Biological network analysis

Rui Benfeitas
NBIS - National Bioinformatics Infrastructure Sweden
Science for Life Laboratory, Stockholm
Stockholm University

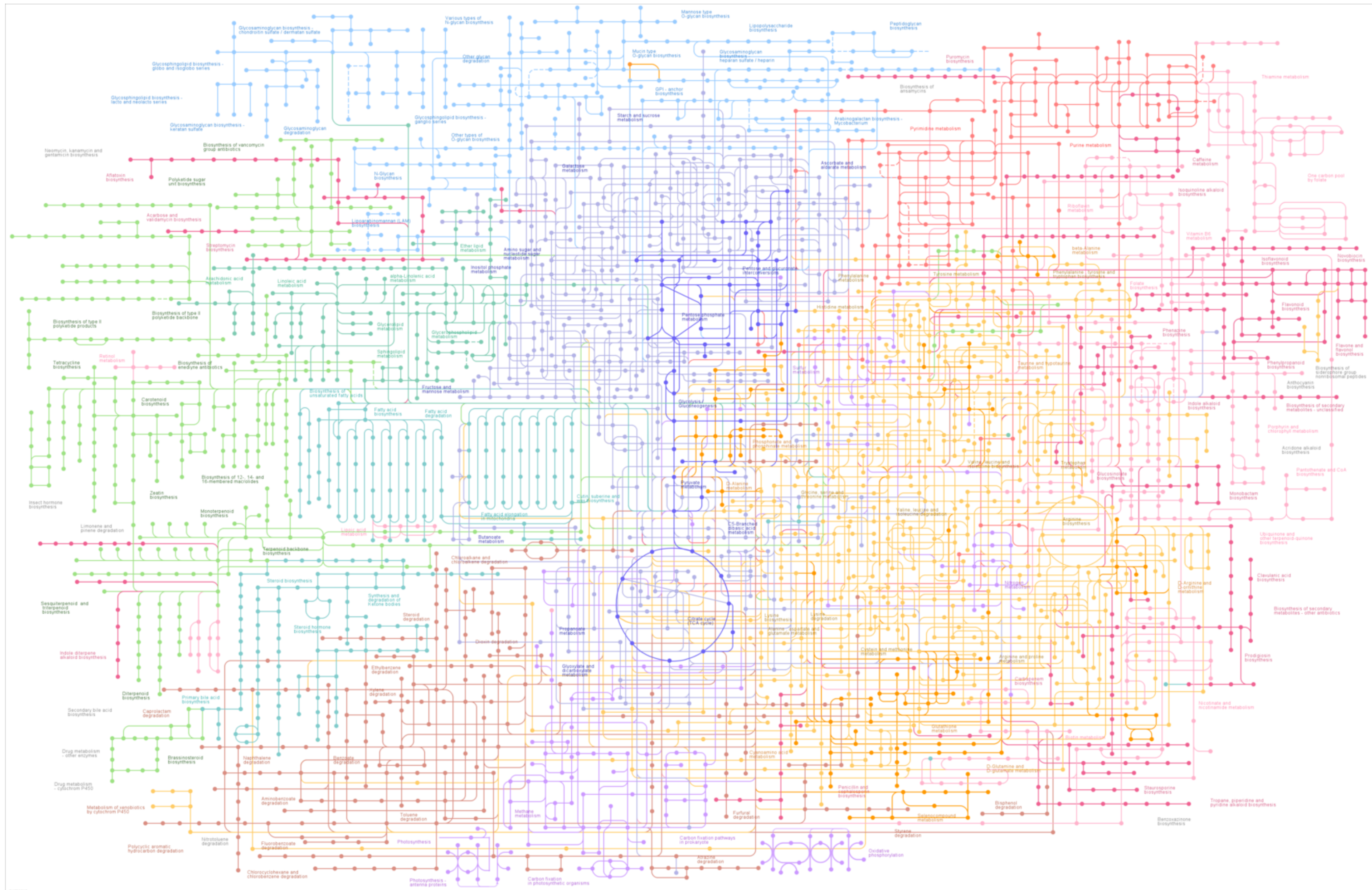
rui.benfeitas@scilifelab.se

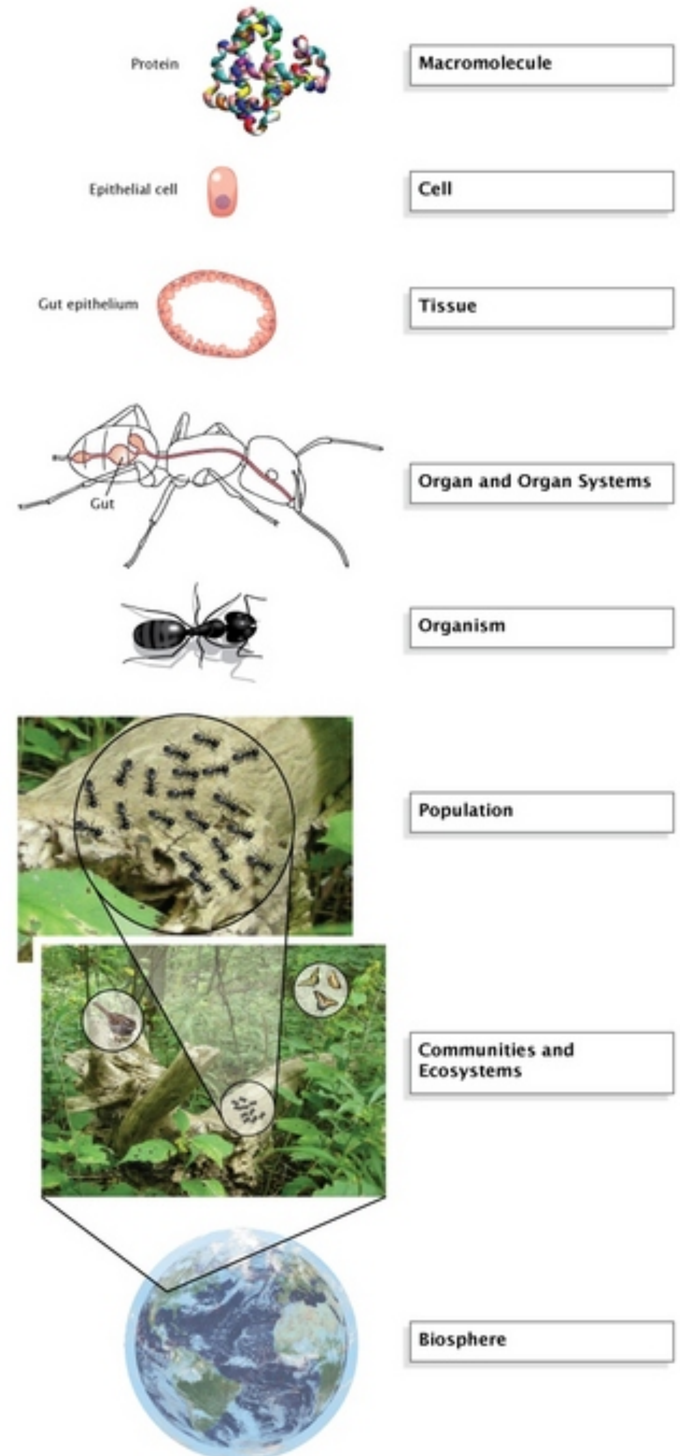
Overview

1. Introduction to network analysis
2. Terminology
3. Network construction
4. Key network properties
5. Community analysis
6. Visualization
7. Workshop

Introduction

Biological complexity under attack

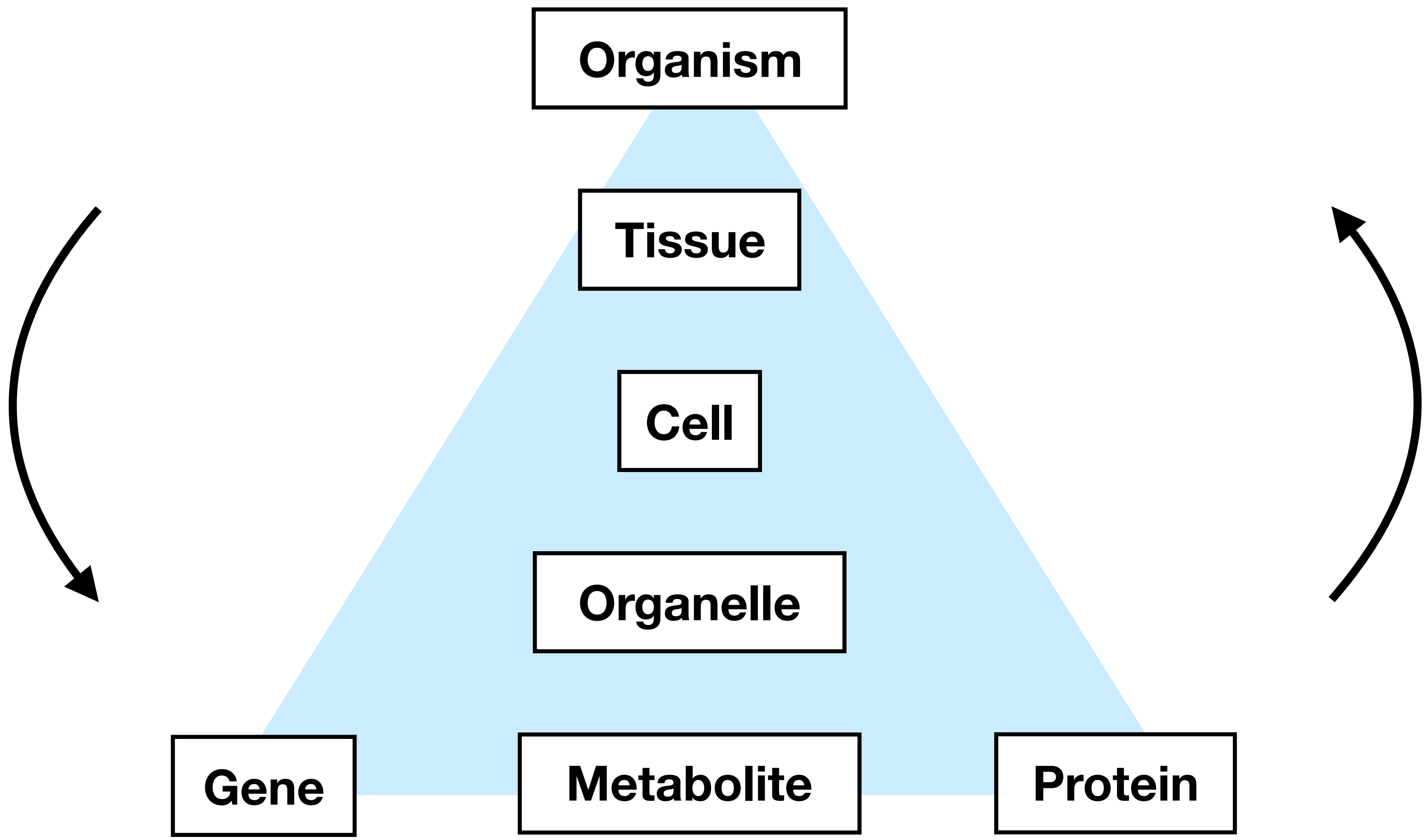




Lobo 2008
genome.jp/kegg/

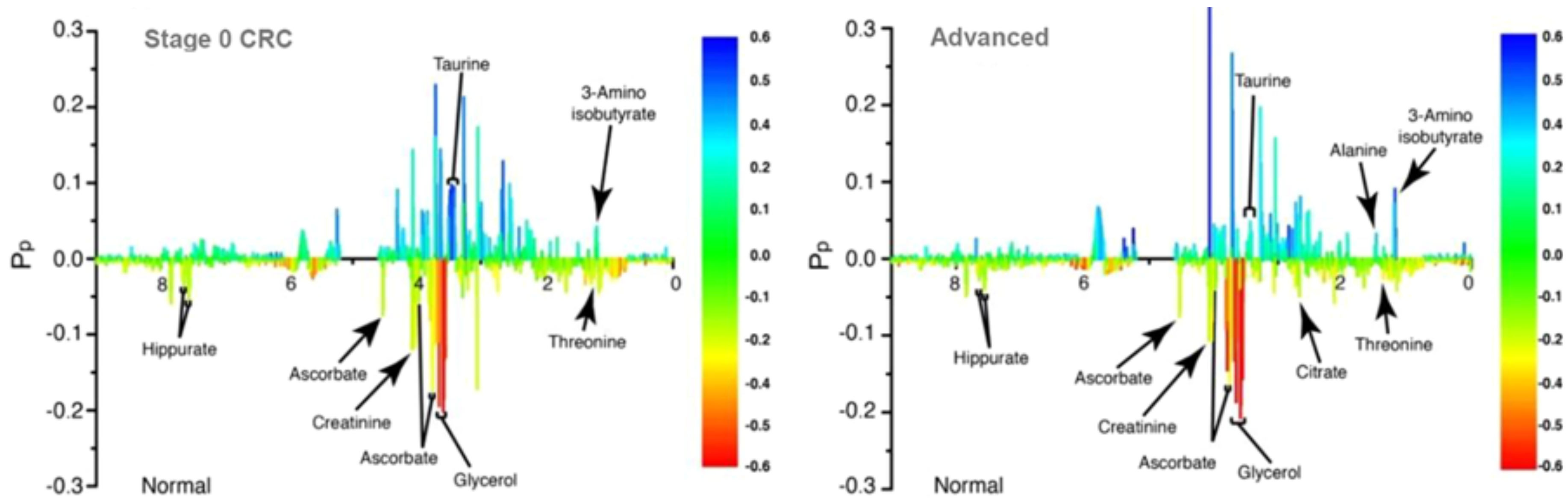
How to tackle biological complexity?

