

Active Networks

T. M. Murali

August 20, 2008

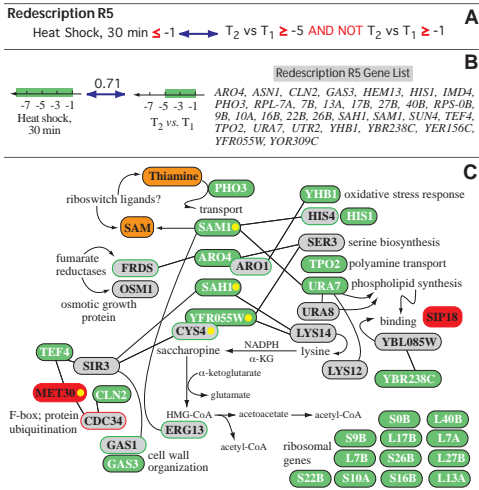
Motivation: Manual Systems Biology

- ▶ Biologists want to study a favourite stress, e.g., oxidative stress or desiccation tolerance.
- ▶ Measure gene expression, apply clustering algorithms, and find genes whose expression level change in response to the stress.

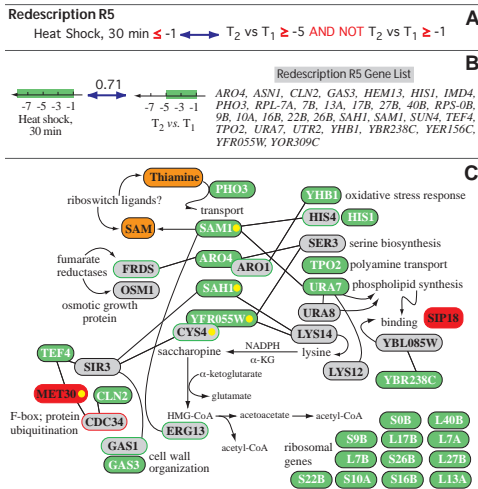
Motivation: Manual Systems Biology

- ▶ Biologists want to study a favourite stress, e.g., oxidative stress or desiccation tolerance.
- ▶ Measure gene expression, apply clustering algorithms, and find genes whose expression level change in response to the stress.
- ▶ Trace genes by hand through databases of protein-protein interactions, gene regulatory networks, metabolic pathways, PubMed searches to build networks activated in response to the stress.

Motivation: Manual Systems Biology



Motivation: Manual Systems Biology



Can we automate this process?

S. cerevisiae Wiring Diagram

- ▶ Physical network
 - ▶ 15,429 **protein-protein interactions** from the Database of Interacting Proteins (DIP).
 - ▶ 5869 **protein-DNA interactions** (Lee et al., Science, 2002).
 - ▶ 6,306 **metabolic interactions** (proteins operate on at least common metabolite) based on KEGG.
- ▶ Genetic network
 - ▶ 4,125 **synthetically lethal/sick interactions** (Tong et al., Science, 2004).
 - ▶ 687 synthetically lethal interactions (MIPS).
- ▶ Overall network has 32,416 (27,604 physical and 4,812 genetic) interactions between 5601 proteins (Kelley and Ideker, Nature Biotech., 2005).

Challenges in Utilising the Wiring Diagram

- ▶ Networks are large; they contain tens of thousands of interactions.
- ▶ High-throughput experiments contain many errors.
- ▶ Networks are incomplete; experiments are expensive and have biases.

Challenges in Utilising the Wiring Diagram

- ▶ Networks are large; they contain tens of thousands of interactions.
- ▶ High-throughput experiments contain many errors.
- ▶ Networks are incomplete; experiments are expensive and have biases.
- ▶ A biologist wants to explore and analyse system of interest.
- ▶ How do we zoom into the appropriate parts of the wiring diagram?

Active Networks

Murali and Rivera, *Journal of Computational Biology*, 2008

Active Network: network of interactions activated in response to a stress or in a particular condition.

