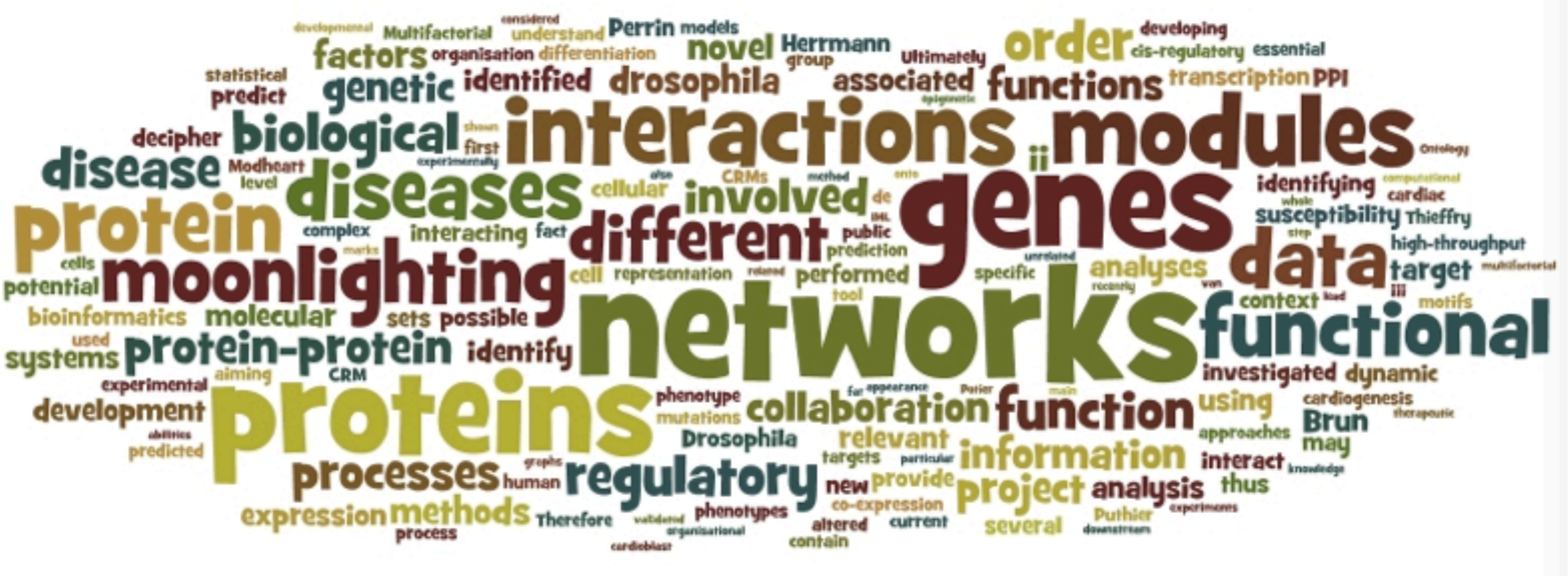
Network Analysis and Application

Yao Fu
10-15-2012



✧ Introduction to Biological Networks

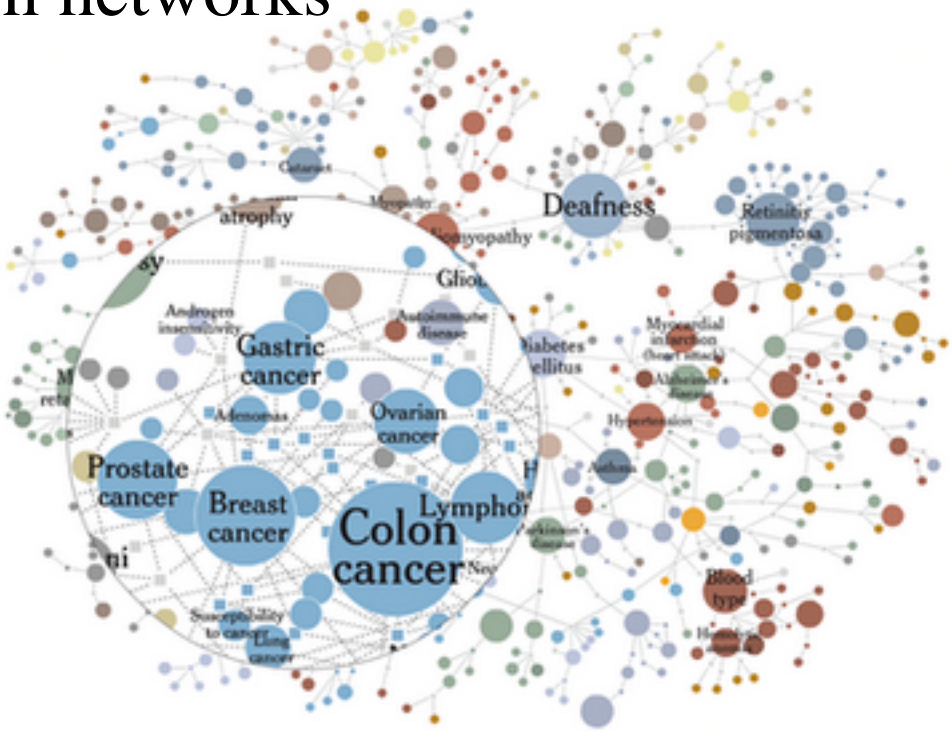
✧ Dynamic properties

✧ Network application

INTRODUCTION TO BIOLOGICAL NETWORKS

- Characterization of **biological processes**

- Network examples
  - Protein-protein interaction networks
  - Metabolic network
  - Transcriptional networks
  - RNA networks
  - Co-expression network
  - Disease gene network
  - Signaling network…



Human Disease Network

Protein Protein interaction Network

- **Nodes** : Protein ; **Edges**: interactions.

- High throughput interaction screening methods:
  - Yeast two hybrid experiments (Y2H)
  - Protein complex purification (PCP)

- Database:

BioGRID 3.1

Database of Interacting Proteins

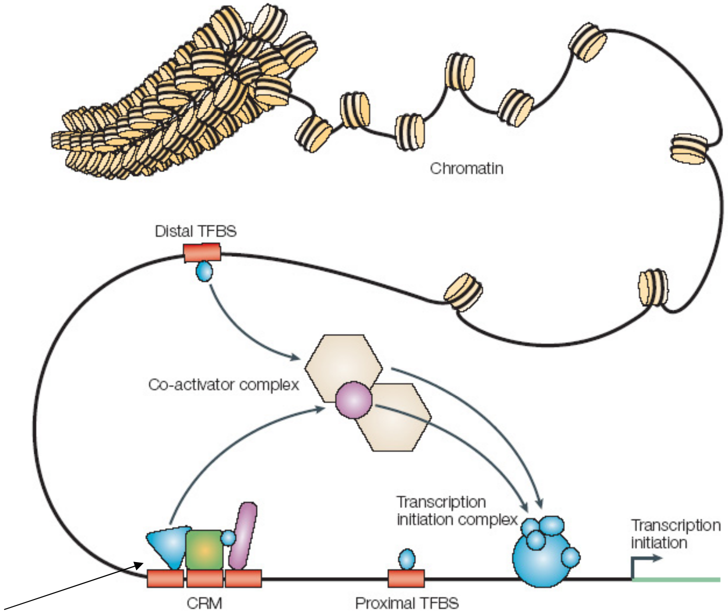
IntAct

MINT

Transcriptional Network

- **Nodes**: Protein (transcription factors or target genes)
- **Edges**: (directed) from a transcription factor to a gene.

