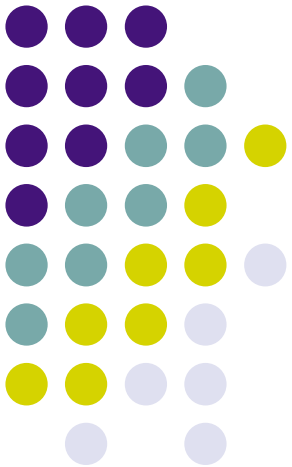
Cytoscape

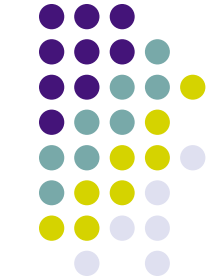
Efficient Rendering and Modeling of Large Biological Networks

Nerius Landys

Trey Ideker group, UCSD



What is Cytoscape?



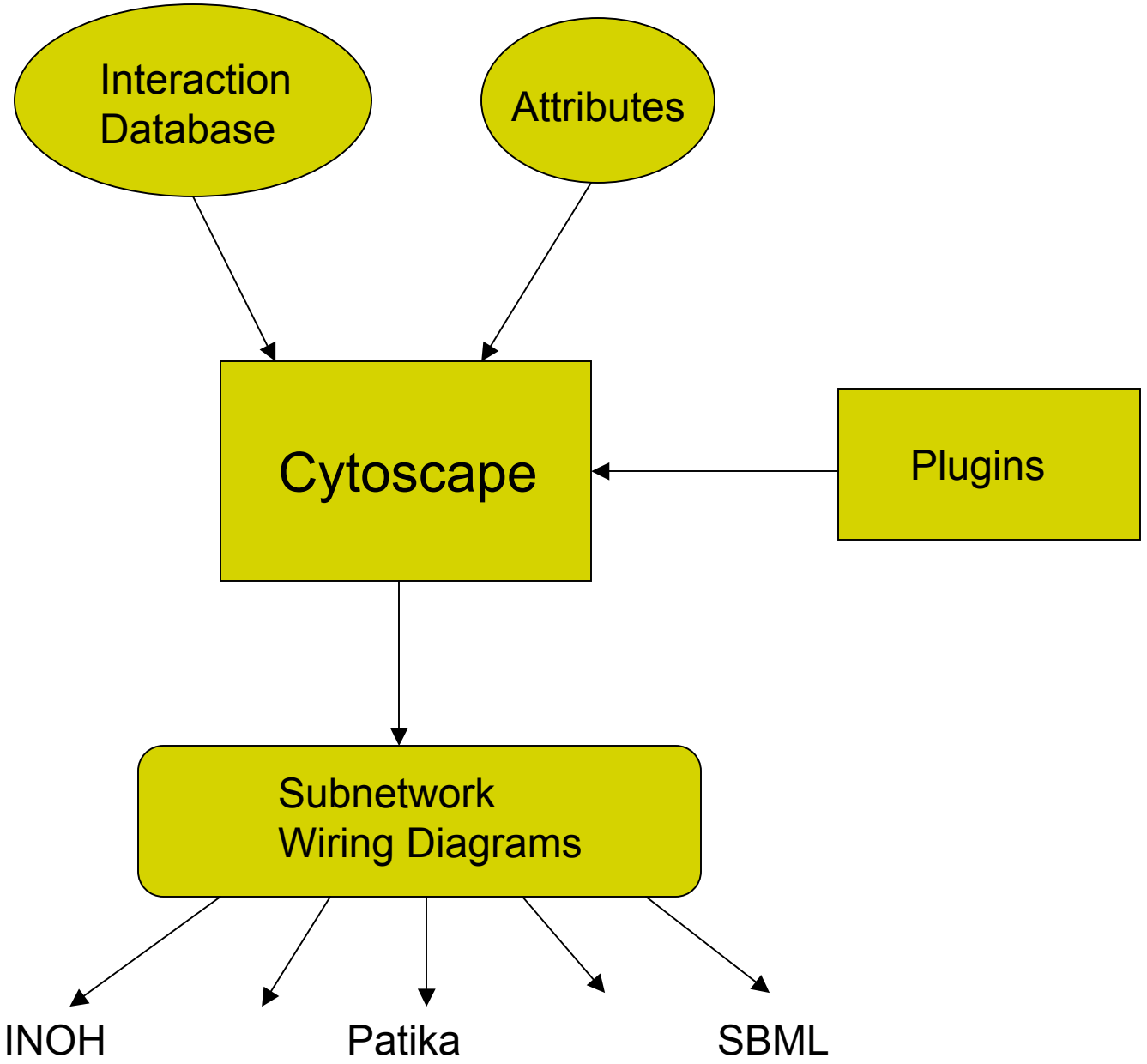
- Cytoscape is a tool for visualizing networks in general (not necessarily biological).
- With Cytoscape, one is able to visualize attributes associated with nodes and edges.
- Cytoscape's open-source nature enables programmers to write custom code against a well-established API to do various computations on networks.
- Cytoscape is a huge collaboration between five labs.