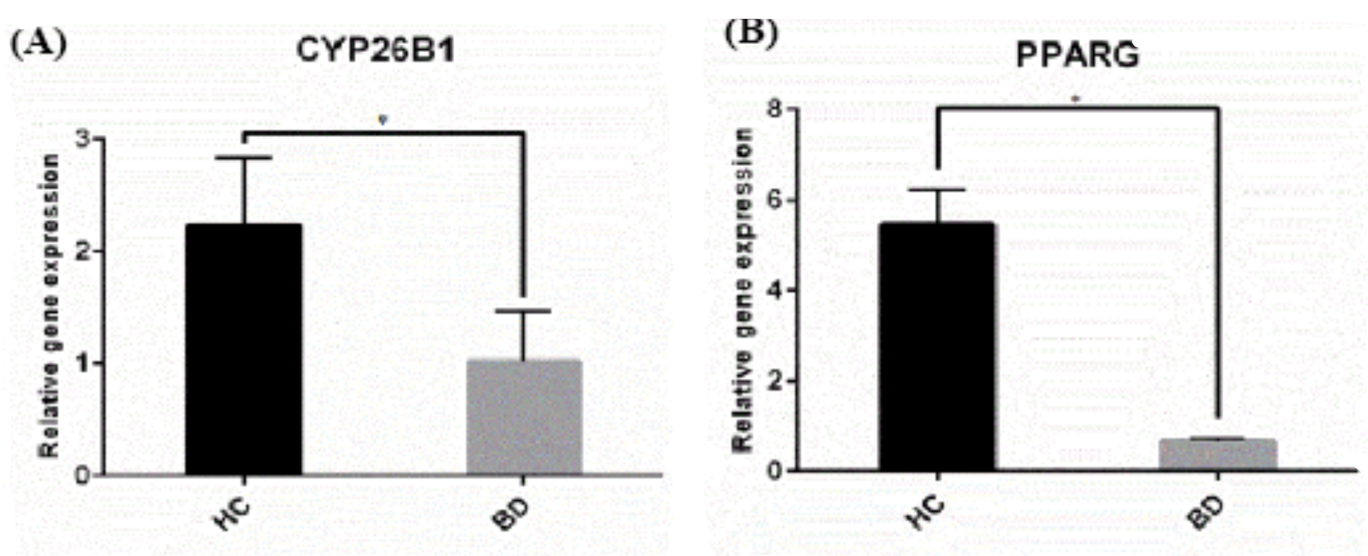
Moving from reductionist approaches towards global characterisations



How to tackle biological complexity?

Moving beyond reductionist approaches

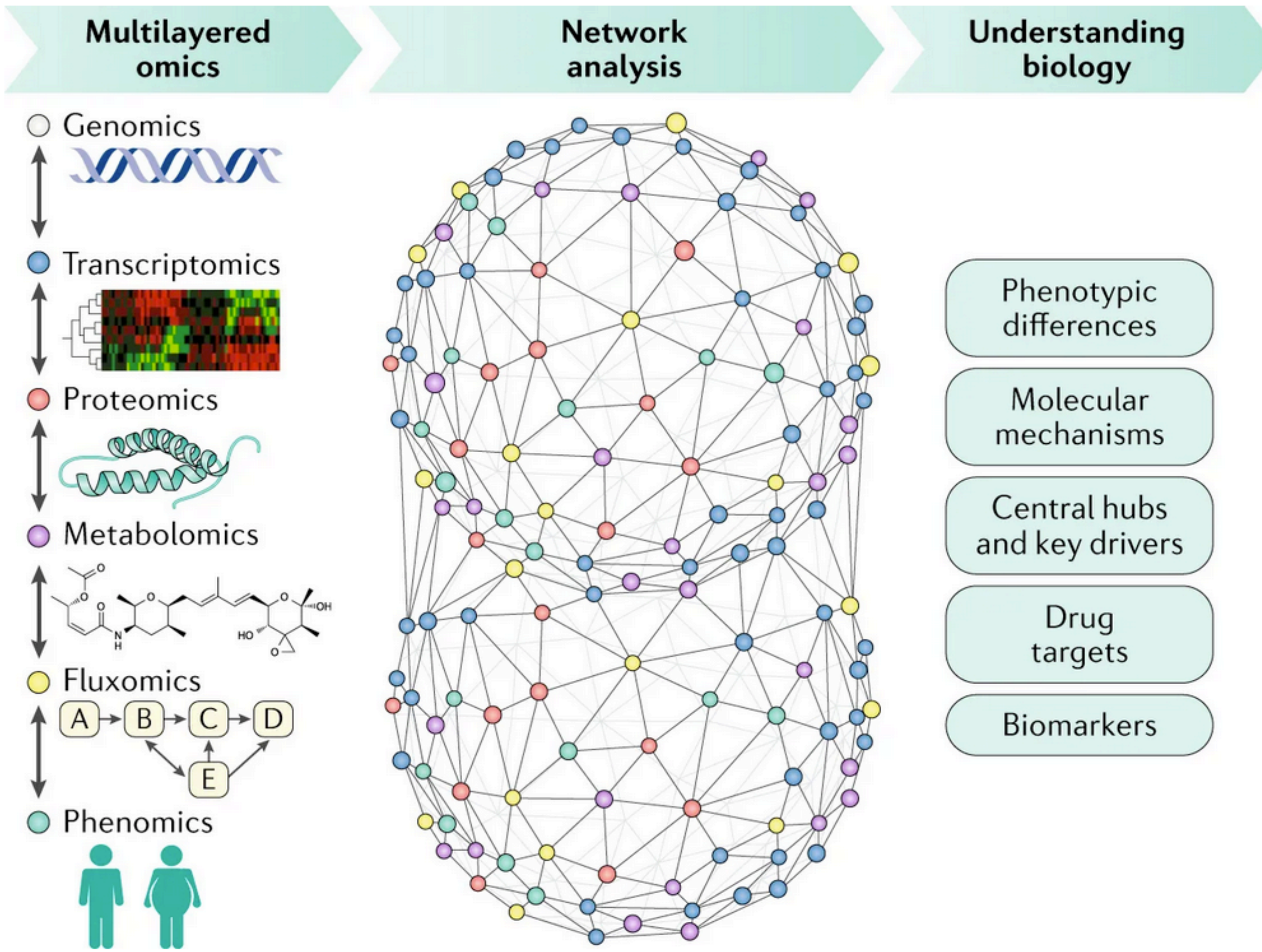




How to tackle biological complexity?

Integrative approaches, and global patterns

- Feature association
- Modeling
- Network analysis



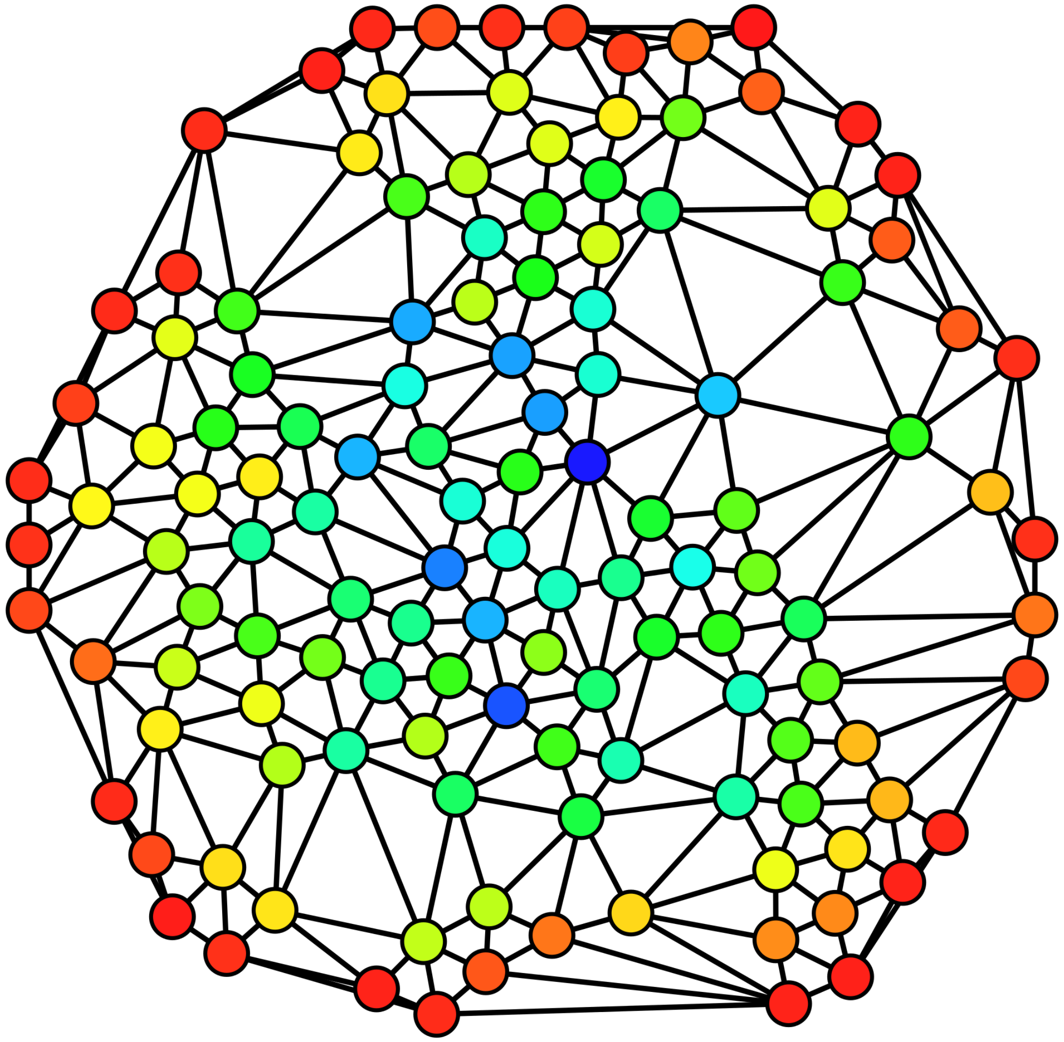
What are networks?

Networks are representations of complex systems

Permit defining and studying global properties of interacting components

Give us insight not easily achieved by other approaches:

- Comprehensive
- Coordinated



What are networks?

?

