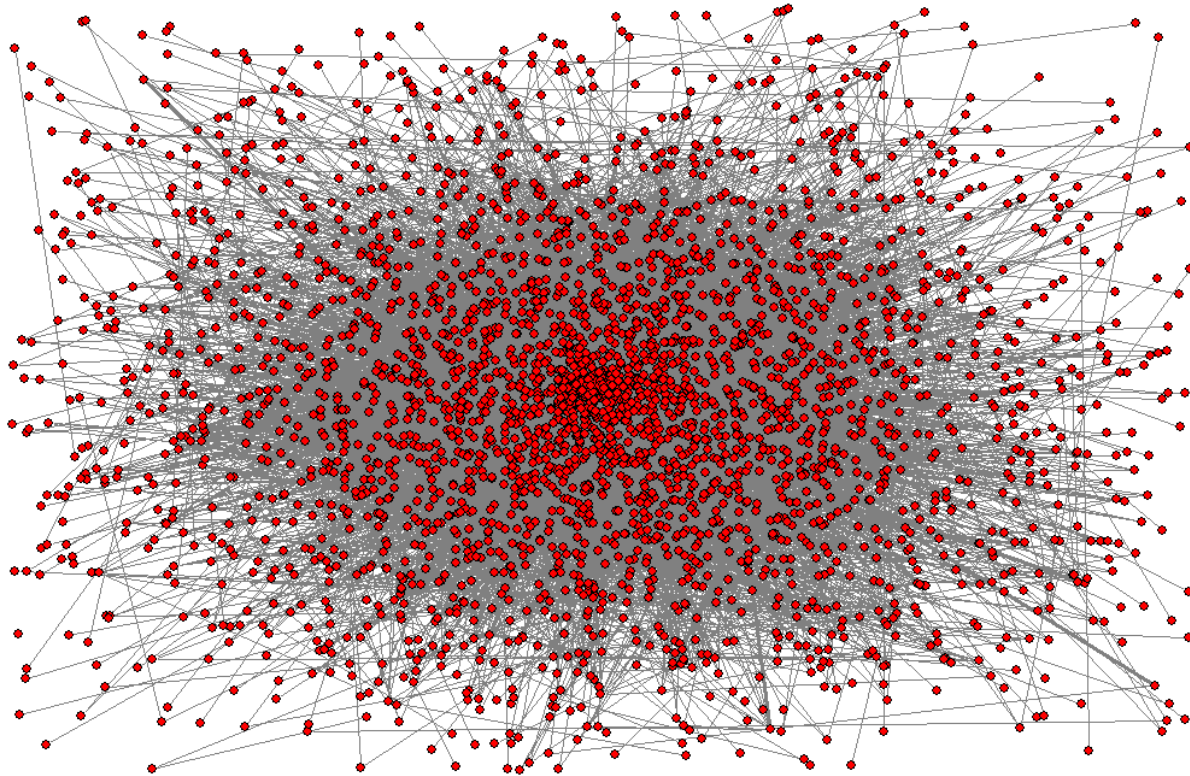


Biological Networks

Giorgio Valentini

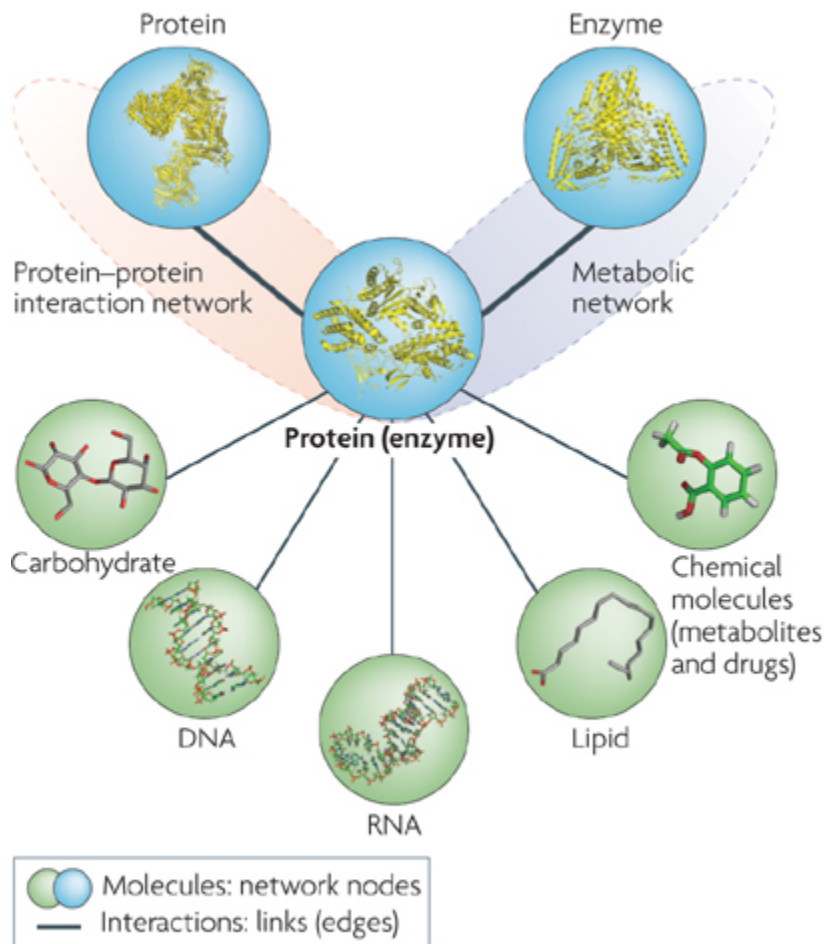
AnacletoLab – DI, Università degli Studi di Milano



