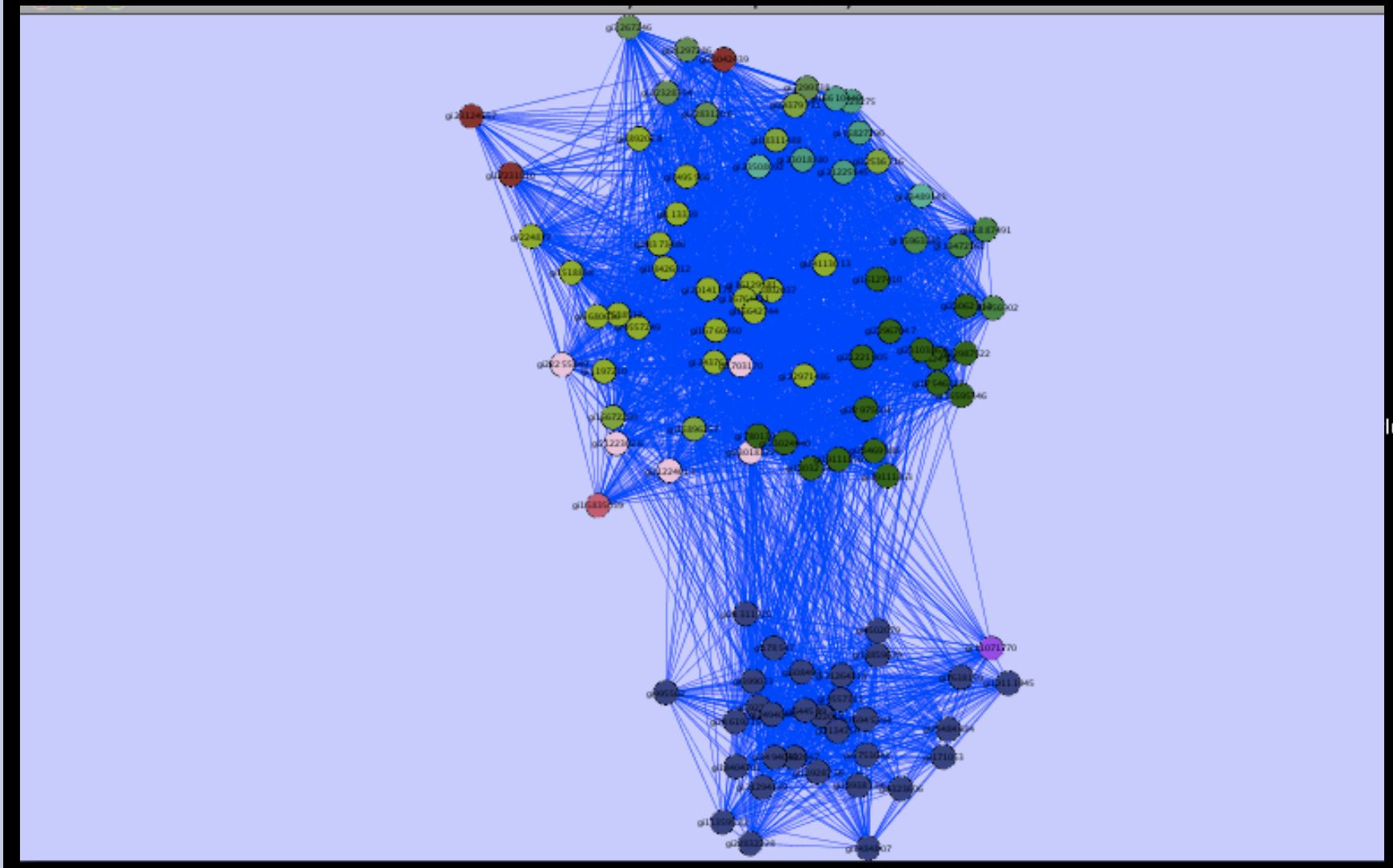
Data mapping

- Avoid cluttering your visualization with too much data
  - Map the data you are specifically interested in to call out meaningful differences
  - Mapping too much data to visual attributes may just confuse the viewer
  - Can always create multiple networks and map different values

Layouts


Root Node
Ancestor5
Ancestor2
Ancestor4
Ancestor0
Ancestor1
Ancestor3

Layouts



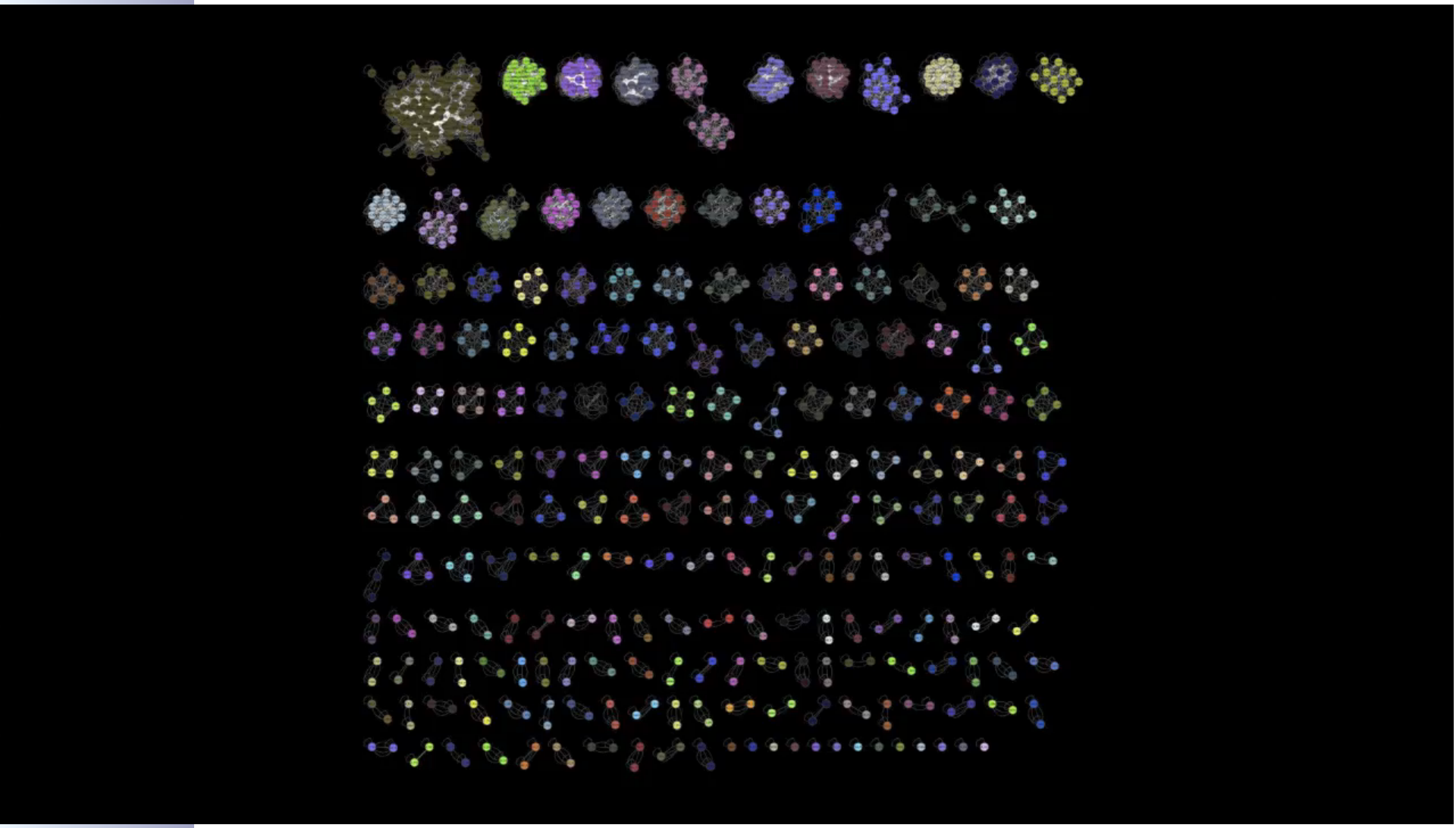
– Many, many others



Layouts

- Use layouts to convey the relationships between the nodes
- Layout algorithms may need to be "tuned" to fit your network
  - Layouts→Settings… menu
- Lots of parameters to change layout algorithm behavior
- Can also consider laying out portions of your network