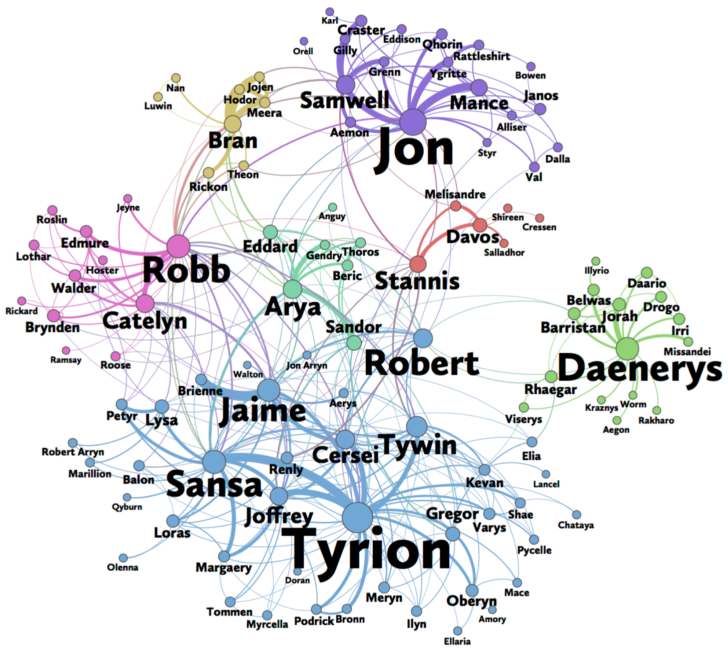
What are networks?

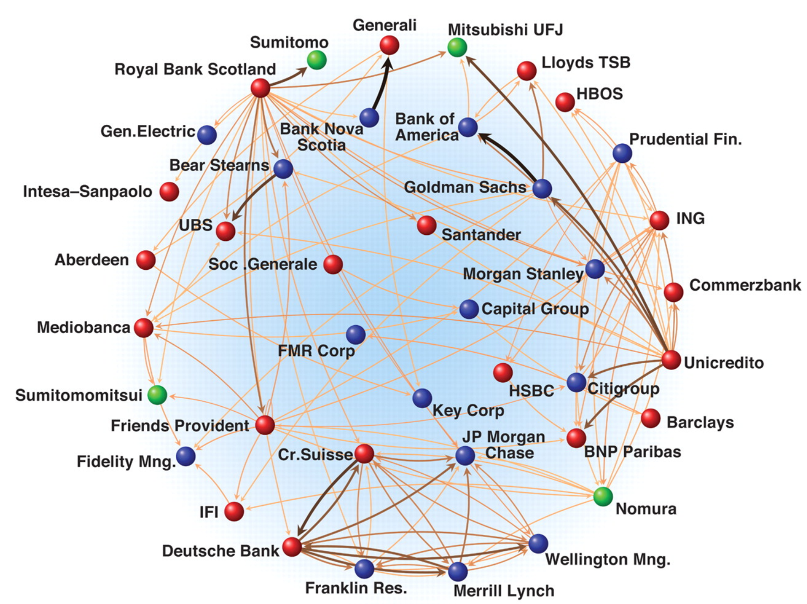
Social

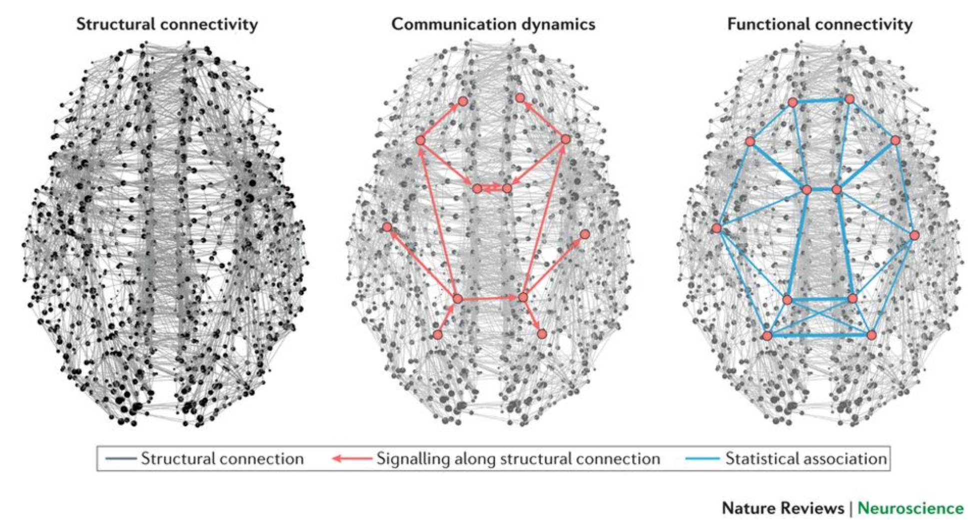
Economic

Communication

Neuronal







What are biological networks?

?

What are biological networks?

Protein - Protein interaction (PPI) networks

Transcription-factor regulatory networks

Gene - gene co-expression networks

Signal transduction networks

Transcription-Factor Regulatory networks

Drug-disease association networks

Aim
Functional characterisations

What are biological networks?

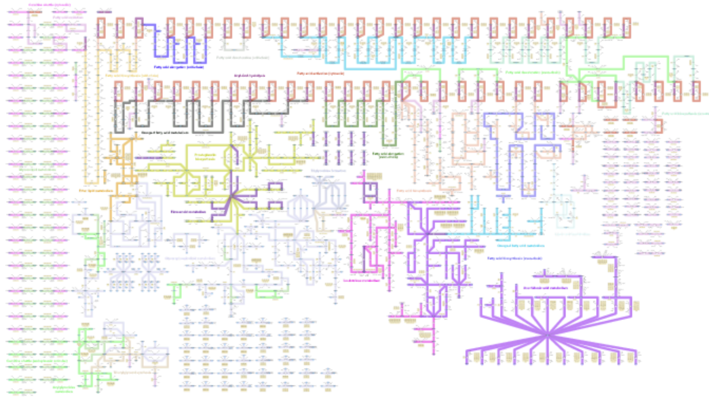
Metabolite - Enzyme - Signal - Genes (GEMs)