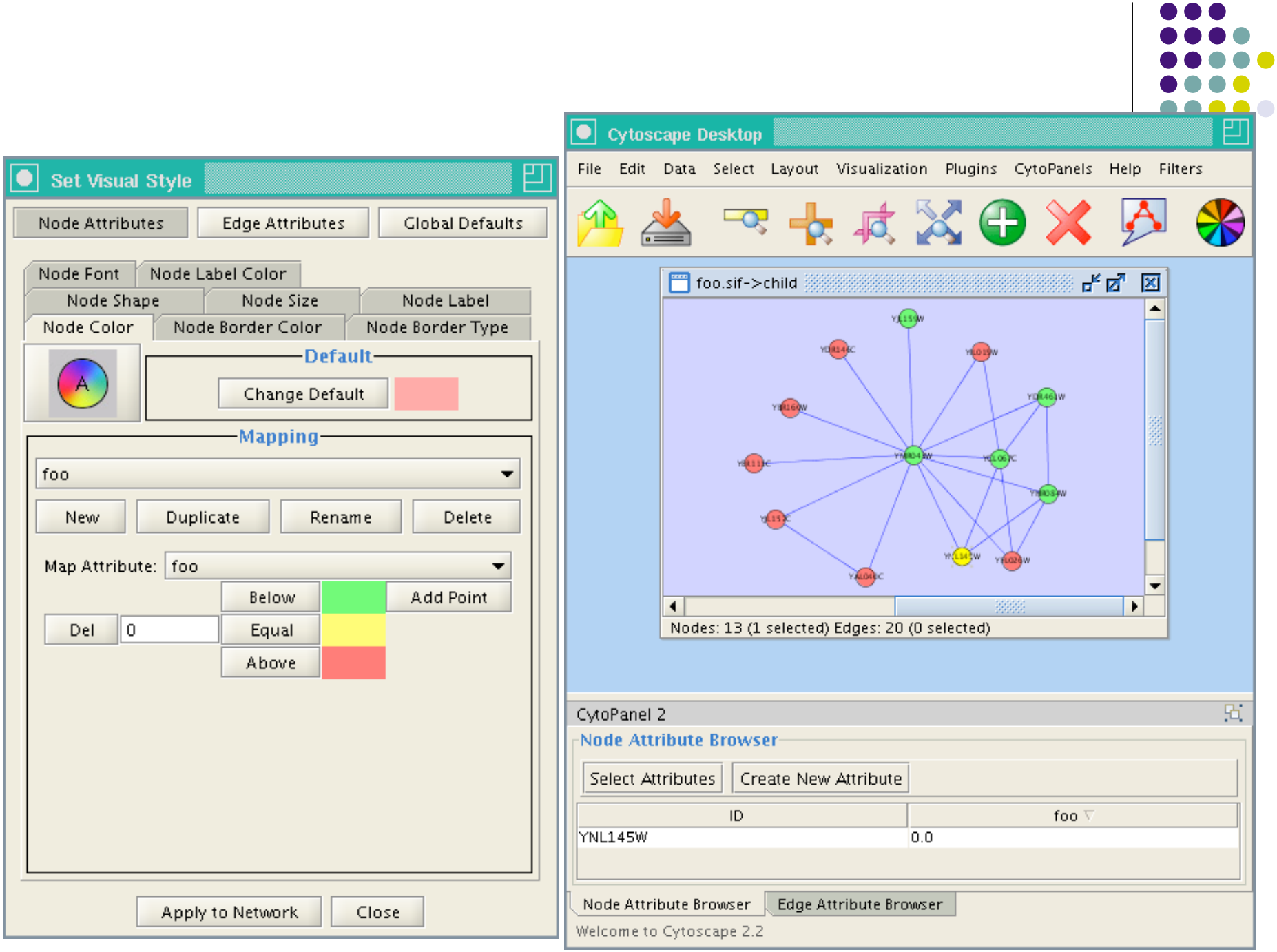
Enabling biological semantics + representation in Cytoscape