Animation





Introduction to Cytoscape

- Overview
- Core Concepts
  - Networks vs. Attributes
  - VizMapper
  - Plugins
- Working with Data
  - Loading network from the Web
  - Importing networks from csv files or Excel
  - Importing attributes from csv files or Excel
  - The attribute browser
- Cytoscape tips & tricks

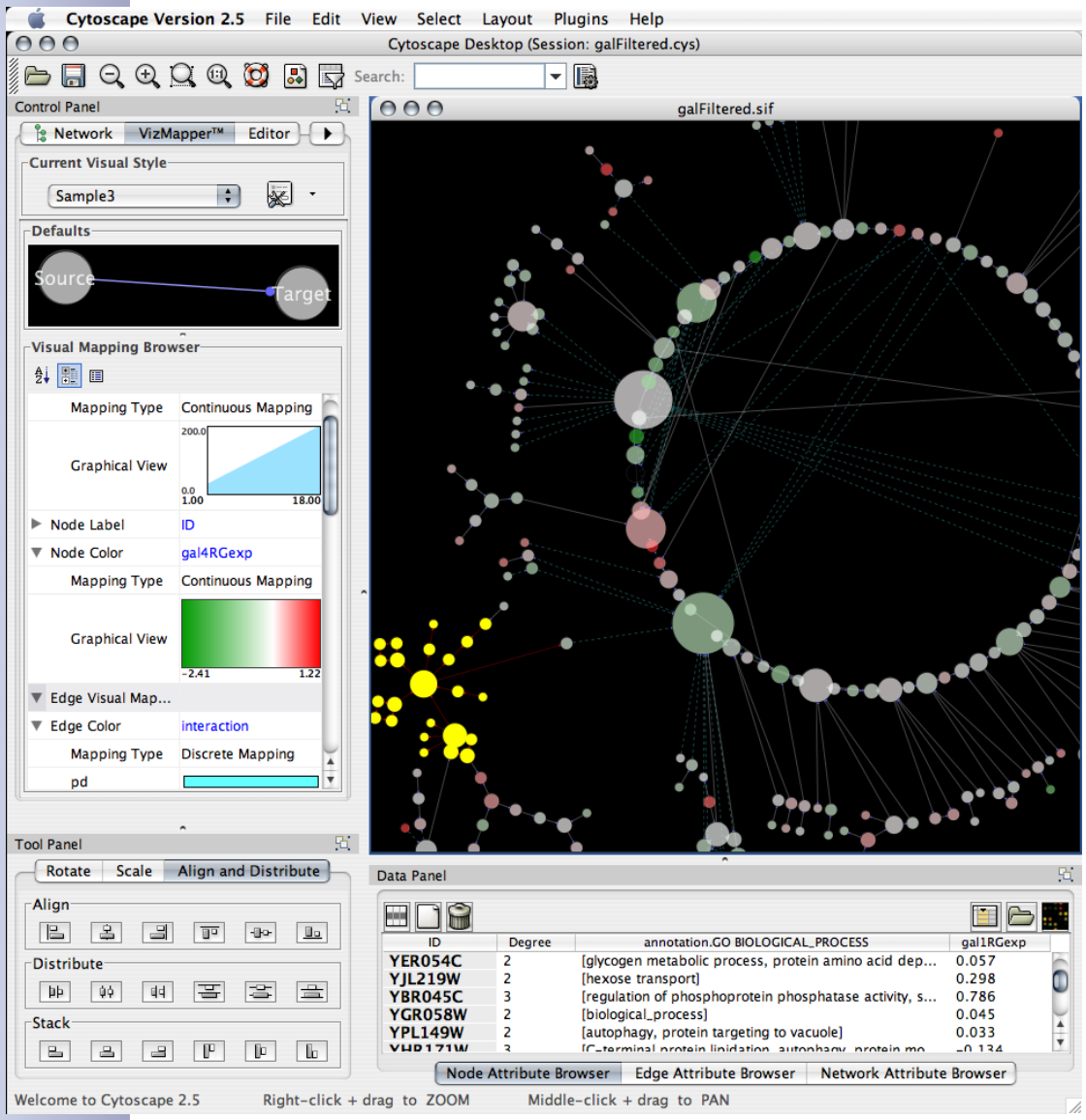


Cytoscape

- About Cytoscape:

  "Cytoscape is an open source bioinformatics software platform for ***visualizing*** molecular interaction networks and ***integrating*** these interactions with gene expression profiles and other state data." – Cytoscape Web Site