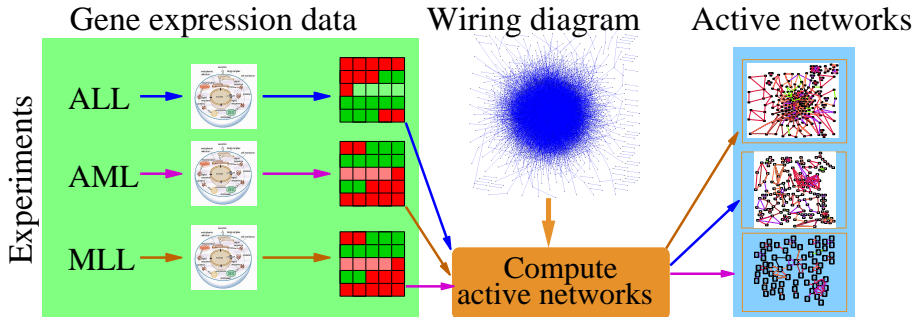
- ▶ Overlay molecular profile for a particular stress on wiring diagram to obtain active network for that stress.

Active Networks

Murali and Rivera, *Journal of Computational Biology*, 2008

Active Network: network of interactions activated in response to a stress or in a particular condition.

- ▶ Overlay molecular profile for a particular stress on wiring diagram to obtain active network for that stress.
- ▶ Why does the SAMBA approach not solve this problem?



Related Research

- ▶ Discovering regulatory and signalling circuits in molecular interaction networks, Ideker et al. ISMB 2002.
- ▶ Discovering molecular pathways from protein interaction and gene expression data, Segal, Wang, and Koller, ISMB 2003.
- ▶ Computational discovery of gene modules and regulatory networks, Bar-Joseph et al., Nature Biotechnology, November 2003.
- ▶ Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data, Tanay et al., PNAS, March 2004.
- ▶ Evidence for dynamically organized modularity in the yeast protein-protein interaction network, Han et al., Nature, July 2004.
- ▶ Genomic analysis of regulatory network dynamics reveals large topological changes, Luscombe et al., Nature, October 2004.
- ▶ Identification of functional modules using network topology and high-throughput data. Ulitsky and Shamir, BMC Systems Biology, 2007.

Related Research

- ▶ **Discovering regulatory and signalling circuits in molecular interaction networks, Ideker et al. ISMB 2002.**
- ▶ Discovering molecular pathways from protein interaction and gene expression data, Segal, Wang, and Koller, ISMB 2003.
- ▶ Computational discovery of gene modules and regulatory networks, Bar-Joseph et al., Nature Biotechnology, November 2003.
- ▶ Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data, Tanay et al., PNAS, March 2004.
- ▶ Evidence for dynamically organized modularity in the yeast protein-protein interaction network, Han et al., Nature, July 2004.
- ▶ Genomic analysis of regulatory network dynamics reveals large topological changes, Luscombe et al., Nature, October 2004.
- ▶ **Identification of functional modules using network topology and high-throughput data. Ulitsky and Shamir, BMC Systems Biology, 2007.**

Overlaying Gene Expression Data

- ▶ Weight of an interaction is the Pearson correlation between the expression profiles of the interacting genes.
- ▶ Weight \equiv “activity” level of the interaction.

Overlaying Gene Expression Data

- ▶ Weight of an interaction is the Pearson correlation between the expression profiles of the interacting genes.
- ▶ Weight \equiv “activity” level of the interaction.
- ▶ Discard interactions based on a threshold.
 - ▶ Unsatisfactory since we test each interaction individually.

Overlaying Gene Expression Data

- ▶ Weight of an interaction is the Pearson correlation between the expression profiles of the interacting genes.
- ▶ Weight \equiv “activity” level of the interaction.
- ▶ Discard interactions based on a threshold.
 - ▶ Unsatisfactory since we test each interaction individually.

We find the most highly active subnetwork.

Defining Highly-Active Subnetworks

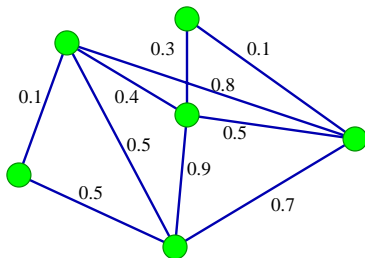
- ▶ How do we measure the activity/weight of a subnetwork?

Defining Highly-Active Subnetworks

- ▶ How do we measure the activity/weight of a subnetwork?
- ▶ Sum or average of edge weights?

Defining Highly-Active Subnetworks

- ▶ How do we measure the activity/weight of a subnetwork?
- ▶ Sum or average of edge weights?
- ▶ The *density* of a network with n nodes is the total weight of the edges divided by n .
- ▶ Problem: Compute the subnetwork with highest density.