We try to be agnostic with respect to semantics, but we try to support community standards.

What is Cytoscape's niche?

- Try to provide a tool to distill abstract biological models (protein wiring diagrams) out of large molecular interaction databases
- Cytoscape is not a general-purpose rendering tool
- Cytoscape's strength: rendering networks (networks in pure mathematical sense).


Interaction Database
Attributes
Cytoscape
Plugins
Subnetwork Wiring Diagrams
INOH
Patika
SBML


Set Visual Style
Node Attributes
Edge Attributes
Global Defaults
Node Font
Node Label Color
Node Shape
Node Size
Node Label
Node Color
Node Border Color
Node Border Type
Default
Change Default
Mapping
foo
New
Duplicate
Rename
Delete
Map Attribute:
foo
Below
Add Point
Del
0
Equal
Above
Apply to Network
Close
Cytoscape Desktop
File
Edit
Data
Select
Layout
Visualization
Plugins
CytoPanels
Help
Filters
foo.sif->child
Nodes: 13 (1 selected) Edges: 20 (0 selected)
CytoPanel 2
Node Attribute Browser
Select Attributes
Create New Attribute
ID
foo
YNL145W
0.0
Node Attribute Browser
Edge Attribute Browser
Welcome to Cytoscape 2.2

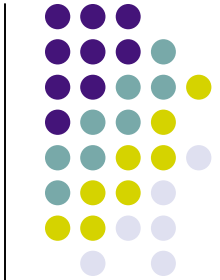
Currently Supported Visual Attributes for Nodes

- Node shape: rectangle, diamond, triangle, ellipse, rounded rectangle, hexagon, octagon, parallelogram
- Node paint
- Node size
- Node border thickness
- Node border paint
- Extensive text label support
- Some amount of custom node features (although it's a real hack at this point)

Currently Supported Visual Attributes for Edges

- Arrowhead (both for target and source): none, delta, disc, tee, diamond
- Arrowhead size
- Arrowhead paint
- Edge thickness
- Edge paint
- Edge dash length
- Edge anchor points (to enable zig-zag or curvy edges)
- Extensive text label support

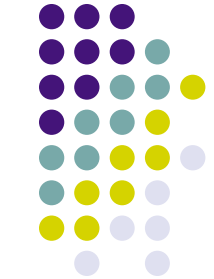
Why reinvent the wheel?



Existing network rendering engines all seem to fall into at least one of the following categories:

1. Proprietary software, costs money, and/or not open source.
2. Low performance: architected heavily towards an object-oriented design (as opposed to procedural); therefore not suitable for some of the optimizations that are going into new engine.

