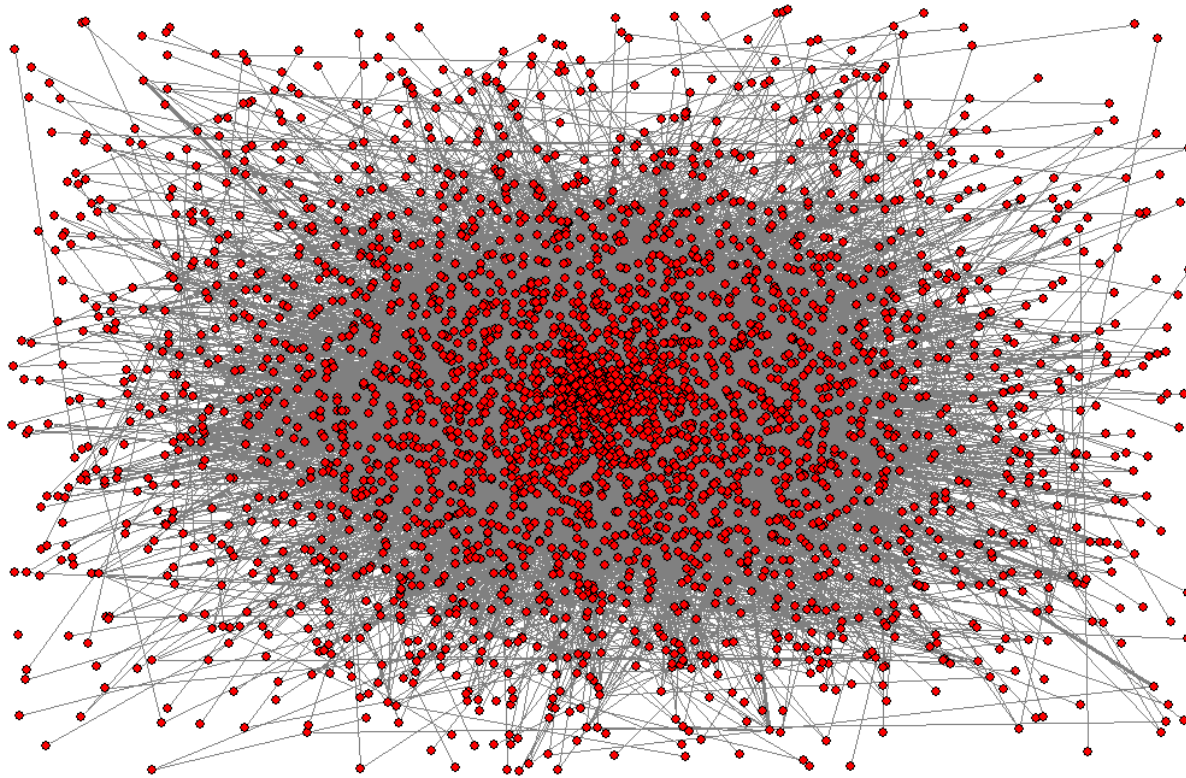


An introduction to biological network analysis



Biological networks

- We cannot understand life by simply characterizing its 'single components', but considering the interactions and the relationships between its components: a systemic approach.
- Biological networks are fundamental tools in the context of 'system biology'
- 'Omics' data (genomic, transcriptomic, proteomic, data) are used to construct biological networks and graph theory and machine learning methods are applied to model and analyze these complex objects.

Biomolecular data and processes can be naturally represented as networks

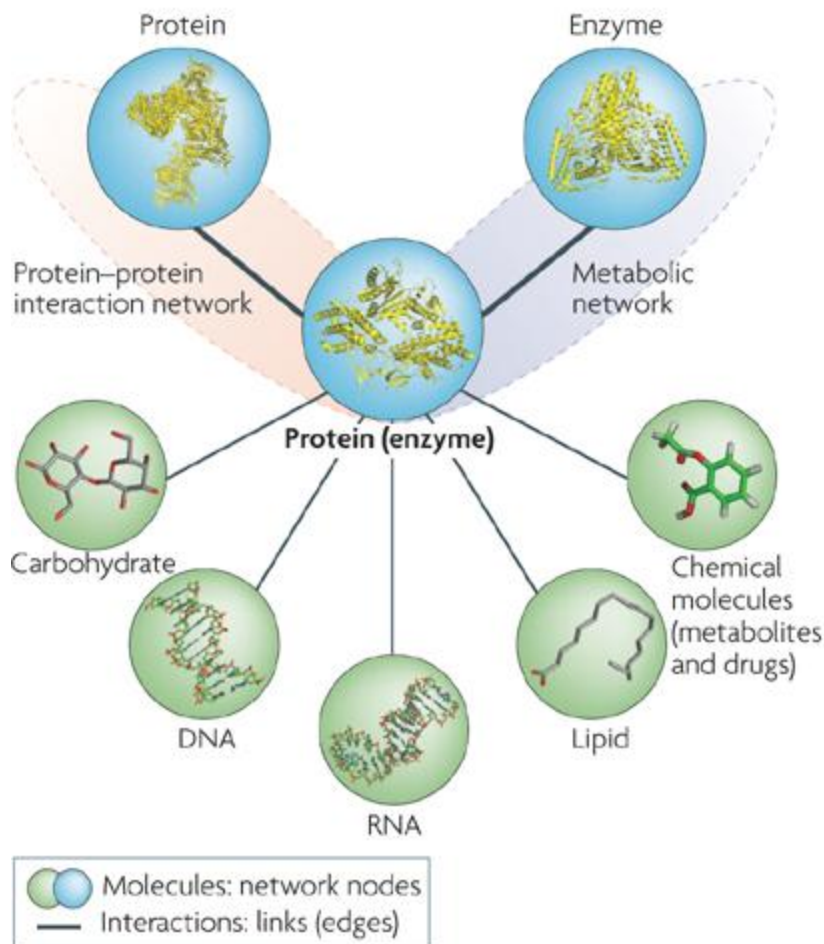
Some very general examples:

- Protein-protein interactions
- Enzyme/ligand interactions
- Gene regulatory networks
- Metabolic pathways
- ...

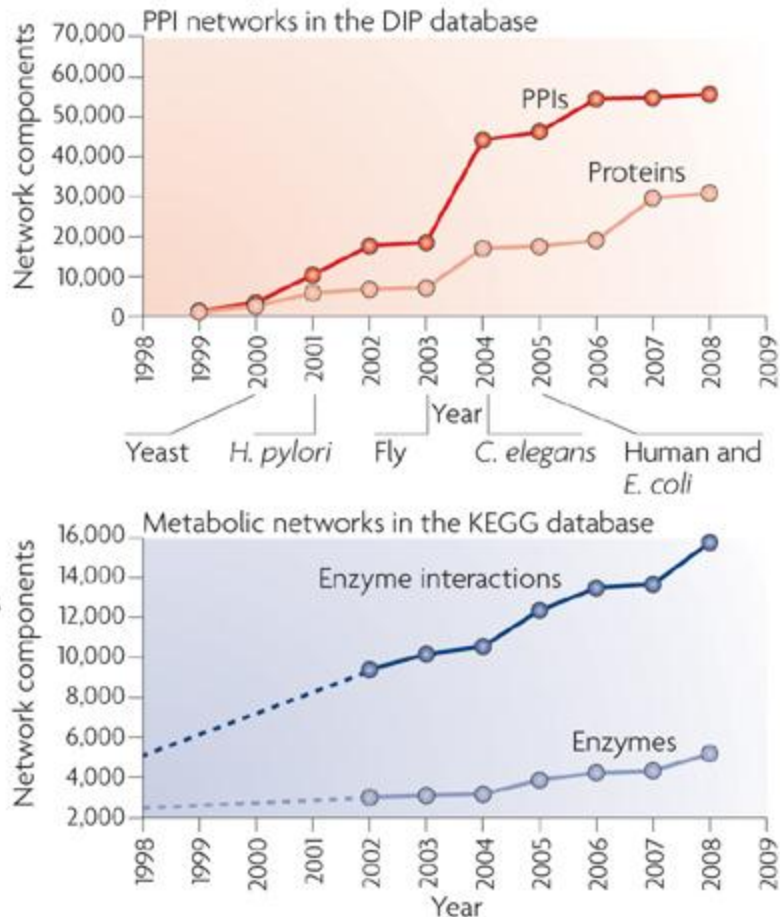
A link between two proteins if...

- The proteins interact physically and form large complexes
- The proteins are enzymes that catalyze two successive chemical reactions in a pathway
- One of the proteins regulates the expression of the other
- ...