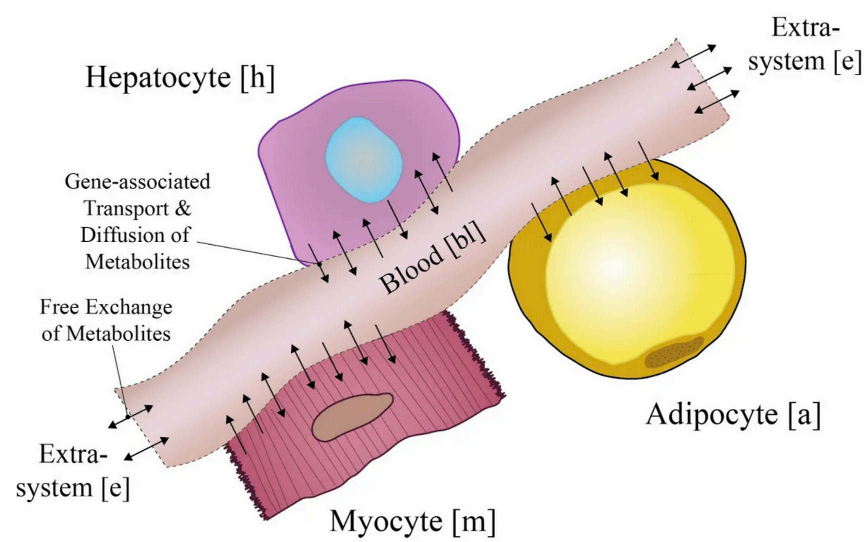
Multi-tissue networks

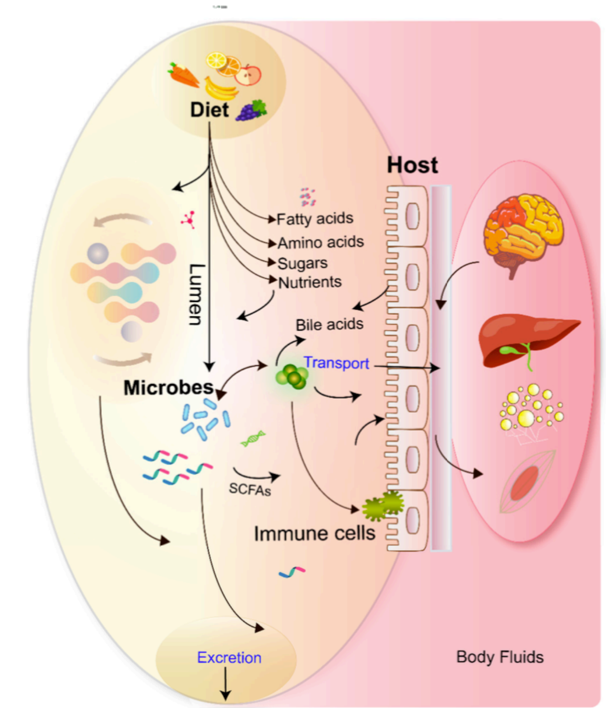
Multi-species networks

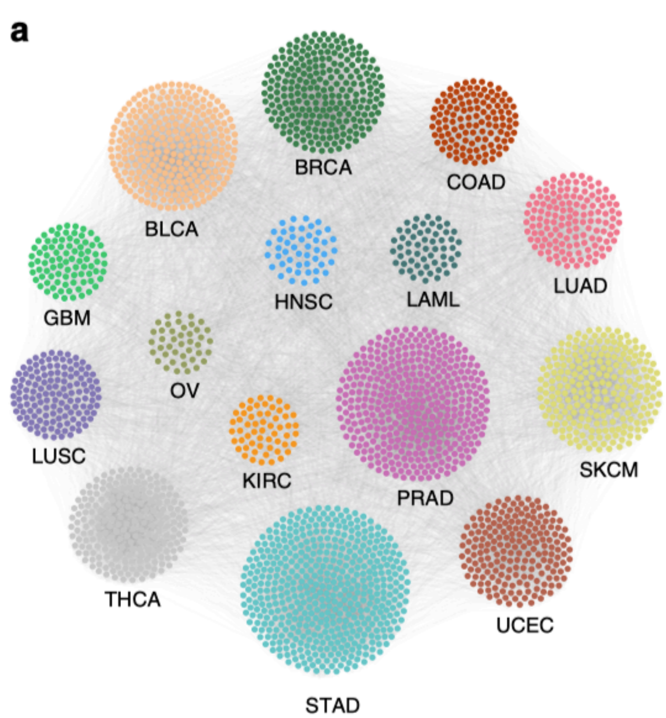
Cancer-disease networks

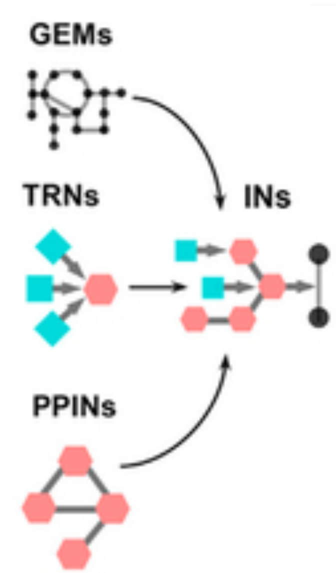
Integrated networks







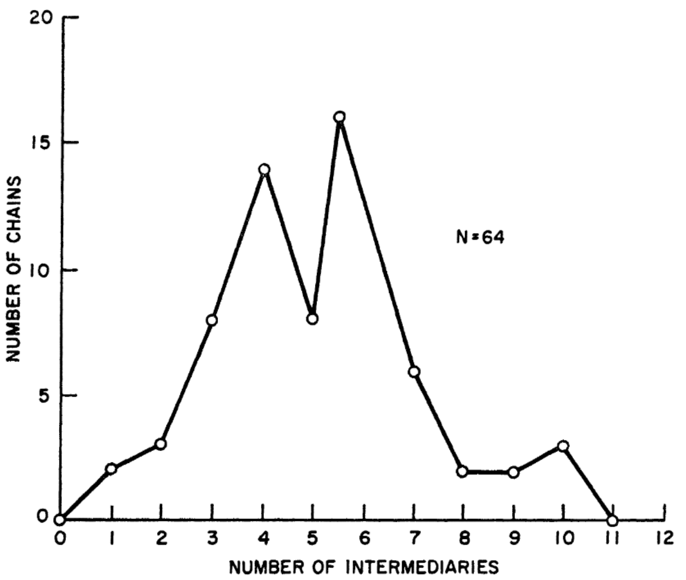




Small world

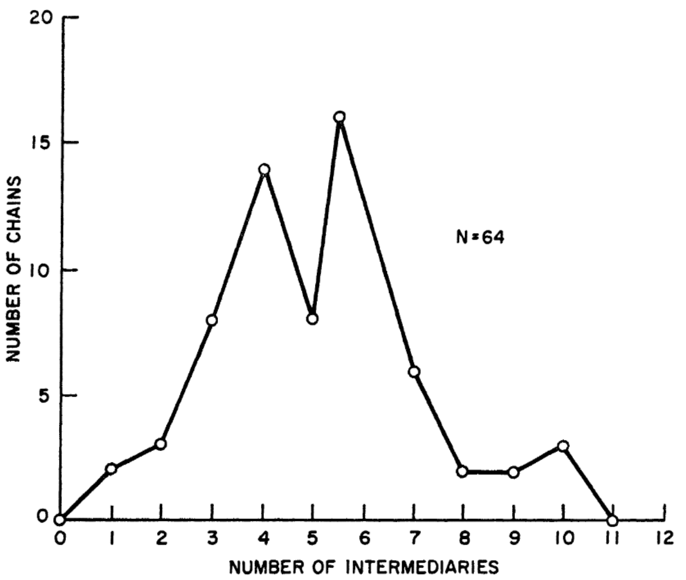
Stanley Milgram (1967) - 6 degrees

- 64 / 296 letters successful

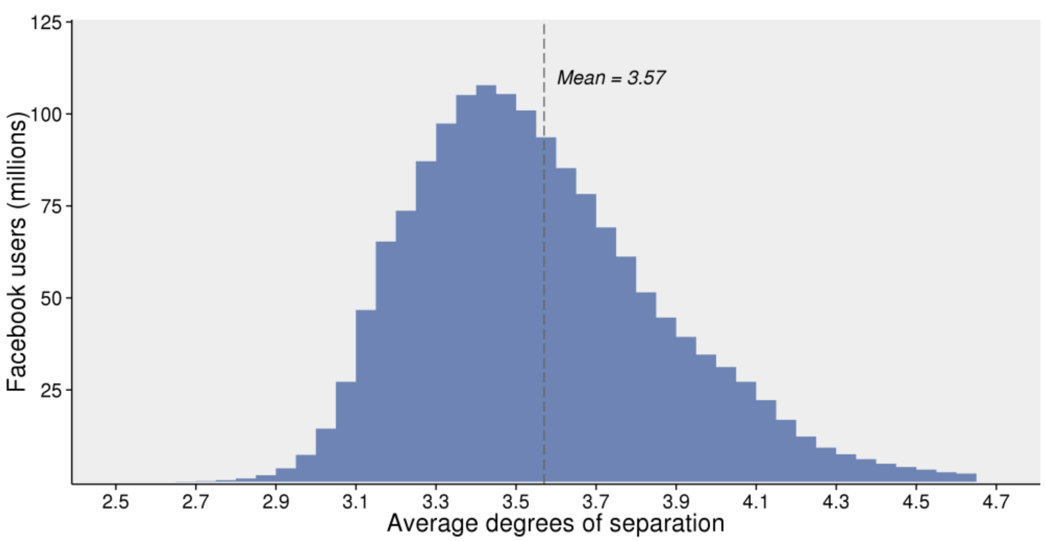


Small world

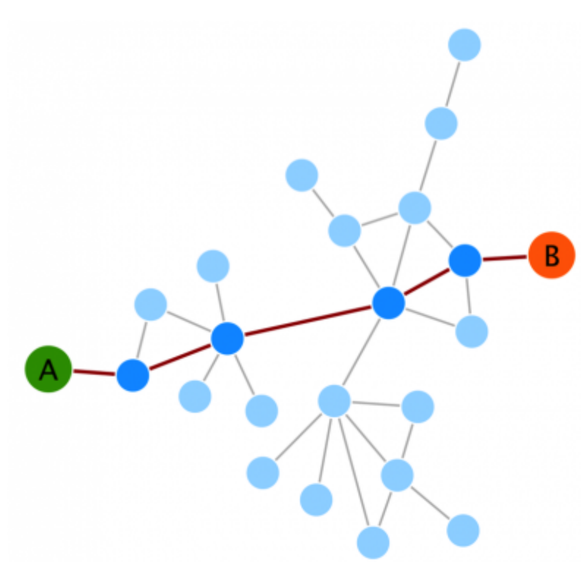
Stanley Milgram (1967) - 6 degrees



Backstrom et al. (2016) - 3.6 degrees



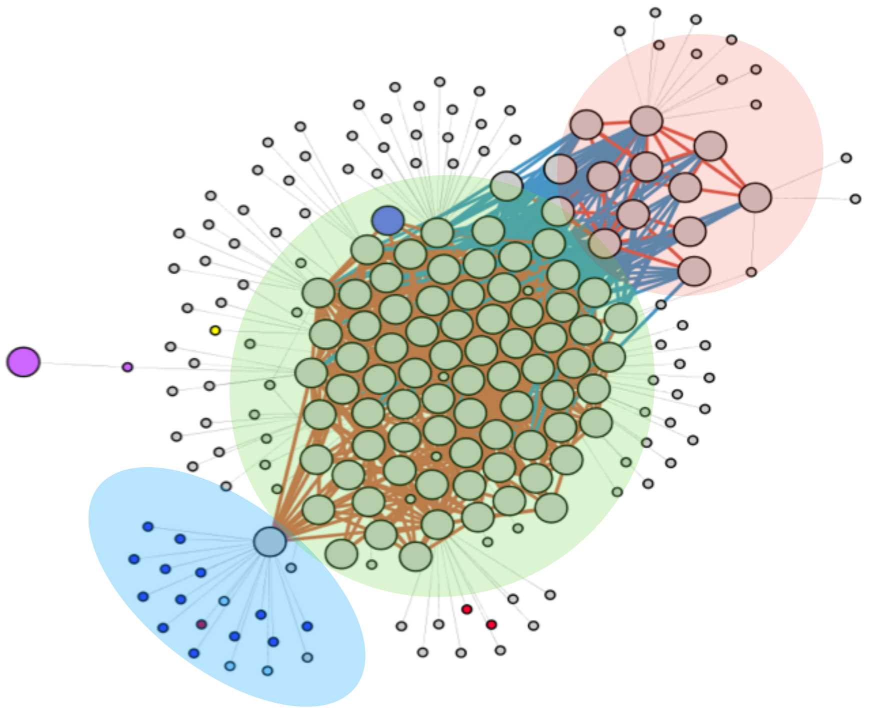
Biological Networks



Why look at network topology?

Use networked systems to identify:

- Identify global / local patterns
- Identify functional properties
- Make predictions



Examples:

- How associated are the elements of my network?
- What are its first-hand associated elements?
- What are the groups of closely-associated elements in my network?
- What are their functional relationships?
- What are the "weakest" links in the network?

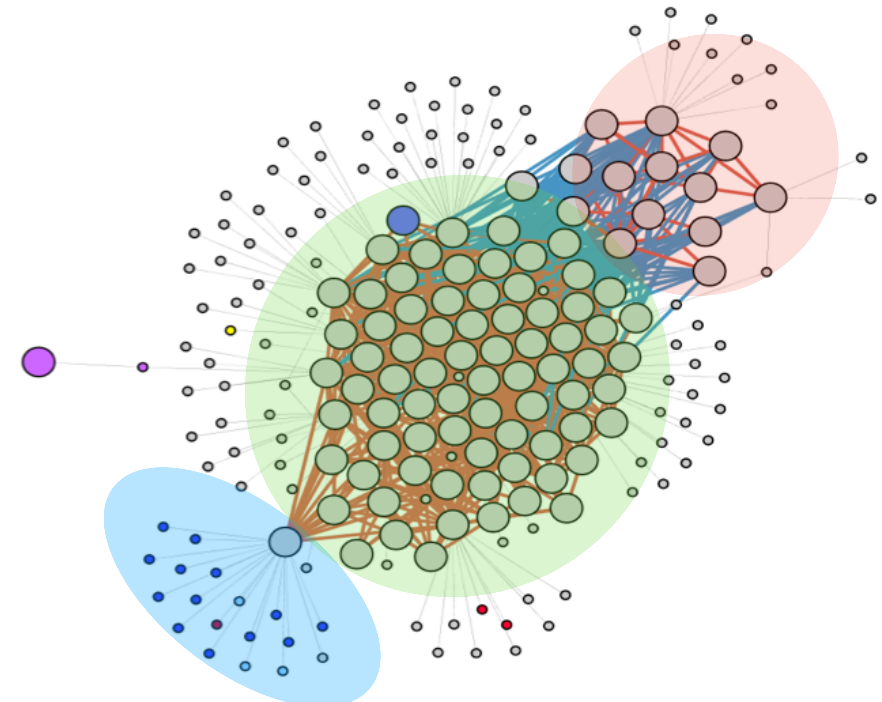
What is my biological network?

Any association matrix may be translated to a network format

Many standard analyses may be employed regardless of data type

…but care must be taken in generating the network

Some of the functional analyses depend on annotation



Limitations

Sample size

False discovery

~requires high throughput

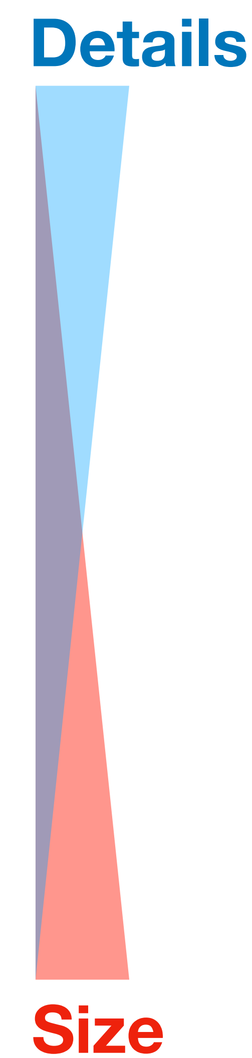
(further discussed in following lectures)

Terminology and initial properties in graph analysis

Motivation

What modeling formalism suits your data and biological question?

	Pros	Cons
Kinetic models	Detailed Quantitative Dynamic / Steady state	Small Requires detailed parameterization
Stoichiometric	Large Semi-quantitative Steady state	Static
Topological	Large Only topological information	No dynamic properties



Adapted from Hartmann et al.

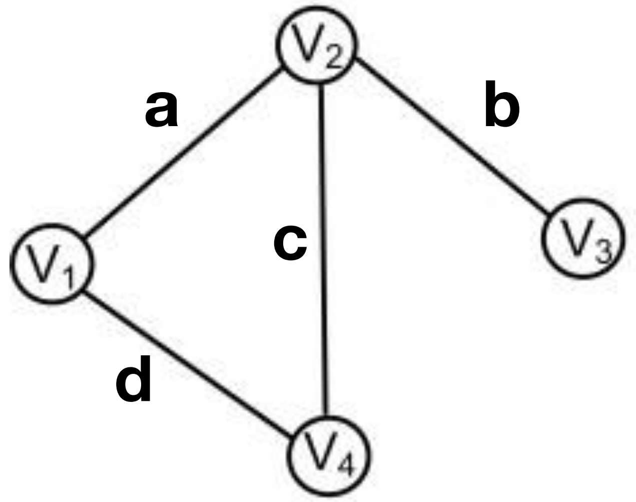
Graphs, nodes, edges

Graph *G* consists of a set of **nodes** (V) interconnected by **edges** (E)