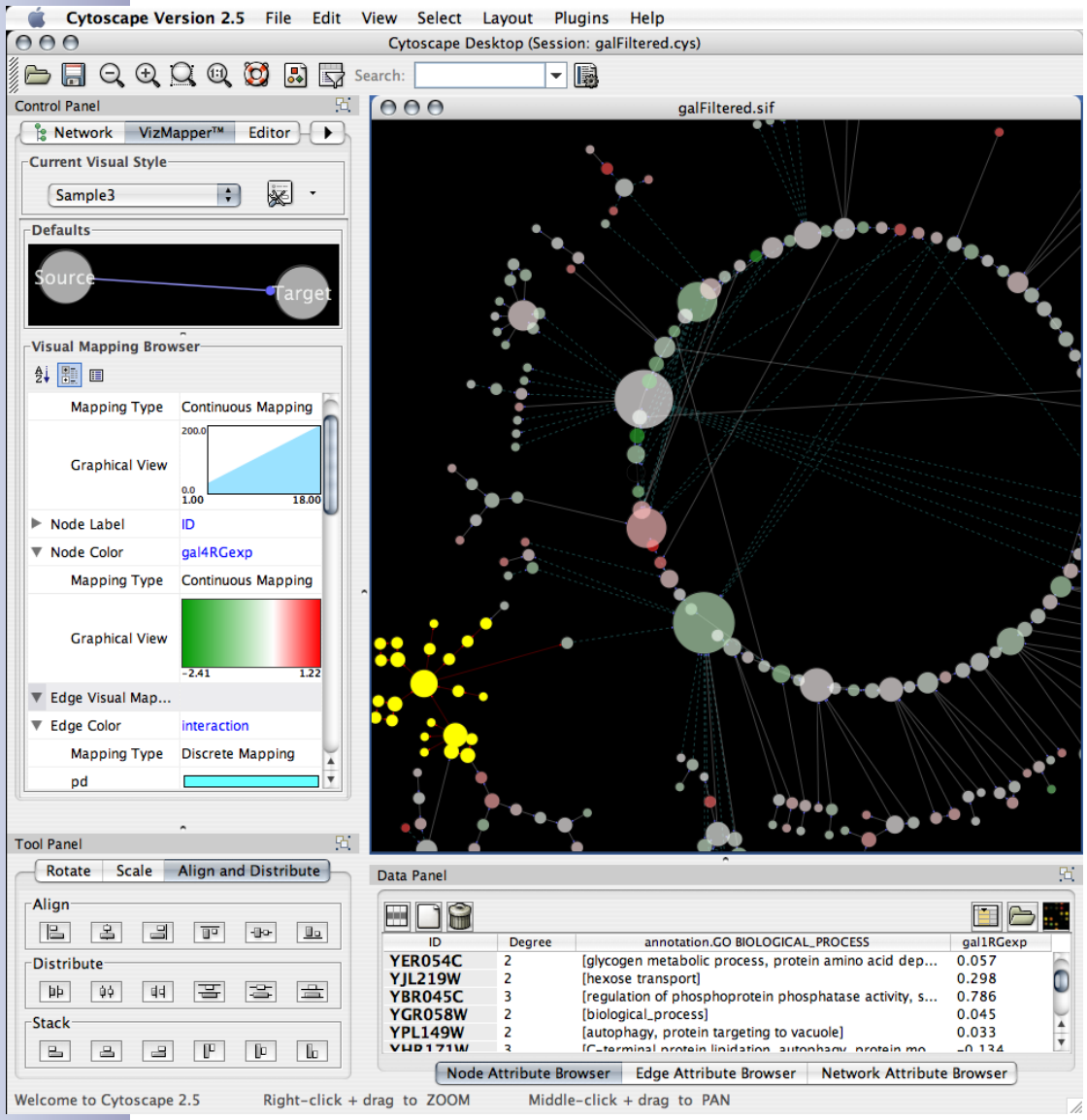
Cytoscape



www.cytoscape.org

- Visualization
- Integration
- Analysis

Cytoscape



- Open source
- Cross platform
- A consortium

Institute for Systems Biology

University of California at San Diego

Memorial Sloan-Kettering Cancer Center

Institut Pasteur

Agilent Technologies

UCSF University of California at San Francisco

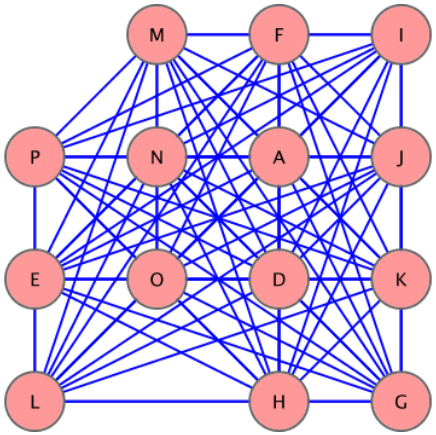
University of Toronto

NCIBI University of Michigan

Unilever

Core Concepts

- Networks and Annotations



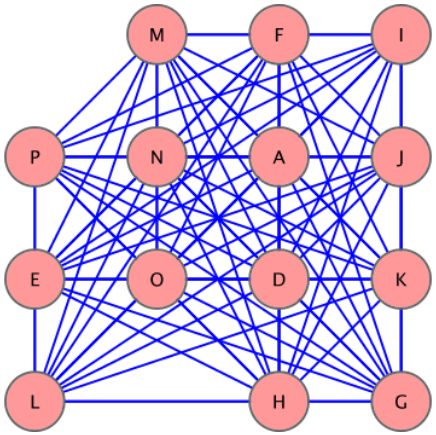
	A	B	C	D	E	F	G	H	I
1	GI_Number	Gene	HPRD_ID	REFSEQ	Entrez_ID	OMIM_ID	SWISSF	Gene_Description	Architecture
2	gi19923959	SDSL	11542	NP_612441.1	113675			SDSL	SP
3	gi14591926	SEC23B	11543	NP_116781.1	10483		Q15437	Protein transport protein SEC23B	GEL
4	gi14602492	SCMH1	11540	AAH09752.1	22955		Q96GD3	SCMH1	SAM
5	gi21361625	EXOC2	11544	NP_060773.3	55770		Q96KP1	SEC5 like 1	IPT
6	gi17998551	SERPINB12	11548	NP_536722.1	89777		Q96P63	Serpin B12	SERPIN
7	gi65287717	EIF2AK4	18447	NP_00101372	440275	609280		Similar to GCN2 eIF2alpha kinase	S_T_Y_Kinase
8	gi4507755	TYROBP	4996	NP_003323.1	7305	604142	O43914	DAP12	ITAM
9	gi5803086	IL24	4995	NP_006841.1	11009	604136	Q13007	Interleukin 24	IL10
10	gi21265034	ADAMTS13	4994	NP_620594.1	11093	604134		ADAMTS 13	TSP1
11	gi1743873	PIP5K1A	4470	AAC50911.1	8394	603275	Q99755	Phosphatidylinositol-4-phosphate 5	PIPKc
12	gi11496982	SVIL	4992	NP_068506.1	6840	604126	O95425	Supervillin	VHP
13	gi31563386	SULT2B1	4991	NP_814444.1	6820	604125		Sulfotransferase family 2B, member 1	
14	gi21040399	RBBP8	4990	AAH30590.1	5932	604124	Q99708	CTIP	LZ
15	gi4758378	FIGF	2102	NP_004460.1	2277	300091	O43915	VEGF D	PDGF
16	gi11321617	DPYSL4	7463	NP_006417.1	10570	608407	O14531	Collapsin response mediator protein	3
17	gi8923202	TASP1	7460	NP_060184.1	55617	608270	Q9H6P5	Threonine aspartase 1	
18	gi5454090	SSR4	2101	NP_006271.1	6748	300090	P51571	Signal sequence receptor delta	SP
19	gi5730045	SLC16A2	2106	NP_006508.1	6567	300095	P36021	X linked PEST containing transporte	TM
20	gi11968027	FTS	7467	NP_071921.1	64400	608483	Q9H8T0	FTS	UBC
21	gi48255885	PRKCI	2105	NP_002731.3	5584	600539	P41743	Protein kinase C, iota type	S_T_kinase
22	gi9558731	RPA4	6591	NP_037479.1	29935			RPA4	TRNAA
23	gi13376812	PPP1R2P9	6593	NP_079486.1	80316			Type 1 protein phosphatase inhibitor	
24	gi15826862		6595	NP_296375.1	90060			JM11 protein	CC
25	gi7661844	CCDC22	6594	NP_054727.1	28952			JM1 protein	CC
26	gi6005794	PRAF2	6596	NP_009144.1	11230			JM4 protein	TM

