

# Biological network analysis

Rui Benfeitas

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

[rui.benfeitas@scilifelab.se](mailto:rui.benfeitas@scilifelab.se)

metabolic  
**ATLAS**



**NBIS**



SciLifeLab



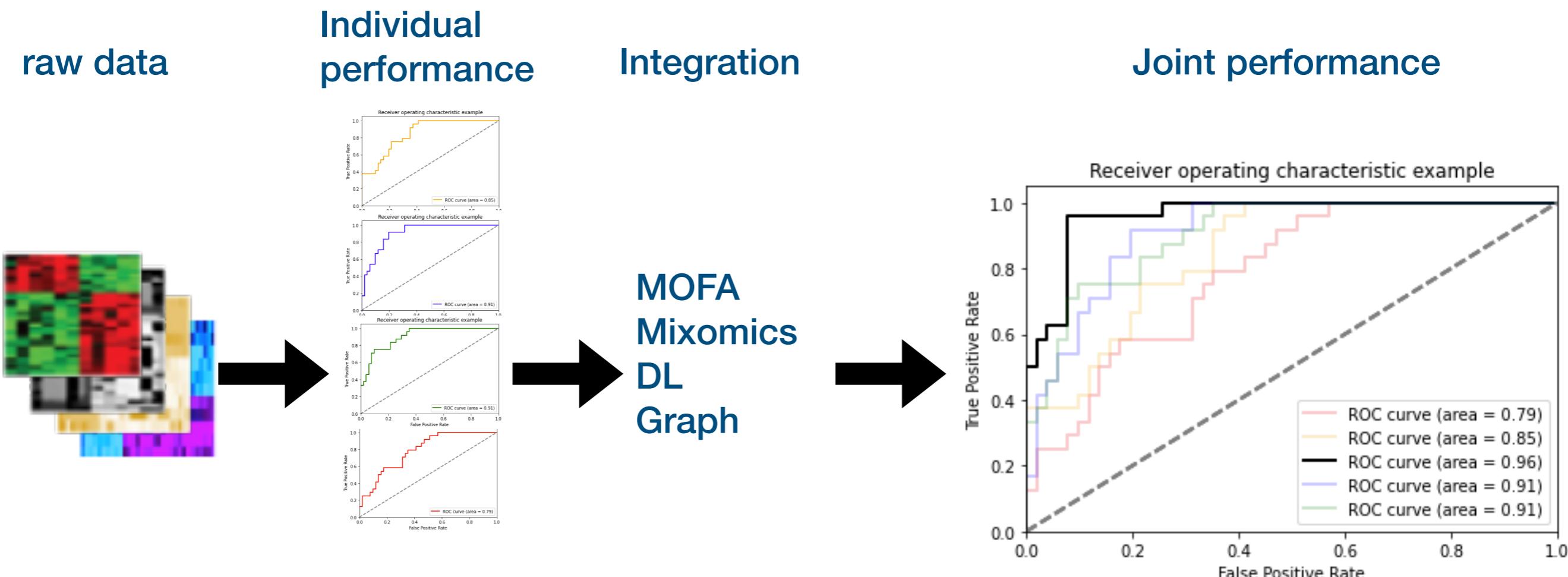
# Overview

---

## 1. Introduction to network analysis

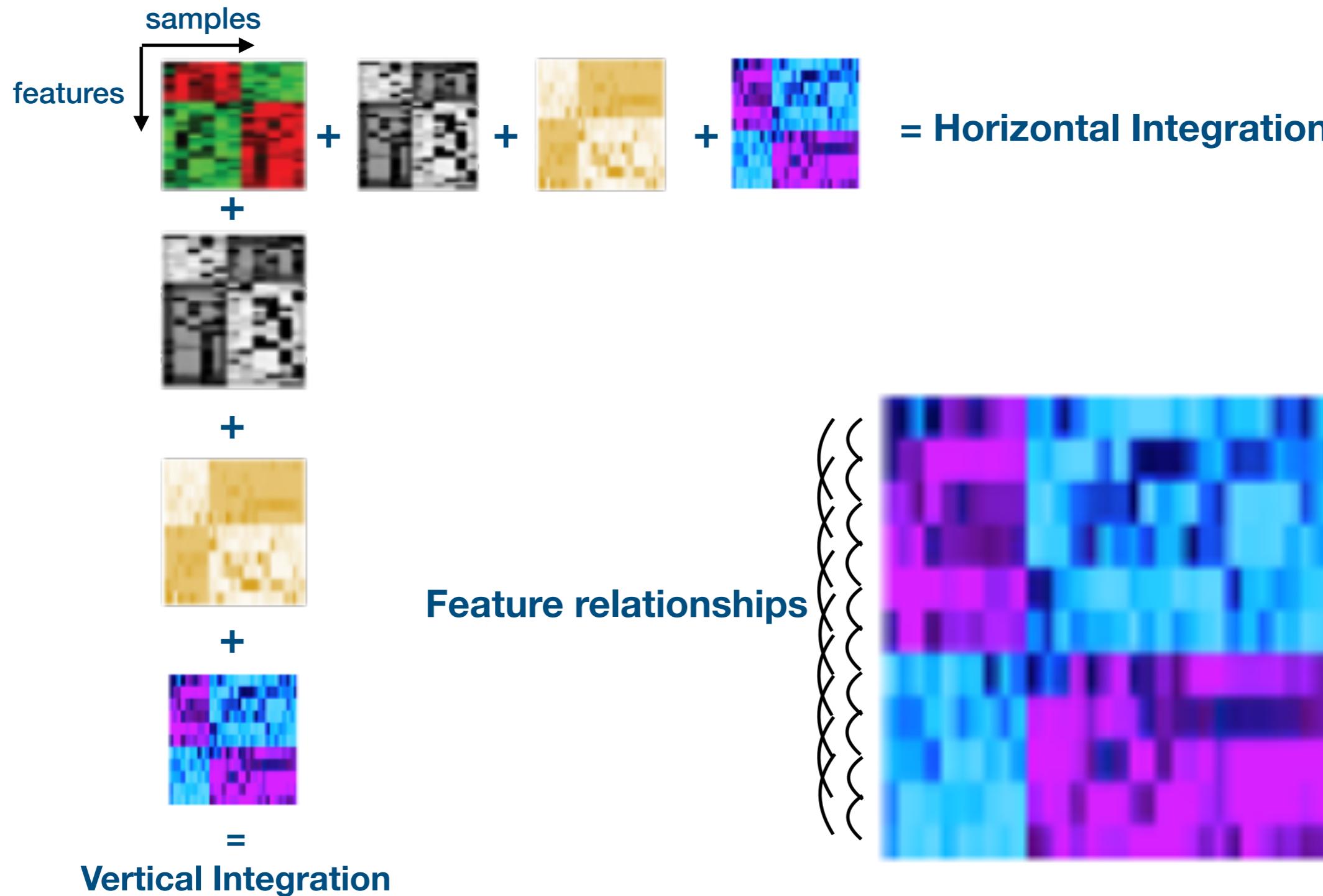
- 2. Terminology
- 3. Network construction
- 4. Key network properties
- 5. Community analysis