$G = (V, E)$

$V=\{v1,v2,v3,v4\}$

$E=\{a,b,c,d\}$



Nodes sometimes called **vertices**

Two connected nodes are called **neighbours, adjacent,** or **end-nodes**

Simple vs multigraphs

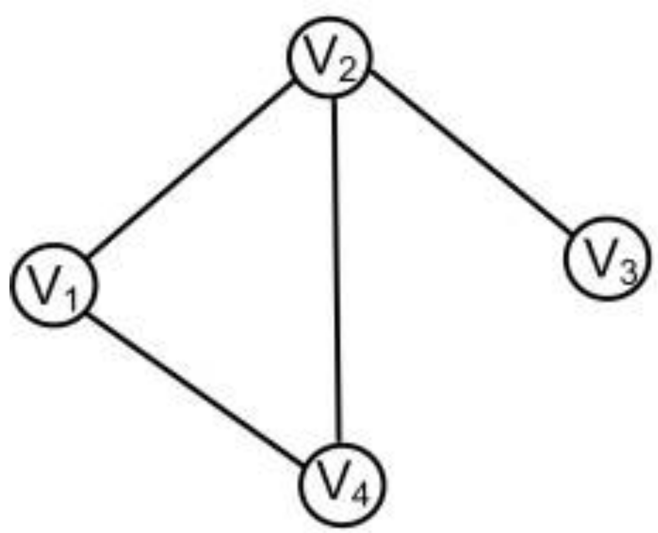
Multigraphs contain parallel edges

Multi-edged connections indicate different properties

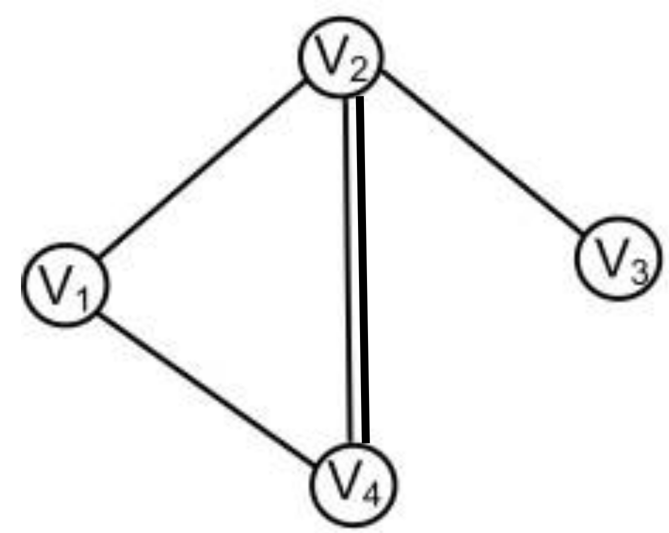
Example: PPI

- Experimental evidence for interaction
- Co-expression

Simple



Multigraphs

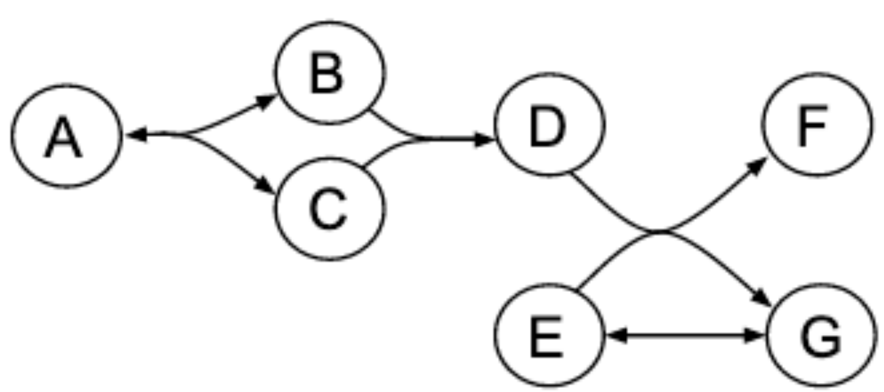


Hypergraphs

Hypergraphs contain edges that connect any number of nodes

Reaction 1: A $\rightarrow$ B + C
Reaction 2: B + C $\rightarrow$ D
Reaction 3: D + E $\rightarrow$ F + G
Reaction 4: E $\rightarrow$ G
Reaction 5: B + C $\rightarrow$ A
Reaction 6: G $\rightarrow$ E

(a) Reaction network



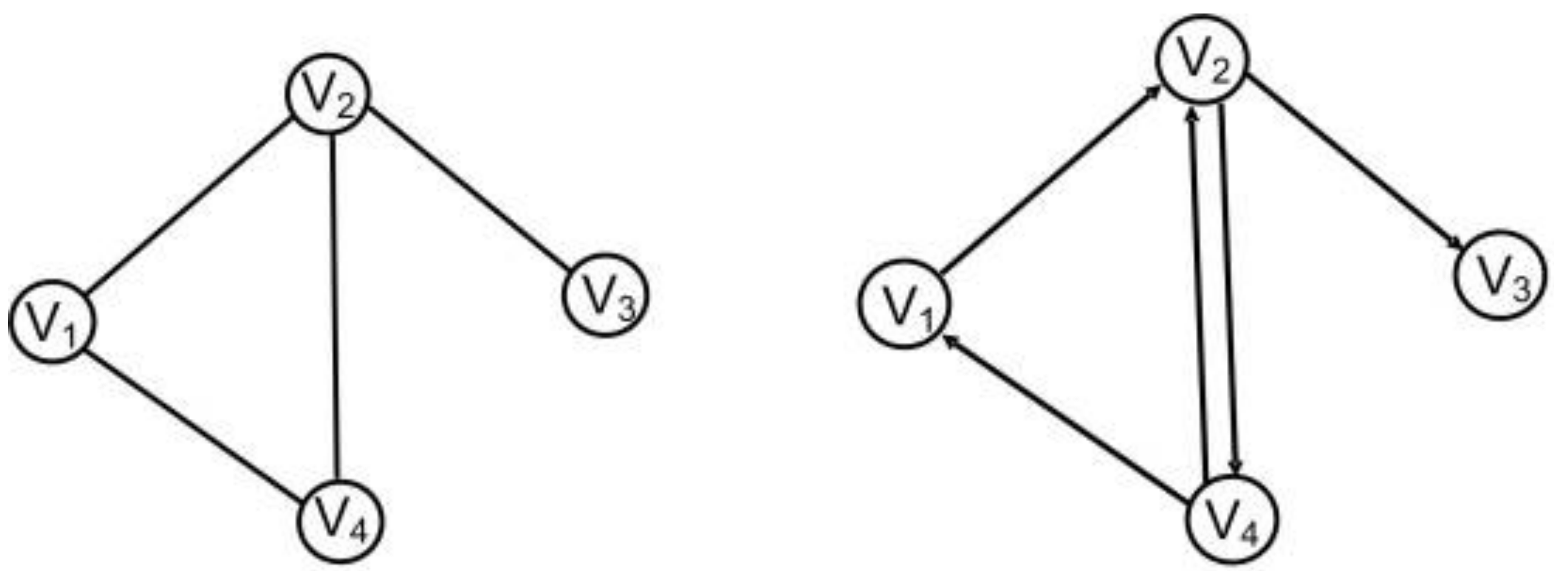
Directed vs undirected graphs

Directed graphs is given by an ordered triple

G=(V,E,f)

f : function mapping an element in ***E*** to the ordered pair of vertices in ***V***

E={ (v1, v2) , (v2, v3), (v2, v4), (v4, v1), (v4, v2)}

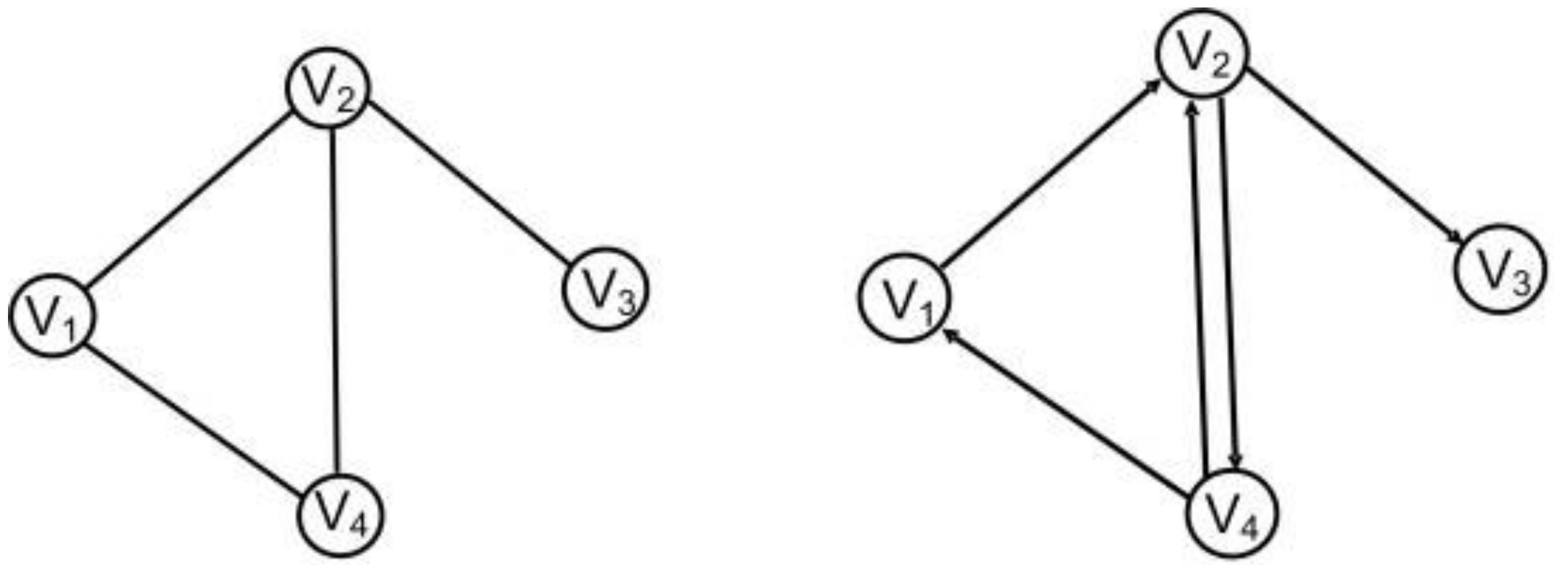


Directed vs undirected graphs

Examples:

- **Undirected graphs:** co-expression networks

- **Directed graphs:** metabolic networks

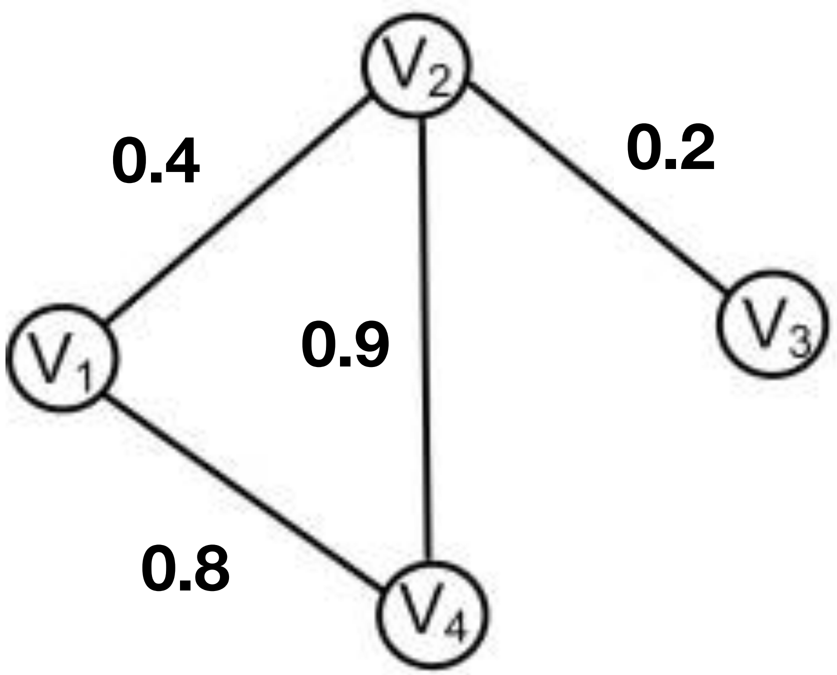


Weighted vs unweighted graphs

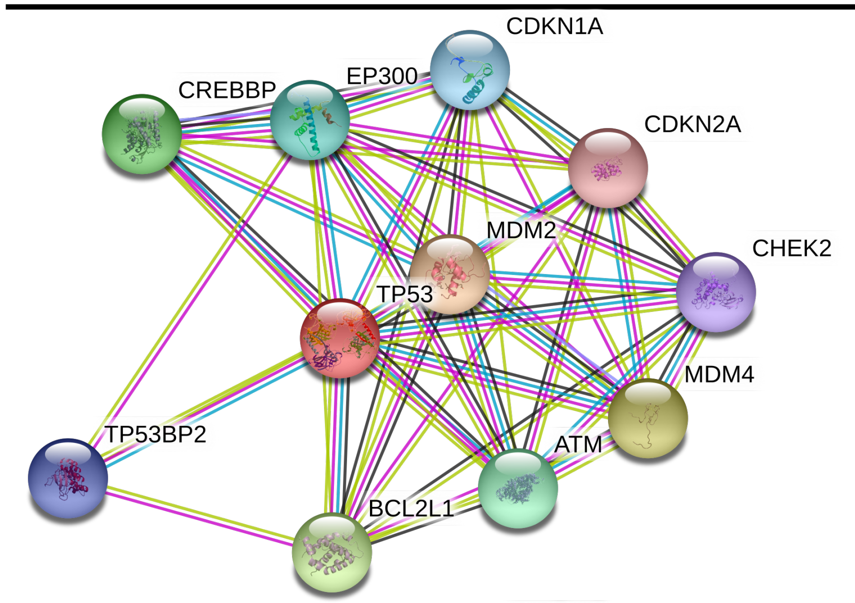
Weighted edges associate a value to an interaction between two nodes. Usually give the confidence in the interaction.