Transcription

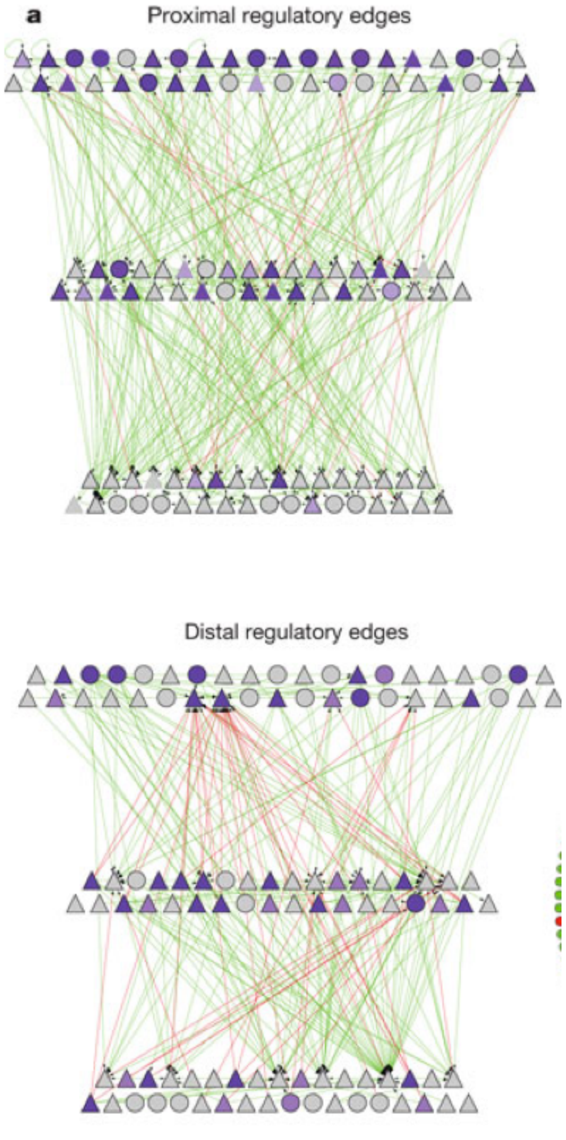


Cis regulatory module

- Experimental
  - Encode Chip-seq data

- Computational
  - Identification of Transcription Factor Binding Sites

* TRANSFAC® is the database on eukaryotic transcription factors, their genomic binding sites and DNA-binding profiles

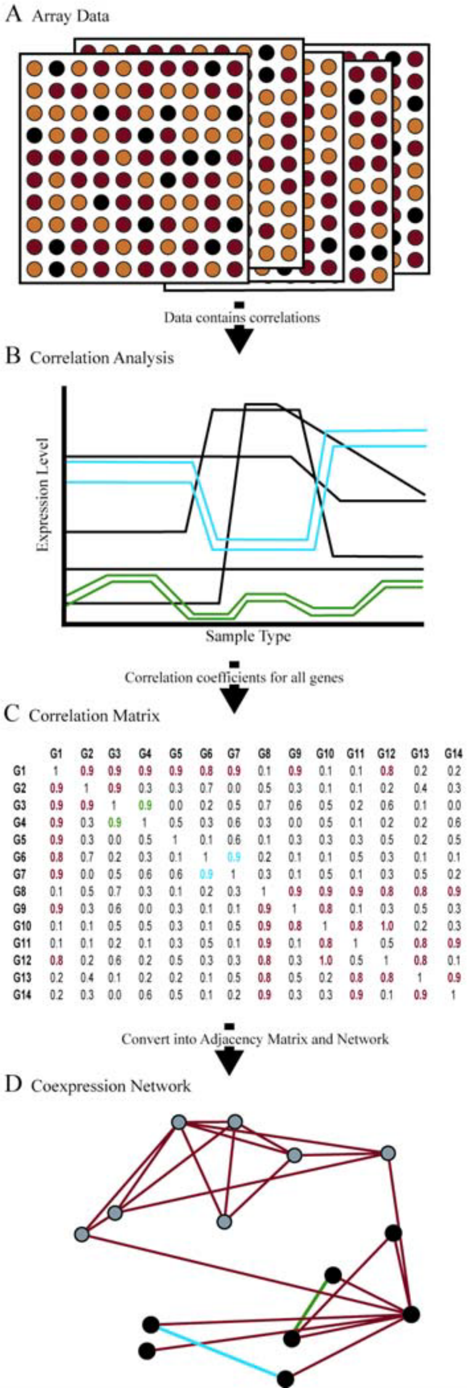


M. Gerstein, et al.,Nature (2012)

Co-expression Network

- Describes the relation between the expression pattern of different genes
- **Nodes**: Genes
- **Edges**: Two vertices are connected if the corresponding genes have similar expression patterns.

Figure 1



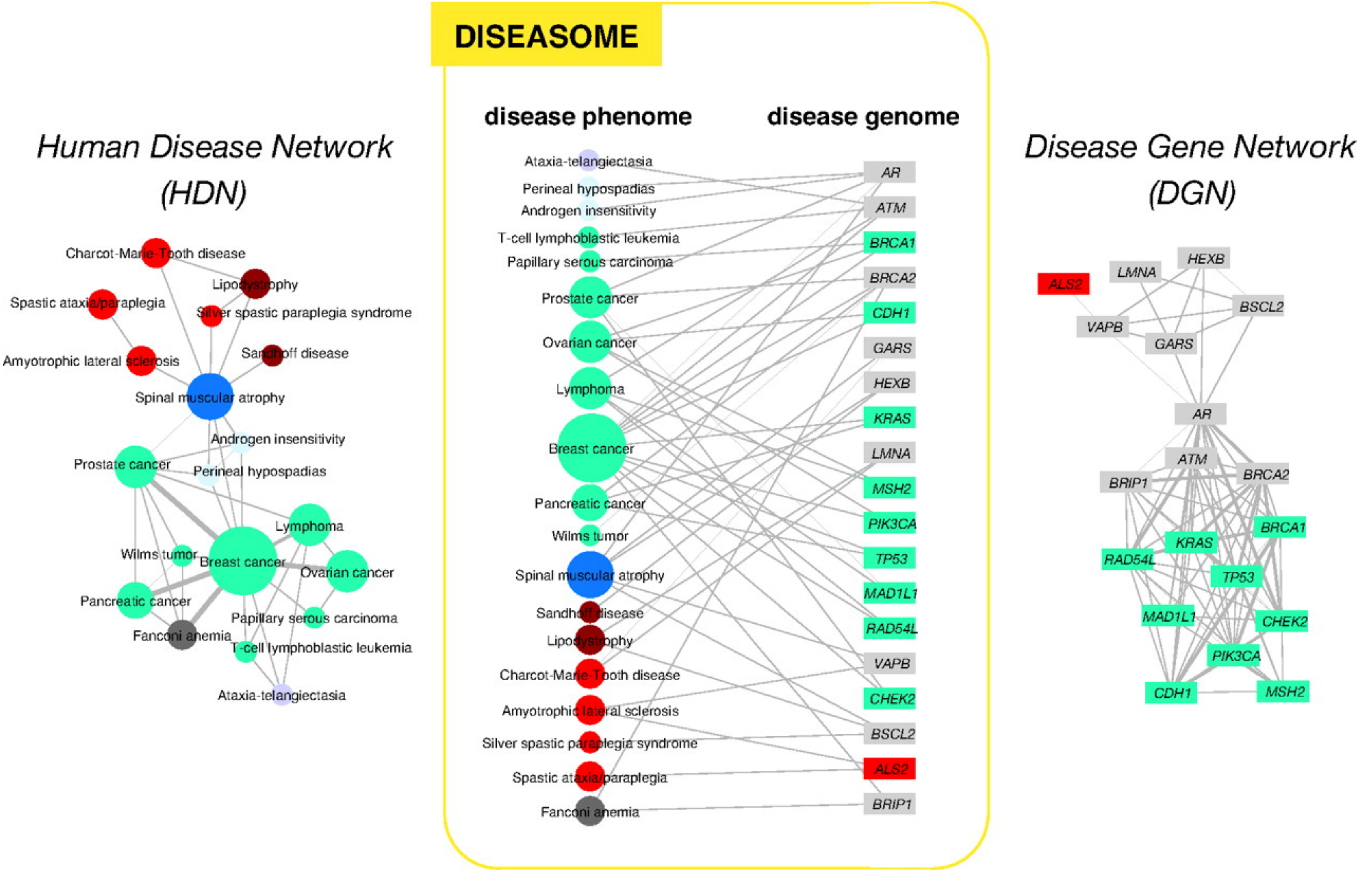
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14
G1	1	0.9	0.9	0.9	0.9	0.8	0.9	0.1	0.9	0.1	0.1	0.8	0.2	0.2
G2	0.9	1	0.9	0.3	0.3	0.7	0.0	0.5	0.3	0.1	0.1	0.2	0.4	0.3
G3	0.9	0.9	1	0.9	0.0	0.2	0.5	0.7	0.6	0.5	0.2	0.6	0.1	0.0
G4	0.9	0.3	0.9	1	0.5	0.3	0.6	0.3	0.0	0.5	0.1	0.2	0.2	0.6
G5	0.9	0.3	0.0	0.5	1	0.1	0.6	0.1	0.3	0.3	0.3	0.5	0.2	0.5
G6	0.8	0.7	0.2	0.3	0.1	1	0.9	0.2	0.1	0.1	0.5	0.3	0.1	0.1
G7	0.9	0.0	0.5	0.6	0.6	0.9	1	0.3	0.1	0.5	0.1	0.3	0.5	0.2
G8	0.1	0.5	0.7	0.3	0.1	0.2	0.3	1	0.9	0.9	0.9	0.8	0.8	0.9
G9	0.9	0.3	0.6	0.0	0.3	0.1	0.1	0.9	1	0.8	0.1	0.3	0.5	0.3
G10	0.1	0.1	0.5	0.5	0.3	0.1	0.5	0.9	0.8	1	0.8	1.0	0.2	0.3
G11	0.1	0.1	0.2	0.1	0.3	0.5	0.1	0.9	0.1	0.8	1	0.5	0.8	0.9
G12	0.8	0.2	0.6	0.2	0.5	0.3	0.3	0.8	0.3	1.0	0.5	1	0.8	0.1
G13	0.2	0.4	0.1	0.2	0.2	0.1	0.5	0.8	0.5	0.2	0.8	0.8	1	0.9
G14	0.2	0.3	0.0	0.6	0.5	0.1	0.2	0.9	0.3	0.3	0.9	0.1	0.9	1

A. Microarray gene expression data.
B. Measure concordance of gene expression with a Pearson correlation.
C. The Pearson correlation matrix is transformed to an adjacency matrix. Binary values in the adjacency matrix correspond to an un-weighted network.
D. The adjacency matrix can be visualized by a graph.