# Introduction



# Introduction

---

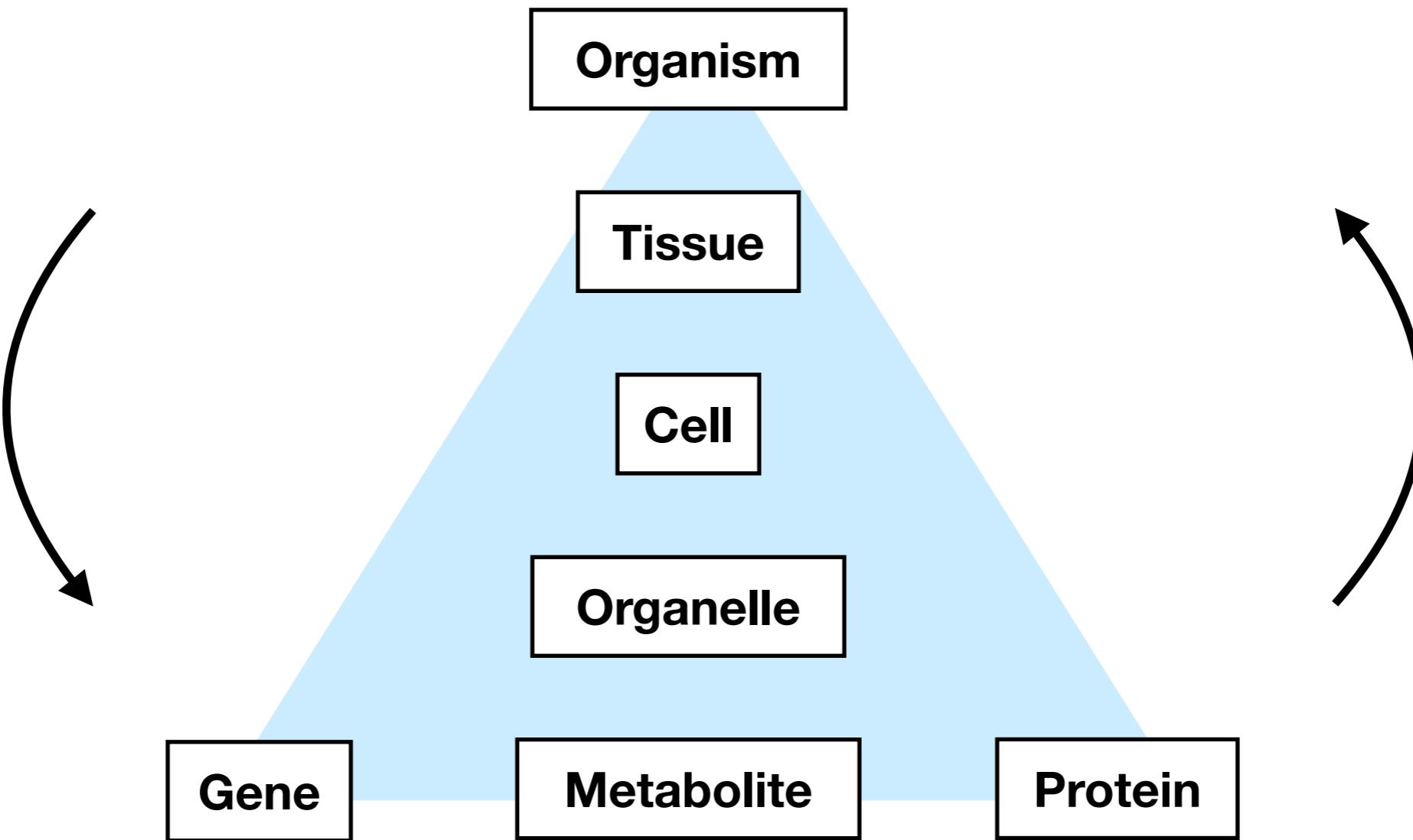


# How to tackle biological complexity?

# **Focus: feature-feature relationships**

# How to tackle biological complexity?

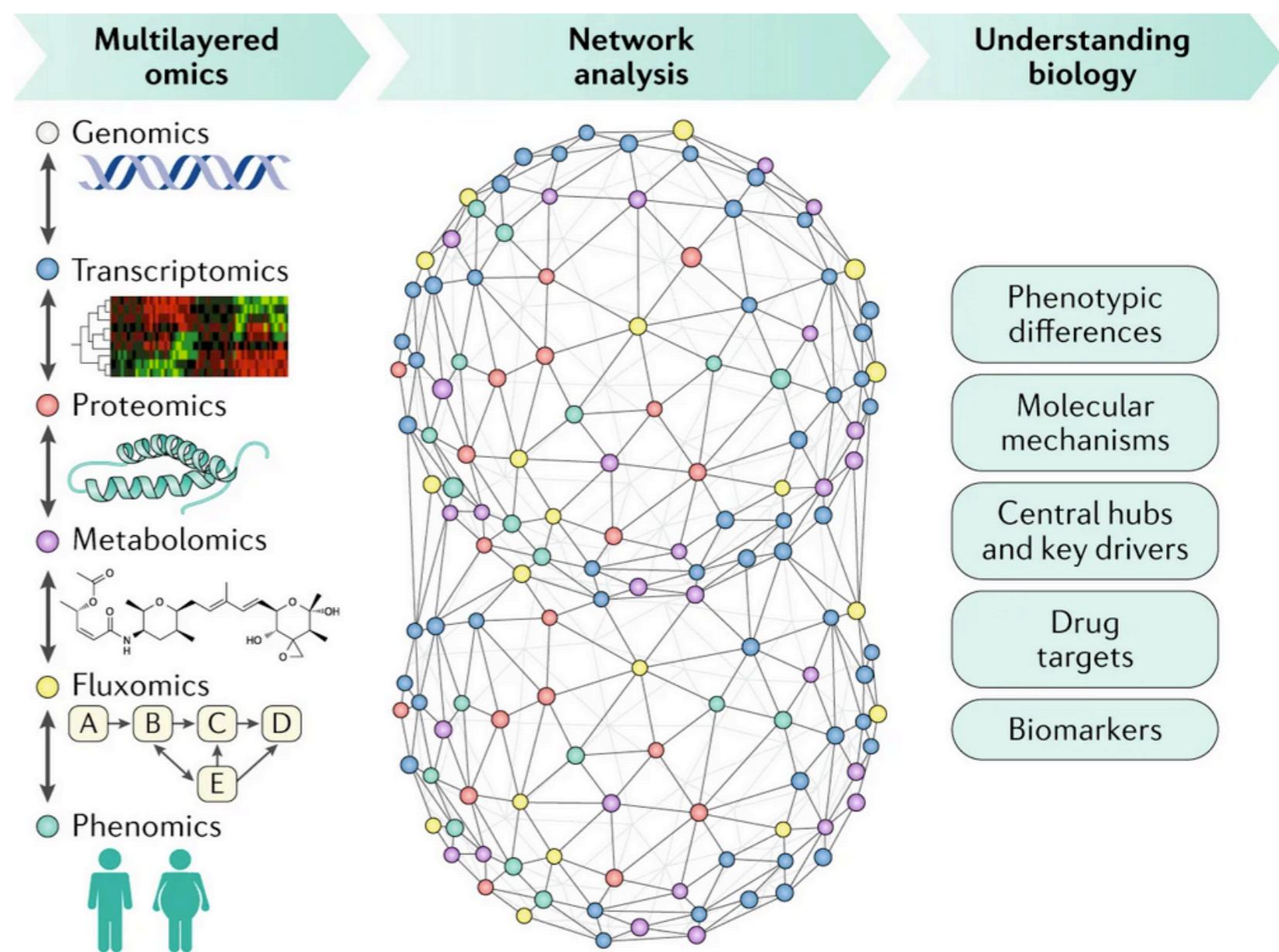
Moving from reductionist approaches towards global characterisations



# How to tackle biological complexity?

Integrative approaches, and global patterns

- Feature association
- Network analysis
- Modeling  
(Genome-scale metabolic modeling)



