Graphical Representation of Cell/Tissue Type Relationships for the Web

Raymond Wan[1] rwan@kuicr.kyoto-u.ac.jp
Hiroshi Mamitsuka[1] mami@kuicr.kyoto-u.ac.jp
Larisa Kiseleva[2] kiseleva-larisa@aist.go.jp
Paul Horton[2] horton-p@aist.go.jp

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, 611-0011, Japan
[2] Computational Biology Research Center, AIST, Tokyo, 135-0064, Japan

Abstract

We describe a web server which allows users to depict the experiments in a microarray data set as a series of minimum spanning trees (MSTs). Our aim is to provide an alternative to hierarchical clustering. This poster describes the similarities and differences between them as well as our web server which implements our method.

We call our system HAMSTER for Helpful Abstraction using Minimum Spanning Trees for Expression Relations.

URL: `http://hamster.cbrc.jp/`.

1. Minimum Spanning Trees

Given a weighted graph $G(V, E)$, a minimum spanning tree (MST) $G_{MST}(V_{MST}, E_{MST})$ is a subgraph which:

- includes all vertices in V (that is, $V_{MST} = V$),
- has exactly $|V| - 1$ edges,
- contains subset of the edges in E such that it is a tree (acyclic), and
- is not always unique if there are edges with equal weights.

One algorithm for determining MSTs is Kruskal's algorithm:

1 Sort all E edges of $G(V, E)$.
2 Consider each of the $V_{MST} = V$ nodes as a tree of one node.
3 Add the shortest edge to G_{MST} if it connects two trees that were not already connected (Repeat $|V| - 1$ times.)

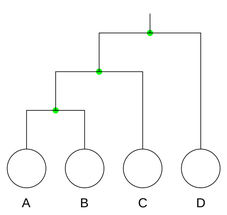
2. Example Comparing Hierarchical Clustering to MSTs

Expt	Gene 1	Gene 2
A	2	1
B	2	2
C	1	3
D	4	4

(a) Sample microarray

	A	B	C	D
A	■	1.000	2.236	3.606
B	■	■	1.414	2.828
C	■	■	■	3.162
D	■	■	■	■

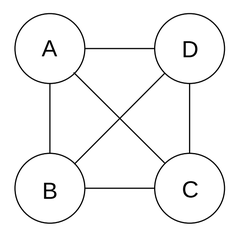
(b) Similarity matrix d



(c) Hierarchical clustering

1.000 A B
1.414 B C
2.236 A C
2.829 B D
3.162 C D
3.606 A D

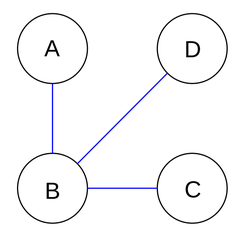
(d) d as a sorted list



(e) Complete graph

1.000 A B
1.414 B C
2.236 A C
2.829 B D
3.162 C D
3.606 A D

(f) Edges used



(g) Final MST

While hierarchical clustering adds $N - 1$ internal nodes, MSTs connect nodes directly and allows for the existence of hubs, spokes, etc. In Figure (g), it is more obvious that experiment B is most similar to the other 3 experiments. Contrast this with Figure (c).

3. Workflow

Input: Microarray data set of N experiments and M genes.
Output: A set of $N - 1$ MSTs.

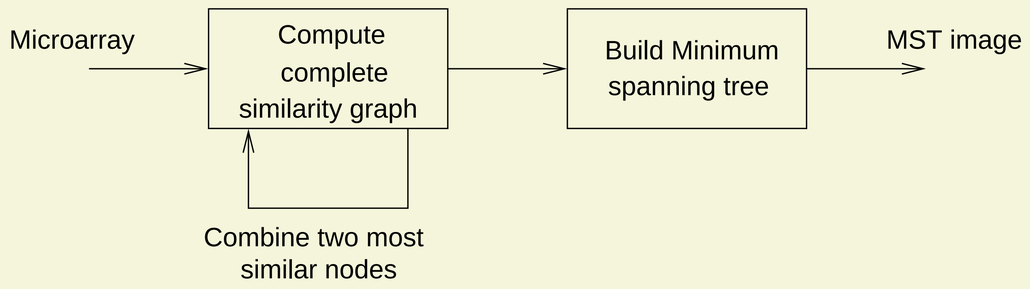


Figure 1: Workflow of our system.

4. Current Features

- Permits users to select genes from the data set either manually or, for three platforms (GPL91, GPL92, and GPL96), via **GO terms**.
- Manually assigned 559 GSMs to one of 87 classes such as **epithelial** and **cerebral**.
- Images are clickable PNG or SVG images which link out to these manually assigned labels or to the original experiment descriptions at NCBI GEO.
- Instead of re-using the similarity matrix d, we **recalculate** similarities between merged nodes.
- Similarity calculated from Pearson and Spearman correlation coefficients as well as distance measures such as Euclidean and Manhattan distance.