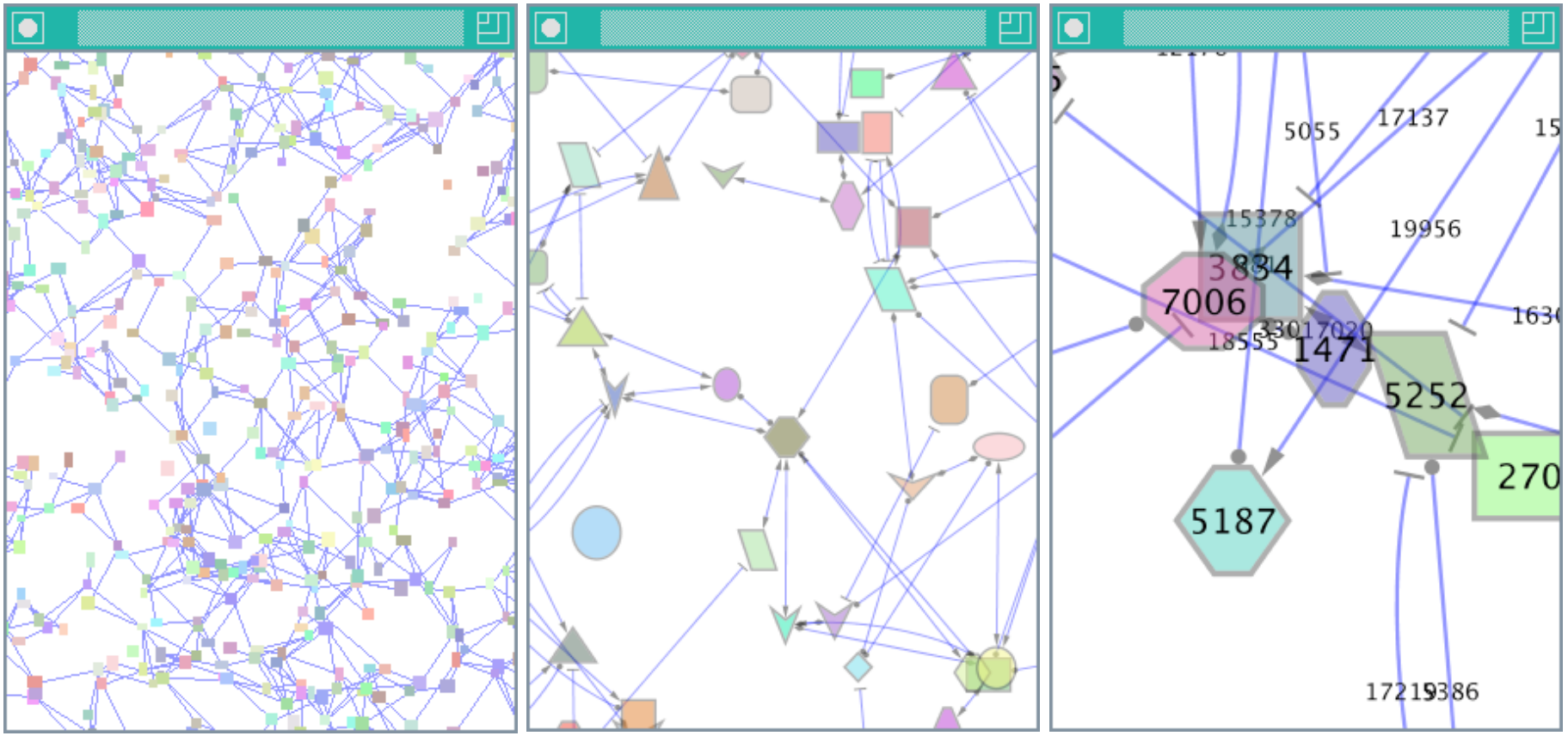
Rendering Engine Goals:



- Most important: 100,000+ nodes and 200,000+ edges no problem
- Maybe even 2,000,000 nodes.
- Simplicity
- Keep semantics (biological or other) out of core
- Enough flexibility to support community standards