Negative weights?

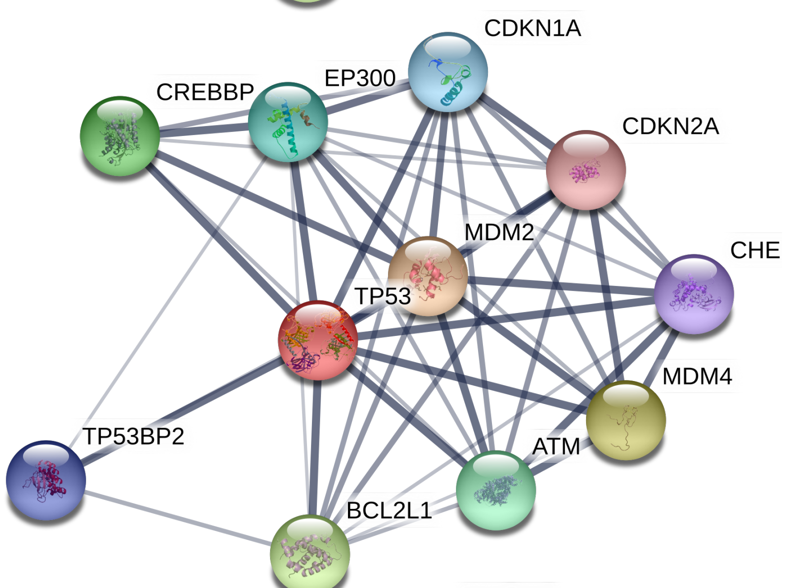
E.g. weighted co-expression networks



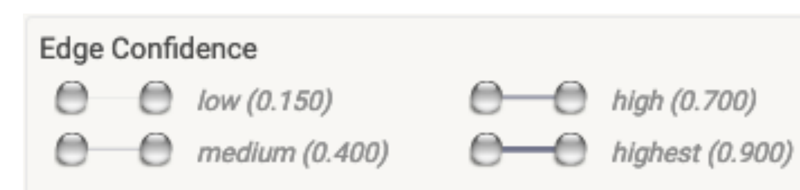
STRING-db.org: TP53

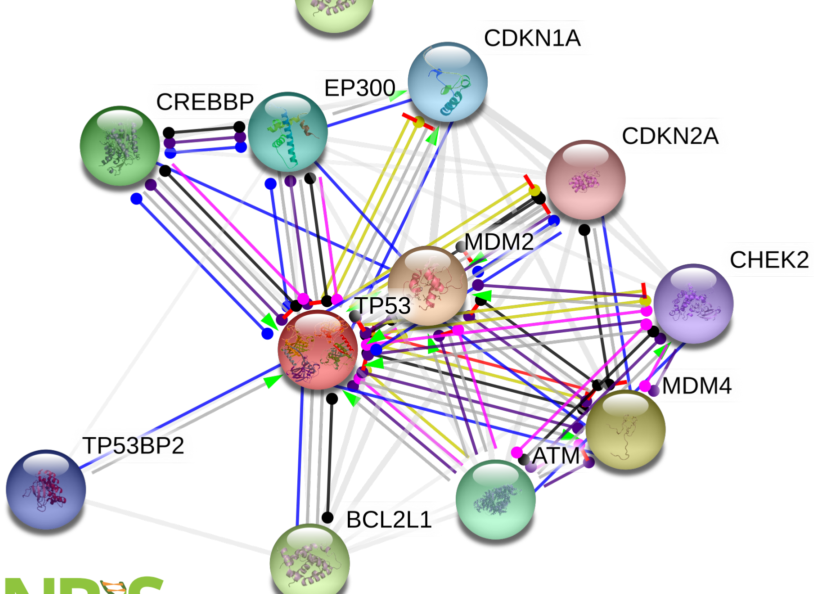


Multi-edged



Weighted multi-edged





Multi-edged directed

Bipartite graphs

A graph

$$G=(V,E)$$

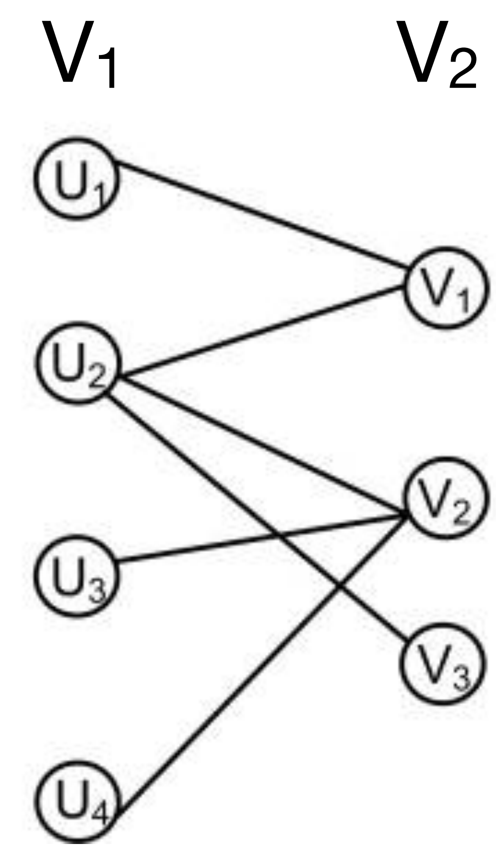
may be partitioned into two sets of nodes (V_1, V_2) such that

$$u \in V_1 \text{ and } v \in V_2$$

or

$$u \in V_2 \text{ and } v \in V_1$$

All e_i has end-nodes in V_1, V_2



A **subgraph** of G will thus be given by

$$G_1 = (V_1, E_1)$$

Finding associations in V_1

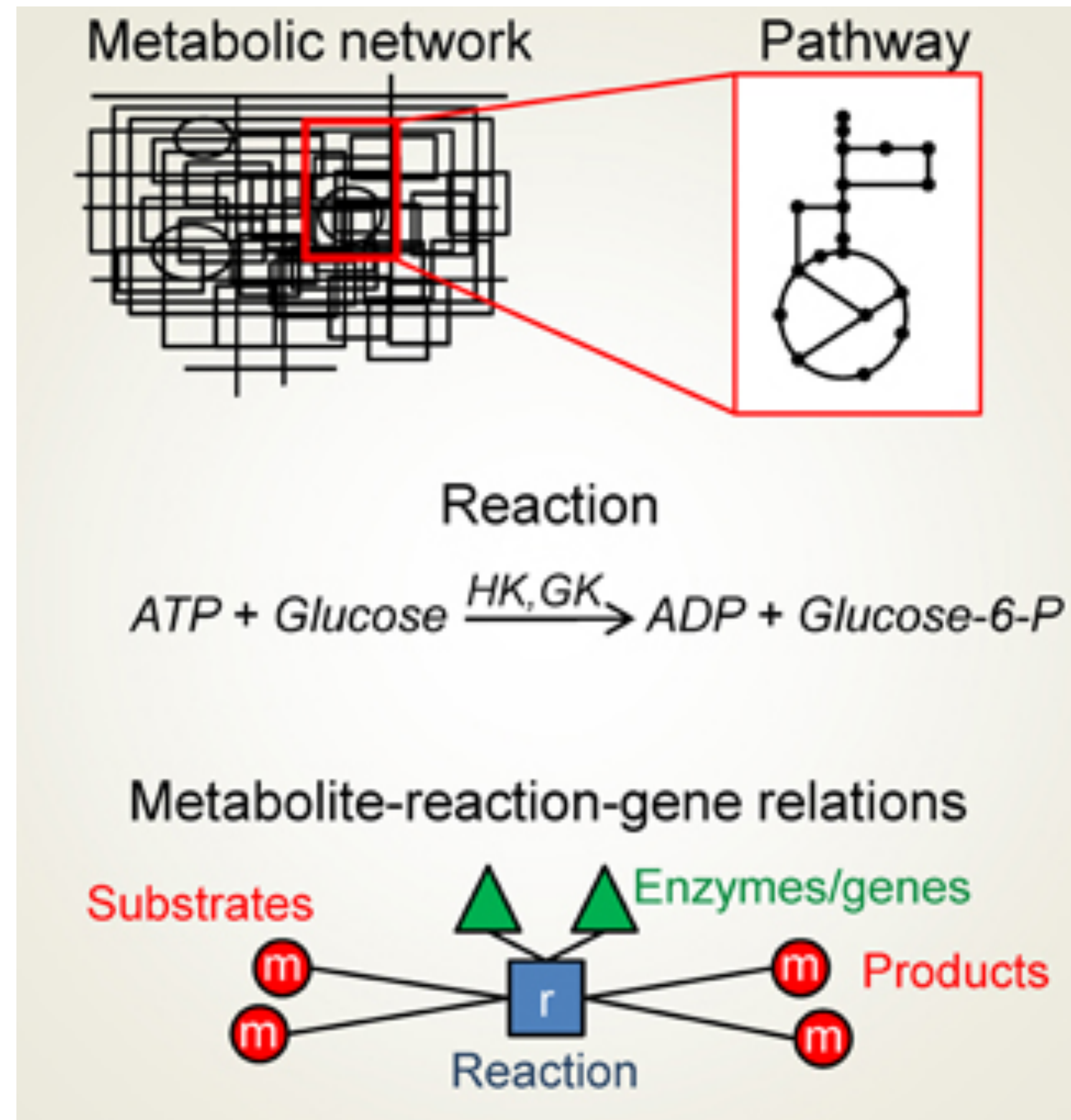
Bipartite and *k*-partite graphs

Example of bipartite graph:

Enzyme - Reaction

Metabolite - reaction - enzyme

k-partite graphs display *k*-types of nodes

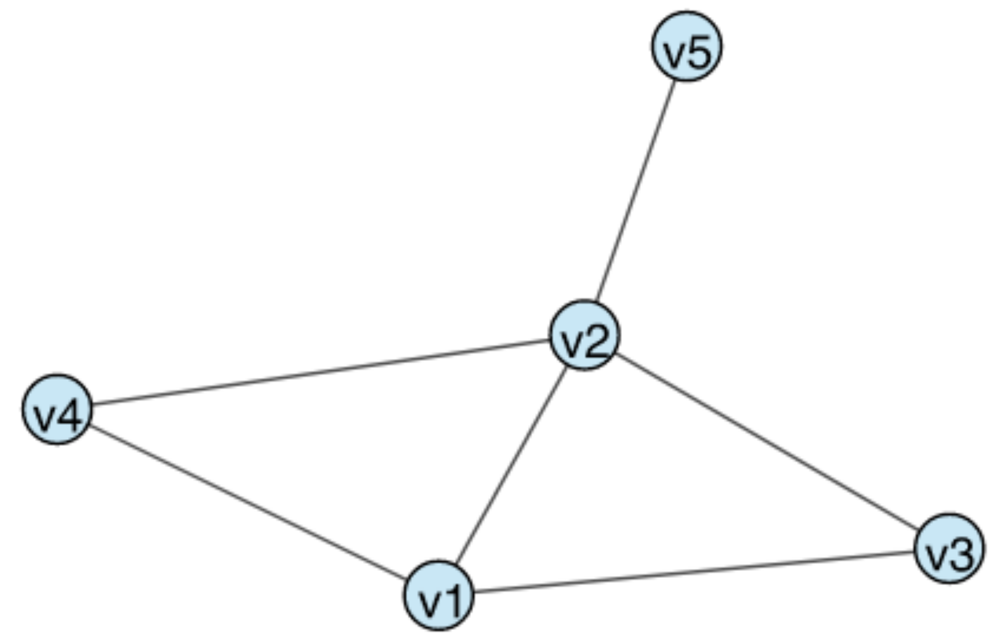


Väremo 2013

Adjacency matrix (undirected graphs)