Networks
e.g, biological pathways

Annotations
e.g., attributes or data

Core Concepts

- Visual Mapping with VizMapper



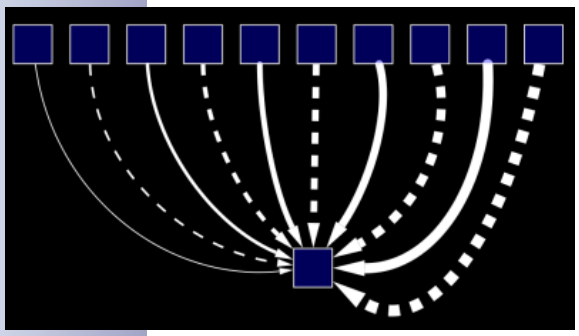
	A	B	C	D	E	F	G	H	I
1	GI_Number	Gene	HPRD_ID	REFSEQ	Entrez_ID	OMIM_ID	SWISSF	Gene_Description	Architecture
2	gi19923959	SDSL	11542	NP_612441.1	113675			SDSL	SP
3	gi14591926	SEC23B	11543	NP_116781.1	10483		Q15437	Protein transport protein SEC23B	GEL
4	gi14602492	SCMH1	11540	AAH09752.1	22955		Q96GD3	SCMH1	SAM
5	gi21361625	EXOC2	11544	NP_060773.3	55770		Q96KP1	SEC5 like 1	IPT
6	gi17998551	SERPINB12	11548	NP_536722.1	89777		Q96P63	Serpin B12	SERPIN
7	gi65287717	EIF2AK4	18447	NP_00101372	440275	609280		Similar to GCN2 eIF2alpha kinase	S_T_Y_Kinase
8	gi4507755	TYROBP	4996	NP_003323.1	7305	604142	O43914	DAP12	ITAM
9	gi5803086	IL24	4995	NP_006841.1	11009	604136	Q13007	Interleukin 24	IL10
10	gi21265034	ADAMTS13	4994	NP_620594.1	11093	604134		ADAMTS 13	TSP1
11	gi1743873	PIP5K1A	4470	AAC50911.1	8394	603275	Q99755	Phosphatidylinositol-4-phosphate 5	PIPKc
12	gi11496982	SVIL	4992	NP_068506.1	6840	604126	O95425	Supervillin	VHP
13	gi31563386	SULT2B1	4991	NP_814444.1	6820	604125		Sulfotransferase family 2B, member 1	
14	gi21040399	RBBP8	4990	AAH30590.1	5932	604124	Q99708	CTIP	LZ
15	gi4758378	FIGF	2102	NP_004460.1	2277	300091	O43915	VEGF D	PDGF
16	gi11321617	DPYSL4	7463	NP_006417.1	10570	608407	O14531	Collapsin response mediator protein 3	
17	gi8923202	TASP1	7460	NP_060184.1	55617	608270	Q9H6P5	Threonine aspartase 1	
18	gi5454090	SSR4	2101	NP_006271.1	6748	300090	P51571	Signal sequence receptor delta	SP
19	gi5730045	SLC16A2	2106	NP_006508.1	6567	300095	P36021	X linked PEST containing transporte	TM
20	gi11968027	FTS	7467	NP_071921.1	64400	608483	Q9H8T0	FTS	UBC
21	gi48255885	PRKCI	2105	NP_002731.3	5584	600539	P41743	Protein kinase C, iota type	S_T_kinase
22	gi9558731	RPA4	6591	NP_037479.1	29935			RPA4	TRNAA
23	gi13376812	PPP1R2P9	6593	NP_079486.1	80316			Type 1 protein phosphatase inhibitor	
24	gi15826862		6595	NP_296375.1	90060			JM11 protein	CC
25	gi7661844	CCDC22	6594	NP_054727.1	28952			JM1 protein	CC
26	gi6005794	PRAF2	6596	NP_009144.1	11230			JM4 protein	TM

Networks

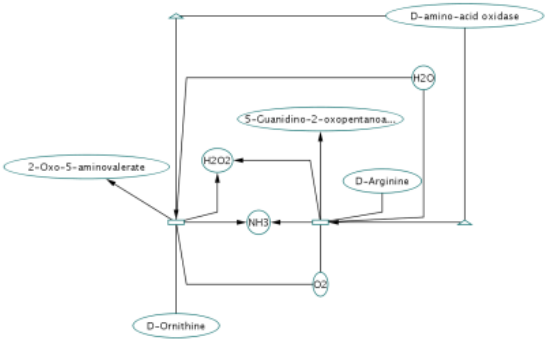
Annotations

VizMapper

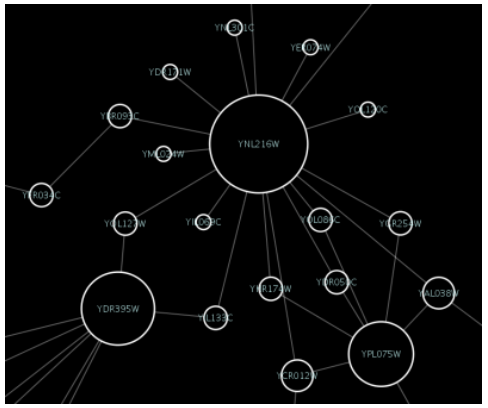
Visual Styles



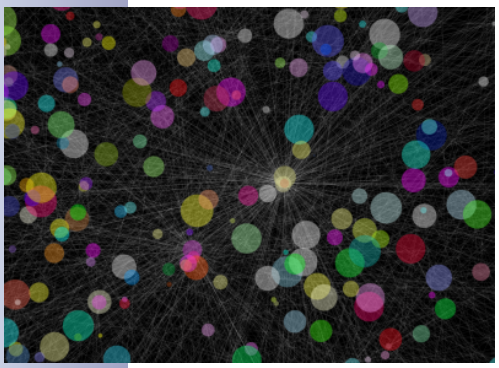
Use specific line types to indicate different types of interactions



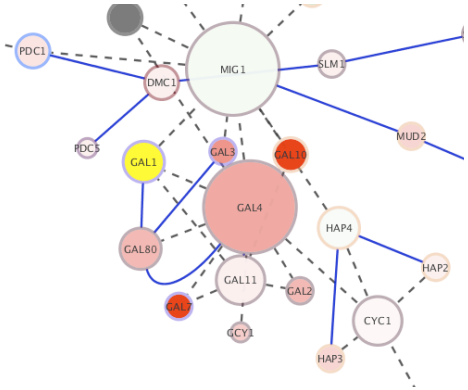
Encode specific physical entities as different node shapes.



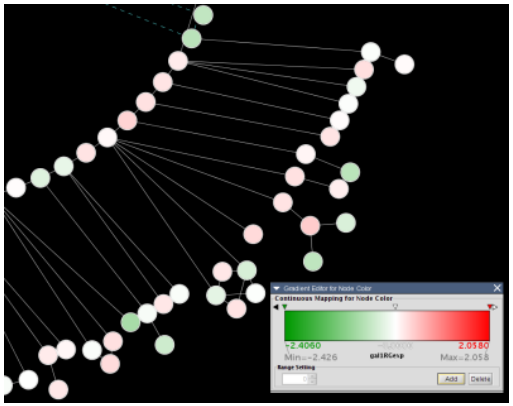
Set node sizes based on the degree of connectivity of the nodes.



Browse extremely dense networks by controlling for the opacity of nodes



Set node font sizes based on the degree of connectivity of the nodes.



Expression data



Core Concepts

- Plugins!

