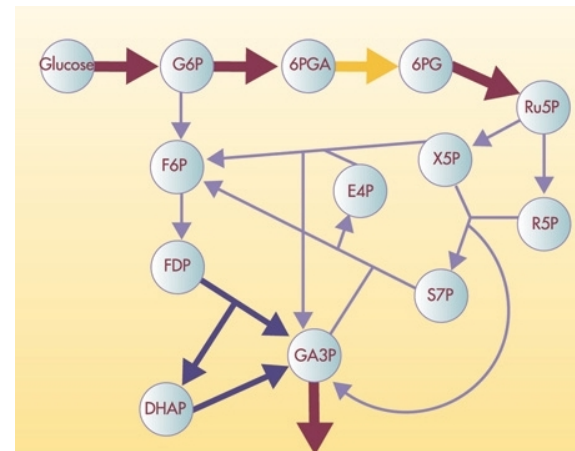


GeneGobi

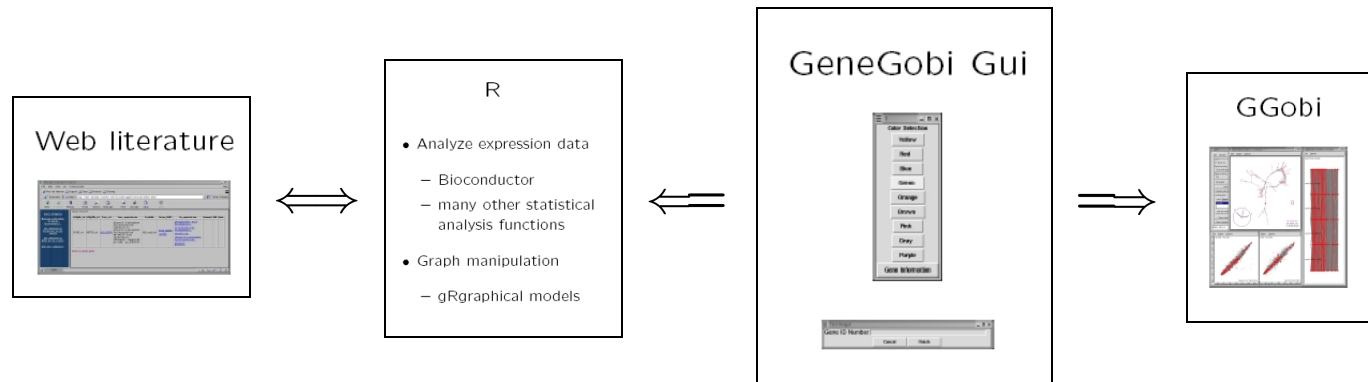
Visual Data Analysis Aid Tools
for the interconnections between genes and metabolic network



Eun-kyung Lee, Dianne Cook

February 18, 2003

Main Purpose :



Help biologists

- explore patterns in gene expression data
- explore the regulatory and metabolic pathway of Arabidopsis
- analyze the connections between microarray data and metabolic pathway visually and interactively
- provide “user-friendly” GUI

Outline

- GGVis : Interactive Graph Visualization with MDS
- Linking expression data and network
- GeneGobi Gui - V0.1

GGVis : GGobi Graph Visualization

- GGobi Plug-in
- use MDS(MultiDimensional Scaling) to construct variety of graph layout
- able to use **Rotation** or **2D Tour** to examine the graph layout
- able to use linked brush(**Brush**) with other plots(scatter plot, parallel coordinate plot) to explore data about edges, and data about nodes.
- Make/Remove node labels with **Identify** freely