# What are networks?

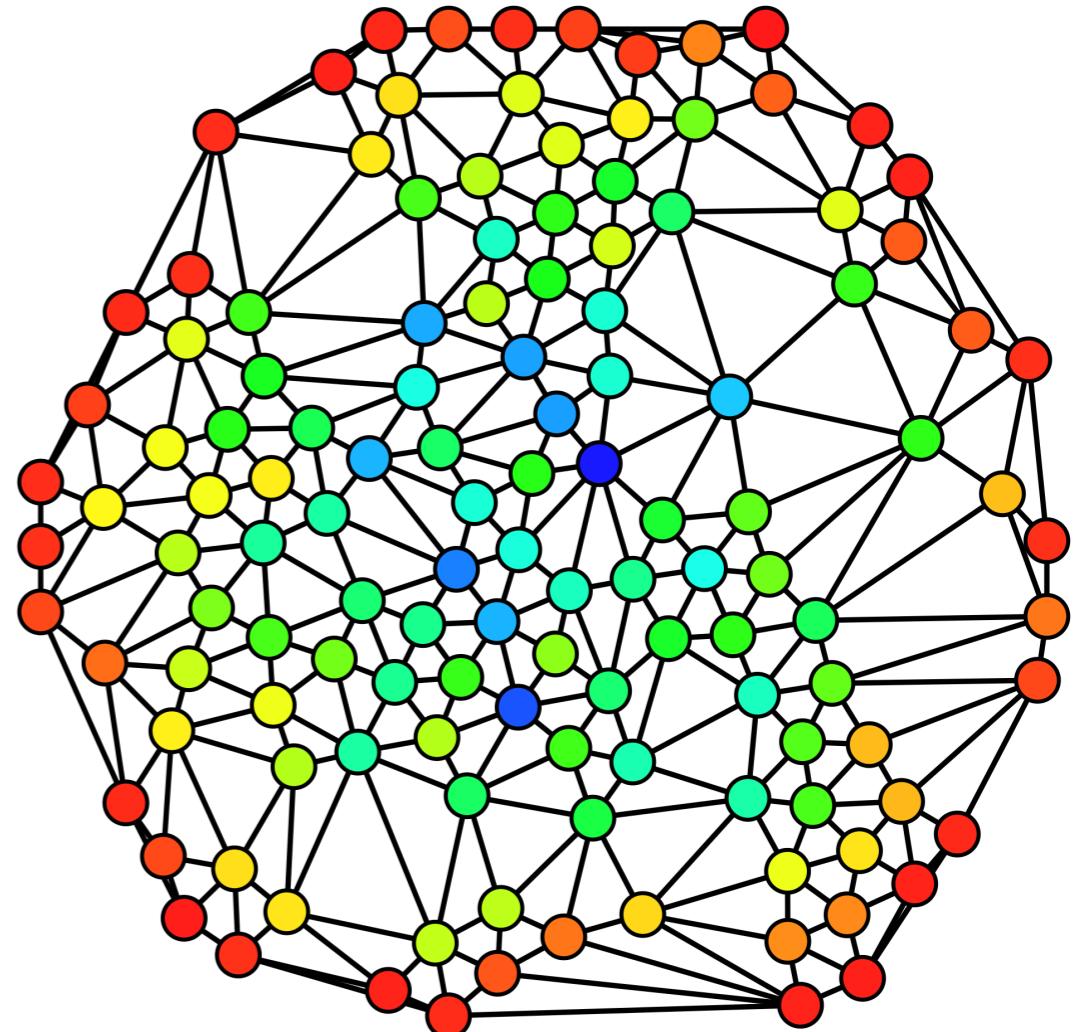
---

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 biological networks?

---

Protein - Protein interaction (PPI) networks

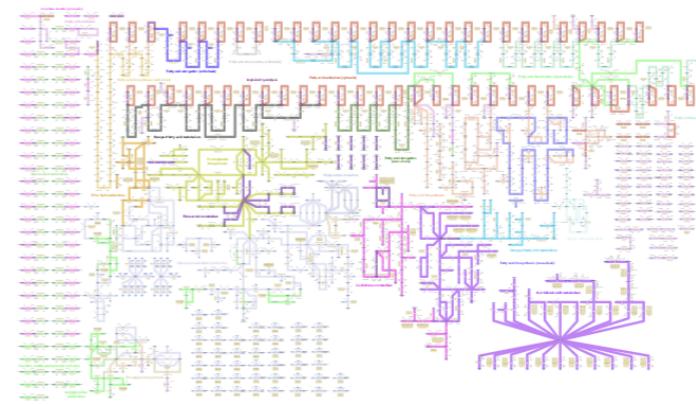
Transcription-factor regulatory networks

Gene - gene co-expression networks

Signal transduction networks

# What are biological networks?

Metabolite - Enzyme - Signal - Genes (GEMs)

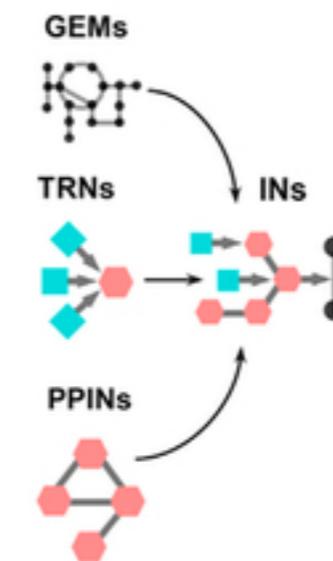
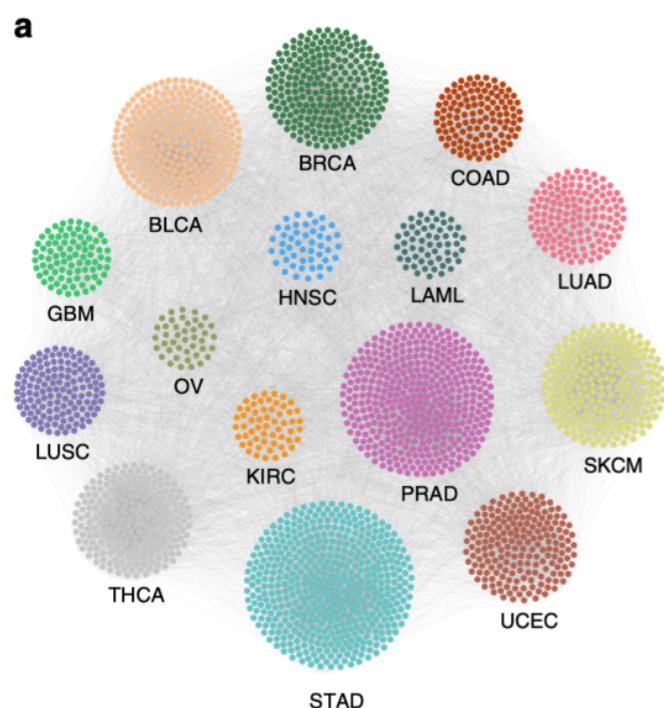
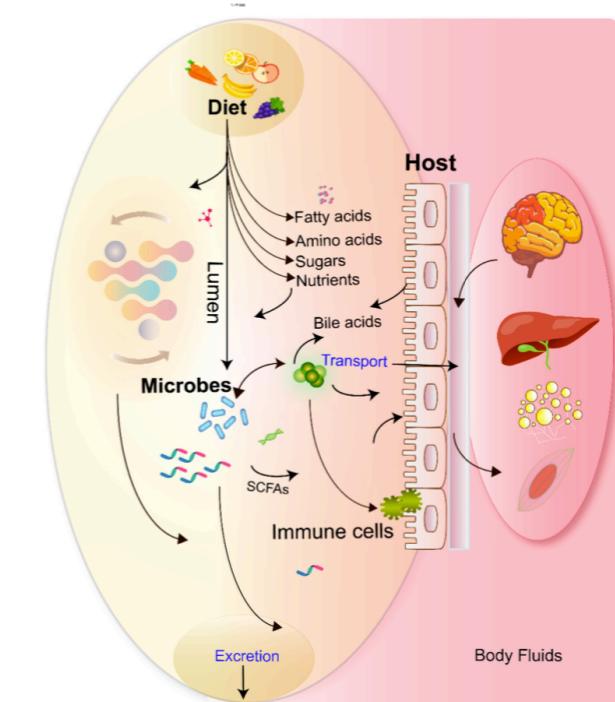
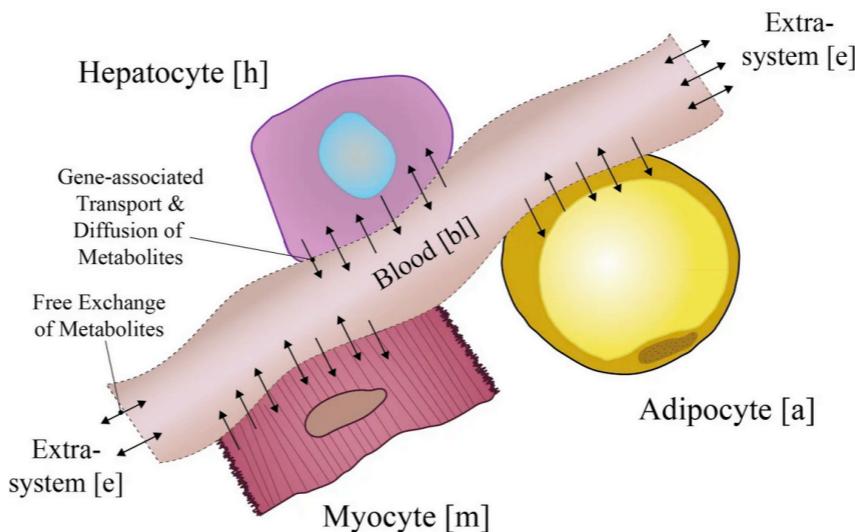