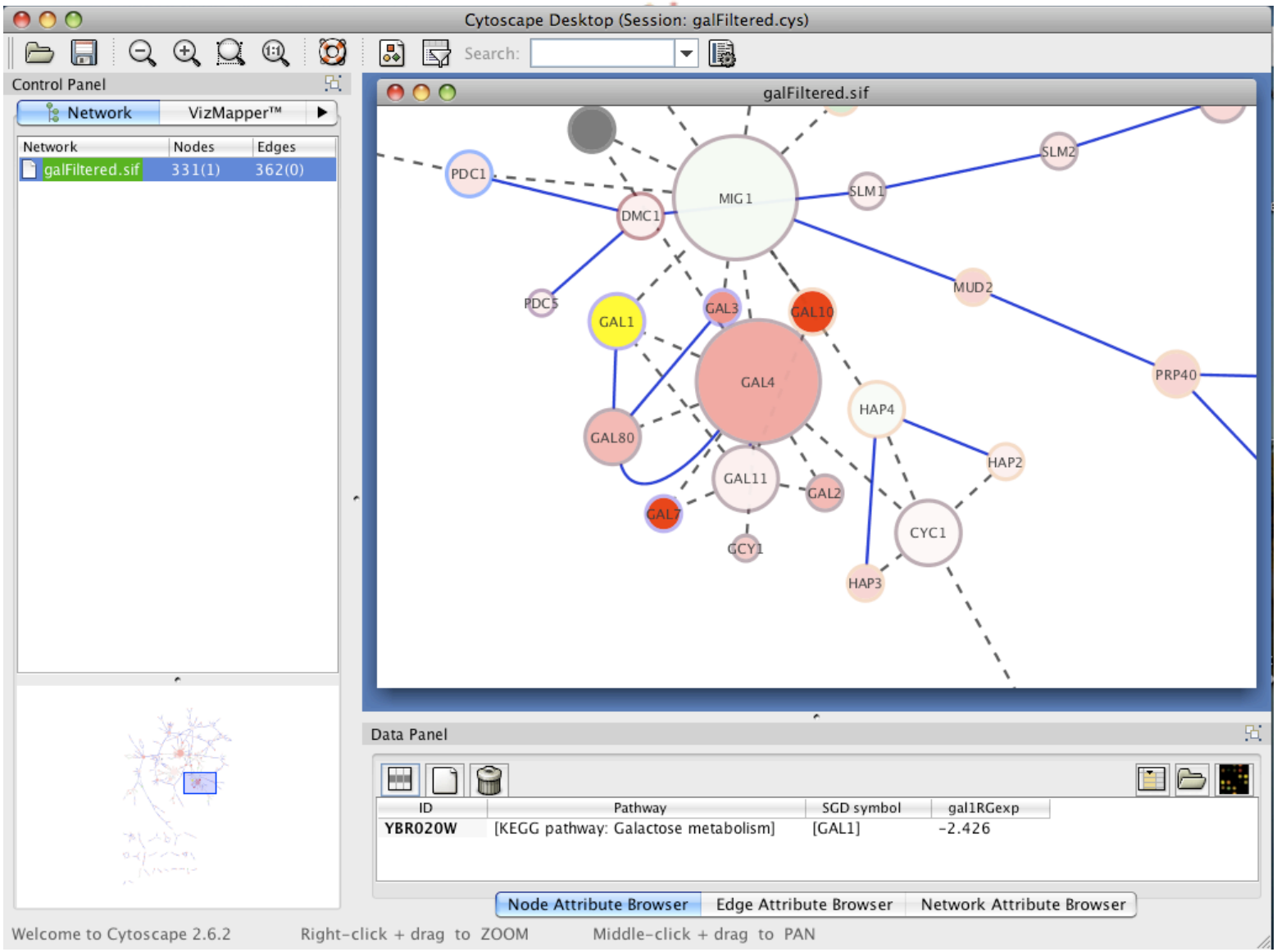
http://cytoscape.org/plugins.html



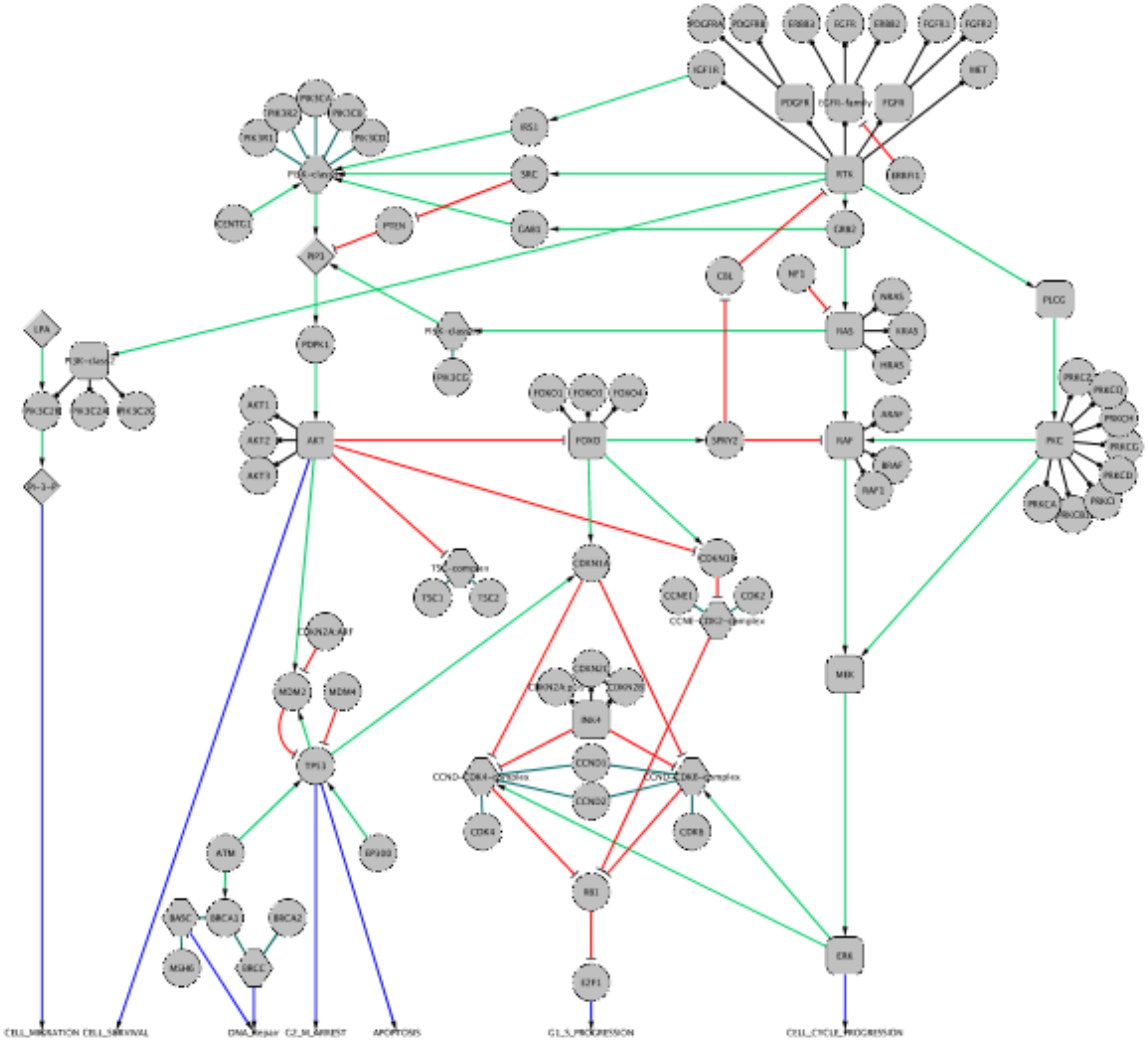
Cytoscape

- Traditional uses
  - Visualizing:
    - PPI
    - Pathways
  - Mapping:
    - Expression profiles
    - “Other state data”