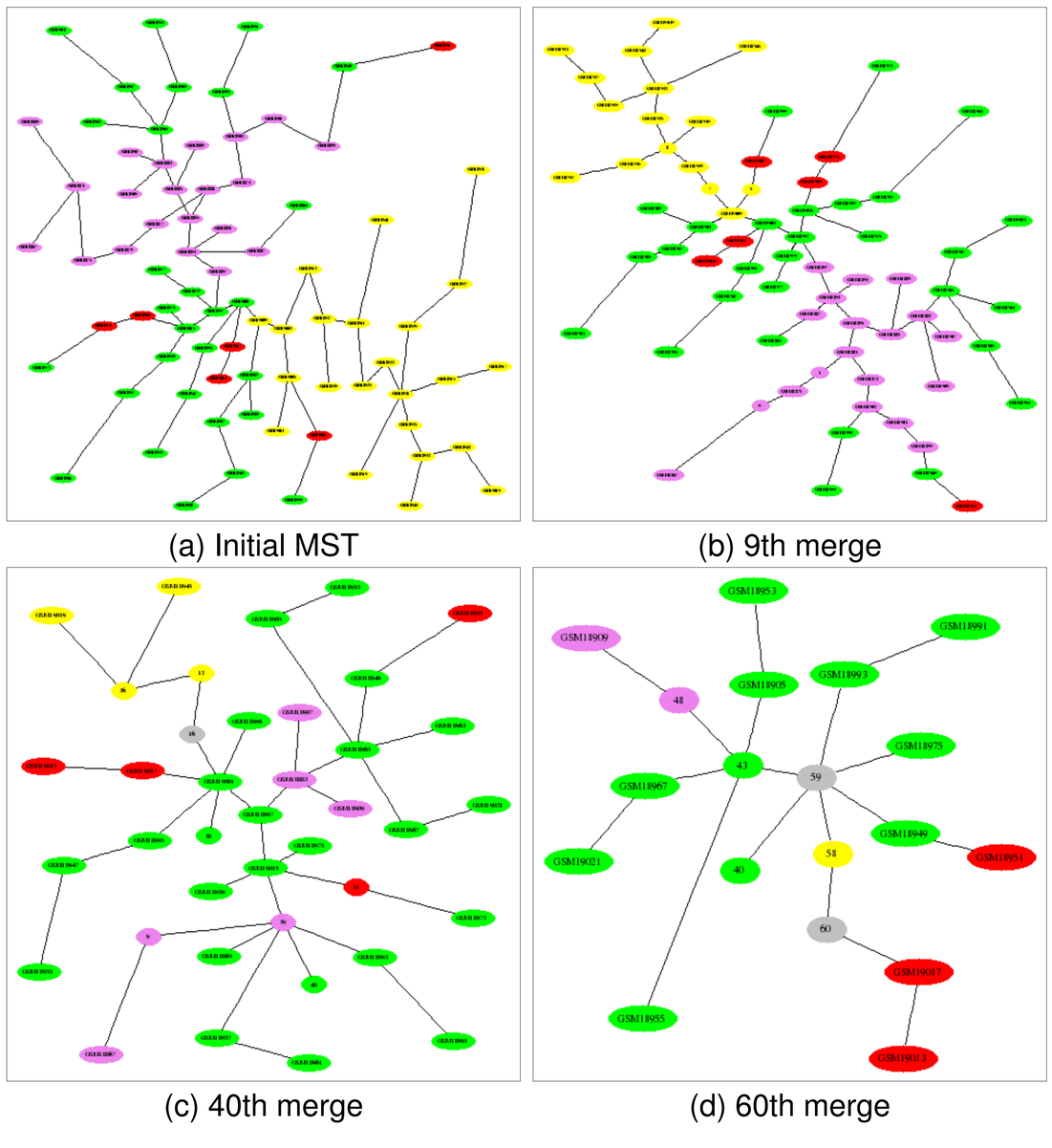
5. Future Work

The system is still being developed. Some future directions for HAMSTER include:

- Scoring mechanism for the MSTs.
- Allow more user options for similarity measures, file formats, etc.
- Better integration with the manually annotated data. Perhaps allow questions such as "how do the available **epithelial** experiments compare with the **cerebral** ones"?

Example using half of GDS596 (GPL96). 79 experiments and 22,283 genes and a similarity measure of $1 - r$, where r is the Pearson correlation coefficient.

Legend: epithelial (29) muscle (6) cerebral (23) immune (21) Total (79)



(a) Initial MST (b) 9th merge

(c) 40th merge (d) 60th merge

Acknowledgements: Hajime Harada (CBRC) maintains the server and helped prepare the annotated dataset. Development of HAMSTER is due to Mason (Embedding Perl code in HTML), GraphViz, and the Perl Graph module available from CPAN.

The 6th Asia Pacific Bioinformatics Conference (**APBC 2008**), January, 14-17, 2008, Kyoto, Japan.