Multi-tissue networks

Multi-species networks

Disease networks

Integrated networks

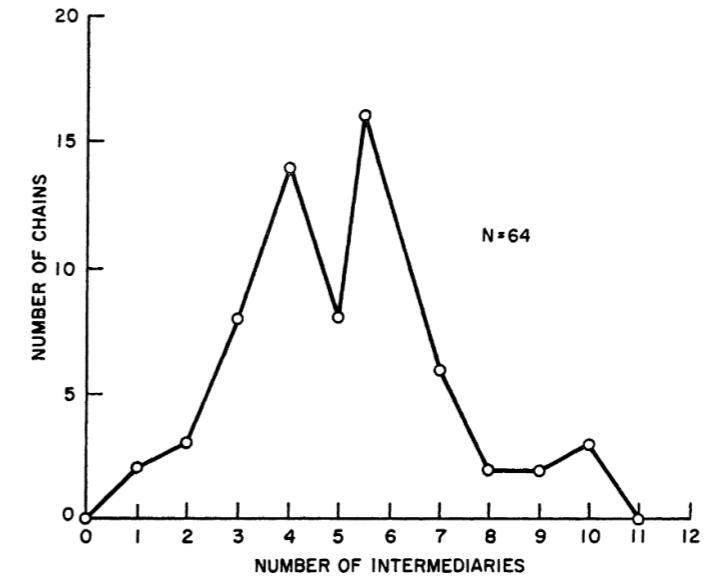


<https://metabolicatlas.org/>  
Bordbar et al 2011  
Sen & Oresic 2019  
Cheng et al 2019  
Lee et al 2016

# Small world

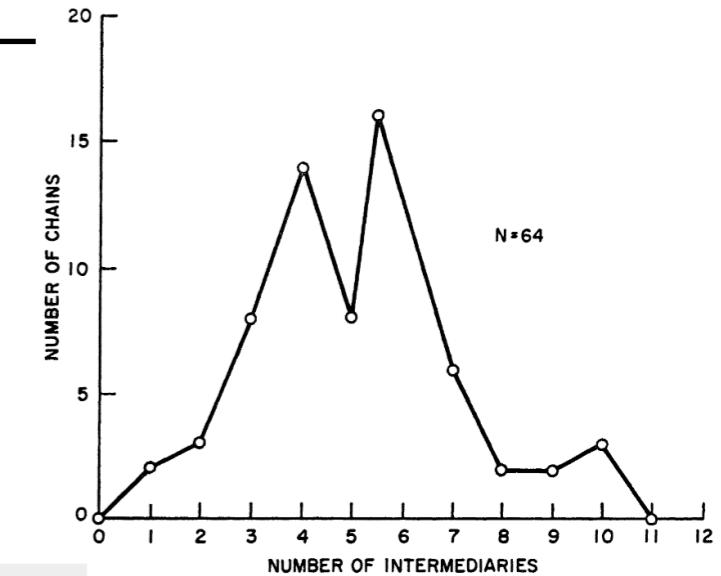
Stanley Milgram (1967) - 6 degrees

- 64 / 296 letters successful

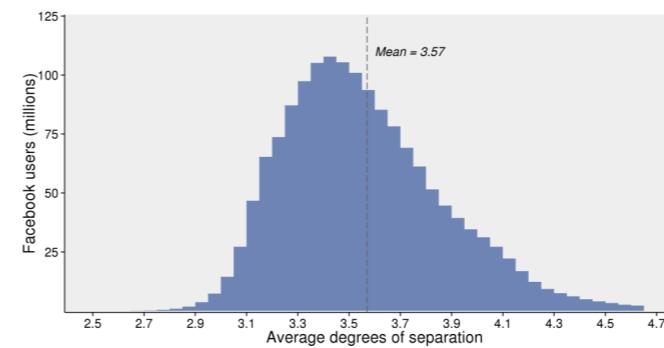


# Small world

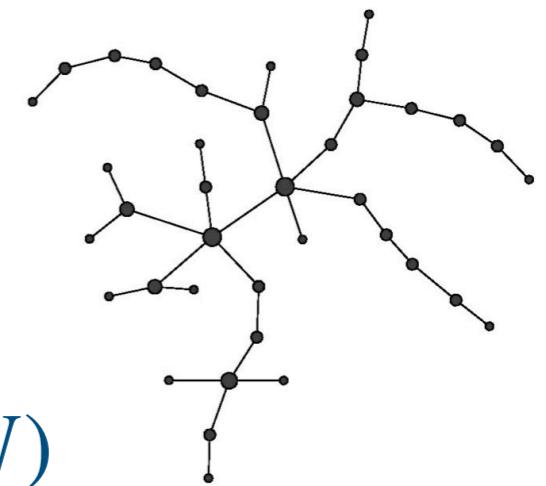
Stanley Milgram (1967): 6 degrees



Backstrom et al. (2016): 3.6 degrees



Random networks:  $l(G) \approx \log N$  for network of size  $N$



Biological Networks: **ultra-small world**  $l(G) \approx \log(\log N)$



# Why look at network topology?

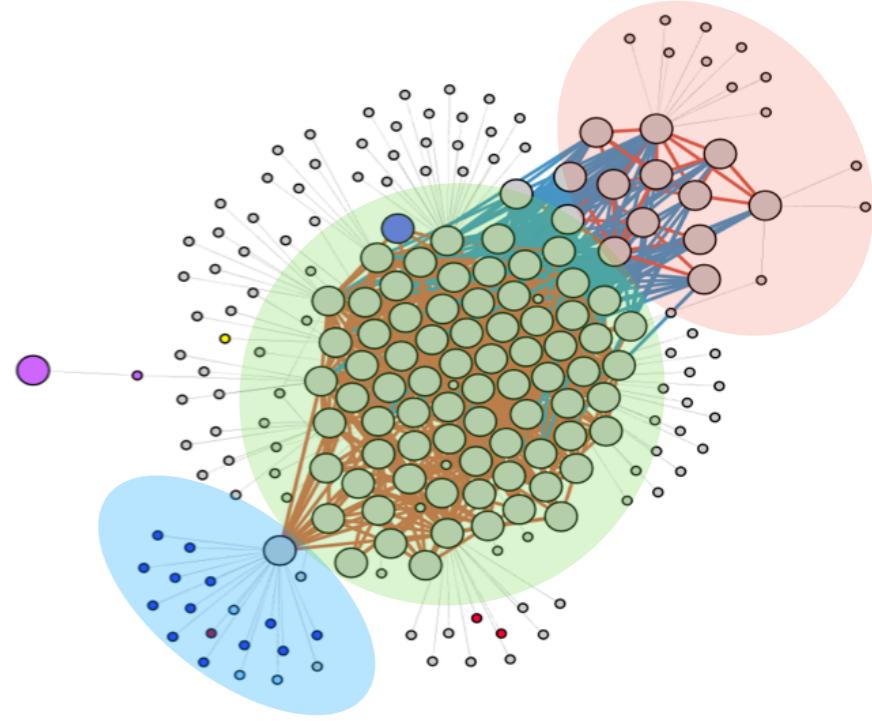
---

Use networked systems to:

- 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 “key” elements in my network?
- What are the “weakest” links in the network?



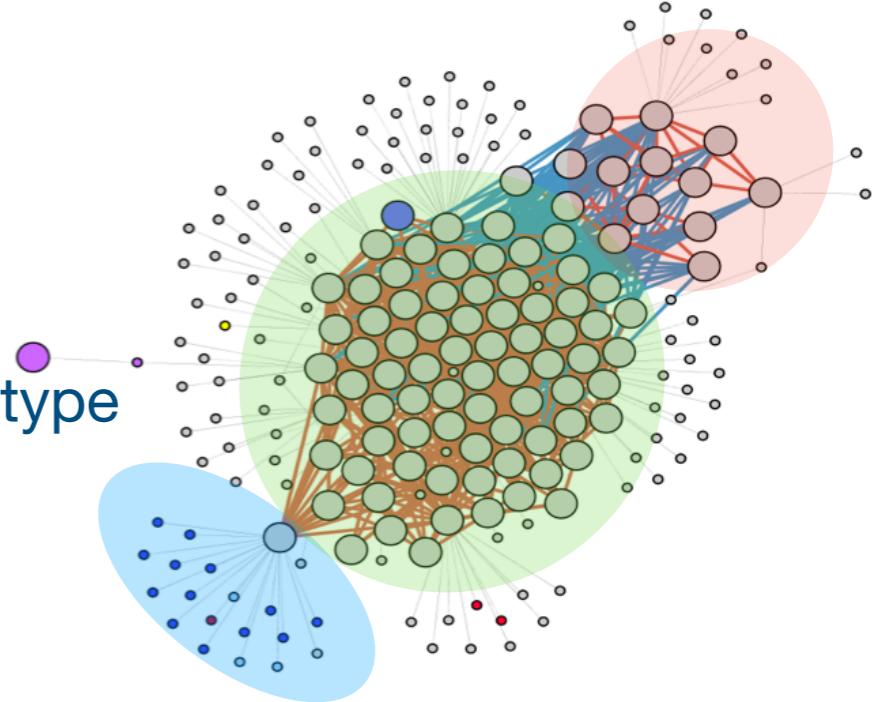
# What is my biological network?