Cytoscape

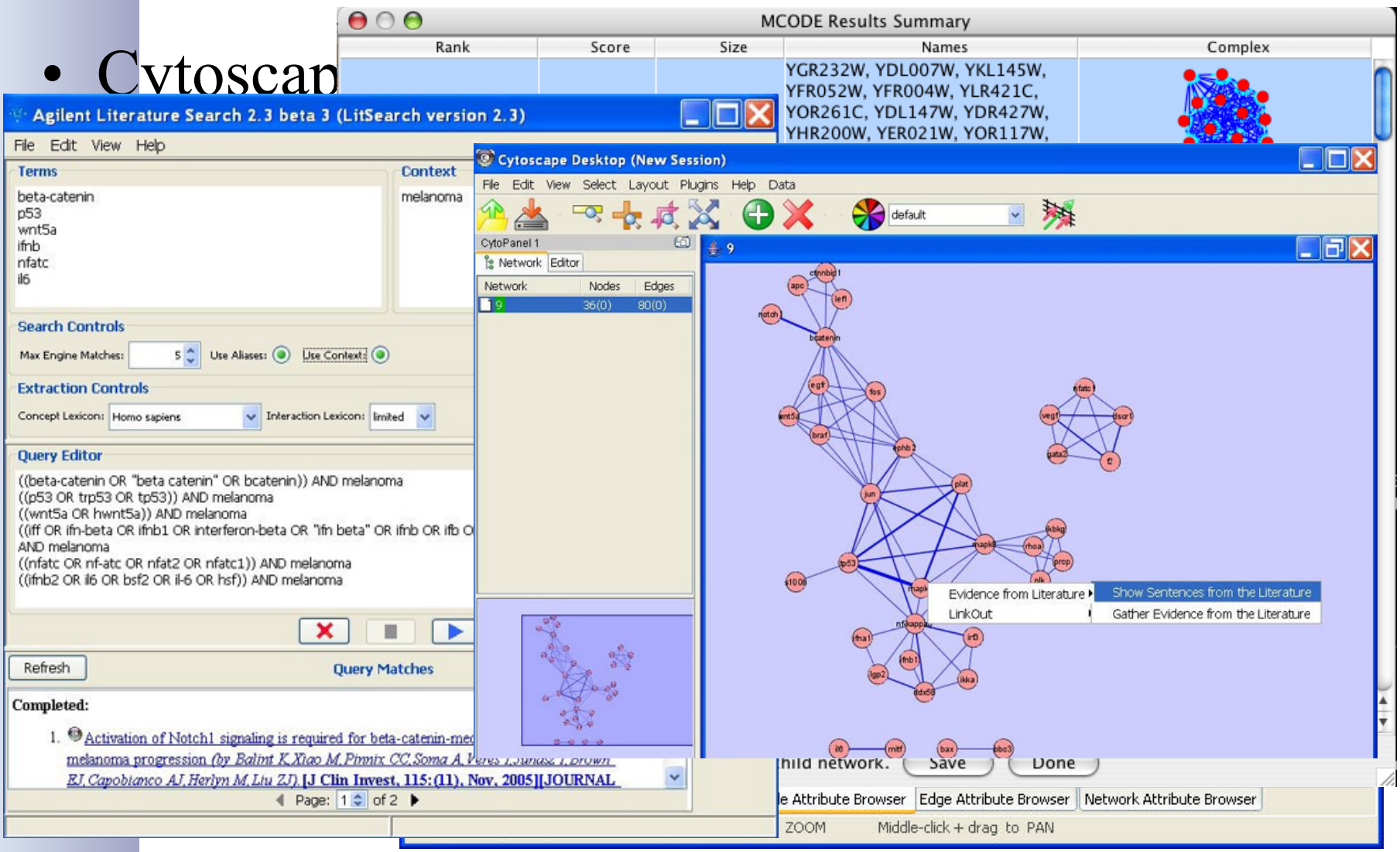

Cytoscape Desktop (Session: galFiltered.cys)
Search:
Control Panel
Network
VizMapper™
Network
Nodes
Edges
galFiltered.sif
331(1)
362(0)
galFiltered.sif
PDC1
DMC1
MIG1
SLM1
SLM2
MUD2
PRP40
PDC5
GAL1
GAL3
GAL10
GAL4
HAP4
GAL80
HAP2
GAL11
GAL2
GAL7
CYC1
GCY1
HAP3
Data Panel
ID
Pathway
SGD symbol
gal1RGexp
YBR020W
[KEGG pathway: Galactose metabolism]
[GAL1]
-2.426
Node Attribute Browser
Edge Attribute Browser
Network Attribute Browser
Welcome to Cytoscape 2.6.2
Right-click + drag to ZOOM
Middle-click + drag to PAN