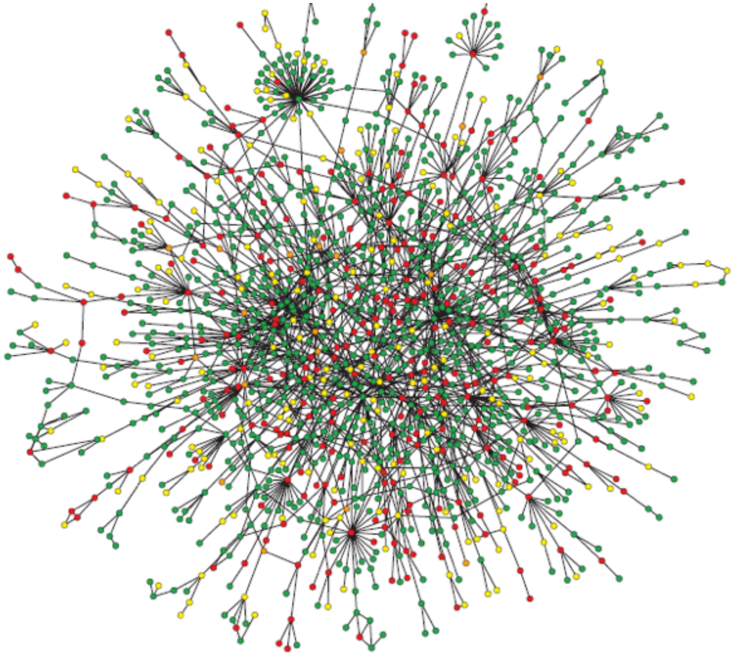
From Steve Horvath

Disease Gene Networks

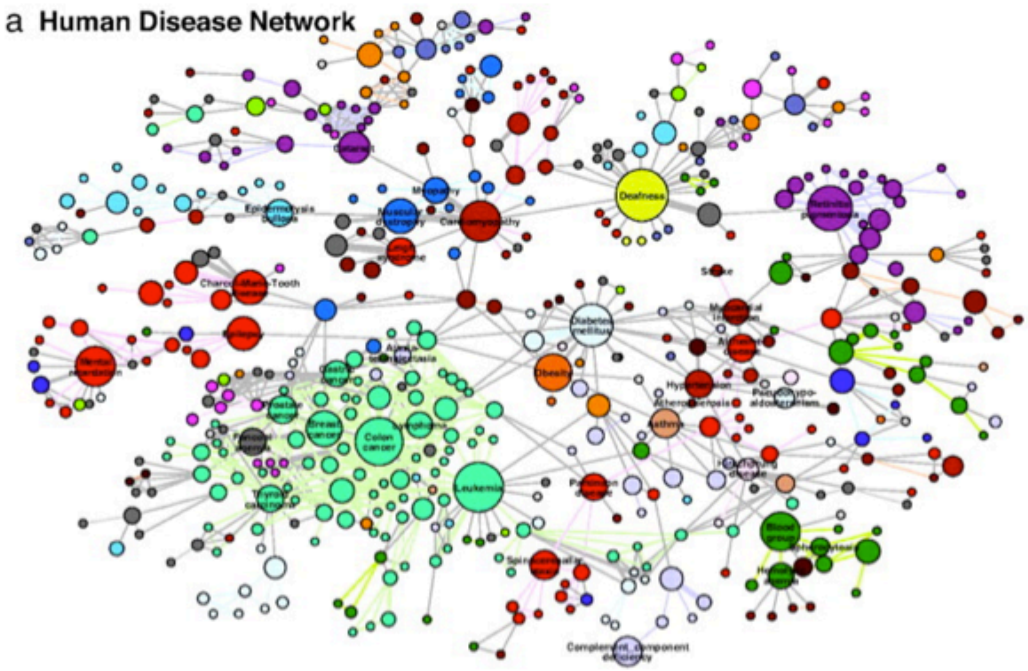
- A bipartite graph with two sets of vertices.
- **Nodes**: One set represents disease; the other represents genes.
- **Edges**: A gene and a disease are connected if the gene is involved in the disease.



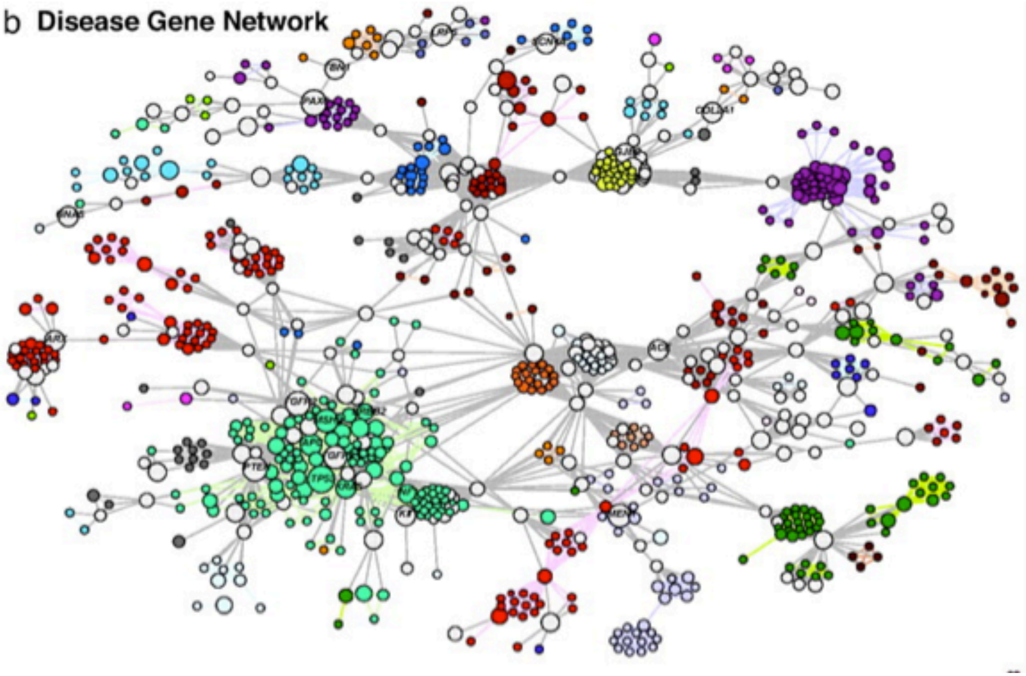
K. Goh, et al., PNAS (2007)



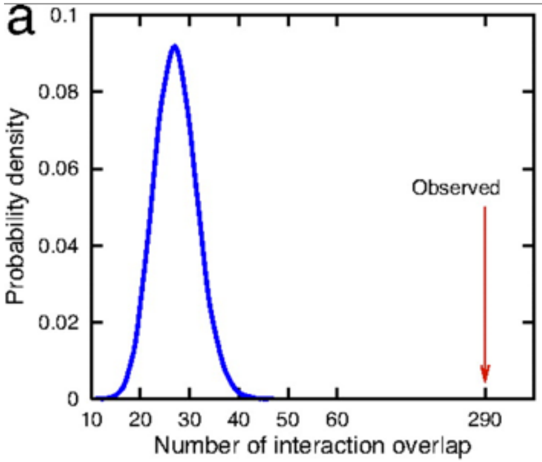
What we make out of the hairball



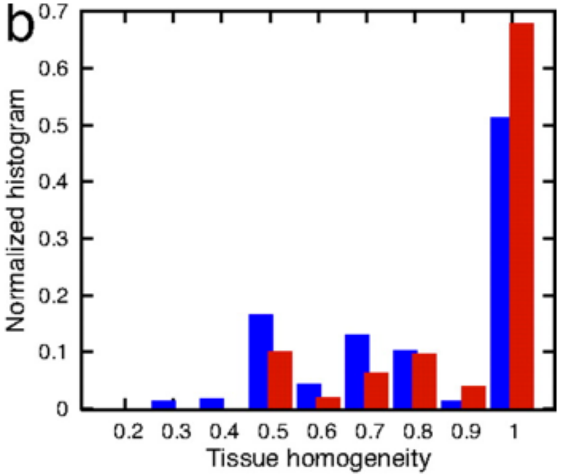
- Genetic origins of most diseases, are shared with other diseases.
- The resulting network is naturally and visibly clustered according to major disorder classes.