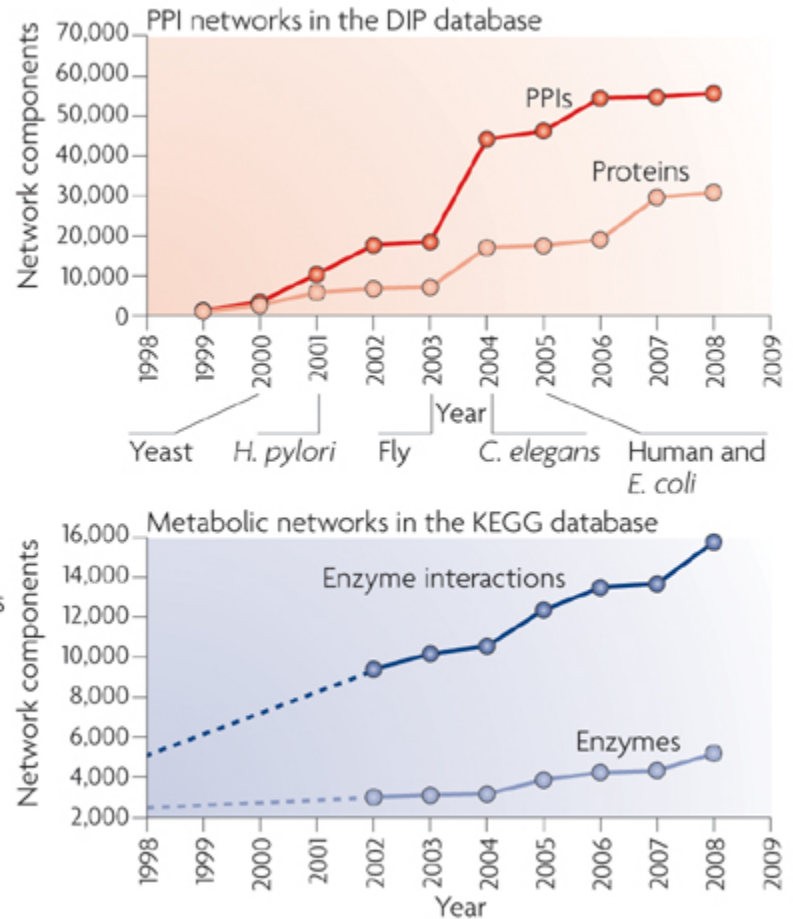
Biological networks

- 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.

a Biomolecular network components



b Accumulation of network components over the past 10 years



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
- ...

Intra-cellular Networks

- Cell functioning is based on a large set of very interconnected relationships between molecules (DNA, RNA, proteins, metabolites)
- Most cellular activities relies and depend on the interactions between different molecules
- The fundamental elements in this complex interaction networks are proteins

Intra cellular Networks

- Metabolic networks
- Transcriptional regulatory networks
- Signaling networks
- Protein – Protein Interaction networks (PPI)
- Protein structural networks
- ...
- They describe the functioning of the cell at different levels and are interlinked

Metabolic Networks

- Elements of metabolic networks:
- *Metabolites*: small molecules as glucose or aminoacids
- *Pathways*: ordered set of biochemical reactions that realize a specific biological function.
- *Pathways steps*: Pathways are structured in steps where a metabolite is transformed into another metabolite.
- *Enzymes* (specialized proteins with catalytic activity) drive each step of each pathway.

Nodes in the network represent metabolites and possibly also enzymes, *edges* biochemical reactions

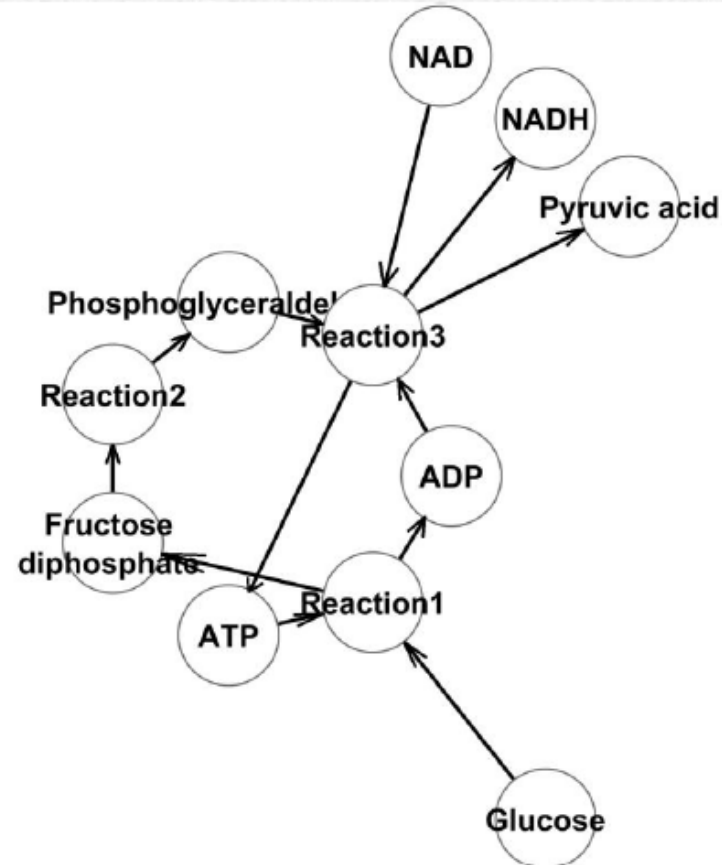
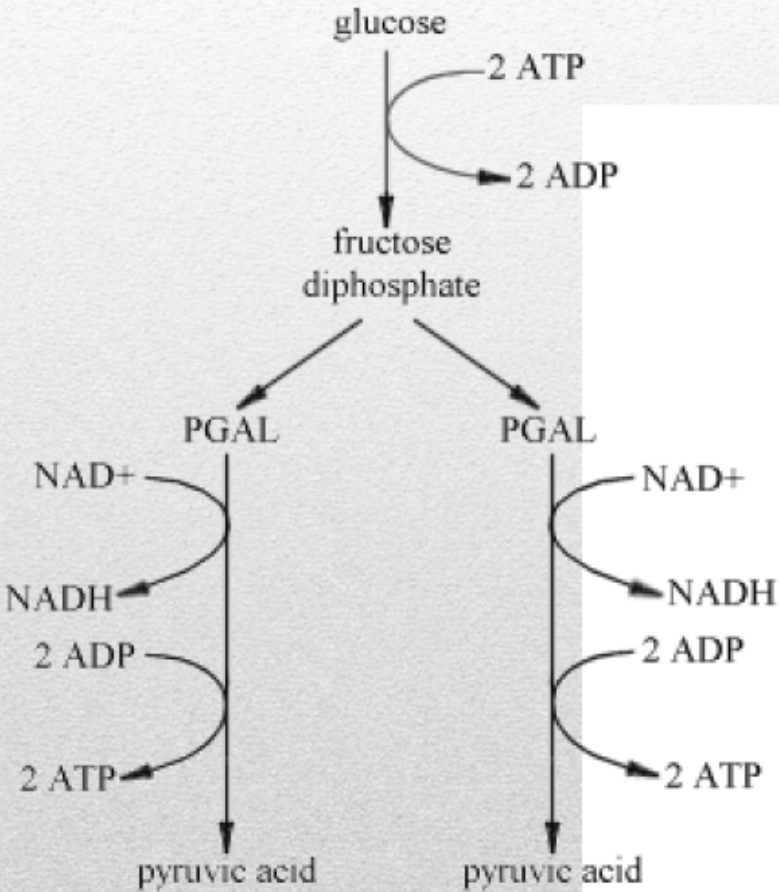
Metabolic Networks