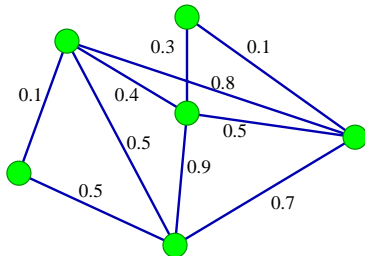
Computing Most Dense Subnetwork

- ▶ $O(n^3)$ time network flow-based approach gives optimal result (Gallo, Grigoriadis, Tarjan, *SIAM J. Comp*, 1989).
- ▶ Can also be solved by linear programming.

Computing Most Dense Subnetwork

► Greedy algorithm:

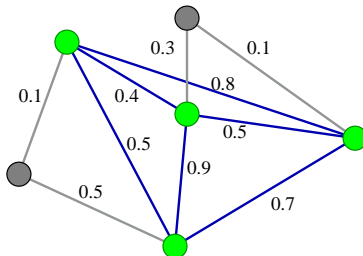
- Weight of a node \equiv total weight of incident edges.
- Repeatedly delete nodes with the smallest weight.
- Keep track of density of remaining network.
- Return the most dense subnetwork.



Computing Most Dense Subnetwork

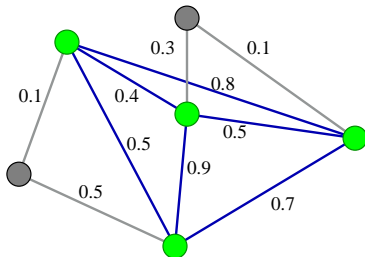
► Greedy algorithm:

- Weight of a node \equiv total weight of incident edges.
- Repeatedly delete nodes with the smallest weight.
- Keep track of density of remaining network.
- Return the most dense subnetwork.



Computing Most Dense Subnetwork

- ▶ Greedy algorithm:
 - ▶ Weight of a node \equiv total weight of incident edges.
 - ▶ Repeatedly delete nodes with the smallest weight.
 - ▶ Keep track of density of remaining network.
 - ▶ Return the most dense subnetwork.
- ▶ Computed subnetwork is at least half as dense as the most dense subnetwork (Charikar, Proc. APPROX, 2000).



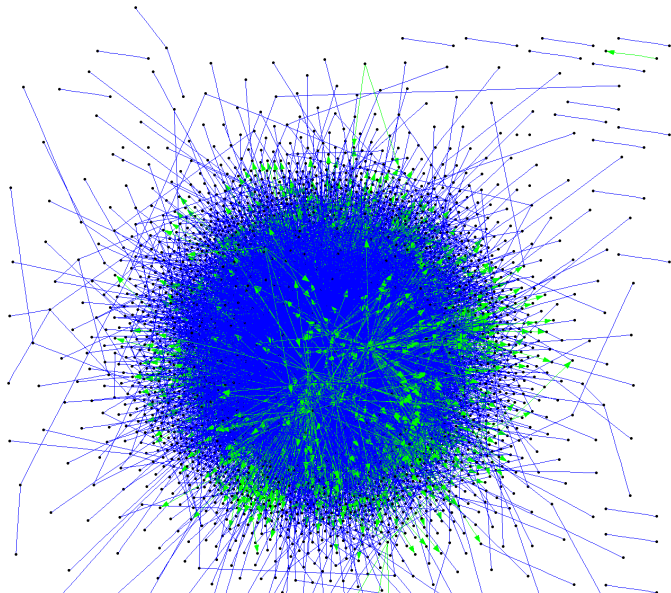
Computing Multiple Dense Subnetworks

- ▶ Repeat
 1. Apply greedy algorithm to compute most dense subnetwork.
 2. Remove edges of computed subnetwork from the network.
- ▶ Until remaining network has density less than the original network.
- ▶ Output is a sequence of decreasingly dense subnetworks that can share nodes but not edges.