Layout : MDS

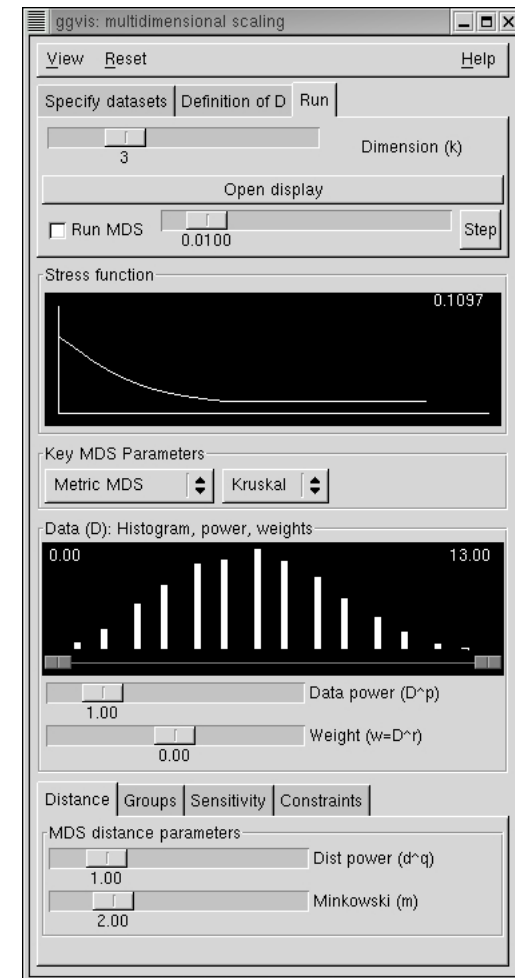
- To map nodes to points(k-dim) by minimizing a cost function
- MDS methods
 - Metric MDS : use the actual magnitudes of the distances
 - Nonmetric MDS : use rank orders only
 - Kruskal : use STRESS as a cost function
 - Classic : use STRAIN as a cost function

GGVis : Structure of Dataset

- network.nodes : biochemicals - proteins, small molecules, etc.
 - IDs or Names
 - *locations*
- network.edges : regulatory or metabolic relationships
 - origin(tail) : destination(head)
 - *distance* - not necessary
- expression : *gene ID*
- Problem
 - cannot use XML data file (we need to figure this out)
 - need to use RGGobi and add data set in RGGobi step by step

GGVis : How to use

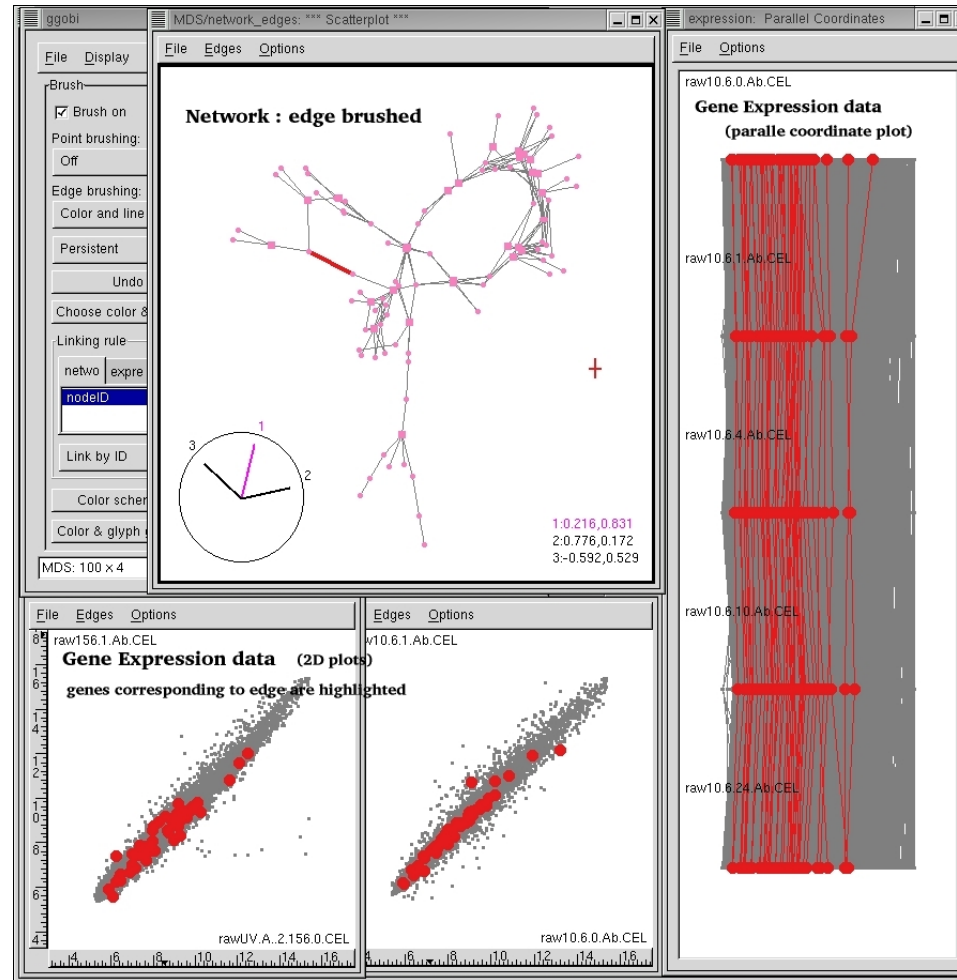
- Open **ggvis(MDS)** in **Tools** menu
- Calculate distance first
 - click **Definition of D** tab
 - check **Complete distance**
 - click **Compute D**
- Calculate MDS
 - click **Run** tab
 - choose **Dimension (k)**
 - click **Open display**
 - turn on edges on MDS plot
 - check **Run MDS**. After optimizing, turnoff **Run MDS**
- Choose **Rotation** in **ViewMode** menu (for $k \geq 3$)