---

Any distance 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

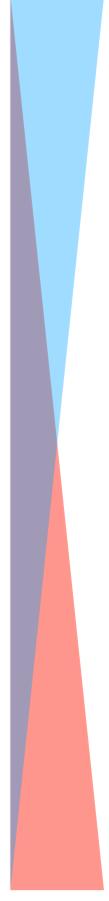


## Limitations:

- Some of the functional analyses depend on annotation
- Sample size
- Effect size
- False discovery

# Motivation

What modeling formalism suits your data and biological question?

	Pros	Cons	Details
<b><u>Kinetic models</u></b>	Detailed Quantitative Dynamic / Steady state	Small Requires detailed parameterization	
<b><u>Stoichiometric GEMs</u></b>	Large Semi-quantitative Steady state	Static	
<b><u>Topological Graphs</u></b>	Large Only topological information	No dynamic properties	 Size

# Overview

---

1. Introduction to network analysis

## **2. Terminology**

3. Network construction

4. Key network properties

5. Community analysis

Original sources of images provided as reference and hyperlinks, where applicable.

# Graphs, nodes, edges

---

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

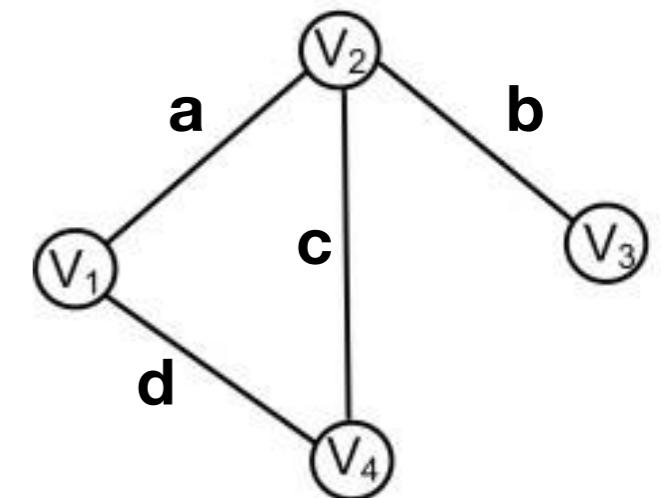
$$G = (V, E)$$

$$V = \{v_1, v_2, v_3, v_4\}$$

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

**Nodes** sometimes called **vertices**

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



# Simple vs multigraphs

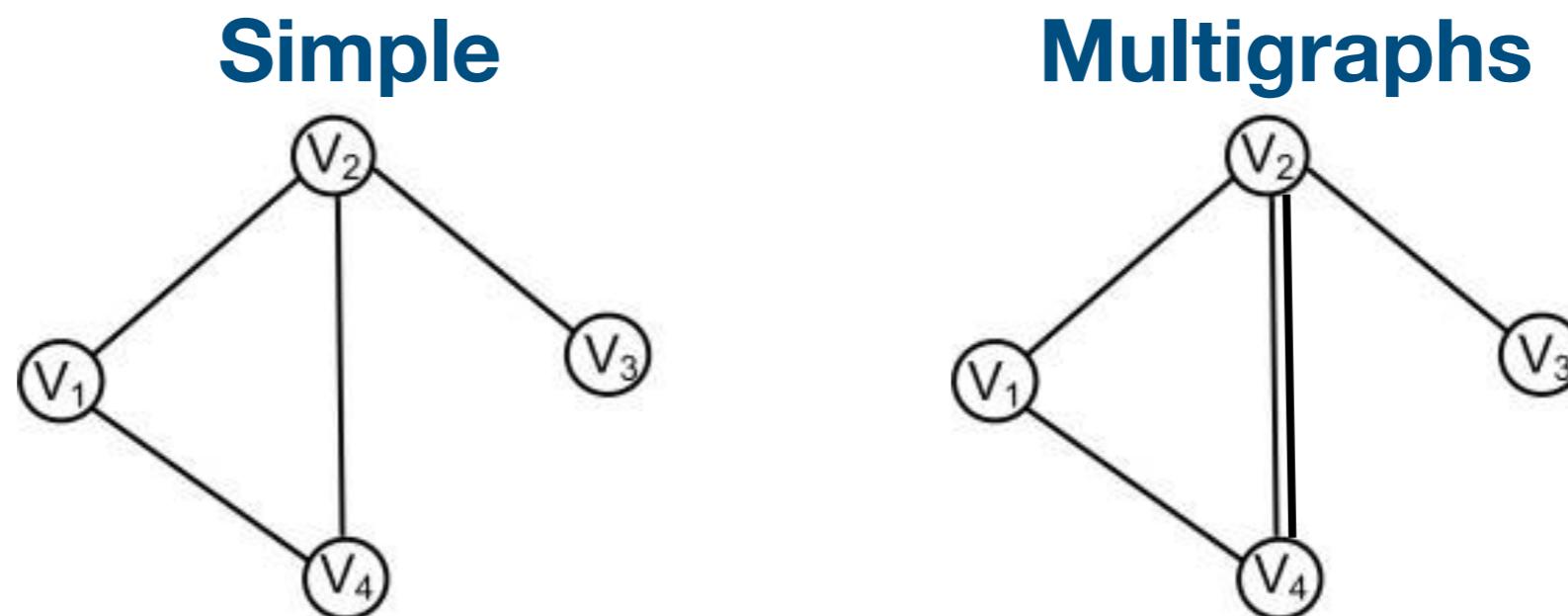
---

**Multigraphs** contain parallel edges

**Multi-edged** connections indicate different properties

Example: PPI

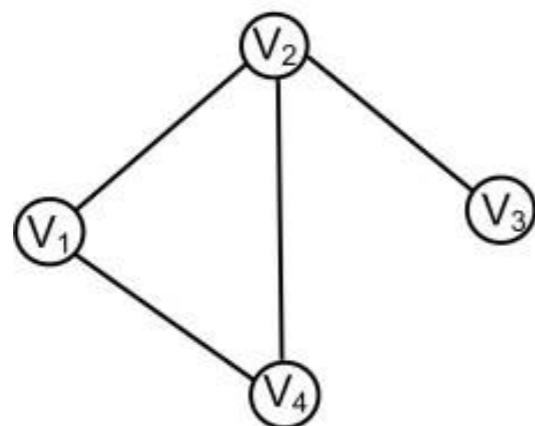
- Experimental evidence for interaction
- Co-expression



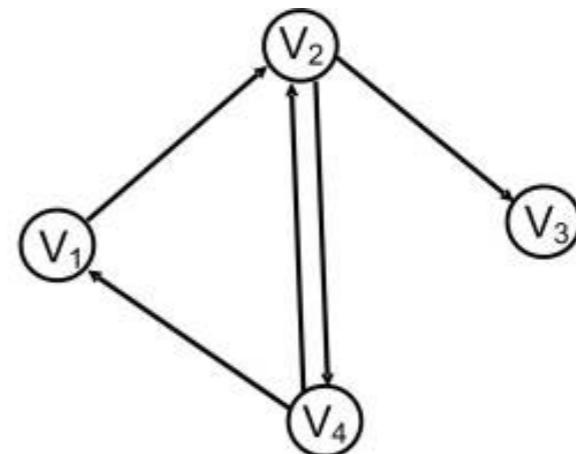
# Directed vs undirected graphs

---

**Undirected graphs:**  
co-expression networks



**Directed graphs:**  
metabolic networks



# Hypergraphs

---

**Hypergraphs** contain edges that connect any number of nodes

**Reaction 1:**  $A \rightarrow B + C$

**Reaction 2:**  $B + C \rightarrow D$

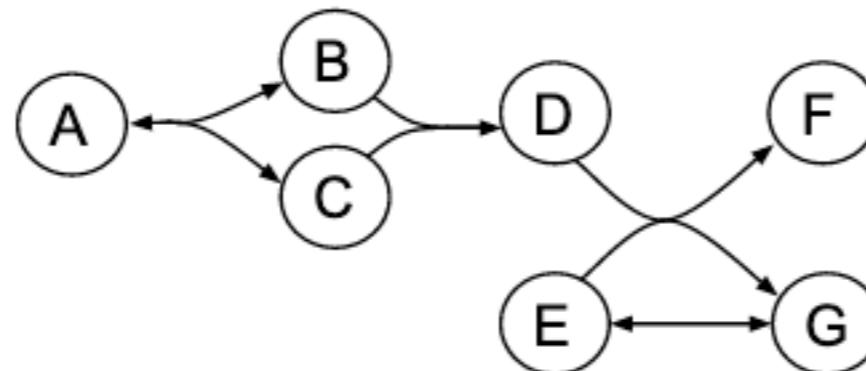
**Reaction 3:**  $D + E \rightarrow F + G$