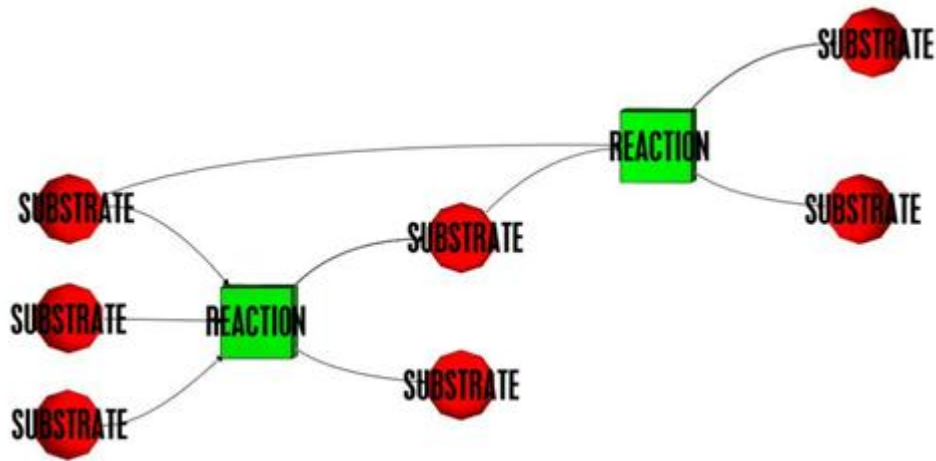
a Biomolecular network components



b Accumulation of network components over the past 10 years

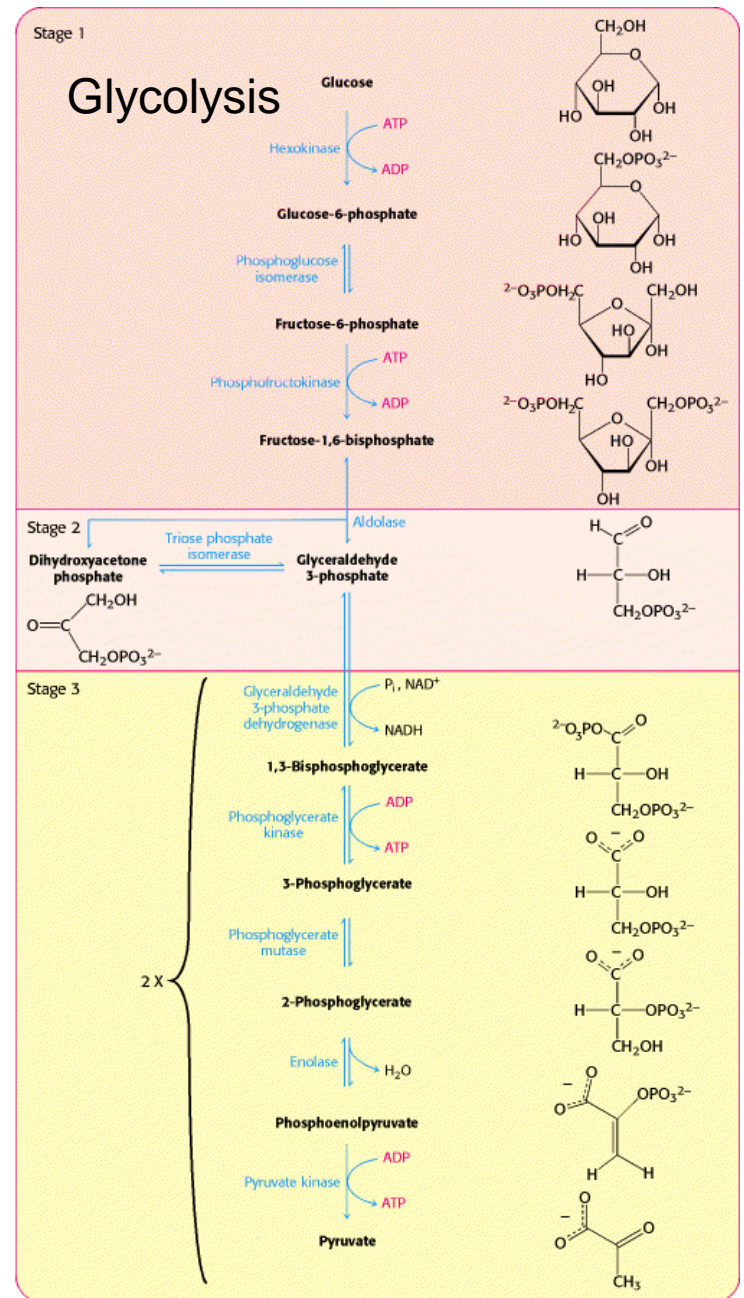


Metabolic Networks



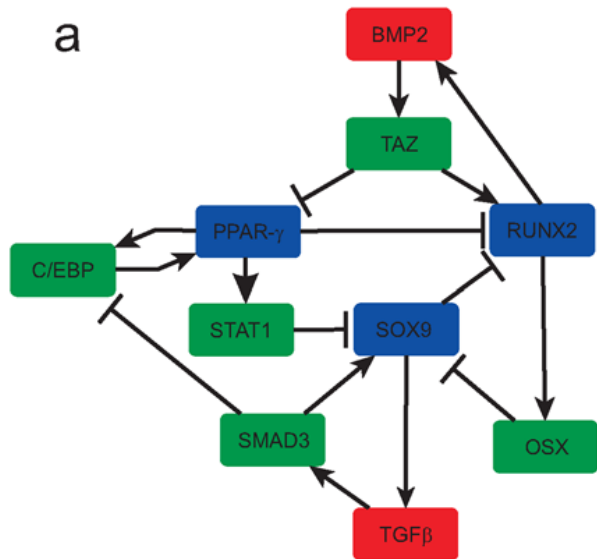
Two types of nodes: enzymes and substrates
 Reactions can be directional or bidirectional