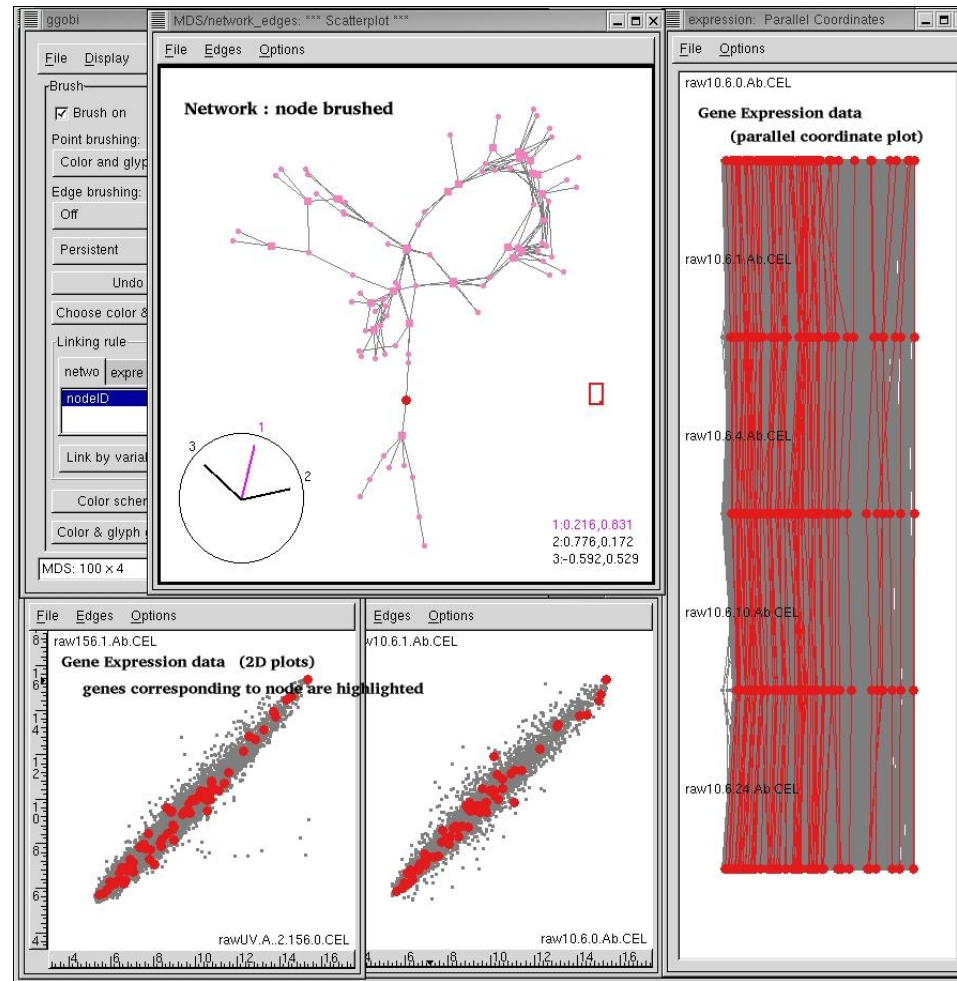
Linked brush : How to link network and gene expression data