- Physical interaction



- Express in same tissue



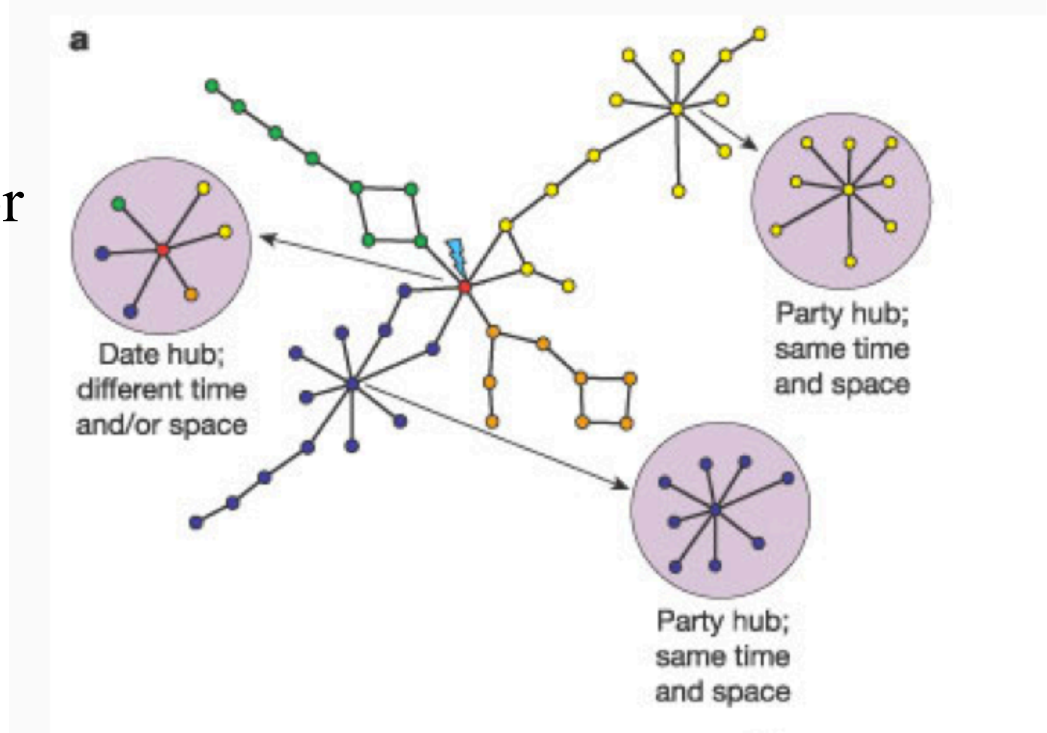
Goh K et al. PNAS 2007;104:8685-8690

DYNAMICS OF NETWORK

- Networks are not static
  - Date and party hubs
- Networks change during evolution
  - Network rewiring

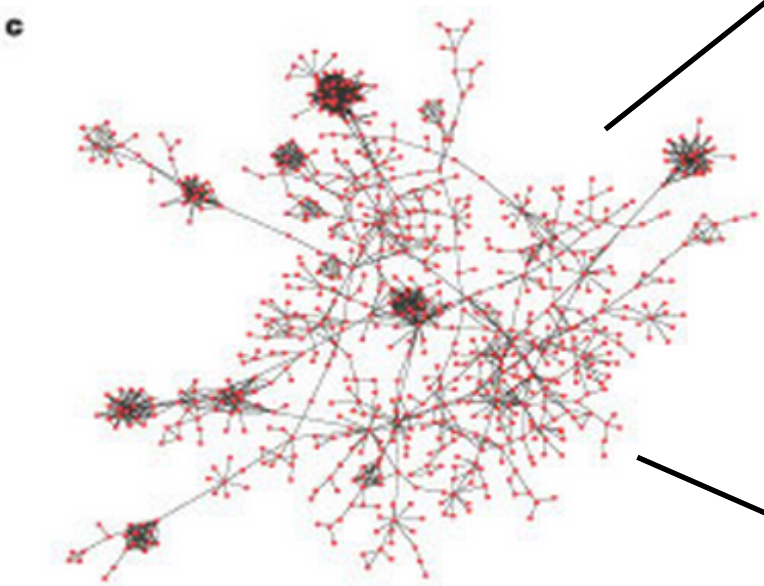
“Party hubs”: interact with their partners simultaneously.

“Date hubs”: bind their partners at different times or locations

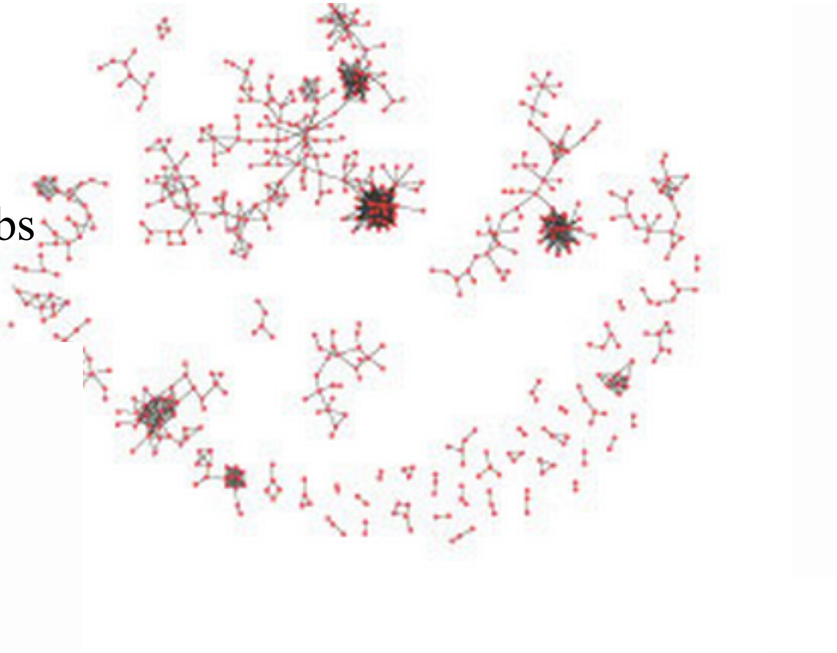


Yeast PPI (color - Mutual similarity in mRNA expression)

Han et al., Nature (2004)



Removal of date hubs



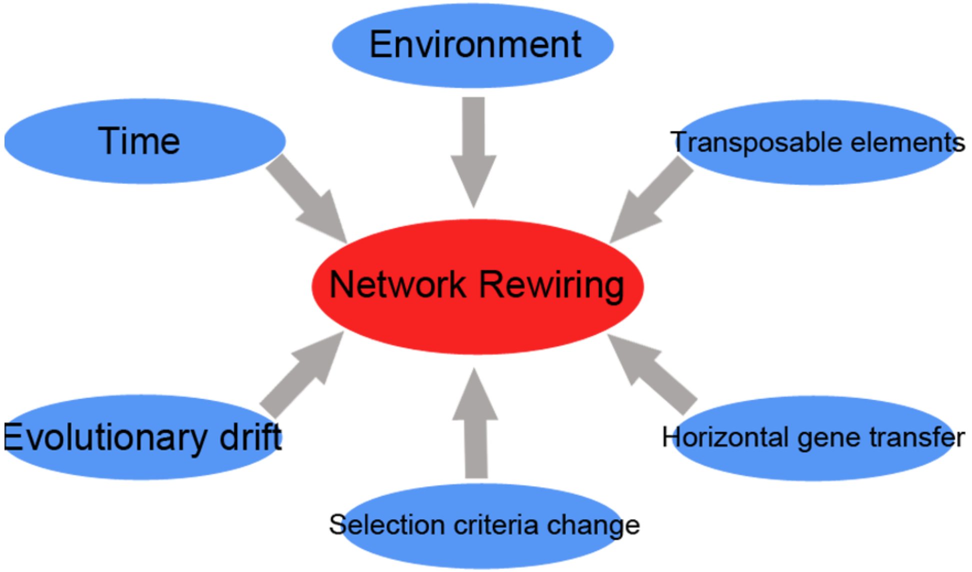
Removal of party hubs

Network rewiring

- How do biological interaction networks change overtime?
- Two ways to look at this:

  - What forces lead to gain and loss of nodes and edges?

  - Evolutionary selective pressure favors the network that represents the most fit phenotype.