Berg et al. Biochemistry
 New York: W. H. Freeman and Co.; c2002

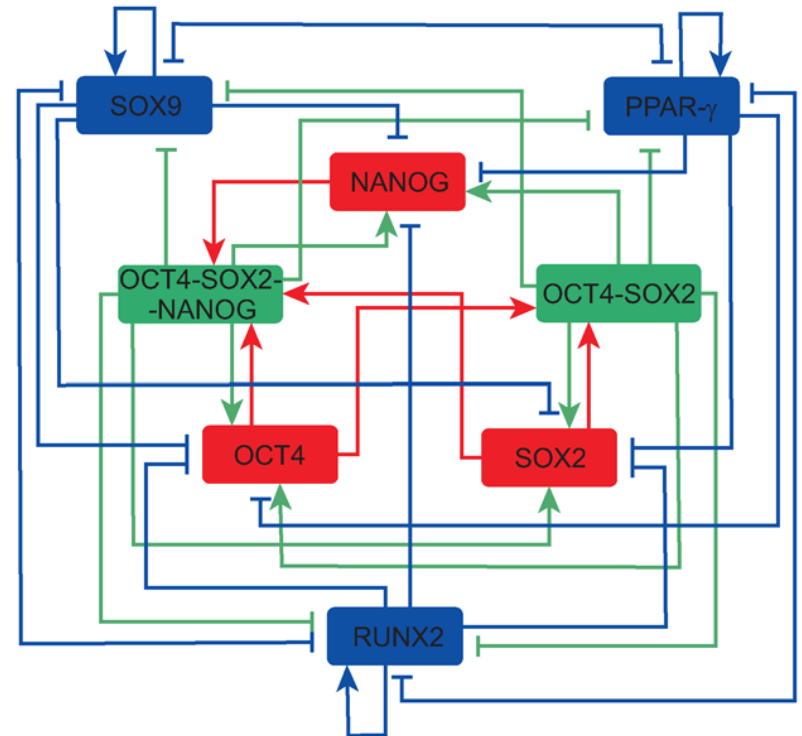


Example of Gene Regulation Networks

Stem cell differentiation regulation



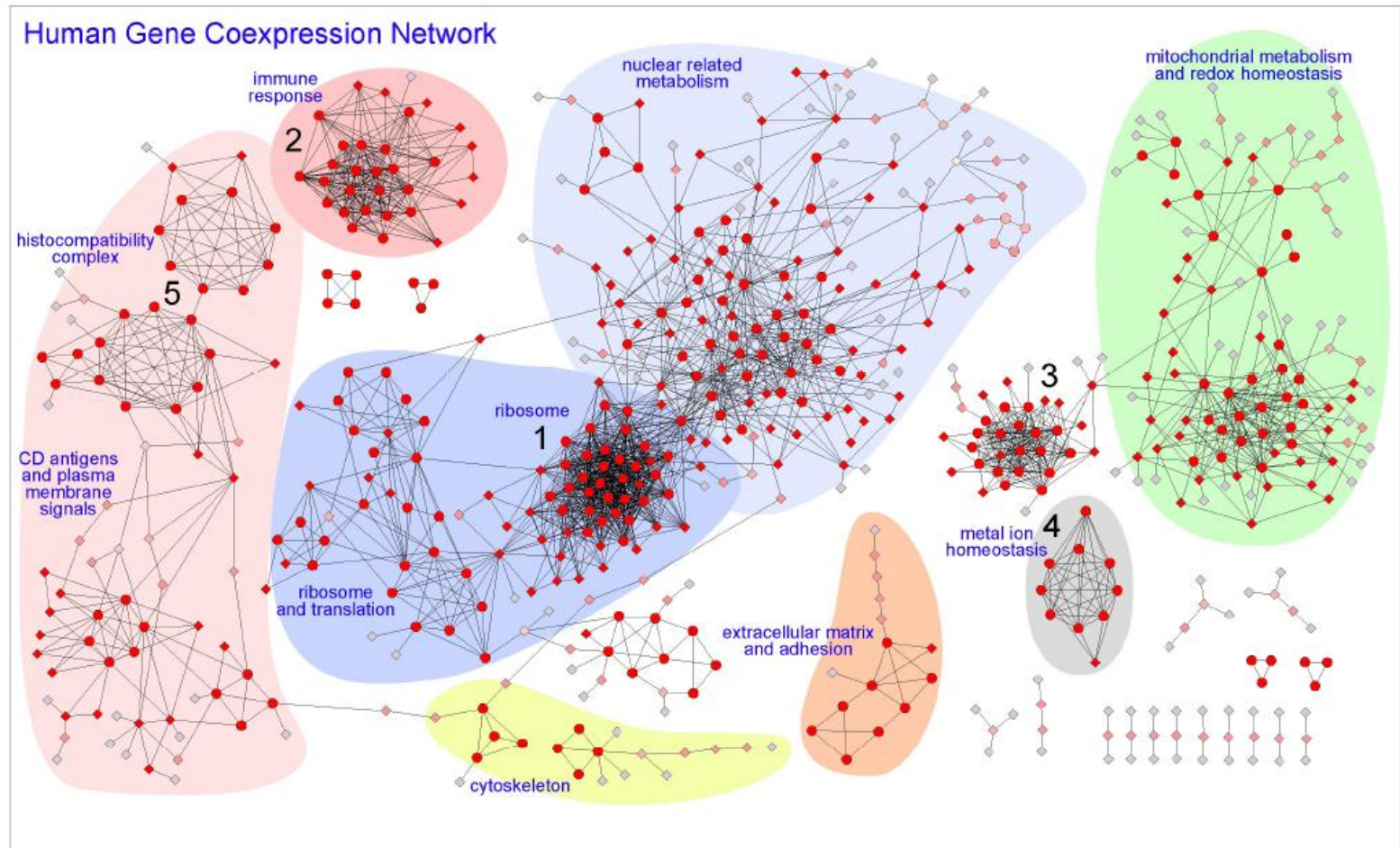
- Nodes are genes and transcription factors
- Interactions can be directional or bidirectional
- Interactions can be activation or inhibition