Optimization #1: Level of Detail

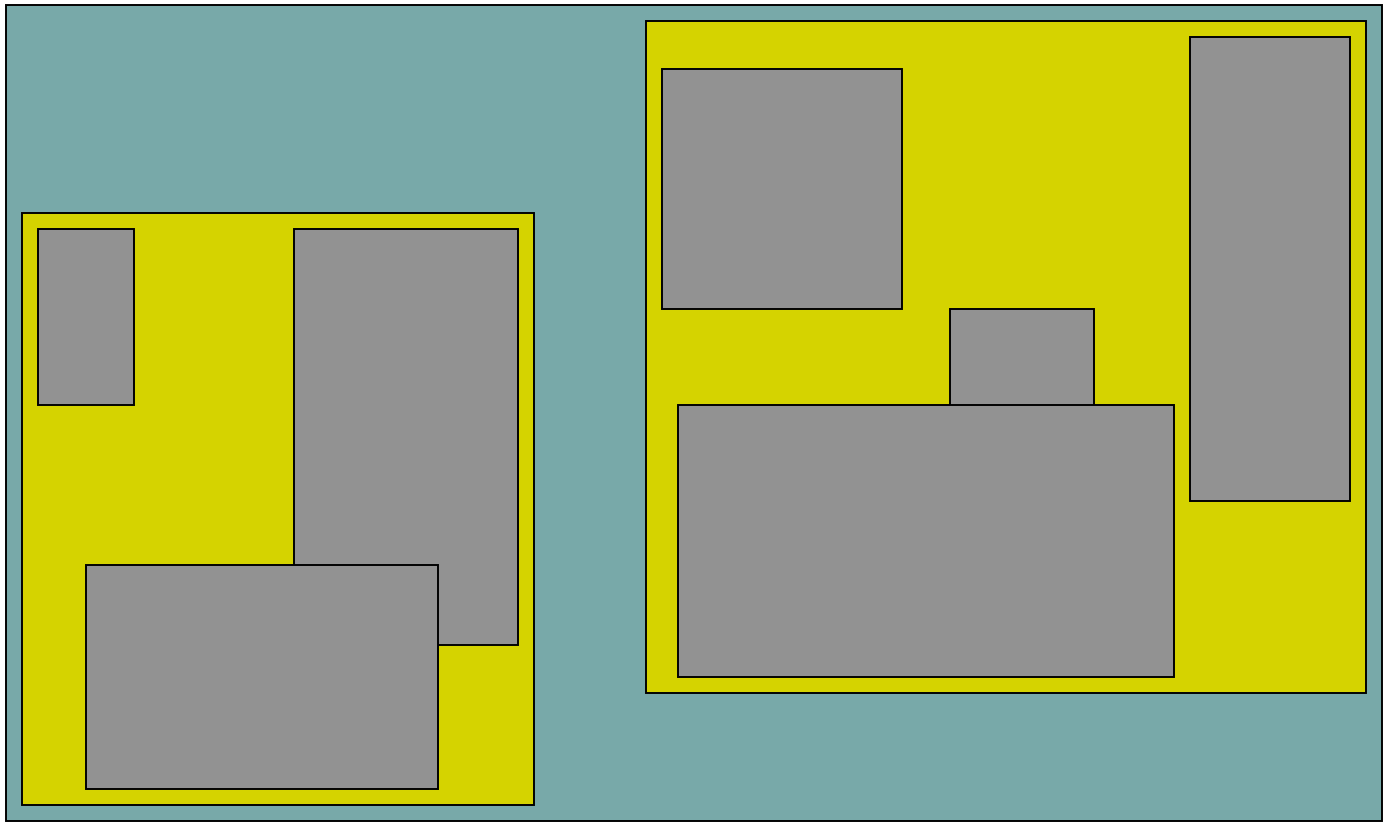
The idea is to vary the level of detail based on how much is being rendered:



Optimization #2: Spacial Indexing of Nodes

- Use spacial index for node positions
- “Give me the node lying directly under the mouse cursor.” Possibly O(log(N)) time for N nodes?
- “Give me the set of nodes intersecting the rectangular viewing area.” Between O(log(N)) and O(sqrt(N)) time possible?

What is an R-tree?