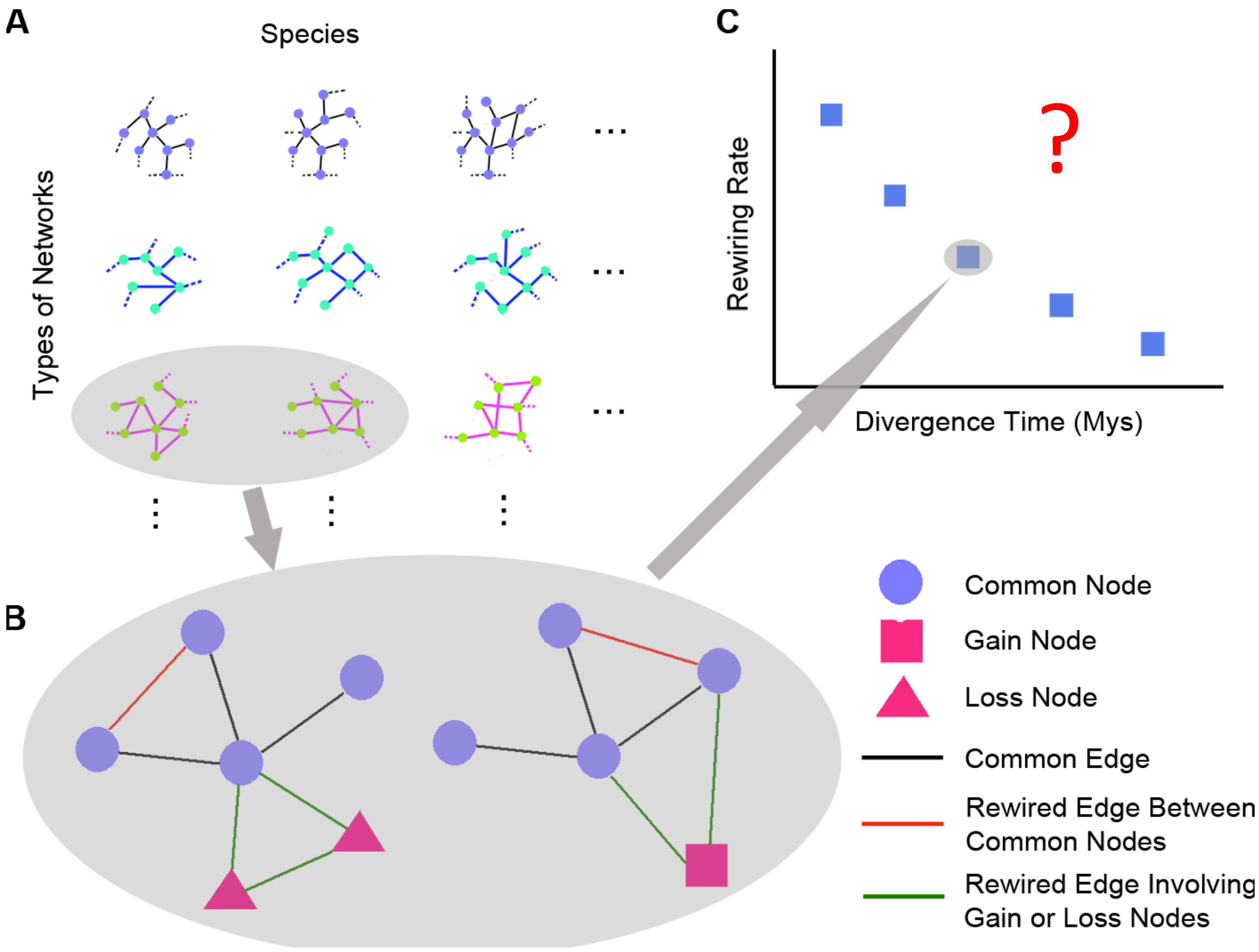


Measuring the Evolutionary Rewiring of Biological Networks

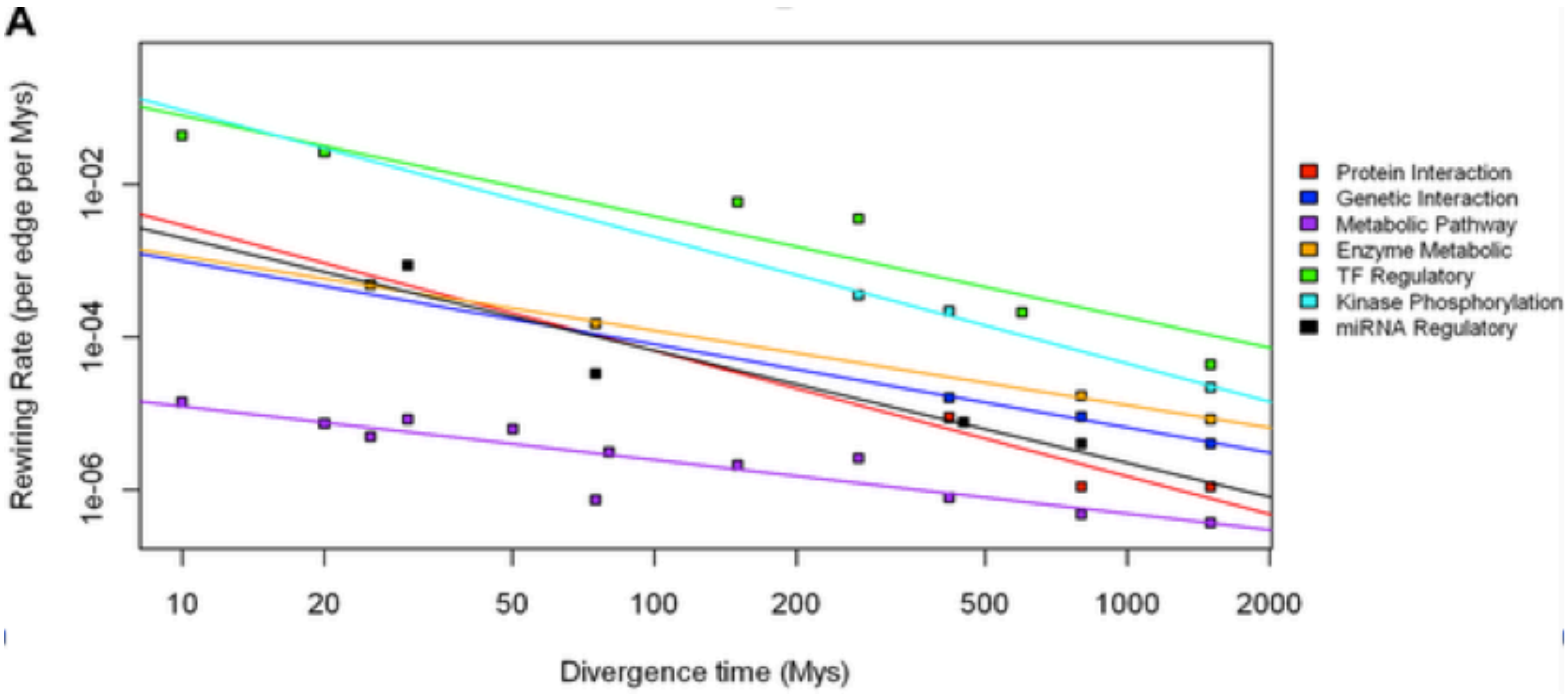
Chong Shou[1], Nitin Bhardwaj[2], Hugo Y. K. Lam[1], Koon-Kiu Yan[2], Philip M. Kim[3], Michael Snyder[4], Mark B. Gerstein[1,2,5*]

- Sought to compare interaction network from various species and reconstruction of evolutionary history

  - Analogous to studying the evolutionary history of genes by comparing versions from different species



$$\text{Rewiring rate} = \frac{\text{R}}{\text{C} \times \text{Time divergence}}$$