MacArthur et al., *PLoS ONE* 3: e3086 (2008)

Hi-Fi human coexpression network

network = intersection with 2 methods and **precision** ≥ 0.60 ($r \geq 0.77$, $N \geq 605$)



Drug-Target Networks

Drugs can be connected to their known protein targets

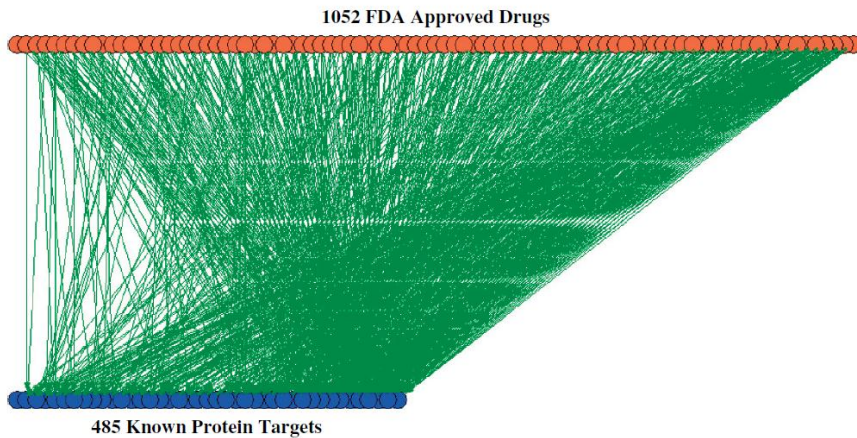
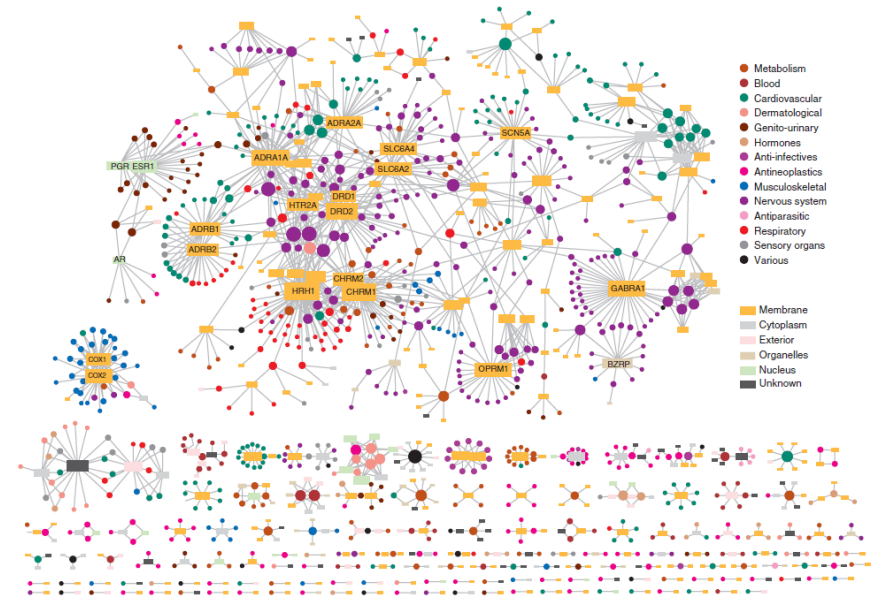


Fig 3. Visualization of the bipartite drug-target network extracted from DrugBank. Orange nodes represent drugs and blue nodes are known biomolecular targets. The network is made of 1537 nodes (1052 drugs and 485 targets) and 1815 interactions extracted from 2240 research articles.