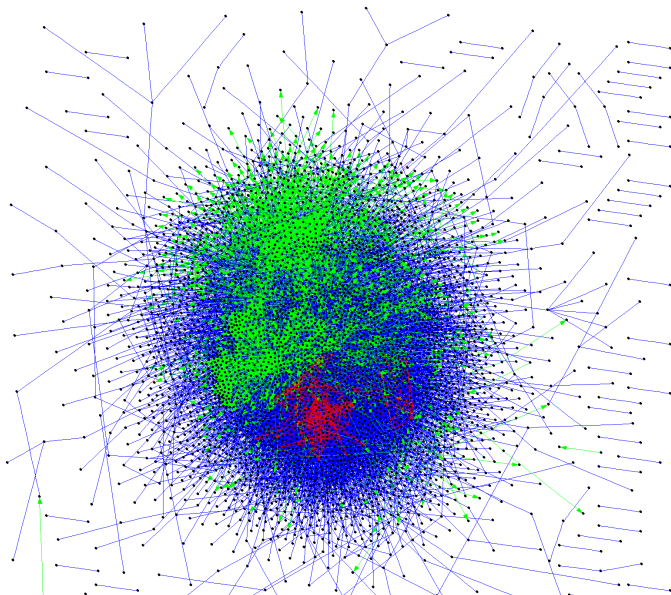
Advantages of Dense Subnetworks

- ▶ Uses no parameters.
- ▶ Avoid inclusion of interactions that appear active due to noise.
- ▶ Relatively weakly correlated interactions can reinforce each other.