- They represent the set of the biochemical reactions that allow the organisms:
 - *To renew their energy*
 - *To respond to external stimuli*
 - *To grow*
 - *To maintain their structure* (life is a struggle against entropy)
 - *To maintain their dynamic equilibrium* (homeostasis)
- In a few words metabolic networks constitute the basic structure underlying life

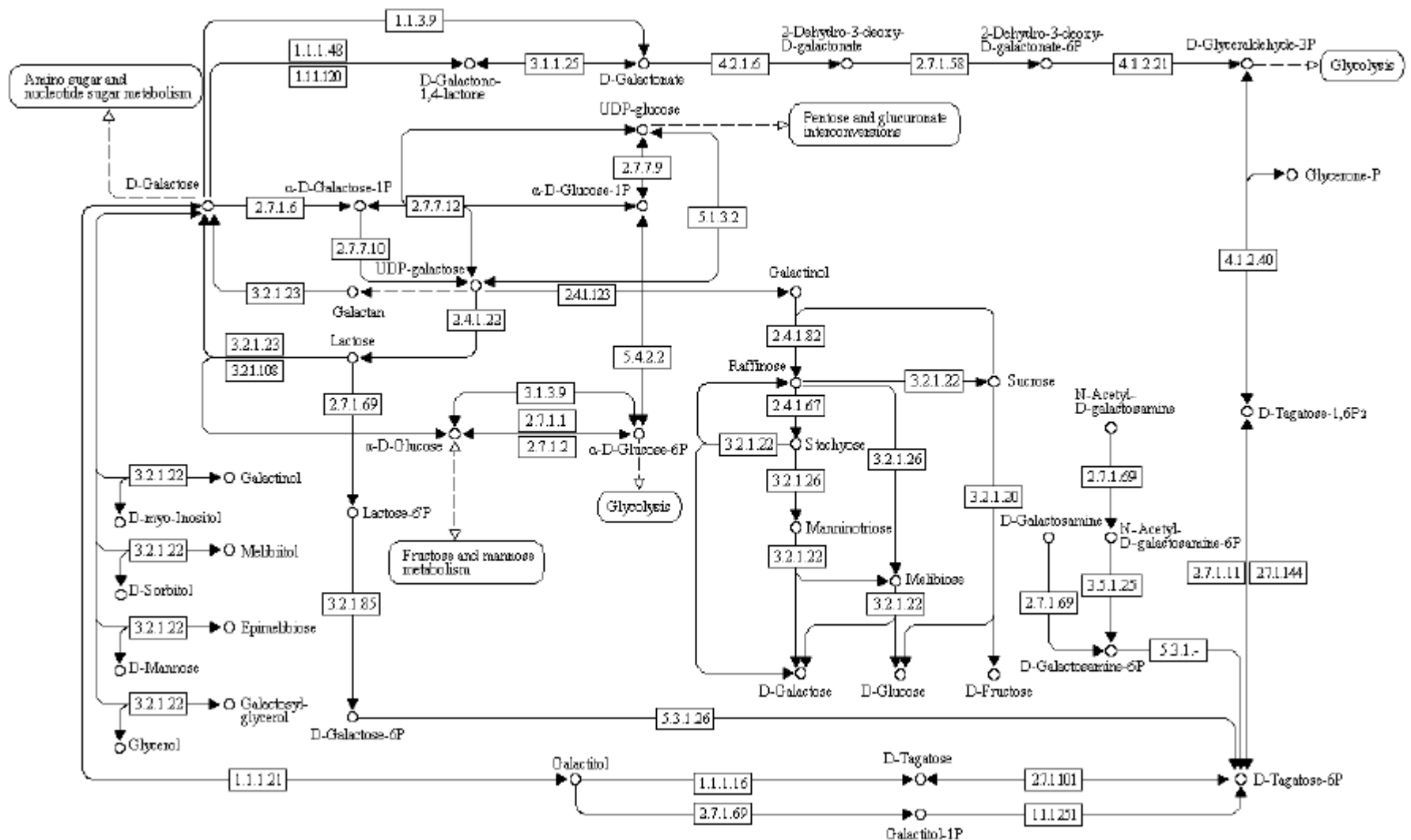
Example of Metabolic Network

Glycolysis

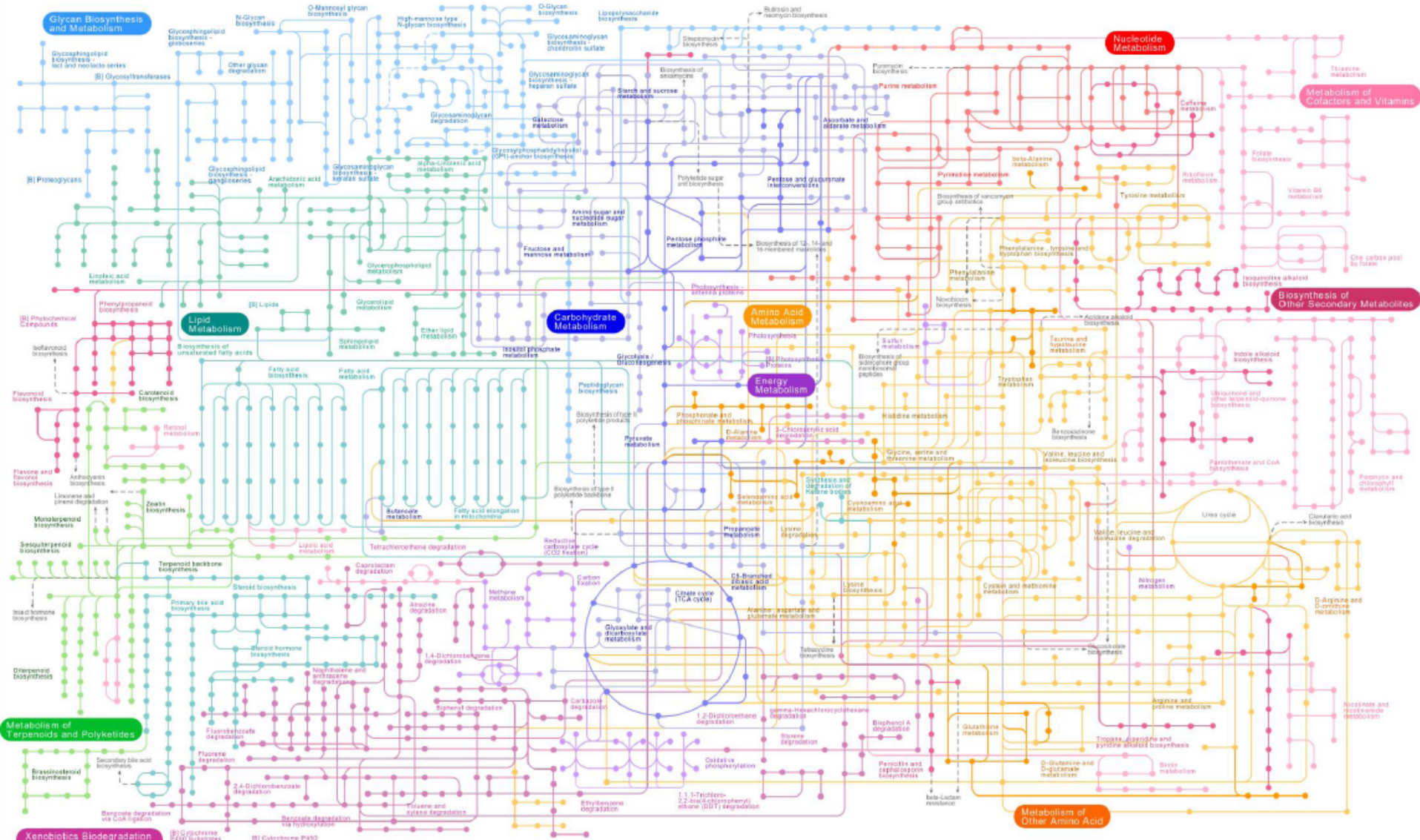
Reactions + metabolites:



There are larger metabolic networks (e.g. galactose metabolic networks)

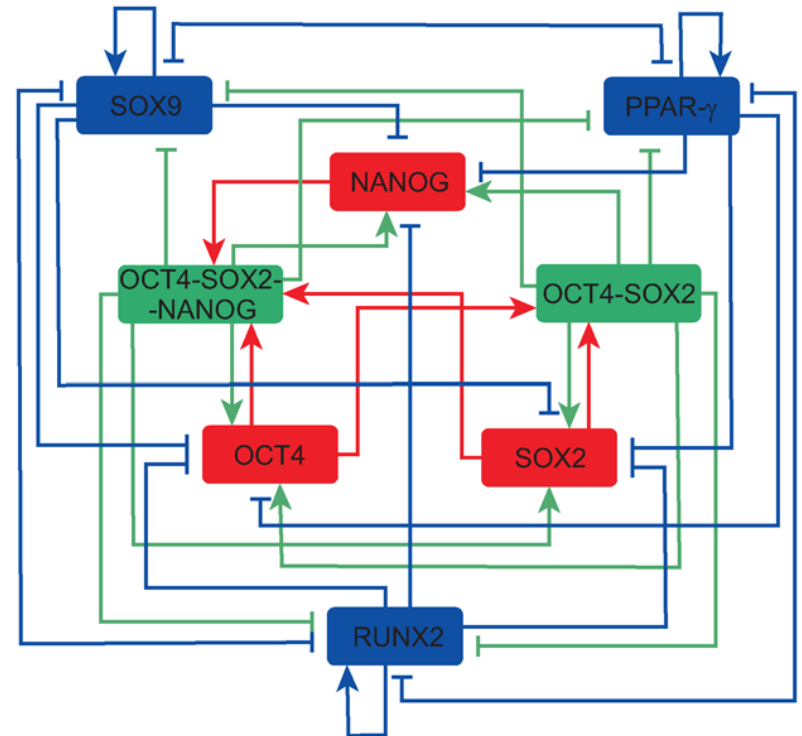
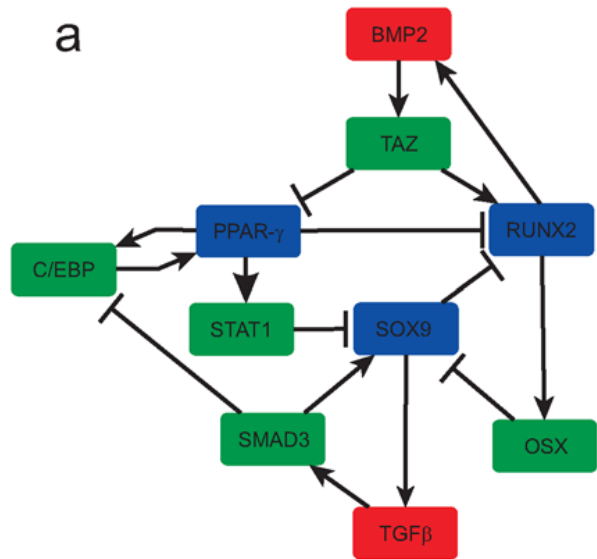


Human metabolic networks (from Reactome)