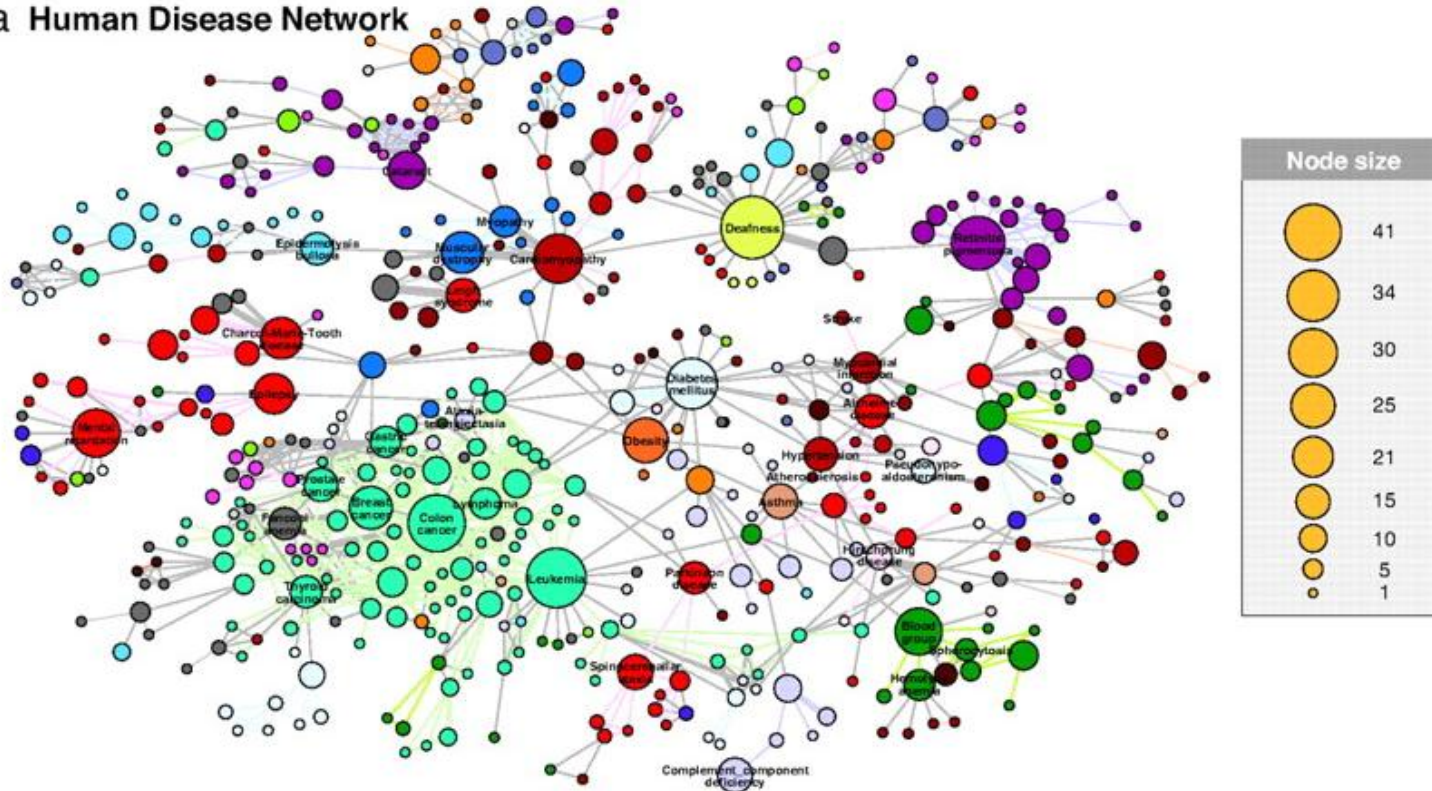
Ma'ayan et al. Mt Sinai J Med (2007) 74:27