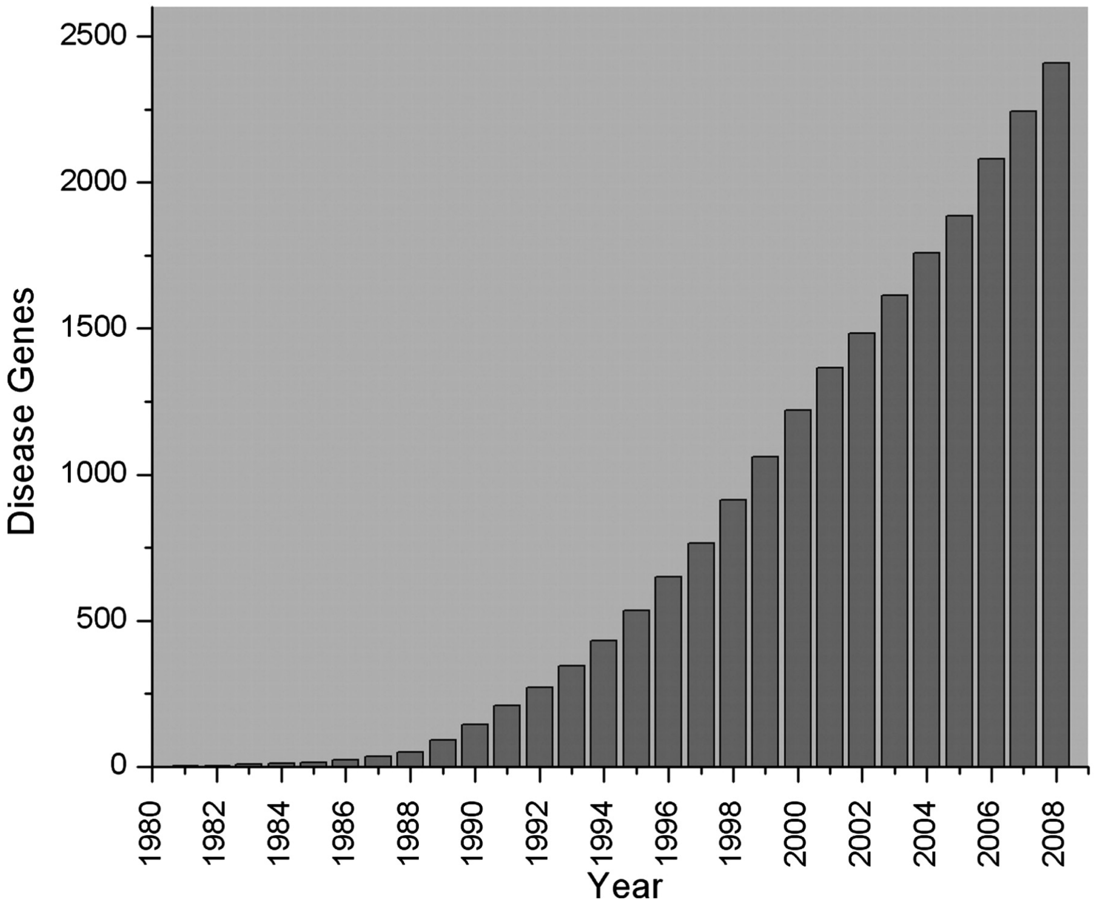
Transcription regulatory > Kinase phosphorylation > Protein interaction > Metabolic interaction

Shou C, Bhardwaj N, Lam HYK, Yan K-K, et al. (2011) Measuring the Evolutionary Rewiring of Biological Networks. PLoS Comput Biol 7(1): e1001050. doi:10.1371/journal.pcbi.1001050
http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1001050

PLoS COMPUTATIONAL BIOLOGY

NETWORK APPLICATION IN DISEASE

Histogram of cumulative growth of disease gene discovery.



The majority of disease genes still remain underneath the tip of the iceberg.

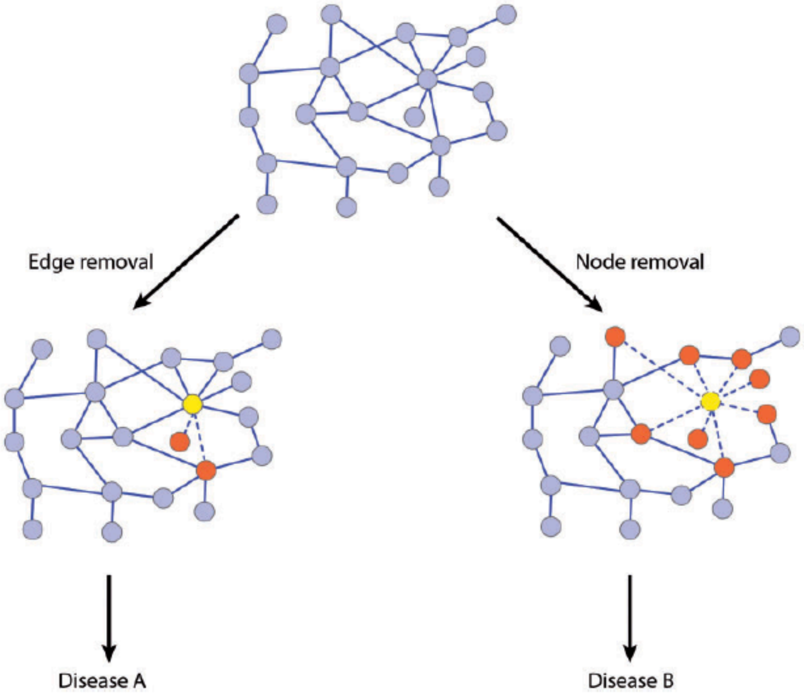
Kann M G Brief Bioinform 2010;11:96-110

GWAS approach

- Genome-wide association studies (GWAS) are powerful in mapping disease genes.

  - Low resolution: multiple genes in the disease-associated loci.

  - Perturbation of molecular networks by the gene causes the disease



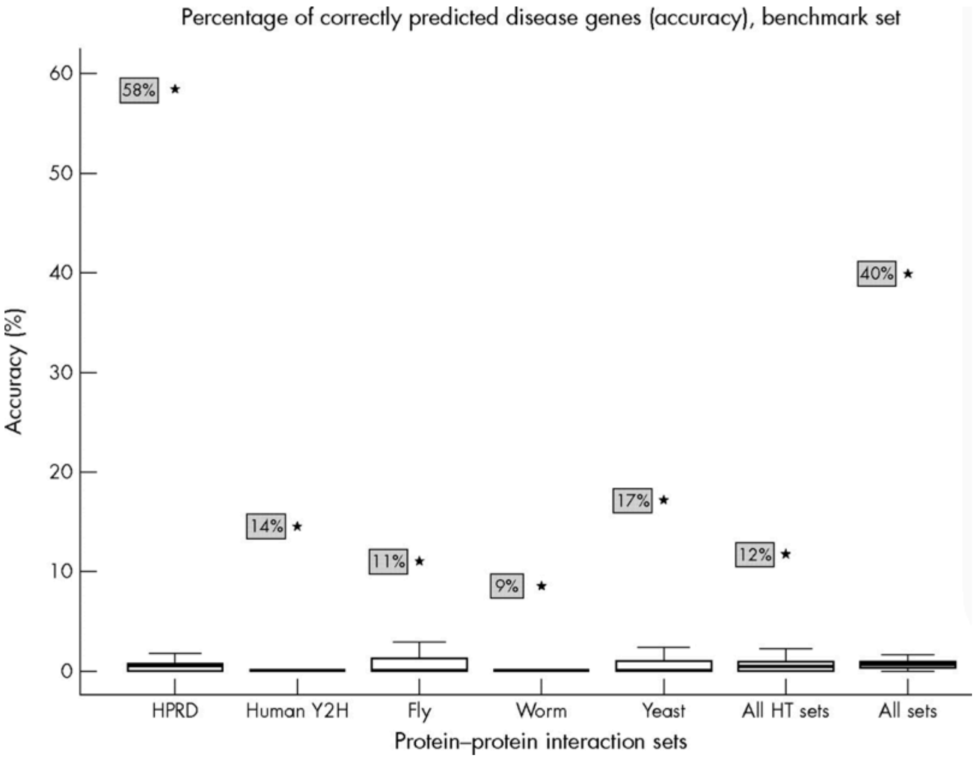
X Wang, et al., 2011

Network Approach

- Basis: genes associated with the same or similar diseases, tend to reside in the same neighborhood in networks and form physical and/or functional modules.

Linkage method

- Predict disease-causing genes in known disease loci by counting the number of known causative genes in direct neighbors.



- 10 fold enrichment in disease genes compared to random.
- Restricting genes to those in same cellular compartment as known disease genes, 1000 fold enrichment.