Gene Regulation Networks

Stem cell differentiation regulation

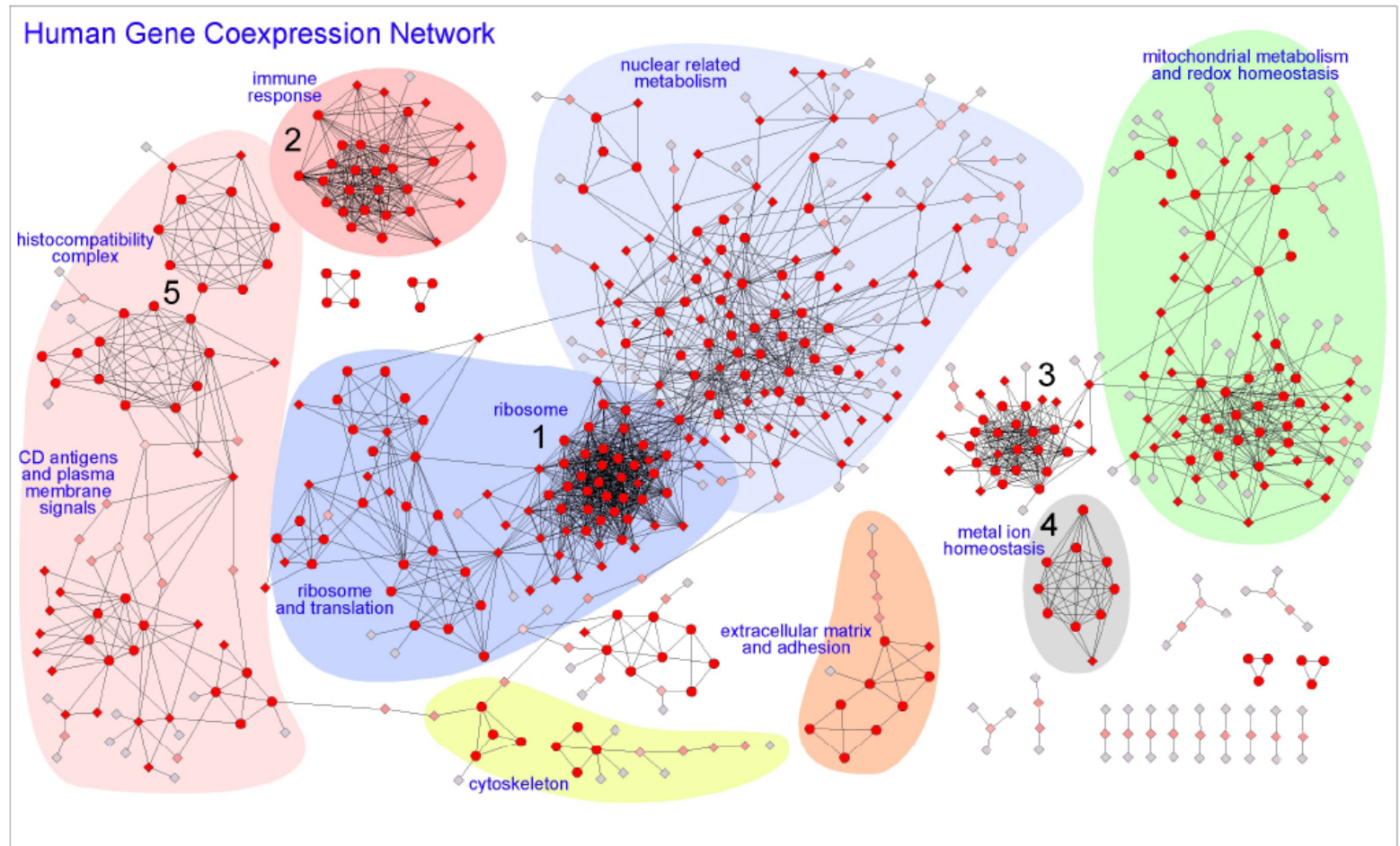


- Nodes are genes and transcription factors
- Interactions can be directional or bidirectional
- Interactions can be activation or inhibition

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

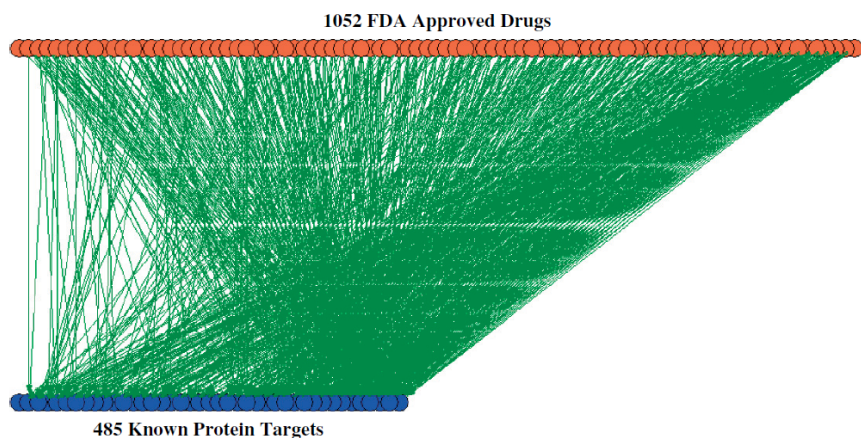
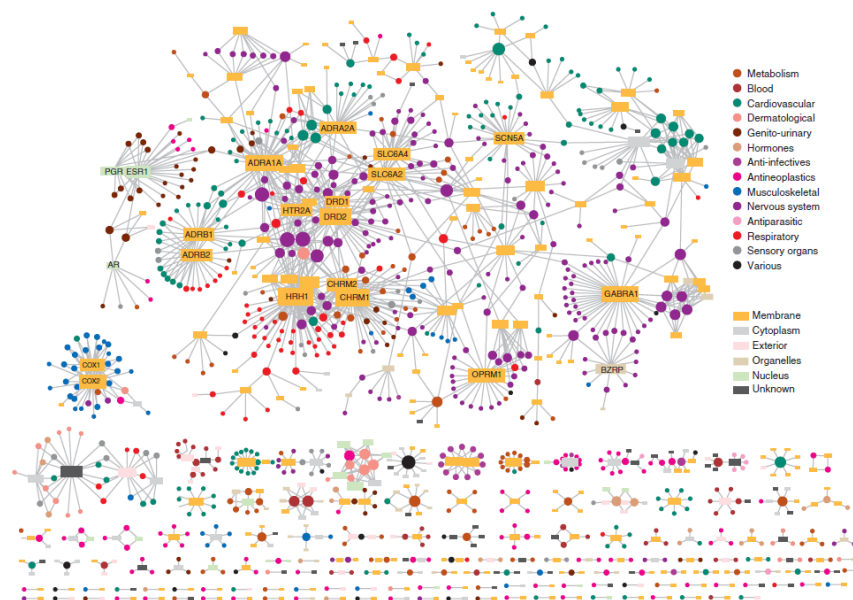


Fig 3. Visualization of the bipartite drug-target network extracted from DrugBank. Orange nodes represent drugs and blue nodes are known biomolecular targets. The network is made of 1537 nodes (1052 drugs and 485 targets) and 1815 interactions extracted from 2240 research articles.