**Reaction 4:**  $E \rightarrow G$

**Reaction 5:**  $B + C \rightarrow A$

**Reaction 6:**  $G \rightarrow E$

(a) Reaction network

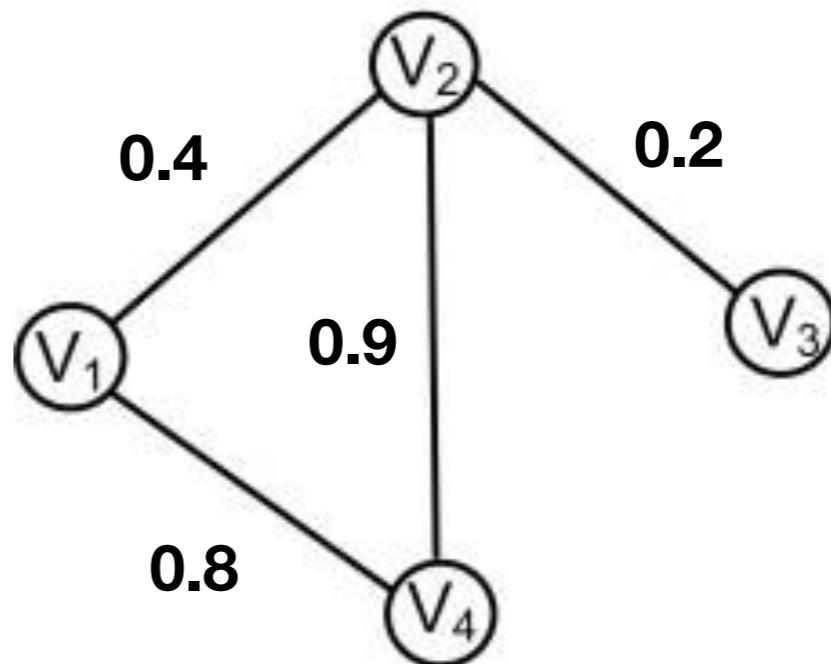


# Weighted vs unweighted graphs

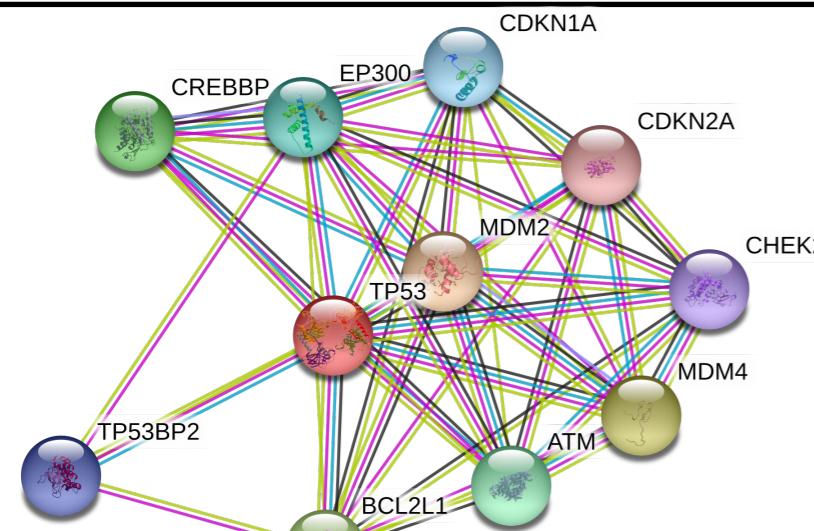
---

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

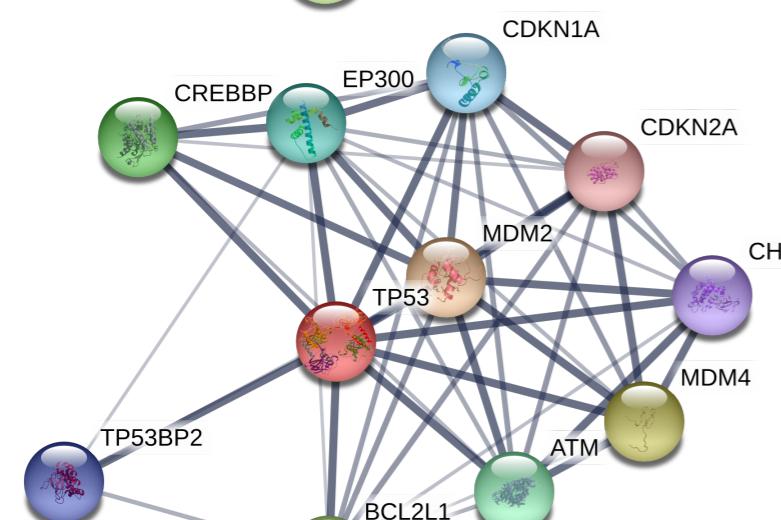
E.g. weighted co-expression networks



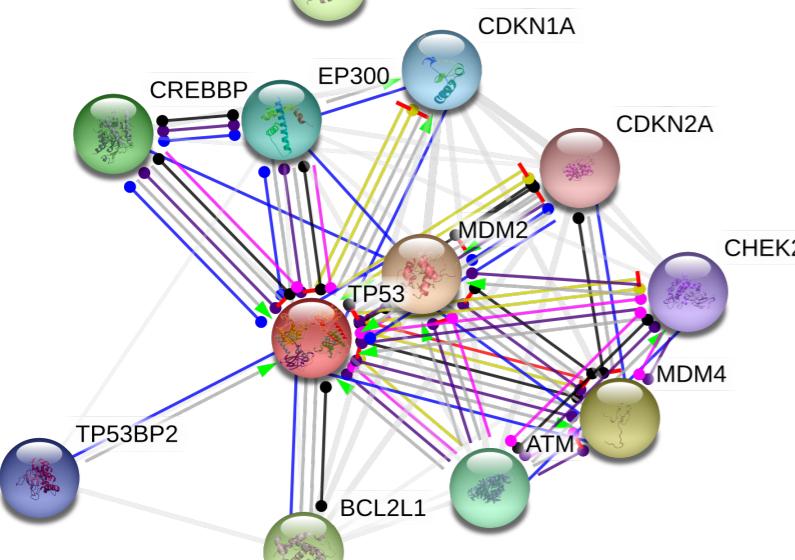
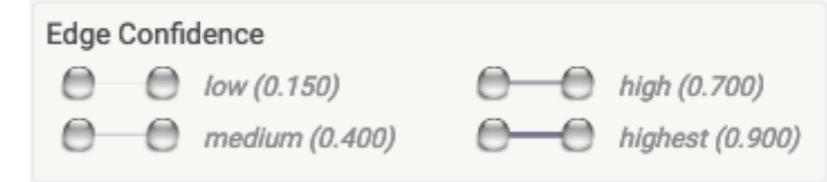
# STRING-db.org: TP53