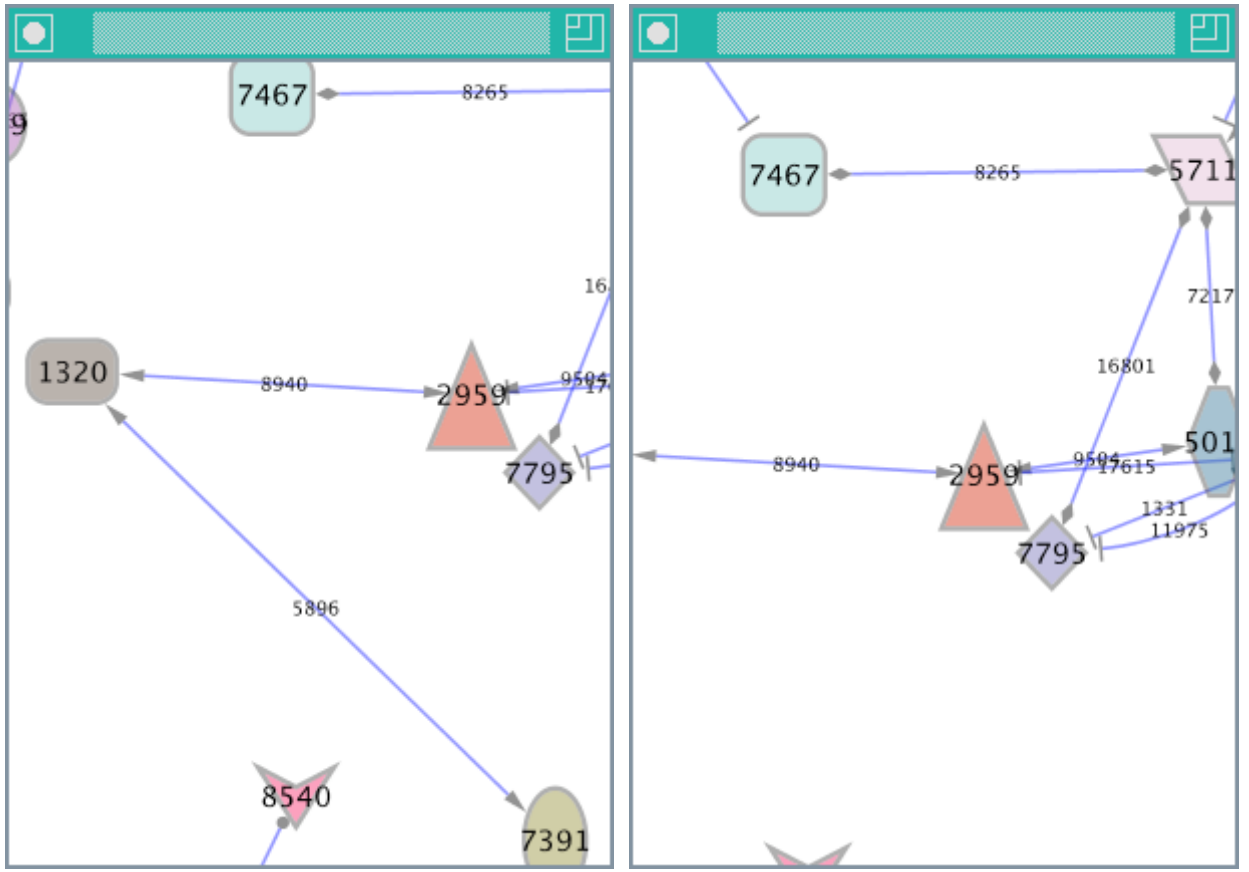
Optimization #2: R-tree (continued)

Which edges to render?

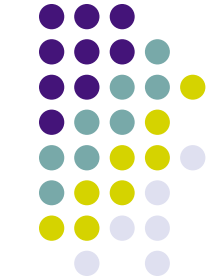
When viewing large networks, render an edge only if at least one of its endpoint nodes is visible.