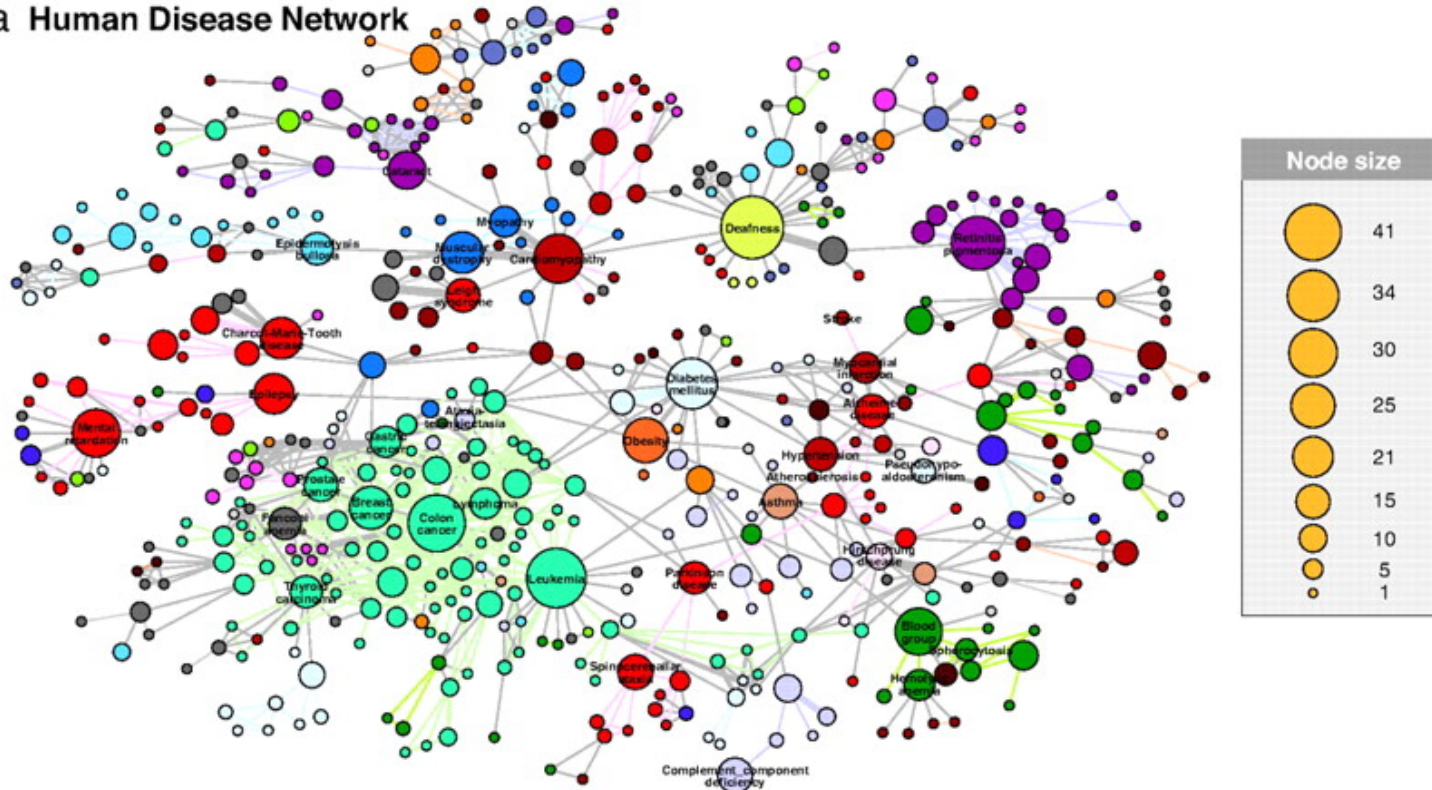
Ma'ayan et al. Mt Sinai J Med (2007) 74:27



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

Disease Networks

a Human Disease Network

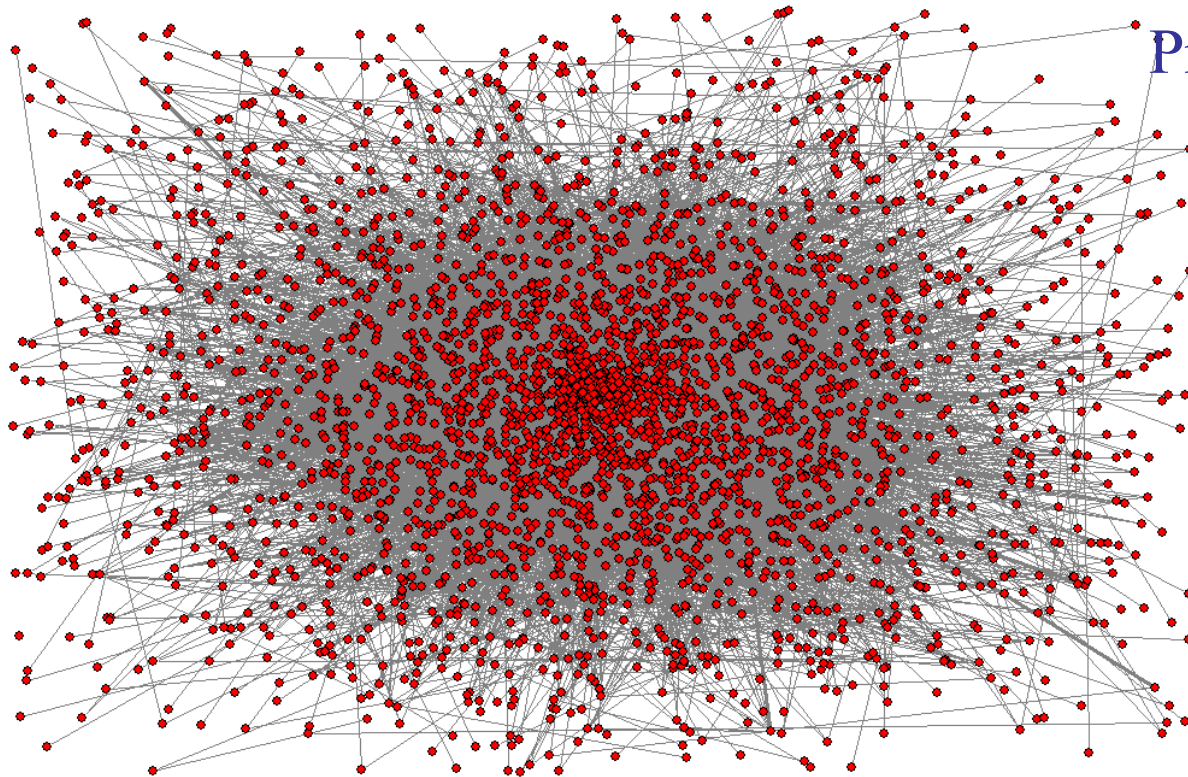


Goh et al. Proc Natl Acad Sci USA. (2007) 104:8685-90

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.