M.Oti et al., J Med. Genet (2006)

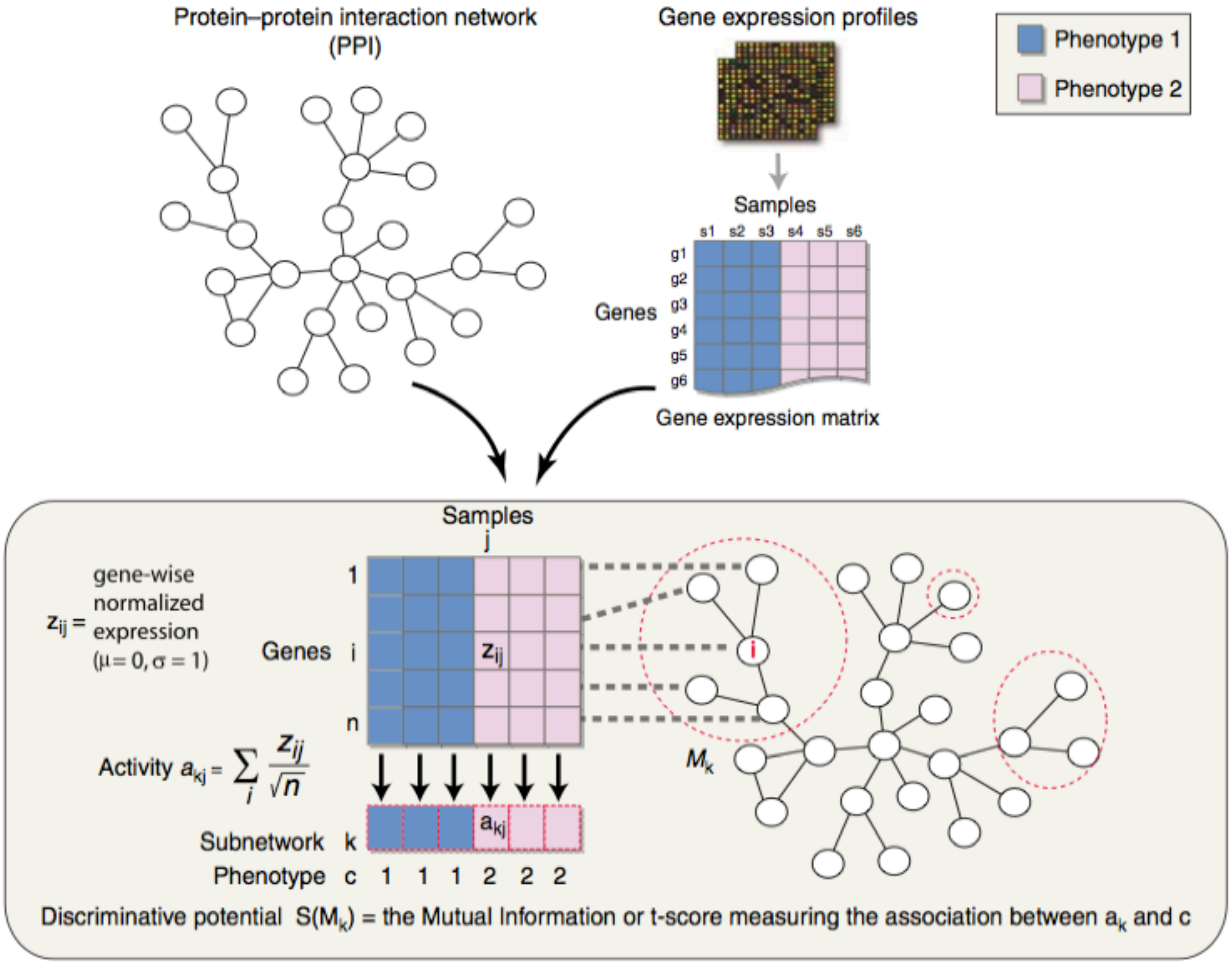
Disease sub-networks method

- Assumption: disease-associated genes form sub-network
- Approach: find a sub-network that contains most of the disease-associated genes.

Network-based classification of breast cancer metastasis

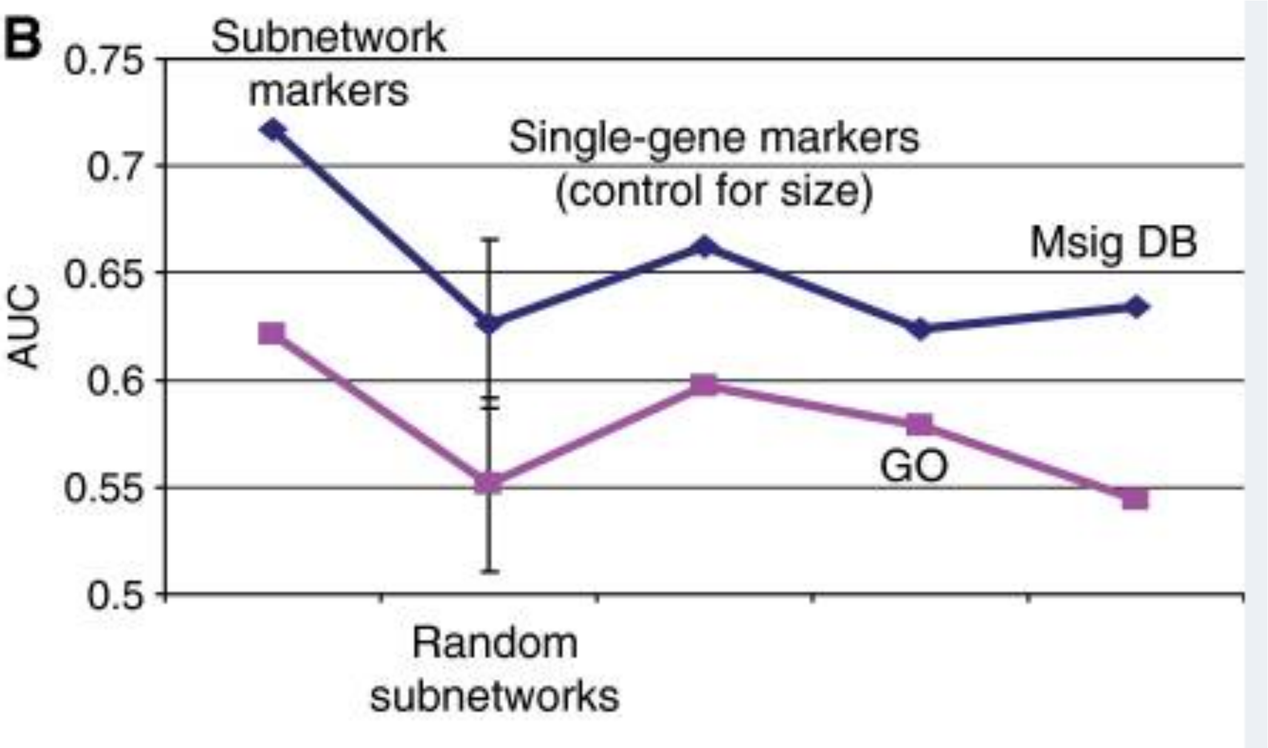
Han-Yu Chuang[1,5], Eunjung

Han-Yu Chuang
Eunjung Lee
Yu-Tsueng Liu
Doheon Lee
Trey Ideker



Disease sub-networks

- Sub-networks markers outperformed single-gene markers.



Diffusion-based methods

- Measure how connected each gene in the network is to the known disease genes (global distance measure).

 - The idea is that genes with greater “reachability” from known disease genes are more likely to be implicated in the disease.

Diffusion methods: random walk

- Start a random walker at one of the known disease gens, let it wonder the network.

1) To start, random pick a disease gene ***S***, with equal probability p^0

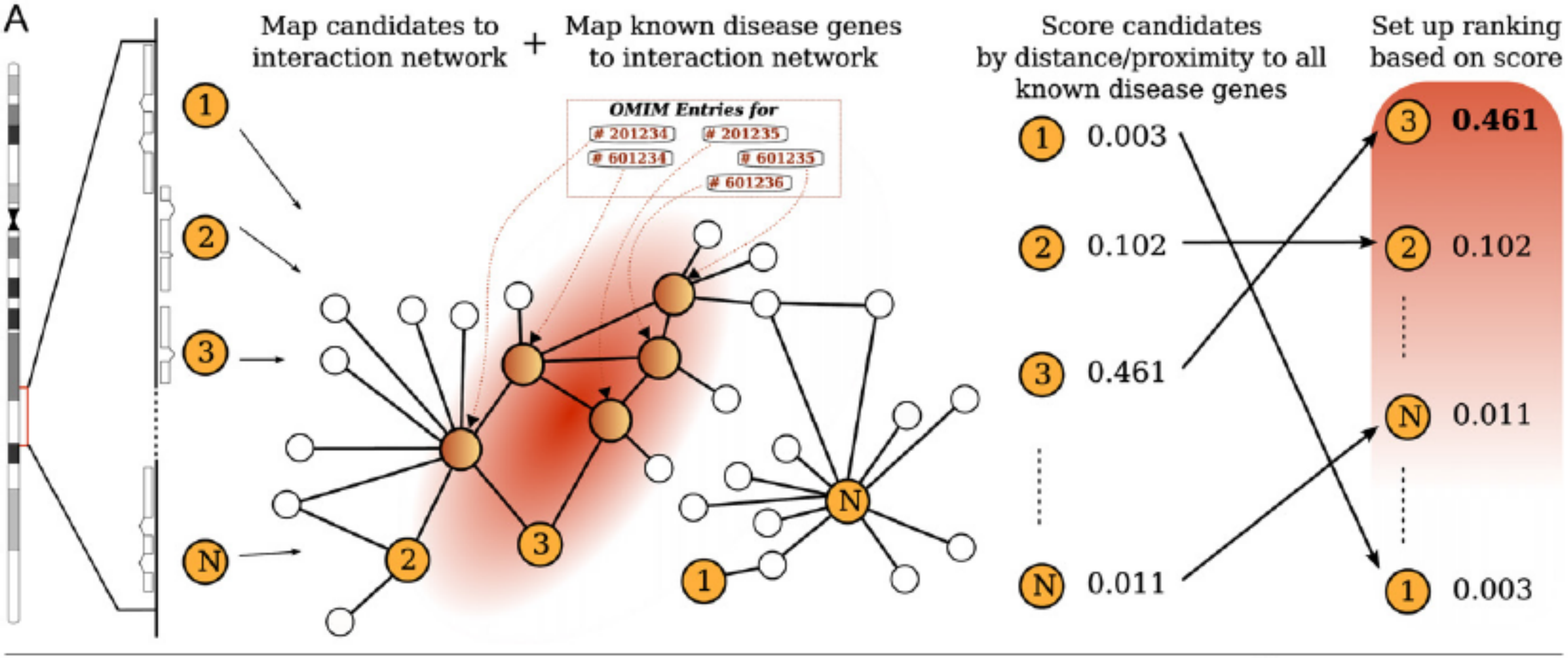
2) From current node, randomly selected a neighbor node.