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

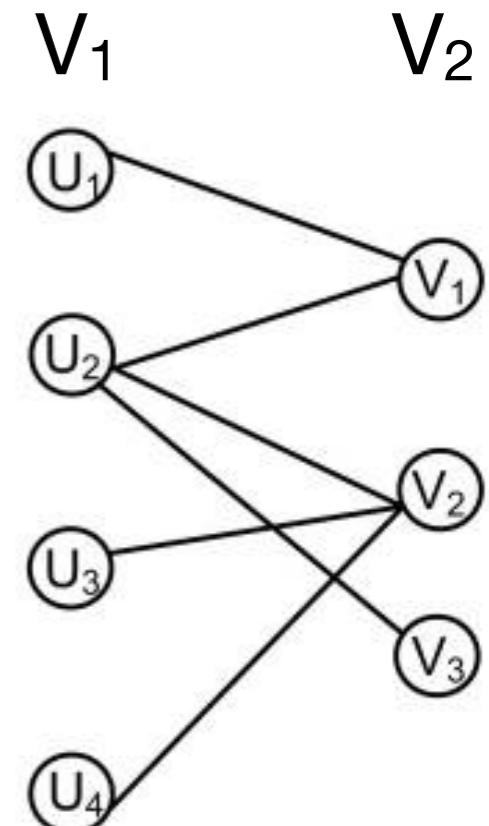
$$u \in V_1 \text{ and } v \in V_2$$

All  $e_i$  has end-nodes in  $V_1, V_2$

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

$$G_1 = (V_1, E_1)$$

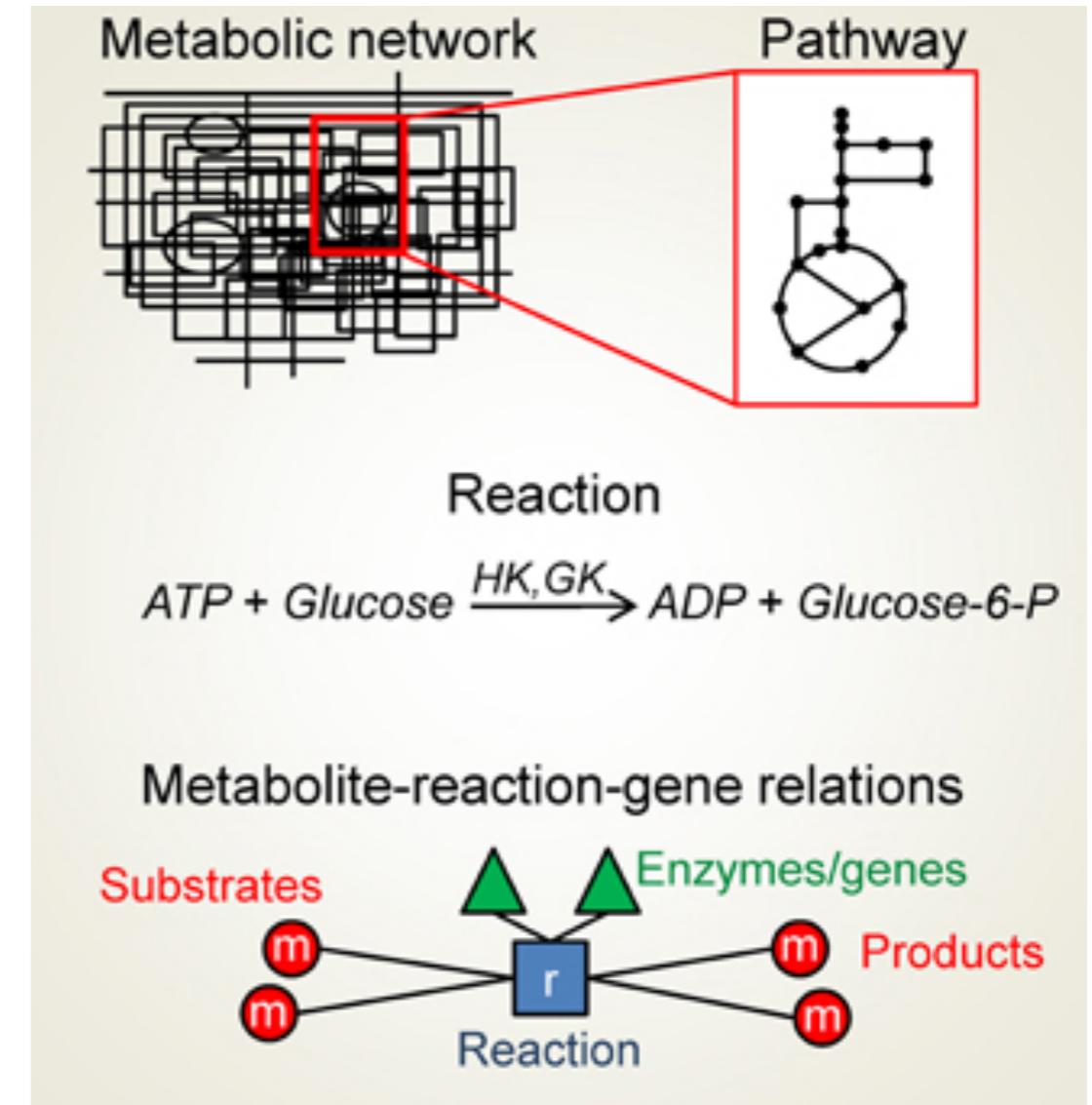
**Examples?**



# Bipartite and $k$ -partite graphs

Bipartite graph:

- Gene - Reaction



Tripartite graph:

- Metabolite - reaction - enzyme

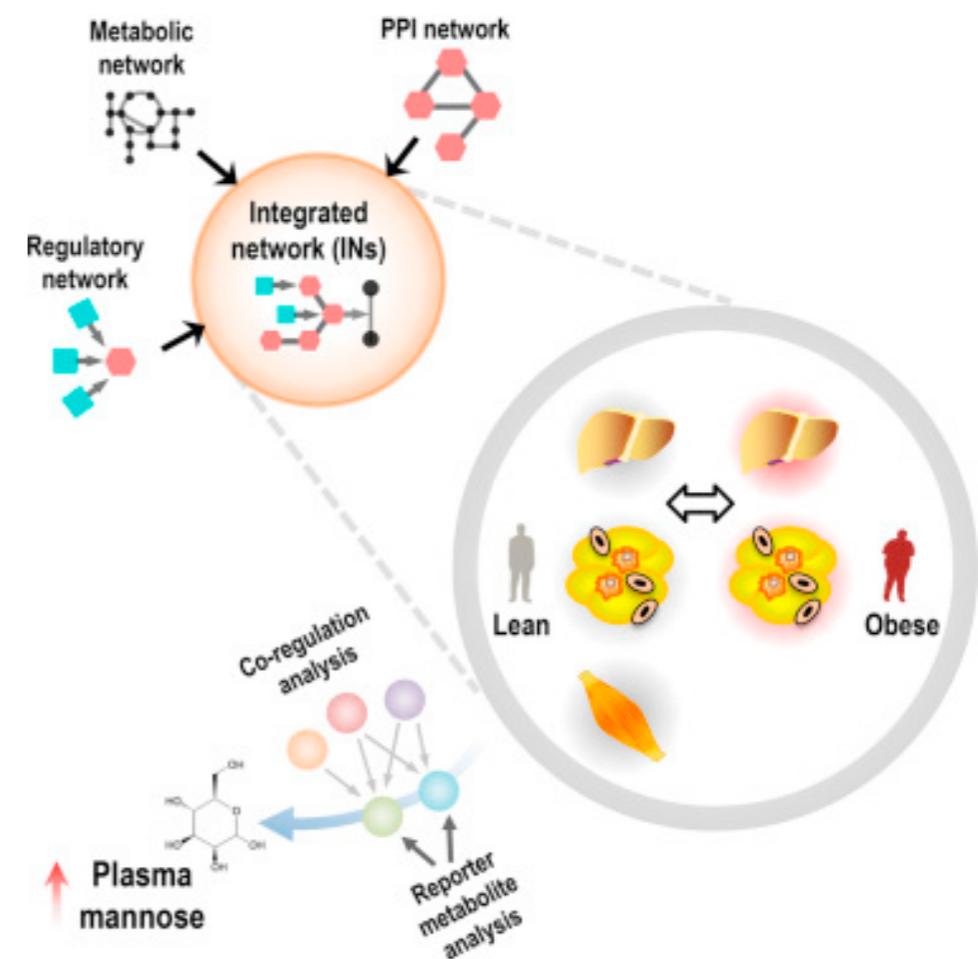
$k$ -partite graphs display  $k$ -types of nodes

# *k*-partite graphs

Multi-modal (*k*-partite) networks may be generated from different sources

- Transcription-factor - Gene (DNAseq)
- Gene-gene (Co-expression, PPI, GEMs)
- Gene-metabolite (GEM)
- Metabolite-metabolite (GEM)