Vertex association (undirected network)

n1	n2
v1	v2
v1	v4
v2	v4
v2	v3
v2	v5
v1	v3



Adjacency matrix is symmetric

	v1	v2	v3	v4	v5
v1	0	1	1	1	0
v2	1	0	1	1	1
v3	1	1	0	0	0
v4	1	1	0	0	0
v5	0	1	0	0	0

Upper triangular

Diagonal

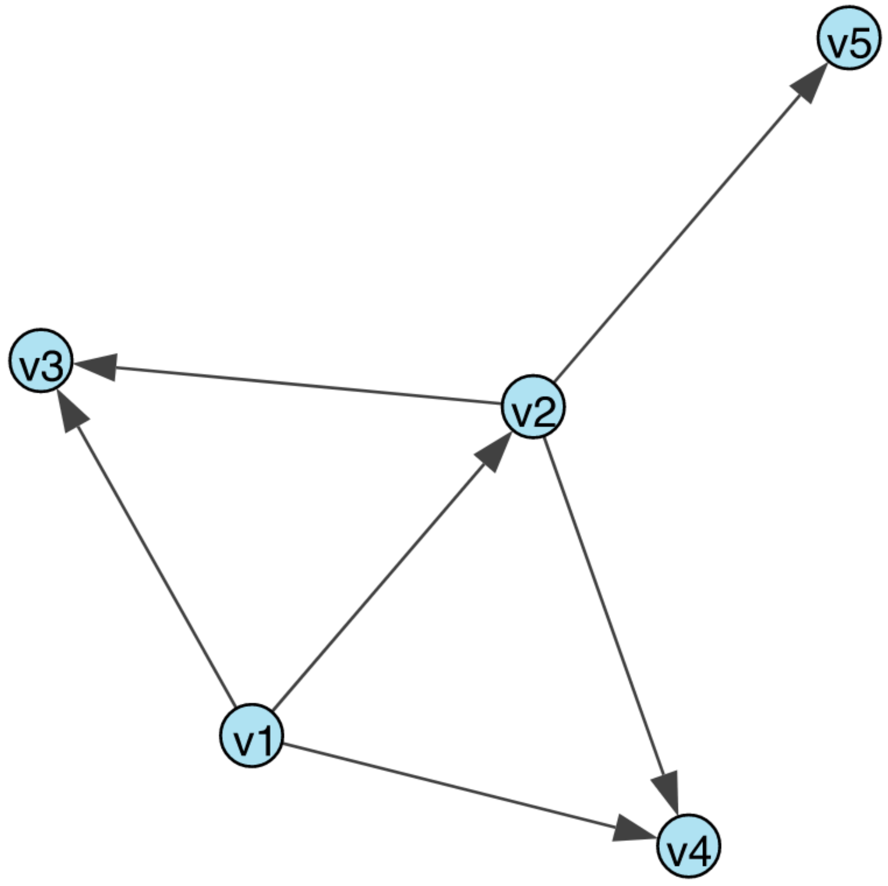
Lower triangular

Adjacency matrix (directed graphs)

Target

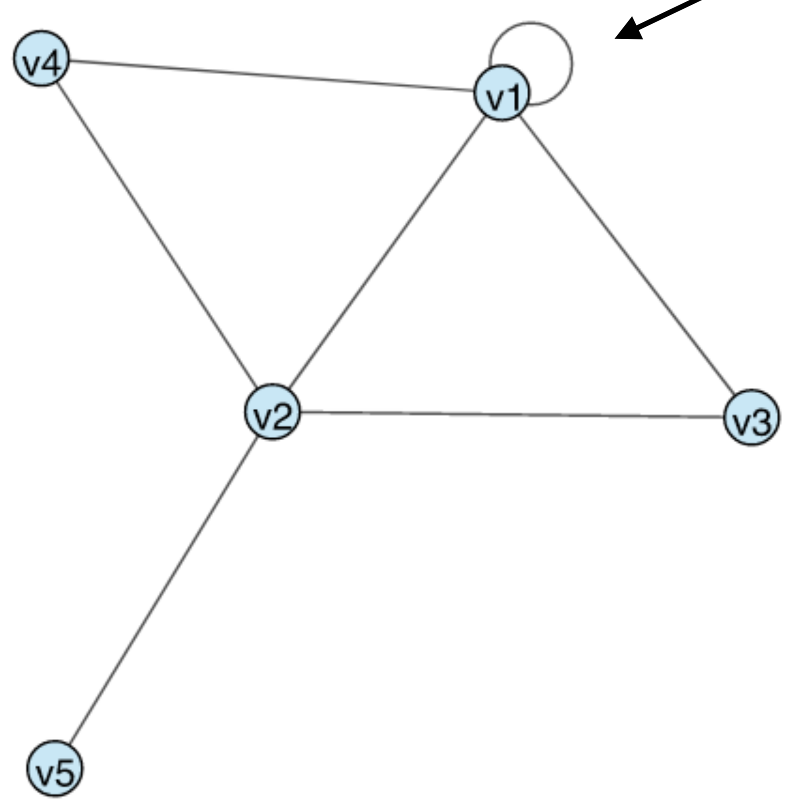
Source

	v1	v2	v3	v4	v5
v1	0	1	1	1	0
v2	0	0	1	1	1
v3	0	0	0	0	0
v4	0	0	0	0	0
v5	0	0	0	0	0



Graphs may contain self-loops

Loop



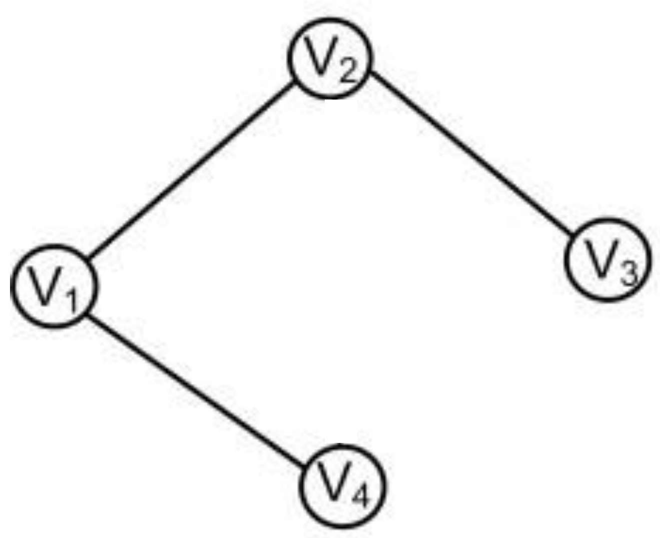
	v1	v2	v3	v4	v5
v1	**1**	1	1	1	0
v2	1	0	1	1	1
v3	1	1	0	0	0
v4	1	1	0	0	0
v5	0	1	0	0	0

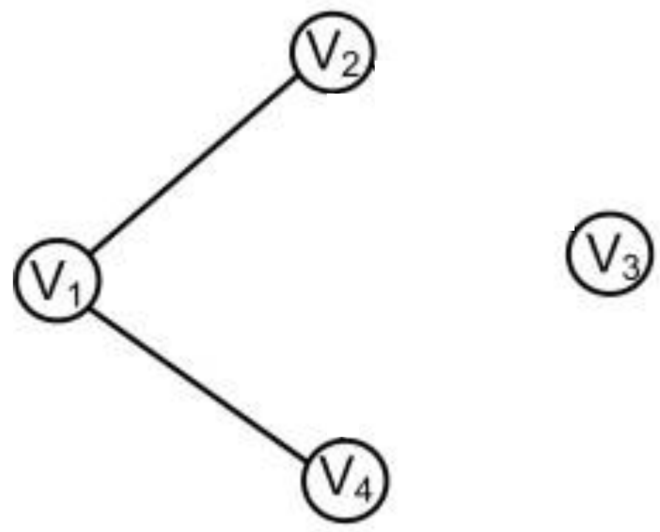
Examples of **self-loops** are auto-regulatory mechanisms in Transcription Factor regulatory networks