Protein interaction networks

- Large scale (genome wide networks):



ProNet (Asthana et al)

Yeast

3,112 nodes

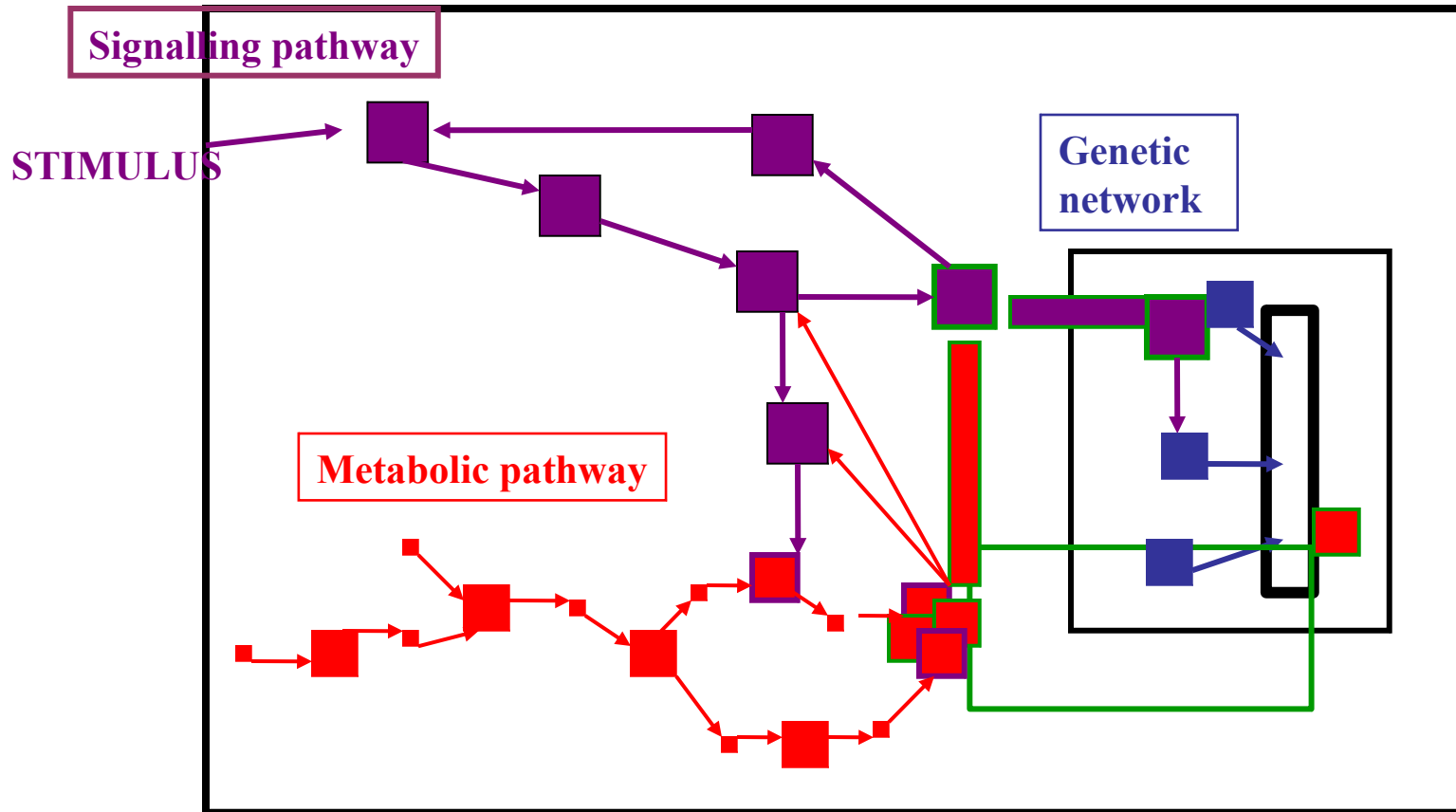
12,594 edges

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)
 - STRING (the largest PPI DB)
 - ...
- High- throughput experiments:
 - Y2H (yeast two-hybrid method)
 - APMS (affinity purification coupled with mass spectrometry)
 - ...

Networks from different sources represent different “views” of the same objects and can be integrated using different algorithms ...

Networks are inter-linked

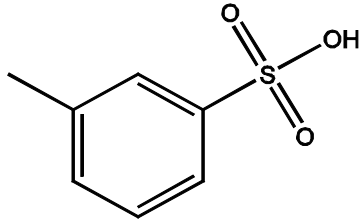


Analyzing Protein Networks

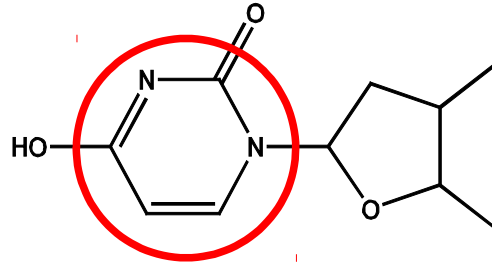
- Predict members of a partially known protein complex/pathway.
- Infer individual genes' functions on the basis of linked neighbors.
- Find strongly connected components, clusters to reveal unknown complexes.
- Find the best interaction path between a source and a target gene.

• Drug repositioning

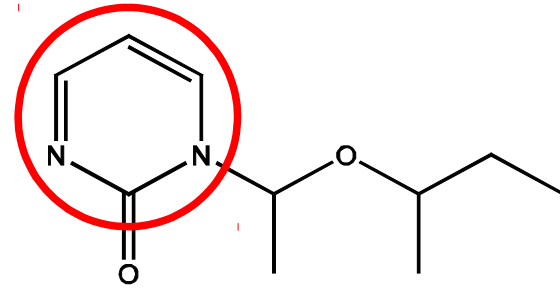
- Given a collection of molecules



(A)



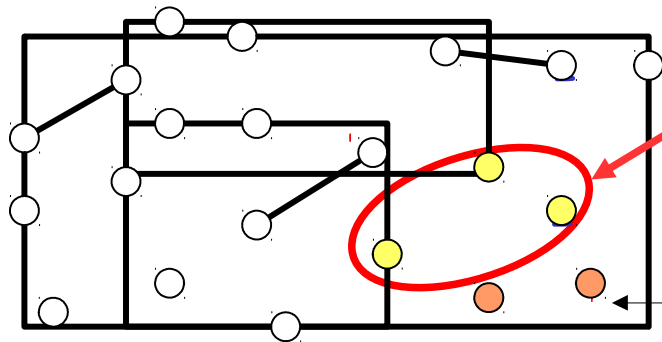
(B)



(C)

Find a meaningful way to express a similarity between them (i.e. binary profiles indicating the presence/absence of **substructures** used as proxy for the computation of a global similarity score between each pair of molecules).