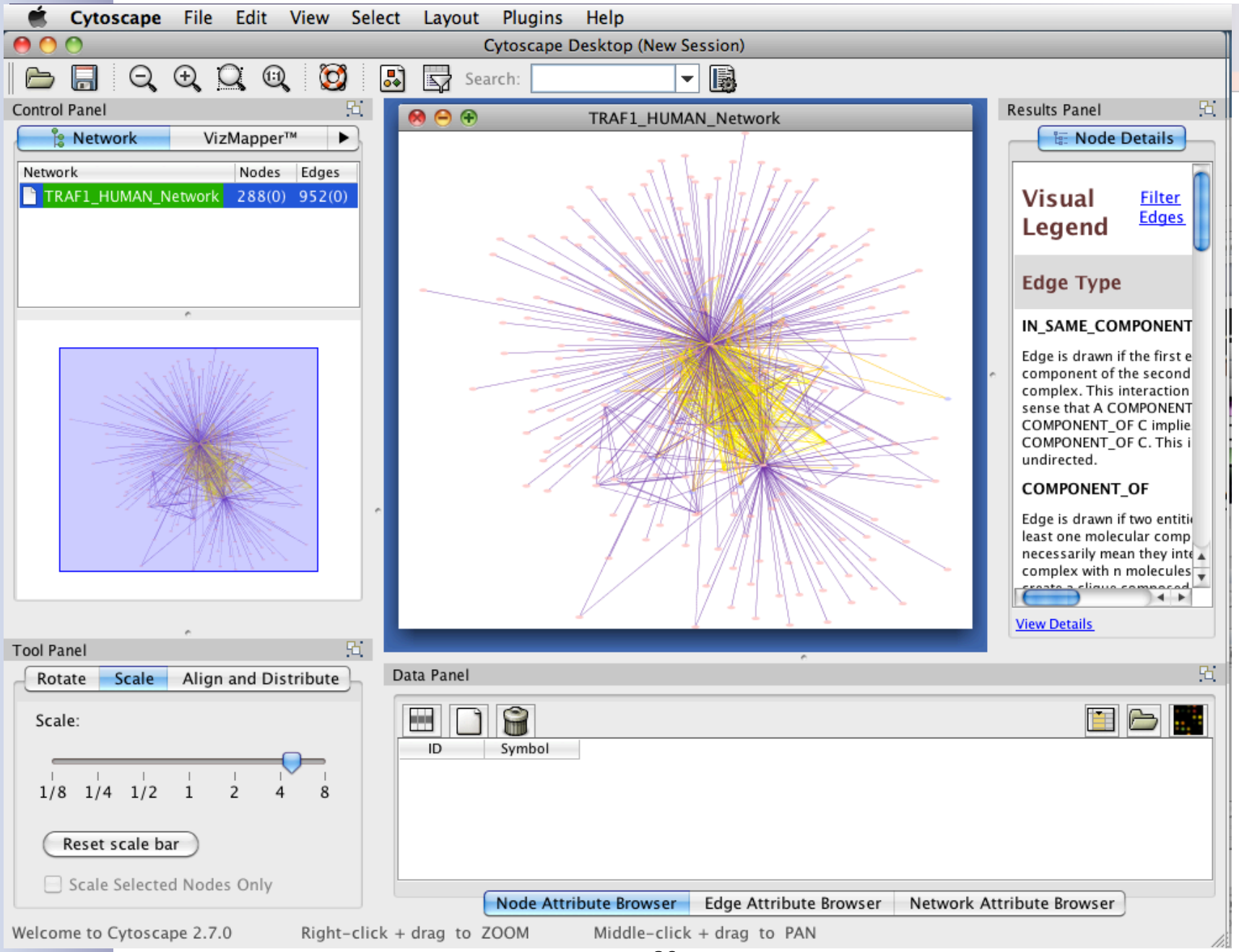
Cytoscape



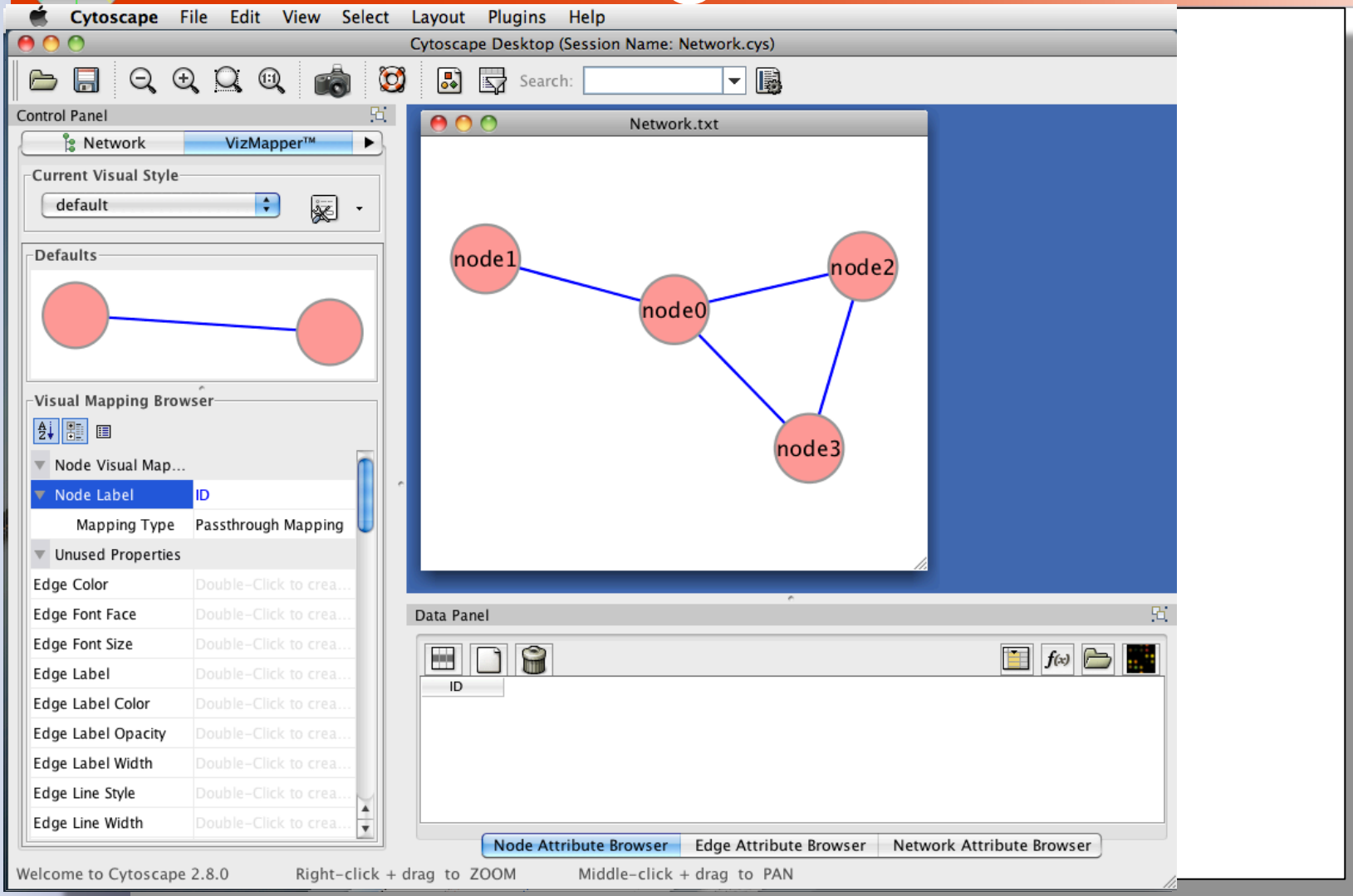
Cytoscape: Platform

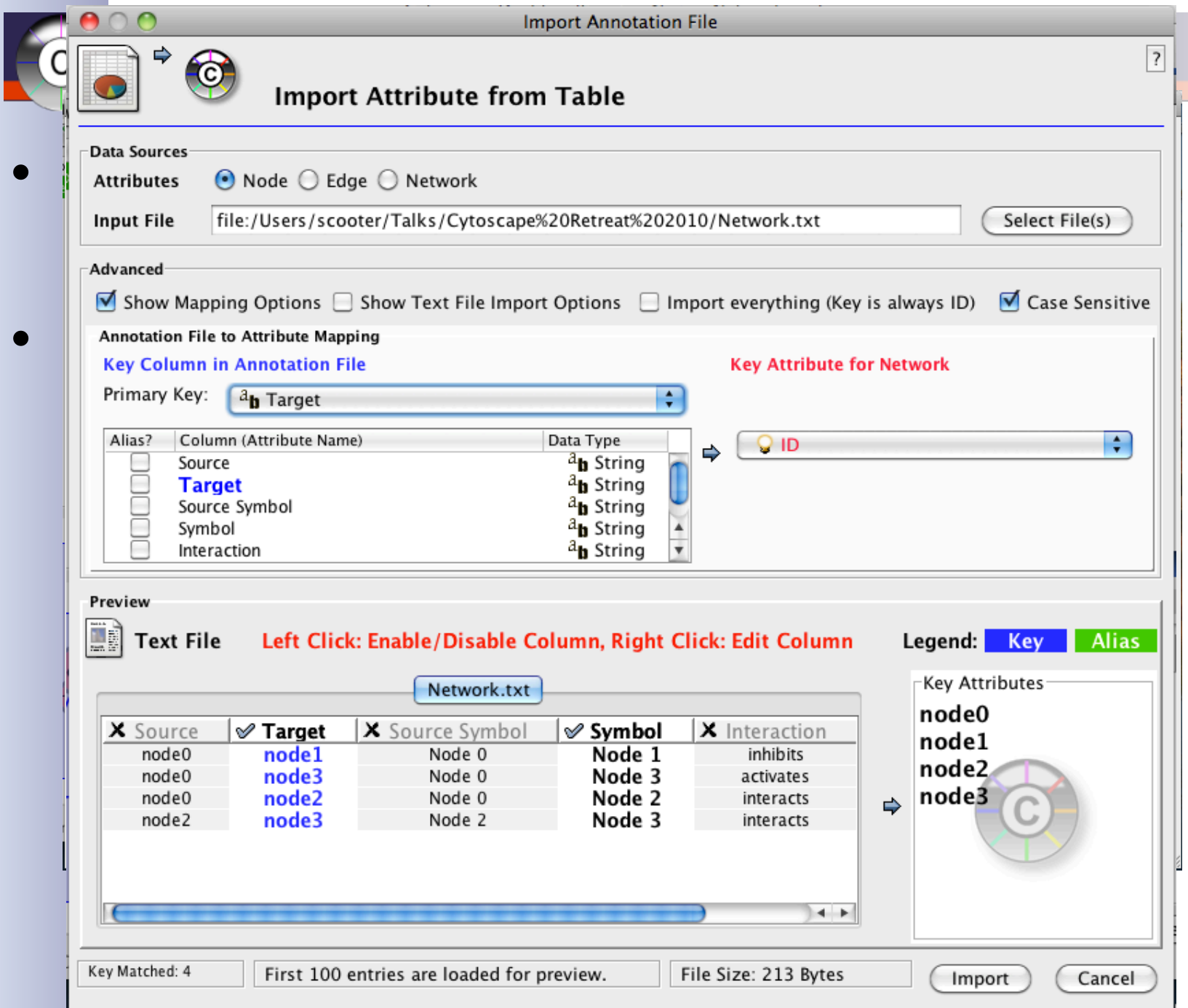
- Cytoscap



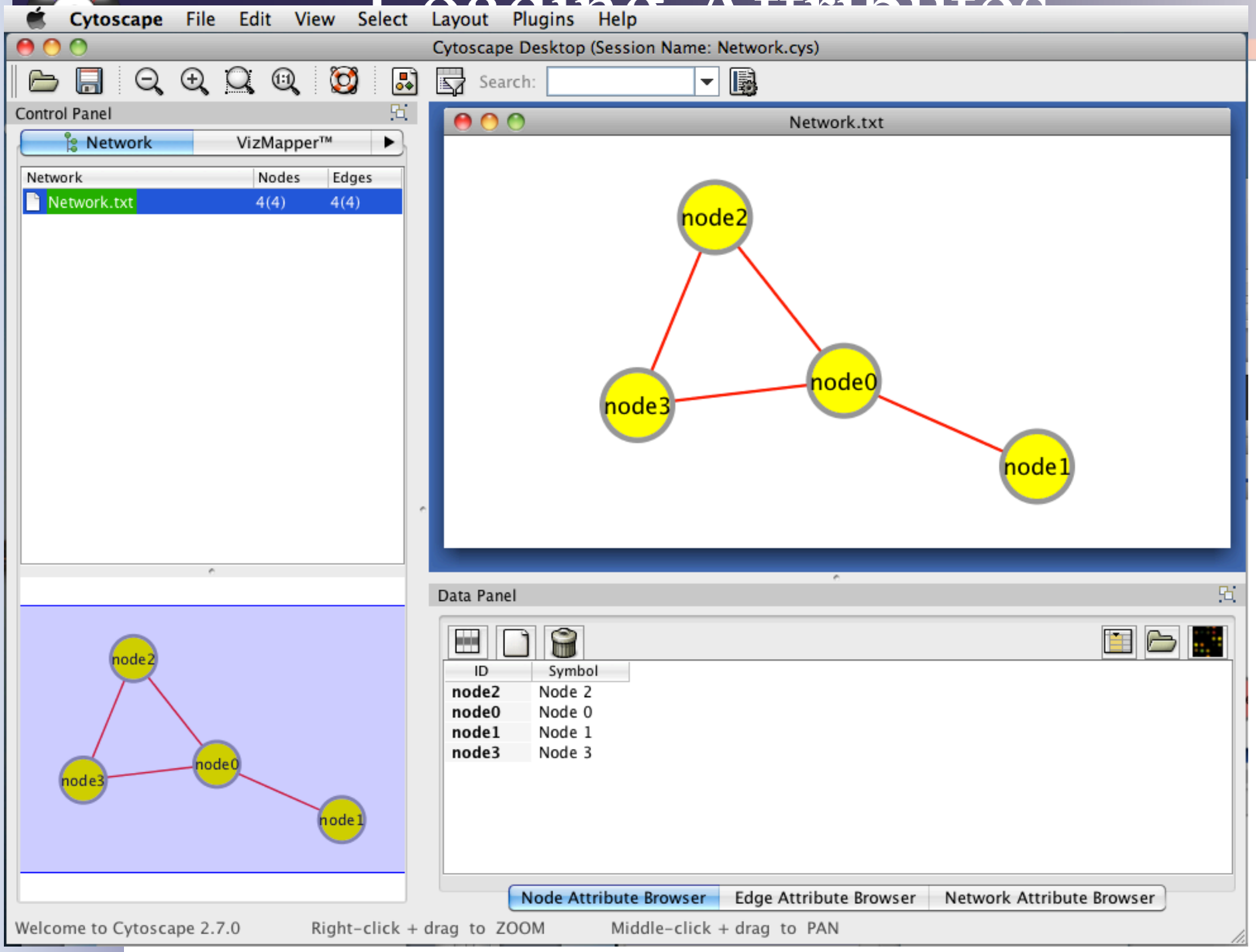

Cytoscape File Edit View Select Layout Plugins Help
Cytoscape Desktop (New Session)
Search:
Control Panel
Network VizMapper™
Network Nodes Edges
TRAF1_HUMAN_Network 288(0) 952(0)
TRAF1_HUMAN_Network
Results Panel
Node Details
Visual Legend
Filter Edges
Edge Type
IN_SAME_COMPONENT
Edge is drawn if the first e
component of the second
complex. This interaction
sense that A COMPONENT
COMPONENT_OF C implie
COMPONENT_OF C. This i
undirected.
COMPONENT_OF
Edge is drawn if two entiti
least one molecular comp
necessarily mean they inte
complex with n molecules
View Details
Tool Panel
Rotate Scale Align and Distribute
Scale:
1/8 1/4 1/2 1 2 4 8
Reset scale bar
Scale Selected Nodes Only
Data Panel
ID Symbol
Node Attribute Browser Edge Attribute Browser Network Attribute Browser
Welcome to Cytoscape 2.7.0
Right-click + drag to ZOOM
Middle-click + drag to PAN

Loading Networks




Import Annotation File
Import Attribute from Table
Data Sources
Attributes
Node
Edge
Network
Input File
file:/Users/scooter/Talks/Cytoscape%20Retreat%202010/Network.txt
Select File(s)
Advanced
Show Mapping Options
Show Text File Import Options
Import everything (Key is always ID)
Case Sensitive
Annotation File to Attribute Mapping
Key Column in Annotation File
Key Attribute for Network
Primary Key:
Target
Alias?
Column (Attribute Name)
Data Type
Source
String
Target
String
Source Symbol
String
Symbol
String
Interaction
String
ID
Preview
Text File
Left Click: Enable/Disable Column, Right Click: Edit Column
Legend:
Key
Alias
Network.txt
Source
Target
Source Symbol
Symbol
Interaction
node0
node1
Node 0
Node 1
inhibits
node0
node3
Node 0
Node 3
activates
node0
node2
Node 0
Node 2
interacts
node2
node3
Node 2
Node 3
interacts
Key Attributes
node0
node1
node2
node3
Key Matched: 4
First 100 entries are loaded for preview.
File Size: 213 Bytes
Import
Cancel

Loading Attributes



Tips & Tricks

- "Root graph"
  - "There is one graph to rule them all…."
  - The networks in Cytoscape are all "views" on a single graph.
  - Changing the attribute for a node in one network *will* also change that attribute for a node with the same ID in all other loaded networks