Example of an Active Network



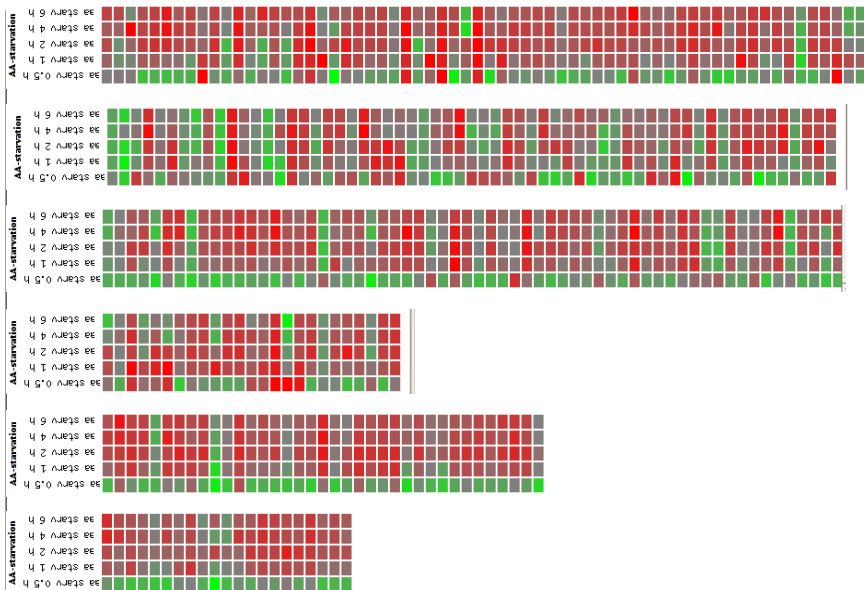
Example of an Active Network



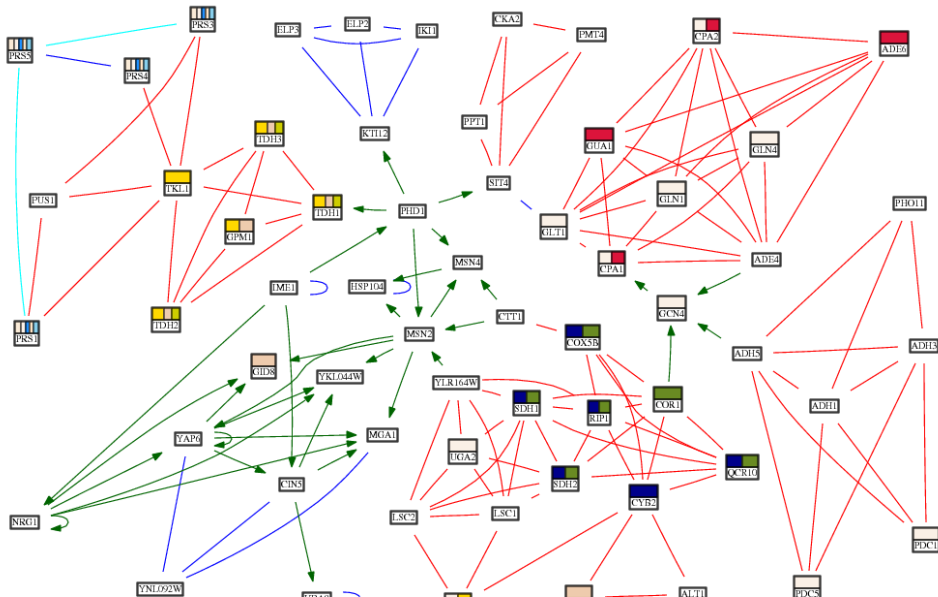
Further Analysis of an Active Network

- ▶ Visualise the network (Graphviz package) and the gene expression profiles.
- ▶ Measure functional enrichment.
 - ▶ Use hypergeometric distribution to calculate the significance of functions enriched in an active network.
 - ▶ Use Bonferroni correction to adjust for testing multiple hypotheses.

AA Starvation: Conventional Analysis



AA Starvation: Active Network



Active Modules

Discovering regulatory and signalling circuits in molecular interaction networks, Ideker et al. ISMB 2002. (Available as plugin for Cytoscape.)

- ▶ Suppose for each gene, we have a p -value that indicates the degree of differential expression of that gene?
- ▶ How do we utilise p -value to compute active networks?

Active Modules

Discovering regulatory and signalling circuits in molecular interaction networks, Ideker et al. ISMB 2002. (Available as plugin for Cytoscape.)

- ▶ Suppose for each gene, we have a p -value that indicates the degree of differential expression of that gene?
- ▶ How do we utilise p -value to compute active networks?
- ▶ Convert each p -value into a z -score: $z(g) = N^{-1}(p(g))$.
- ▶ Define score for a network that combines these z -scores.
- ▶ Compute the sub-network of highest score using simulated annealing.

Active Modules

Discovering regulatory and signalling circuits in molecular interaction networks, Ideker et al. ISMB 2002. (Available as plugin for Cytoscape.)

- ▶ Suppose for each gene, we have a p -value that indicates the degree of differential expression of that gene?
- ▶ How do we utilise p -value to compute active networks?
- ▶ Convert each p -value into a z -score: $z(g) = N^{-1}(p(g))$.
- ▶ Define score for a network that combines these z -scores.
- ▶ Compute the sub-network of highest score using simulated annealing.
- ▶ Suppose $G = (V, E)$ is a network of interactions.

$$z(G) = \frac{\sum_{v \in V} z(v)}{\sqrt{|V|}}$$

Concerns with Previous Approaches

- ▶ Wiring diagrams are very incomplete.
- ▶ How does incompleteness affect the results provided by the previous algorithms?

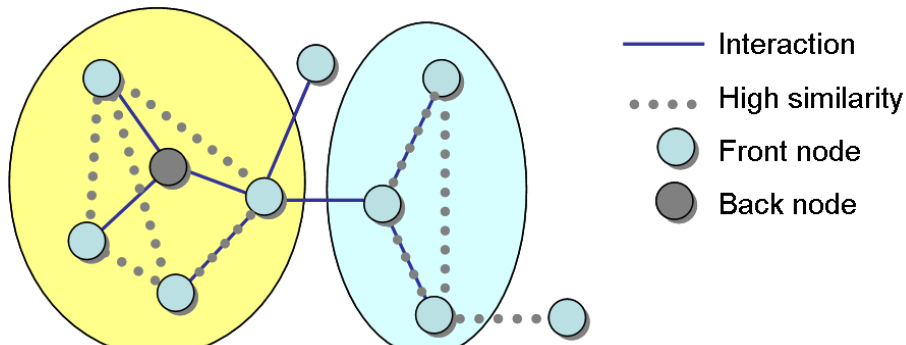
Concerns with Previous Approaches

- ▶ Wiring diagrams are very incomplete.
- ▶ How does incompleteness affect the results provided by the previous algorithms?
- ▶ Can we compute active networks even though wiring diagrams are incomplete?

The MATISSE Algorithm

Identification of functional modules using network topology and high-throughput data. Ulitsky and Shamir, *BMC Systems Biology*, 2007.

- ▶ Use wiring diagram as a constraint network and gene expression data to measure similarity between genes.
 1. Convert gene expression data into similarities between pairs of genes.
 2. Find connected sub-networks (modules) of the wiring diagram such that genes in the module have high expression similarity.



MATISSE on Human Cell Cycle Arrest Data