3) In every time step, restart the walk from disease nodes, with probability ***r.***

$$p^{t+1} = (1 - r) W p^t + r p^0$$

p^t : probability vector of random walker being at each node at time t.

W : Matrix of transition probabilities.



Kohler S and Robinson PN, AJHG, 2011

Comparison of methods

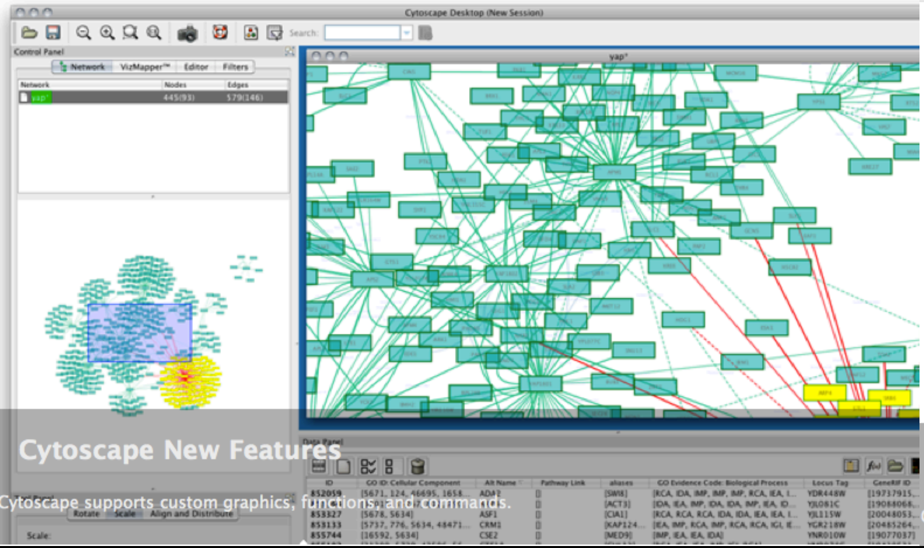
Not surprising that a comparative study of these methods, ranking them by predictive power, produces the following rankings:

1. Diffusion-based methods
2. Disease sub-networks
3. Linkage methods

Navlakha, S. et al., Bioinformatics, 2010

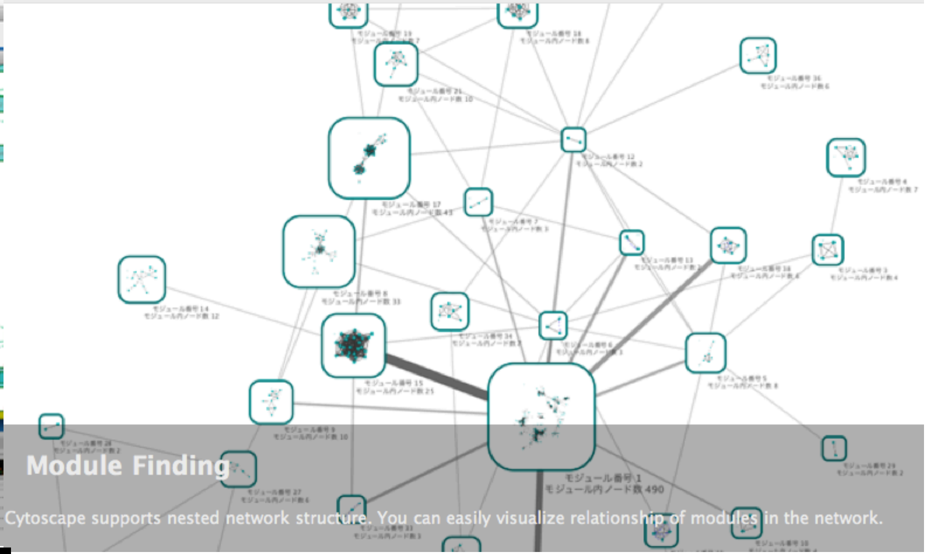
Cytoscape

Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization



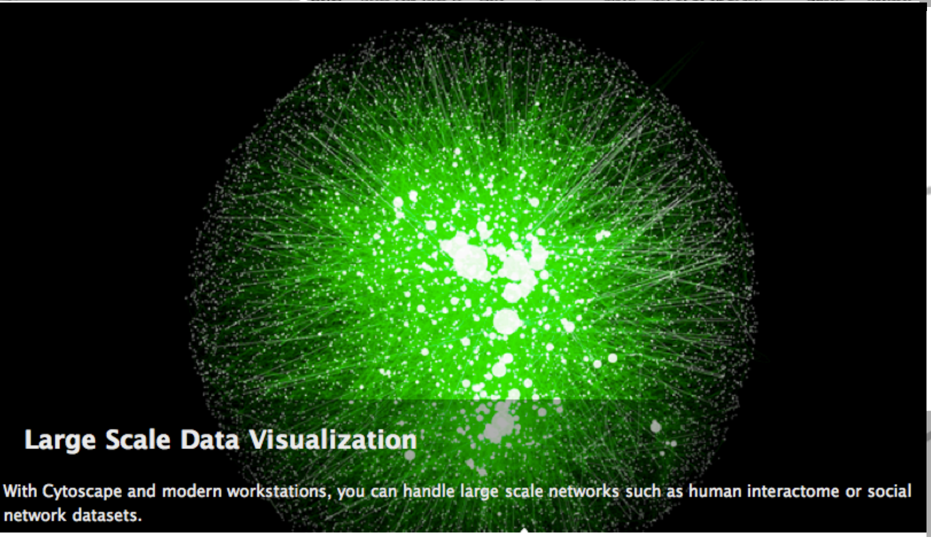
Cytoscape New Features

Cytoscape supports custom graphics, functions, and commands.



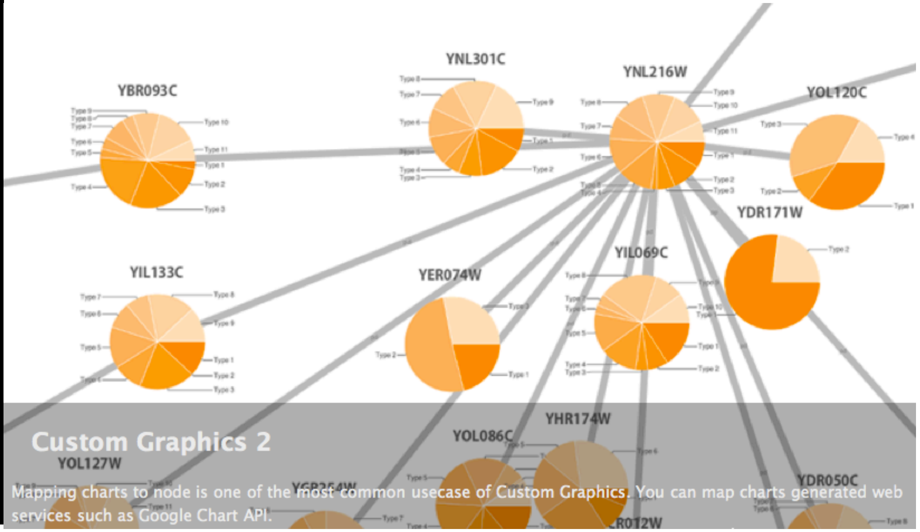
Module Finding

Cytoscape supports nested network structure. You can easily visualize relationship of modules in the network.



Large Scale Data Visualization

With Cytoscape and modern workstations, you can handle large scale networks such as human interactome or social network datasets.



Custom Graphics 2

Mapping charts to node is one of the most common usecase of Custom Graphics. You can map charts generated web services such as Google Chart API.