- network edge - gene expression data : link by ID
 - Point brushing : Off
 - Edge brushing : Color and line
 - Linking rule : Link by ID
- network node - gene expression data : link by variable
 - Point brushing : Color and glyph
 - Edge brushing : Off
 - Linking rule : Link by variable - choose nodeID variable

network edge - gene expression data : Link by ID



network node - gene expression data : Link by variable



GeneGobi GUI : get selected gene information by color

The screenshot displays the GeneGobi GUI interface. At the top, a window titled '4' shows gene information for Affy ID 12115_at and Locus ID At4g22470. The Tair Annotation is 'extensin - like protein; hybrid proline-rich protein, Zea mays, PIR2:JQ1663'. The Predicted Subcellular Location is 'Secreted Proteins'. The Go Annotation is 'endomembrane system;'. Below this is a 'Test Widget' with a 'Gene ID Number' input field and 'Cancel' and 'Finish' buttons.

On the left, a 'Windows of GeneGobi GUI' panel contains a 'Color Selection' section with buttons for Yellow, Red, Blue, Green, Orange, Brown, Pink, Gray, and Purple. A 'Gene Information' button is also present.

In the center, a window titled '2' displays a list of genes under the heading 'Green'. The list includes the following Affy IDs and Locus IDs:

Affy ID	Locus ID
129	12115_at
178	12164_at
581	12575_s_at
789	12785_at
880	12899_i_at
1182	13219_at
1295	13347_at
1532	13585_at
1665	13804_at
1670	13809_at
1674	13813_at
1912	14053_at
1926	14067_at
2064	14318_at
2567	14855_at
2775	15065_at
2964	15265_at
3109	15411_at
3228	15532_r_at
3240	15544_at
3541	15865_at
3634	15962_at
4118	16473_at
4142	16499_at
4215	16579_at
4312	16686_at
4389	16763_at
4465	16843_at

On the right, a window titled 'plot ***' shows a network graph with nodes and edges. Below it, a window titled 'raw10.6.1.Ab.CEL' shows a scatter plot of gene expression data. The x-axis is labeled 'raw10.6.0.Ab.CEL' and the y-axis is labeled 'raw10.6.1.Ab.CEL'. The plot shows a strong positive correlation between the two datasets, with a cluster of green points.

GeneGobi : List of functions

- `get.selected.gobiG`, `get.colored.gobiG`
: get selections, operated from GeneGobi GUI
- `Select.kNN.gobiM`, `Select.kNN.interactive.gobiM`
: display the k nearest genes to selected gene.
- `set.brush.gobiG`, `release.brush.gobiG`
: control brush from R
- `set.identify.gobiM`, `release.identify.gobiM`
: control identify from R
- `setup.gobiforest.wtree`, `setup.gobiforest`, `setup.gobiparcoords`, `color.click.forest.tr`
: link between random forest and GGobi

- Need some routines for clustering, classification.

Future work

- Put Lixia's functions into GeneGobi library
- Develop more GUI for GeneGobi functions.
- Try statistical cluster analysis and compare to clusters from expert knowledge
- Find out what biologists are interested in and develop data analysis tools for the answer.