Yildirim et al. Nat Biotechnol. (2007) 25:1110

Disease Gene Networks

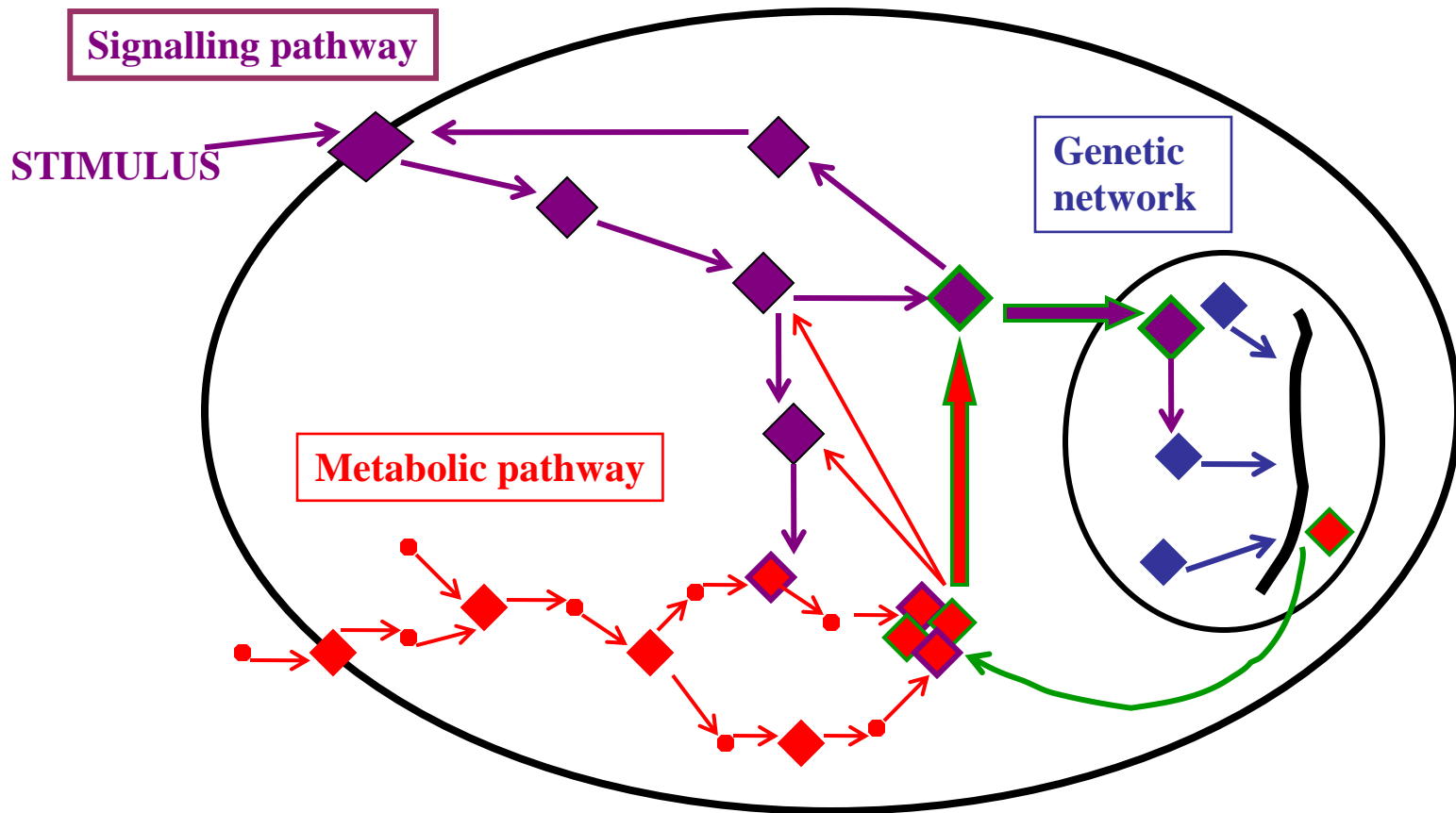
a Human Disease Network



Goh et al. Proc Natl Acad Sci USA. (2007) 104:8685-90

Each node corresponds to a distinct disorder, colored based on the disorder class. The size of each node is proportional to the number of genes in the corresponding disorder, and the link thickness is proportional to the number of genes shared by the disorders connected by the link.

Networks are inter-linked



Interactions → Networks → Pathways

- A collection of interactions defines a network
- Pathways are subsets of networks
 - All pathways are networks of interactions, however not all networks are pathways!
 - Difference in the level of annotation or understanding
- We can define a pathway as a biological network that relates to a **known** physiological process or complete function

Sources for interaction data

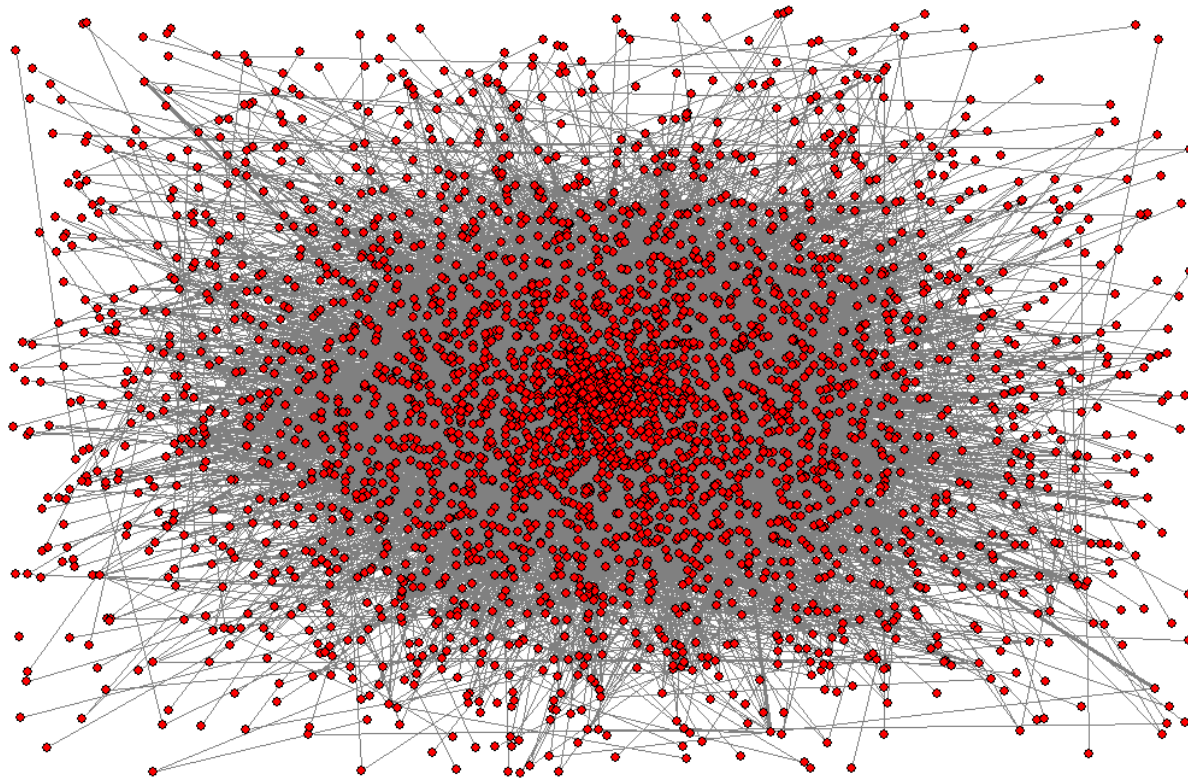
- Literature: research labs have been conducting small-scale experiments for many years!
- Interaction databases:
 - MIPS (Munich Information center for Protein Sequences)
 - BIND (Biomolecular Network Interaction Database)
 - GRID (General Repository for Interaction Datasets)
 - DIP (Database of Interacting Proteins)
 - ...
- High- throughput experiments:
 - Y2H (yeast two-hybrid method)
 - APMS (affinity purification coupled with mass spectrometry)
 - ...