  - There is no way to "copy" a node and keep the same ID
  - Make a copy of the session

Tips & Tricks

- Network views
  - When you open a large network, you will not get a view by default
  - To improve interactive performance, Cytoscape has the concept of "Levels of Detail"
    - Some visual attributes will only be apparent when you zoom in
    - The level of detail for various attributes can be changed in the preferences
    - To see what things will look like at full detail:
      - View→Show Graphics Details



Tips & Tricks

- Sessions
  - Sessions save pretty much everything:
    - Networks
    - Properties
    - Visual styles
    - Screen sizes
  - Saving a session on a large screen may require some resizing when opened on your laptop



Tips & Tricks

- Logging
  - By default, Cytoscape writes it's logs to the Error Dialog: Help→Error Dialog
  - Can change a preference to write it to the console
    - Edit→Preferences→Properties…
    - Set logger.console to true
    - Don't forget to save your preferences
    - Restart Cytoscape
  - (can also turn on debugging: cytoscape.debug, but I don't recommend it)



Tips & Tricks

- Memory
  - Cytoscape uses lots of it
  - Doesn't like to let go of it
  - An occasional restart when working with large networks is a good thing
  - Destroy views when you don't need them
  - Java doesn't give us a good way to get the memory right at start time
    - Cytoscape 2.7 does a much better job at "guessing" good default memory sizes than previous versions



Tips & Tricks

- .cytoscape directory
  - Your defaults and any plugins downloaded from the plugin manager will go here
  - Sometimes, if things get really messed up, deleting (or renaming) this directory can give you a "clean slate"
- Plugin manager
  - "Outdated" doesn't necessarily mean "won't work"
  - Plugin authors don't always update their plugins immediately after new releases
  - Click on "Show outdated plugins" to see the entire list of plugins.



Examples/Demos

- clusterMaker
  - Clustering and cluster visualizations
- BiNGO
  - Gene Ontology overrepresentation analysis
- structureViz
  - Bringing protein structure to networks
- Mosaic
  - Detangling network hairballs into pathways