Side effects:

- Edges which intersect viewing area may not necessarily be visible
- Edges may suddenly disappear during pan operations
- On the plus side: cobwebs of dense edges will not interfere with clarity of diagram

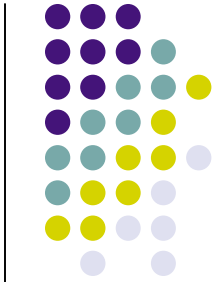


Optimization #2: R-tree (continued)

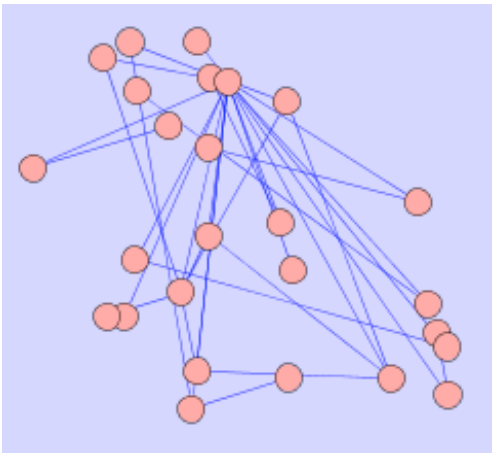


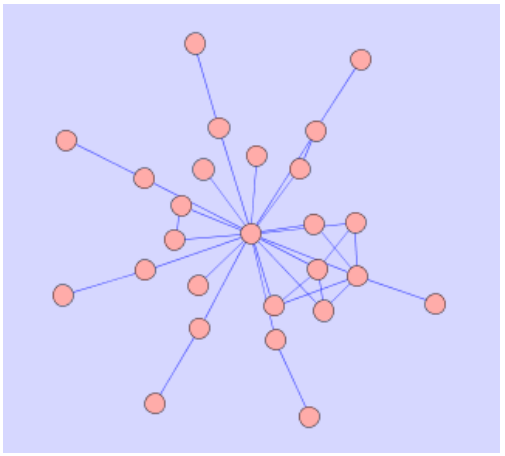
- Viewing a network of 2,000,000 nodes.
- Being zoomed in so that 10 or so nodes are currently visible.
- Panning operation
- With R-tree, 30 frames per second (bottleneck is rendering itself)
- Without R-tree, 1 frame per second (bottleneck is iterating over 2,000,000 nodes on every frame being rendered)

Optimization #2: R-tree (continued)



Layout algorithms:





Spacial index opens door for certain set of large network layout algorithms.

- Optimize topological queries (“give me all node neighbors”)
- Optimize spacial queries (“give me all nodes within two meters of this node”)
- Possible to implement layout algorithms suitable for large networks?
- Divide and conquer?
- O(N*log(N)) time?
- Plan to devote some amount of research in this area

Optimization #2: R-tree (continued)

Drawbacks of using spacial indexing for storing node extents:

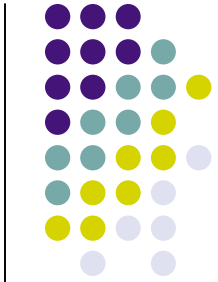
Insertions and deletions become more expensive. In particular, creating a network of N nodes takes approximately O(N*log(N)) time instead of O(N) time.

Imagine an operation that modifies the position of every node, such as a global rotation. This operation may become sluggish with an R-tree.

Optimization #3: Reduce Memory Footprint

- Nodes and edges can have numerous visual attributes
- Node shape, color, edge type, etc.
- Do not store all of these attributes per-node and per-edge
- Only store visual information if it differs from a default value
- With large networks, avoid setting any visual attributes and simply use default values

Summary



- Cytoscape's niche will be the ability to vizualize and compute on large networks.
- Cytoscape will strive to support, in one way or another, standards that are decided upon in conferences such SBGN.
- The new rendering engine will have a well-defined API that will enable other applications (other than Cytoscape) to use it effectively.

Acknowledgements

UCSD Department of Bioengineering
Trey Ideker
Keiichiro Ono
Mike Smoot
Nerius Landys
Chris Workman
Ryan Kelly

Agilent Technologies
Annette Adler
Allan Kuchinsky
Aditya Vailaya

Memorial Sloan Kettering Cancer Center
Chris Sander
Ben Gross
Ethan Cerami
Gary Bader

Institut Pasteur
Benno Schwikowski
Melissa Cline

Institute for Systems Biology
Leroy Hood
Rowan Christmas
Iliana Avila-Campillo
Rich Bonneau

Funding for Cytoscape is provided by a federal grant from the U.S. National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under award number GM070743-01 and the U.S. National Science Foundation (NSF) . Corporate funding is provided through a contract from Unilever PLC.