Integration of multiple sources of data to construct interaction networks

- Use other data sources to create an integrated protein network.
- Other sources include:
 - Genome data:
 - Existence of genes in multiple organisms
 - Locations of the genes
 - Bio-image data
 - Gene Ontology annotations
 - Microarray experiments
 - Sub-cellular localization data

Protein interaction networks

- Large scale (genome wide networks):



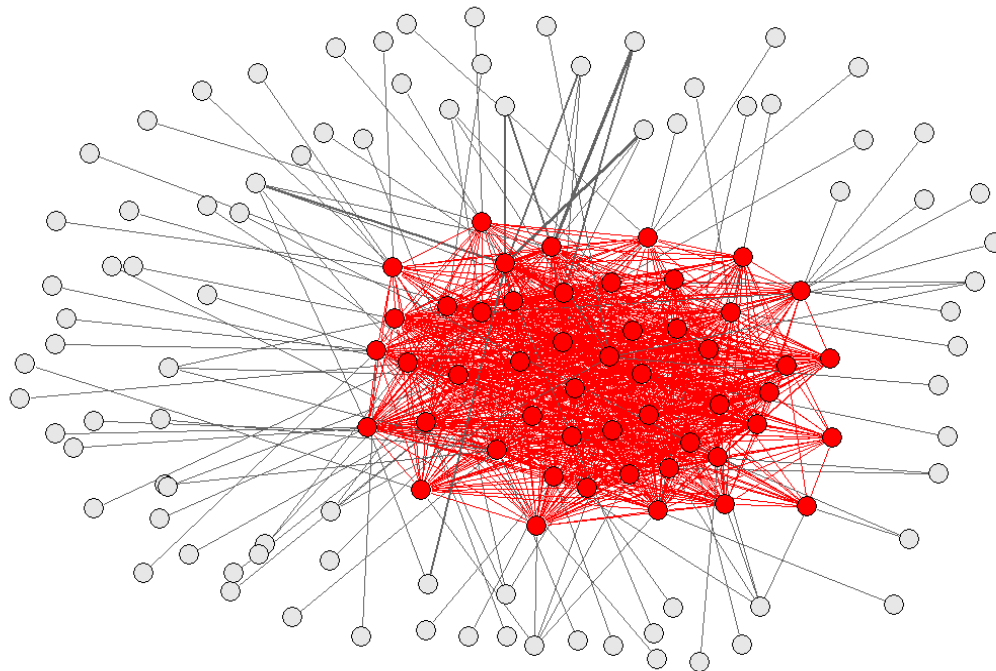
ProNet (Asthana et al.)
Yeast
3,112 nodes
12,594 edges

Analyzing Protein Networks

- Predict members of a partially known protein complex/pathway.
- Infer individual genes' functions on the basis of linked neighbors.
- Find strongly connected components, clusters to reveal unknown complexes.
- Find the best interaction path between a source and a target gene.

Gene Function Prediction (GFP)

- Having a set V of genes, a subset V_C of genes are “a priori” known to belong to a given functional class C (i.e. a Gene Ontology class)
- Can we predict whether other genes in the set $V \setminus V_C$ may belong to C ?



Disease gene prioritization

- Having a set V of genes, a subset V_C of genes are “a priori” known to be associated to a given disease C
- Can we rank genes in the set $V \setminus V_C$ with respect to their likelihood of being associated to C ?