Nodes: drugs
Edges: similarity between drugs



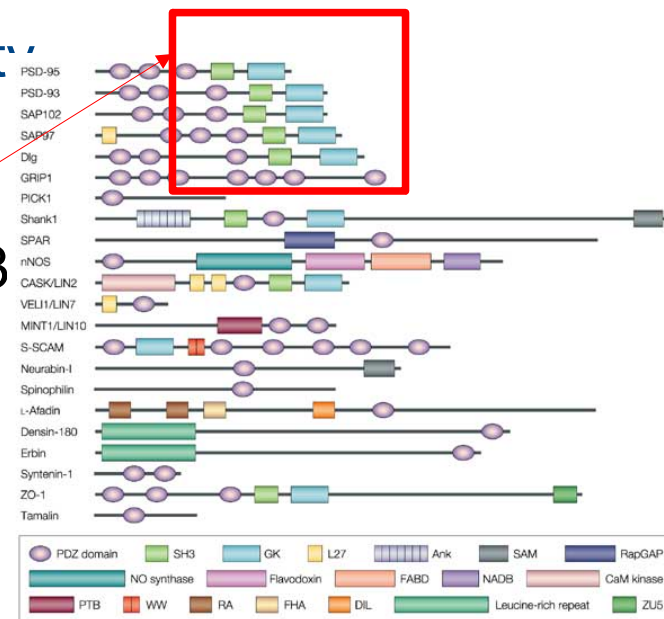
The **most similar** nodes (drugs) are candidates for the development of novel anticonvulsant drugs

Seed node, a marketed drug (i.e. anticonvulsant)

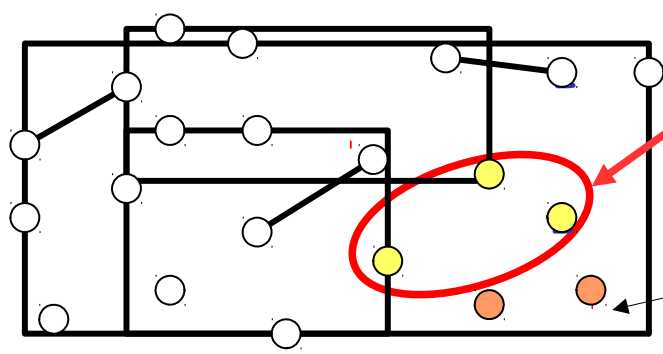
Automated Function Prediction

Given a collection of proteins.

Find a meaningful way to express a similarity between them (i.e. binary profiles indicating the presence/absence of **protein domains**, **3** structure signatures, presence/absence of catalytic groups **used as proxy for the computation of a global similarity score between each pair of proteins**).



Nature Reviews | Neuroscience

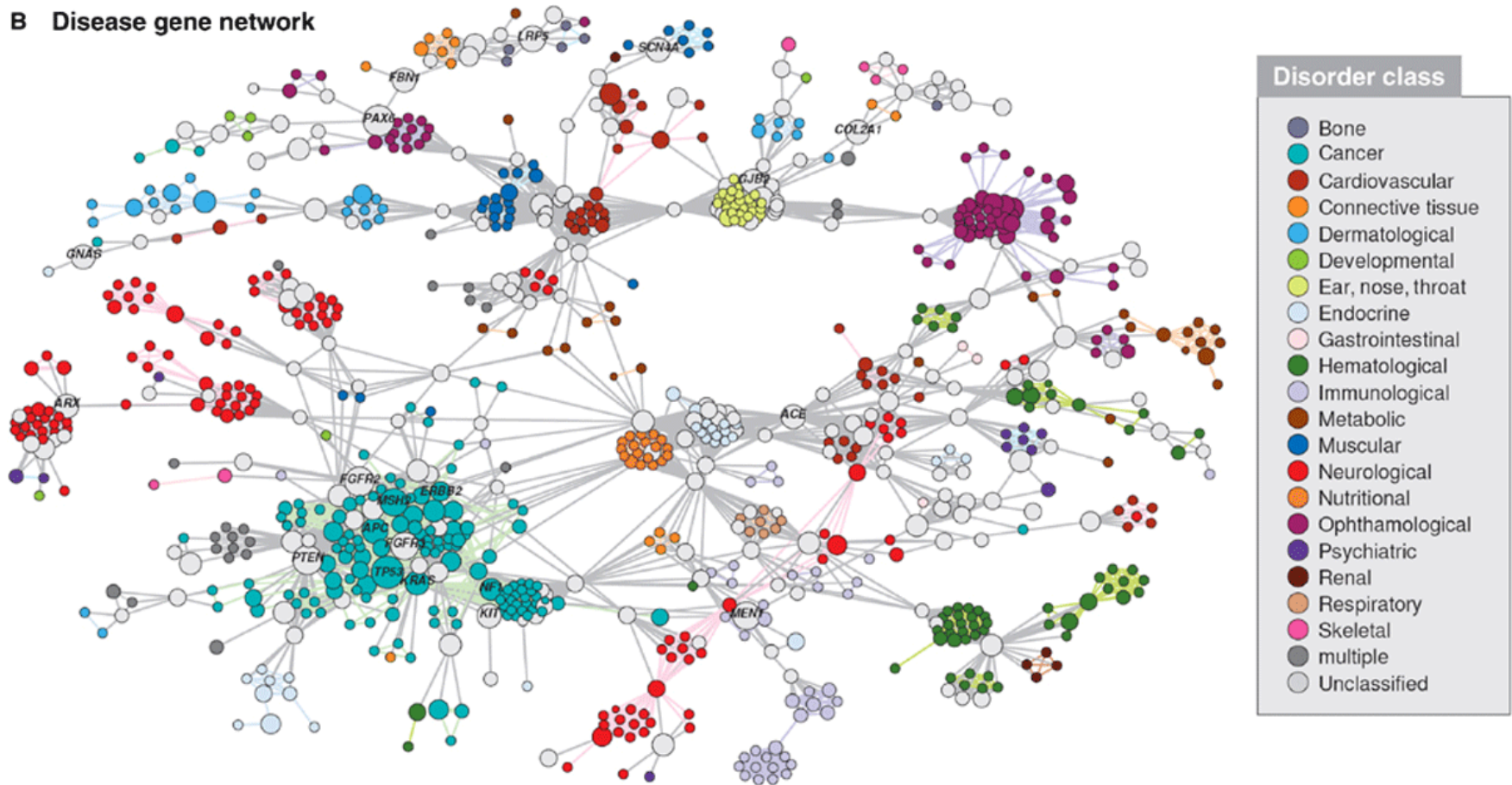


The **most similar** nodes (proteins) are candidates for the association to the functional term associated to the seeds

Seed node, associated to a **functional vocabulary term** (i.e. Gene Ontology)

Disease gene prioritization

B Disease gene network



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 ?