Connected vs disconnected networks

Connected network: there is at least 1 path connecting all nodes in a network

Disconnected network: some of the nodes are unreachable

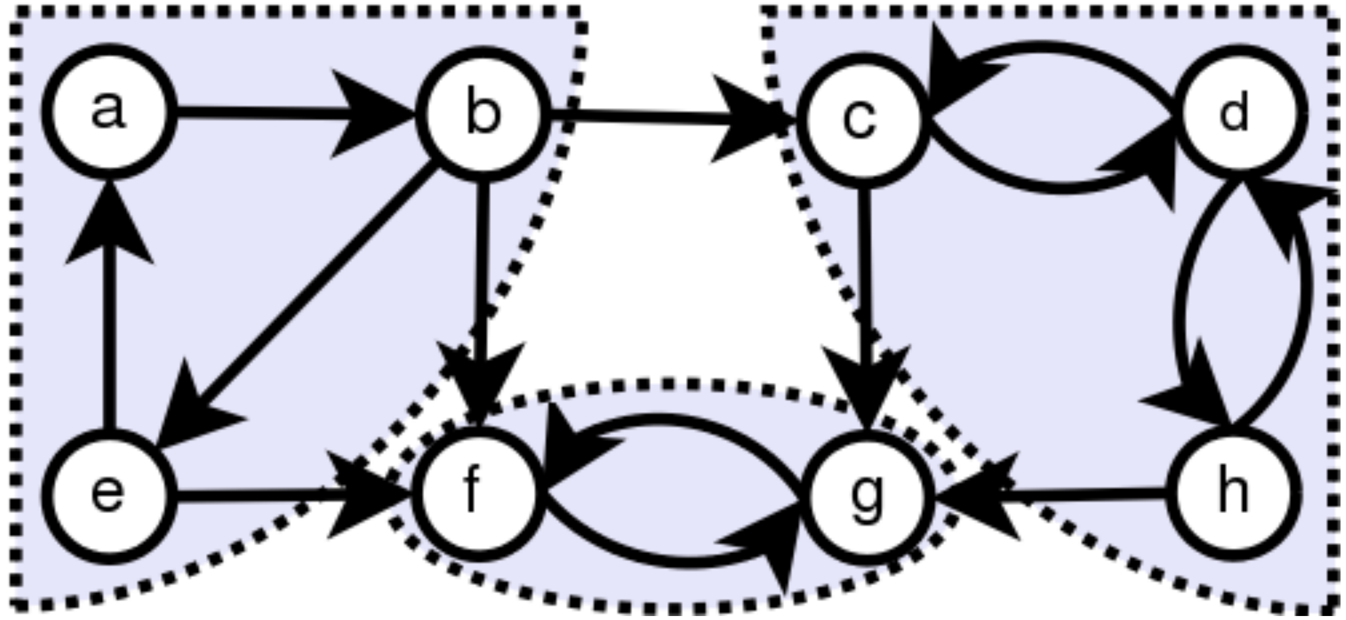




Connected components

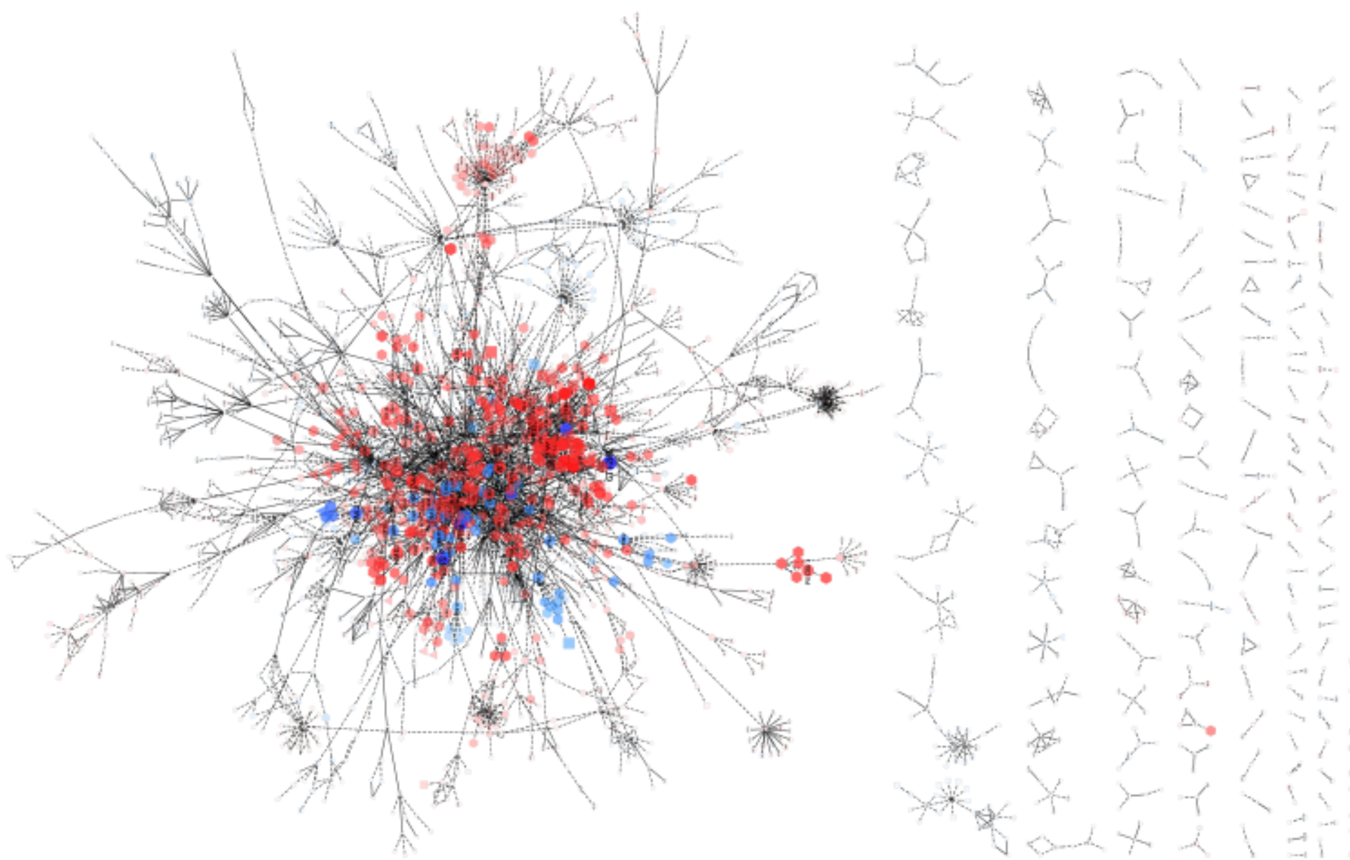
Connected components are those where all nodes of each subgraph are connected.

Weak vs **strong** components



Connected components

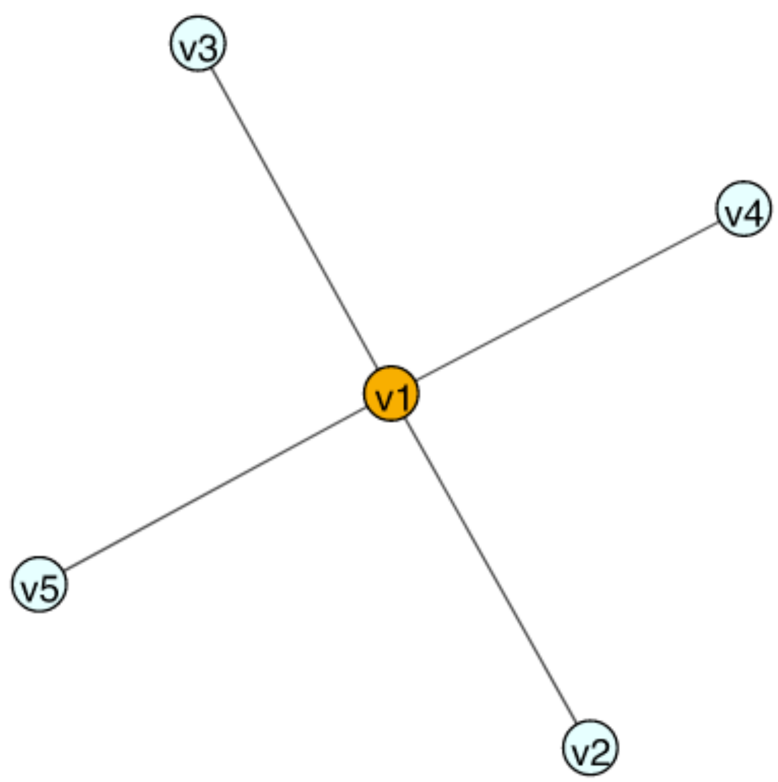
In biological networks, often the most insightful properties come from the **largest connected component**

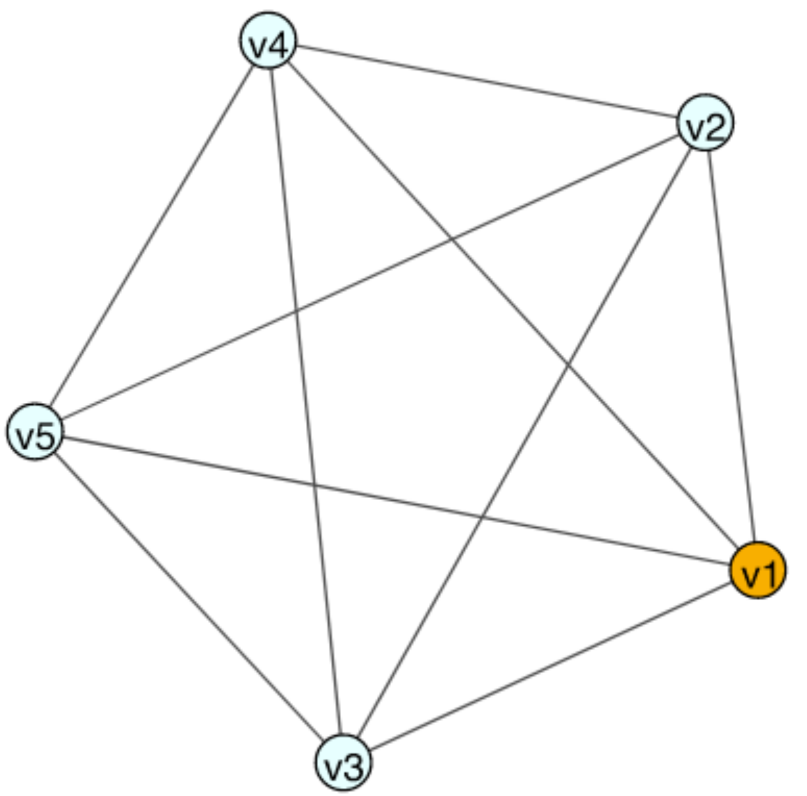


Stars vs cliques

A **star** forms a **complete <u>bipartite</u> subgraph**

A **clique** is a subgraph where all nodes are adjacent (i.e. **complete graph**)

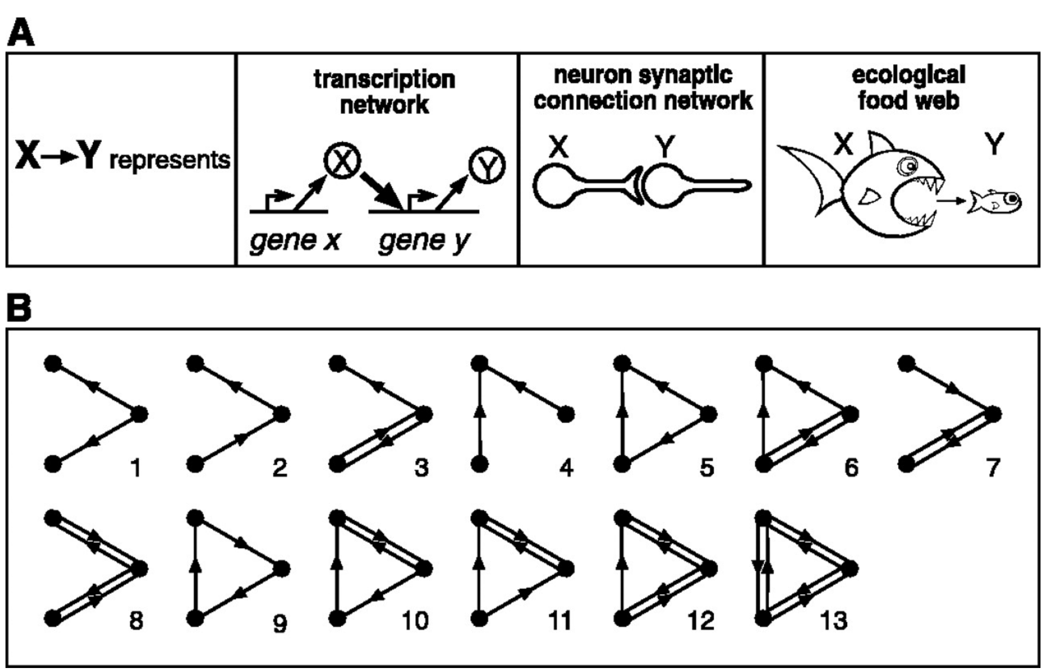


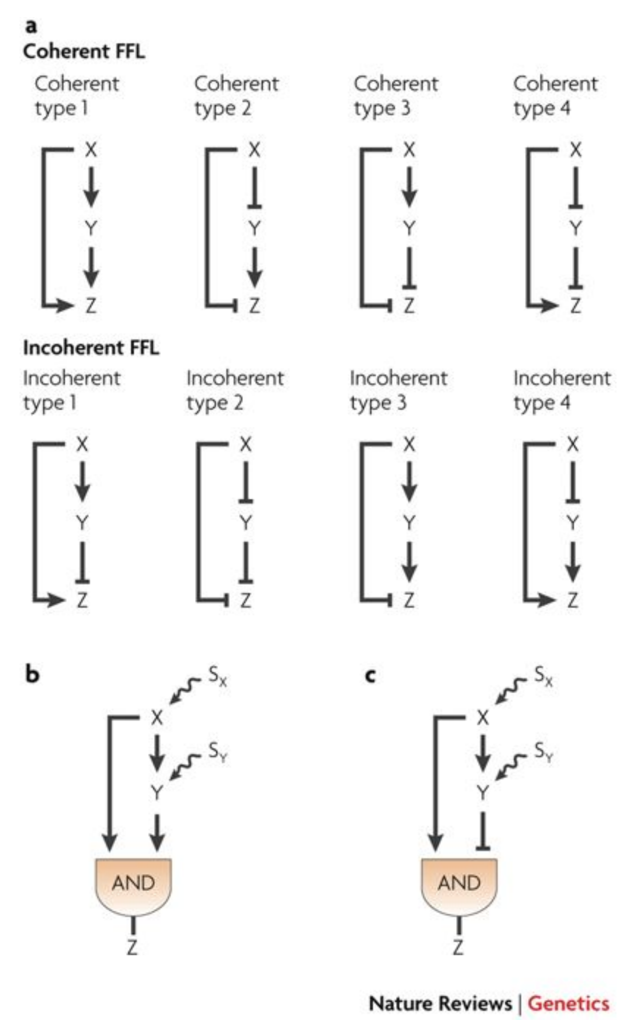


Motifs

Subgraphs are characterised by different motifs

Exploring prevalent motifs may allow us to understand the evolutionary advantage of a given architecture





Milo 2002
Alon 2007

Additional reading

- Network Science - A fascinating textbook on graph theory and network analysis.
- Communication dynamics in complex brain networks - Interesting discussion about whether and how network topology may be applied to study the brain networks.
- A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models - General review and discussion on methods to use in genome-scale metabolic models.
- Analysis of Biological Networks - General introduction into biological networks, network notation, and analysis, including graph theory.
- Multi-omics approaches to disease - Introduction to how integrative approaches may be applied in disease

Additional references displayed as hyperlinks in each slide.