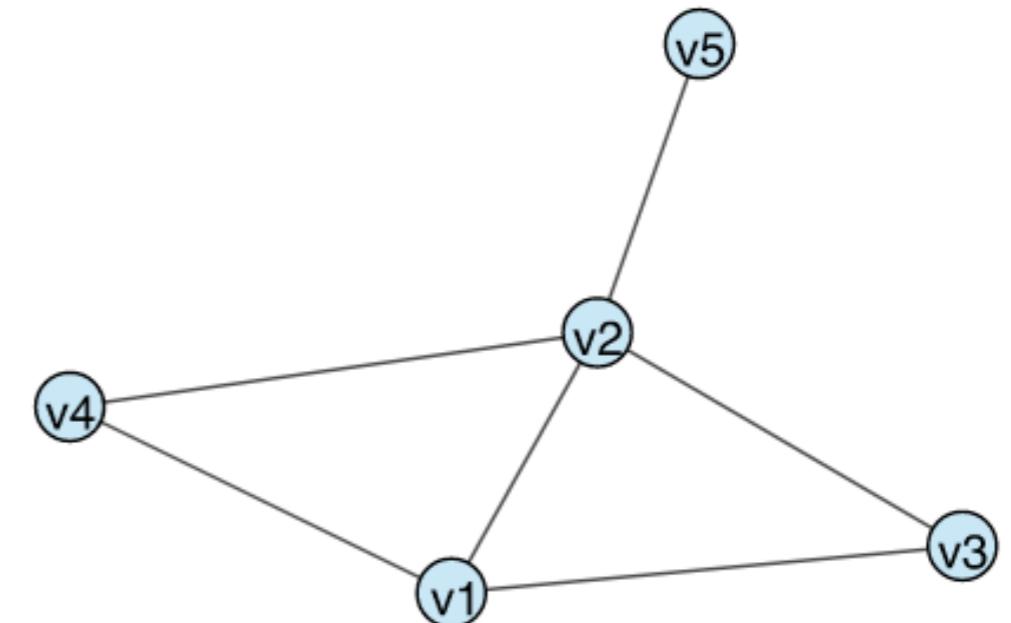
## Integrated Networks



# 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**

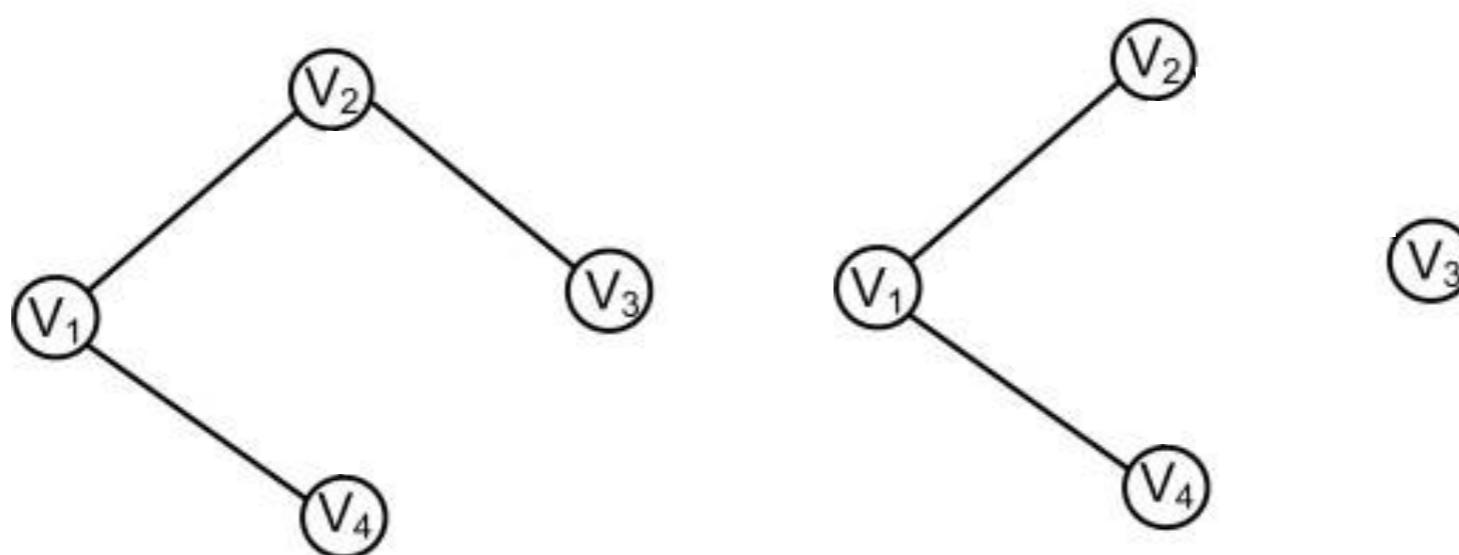
	v1	v2	v3	v4	v5
v1	0	1	1	1	0
v2	1	0	1	1	1
v3	1	1	0	0	0
v4	1	1	0	0	0
v5	0	1	0	0	0

# 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

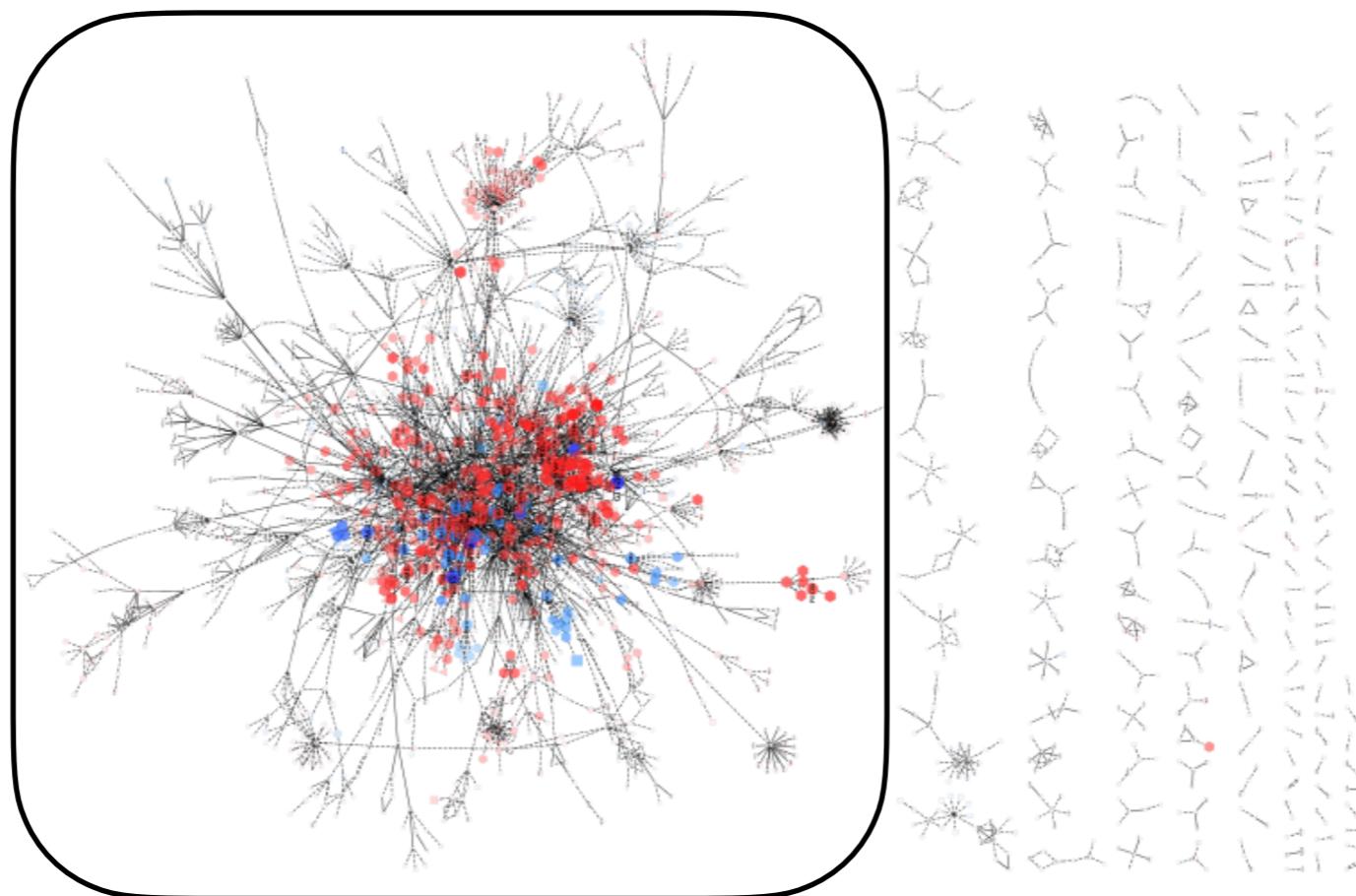


# Connected components

---

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

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

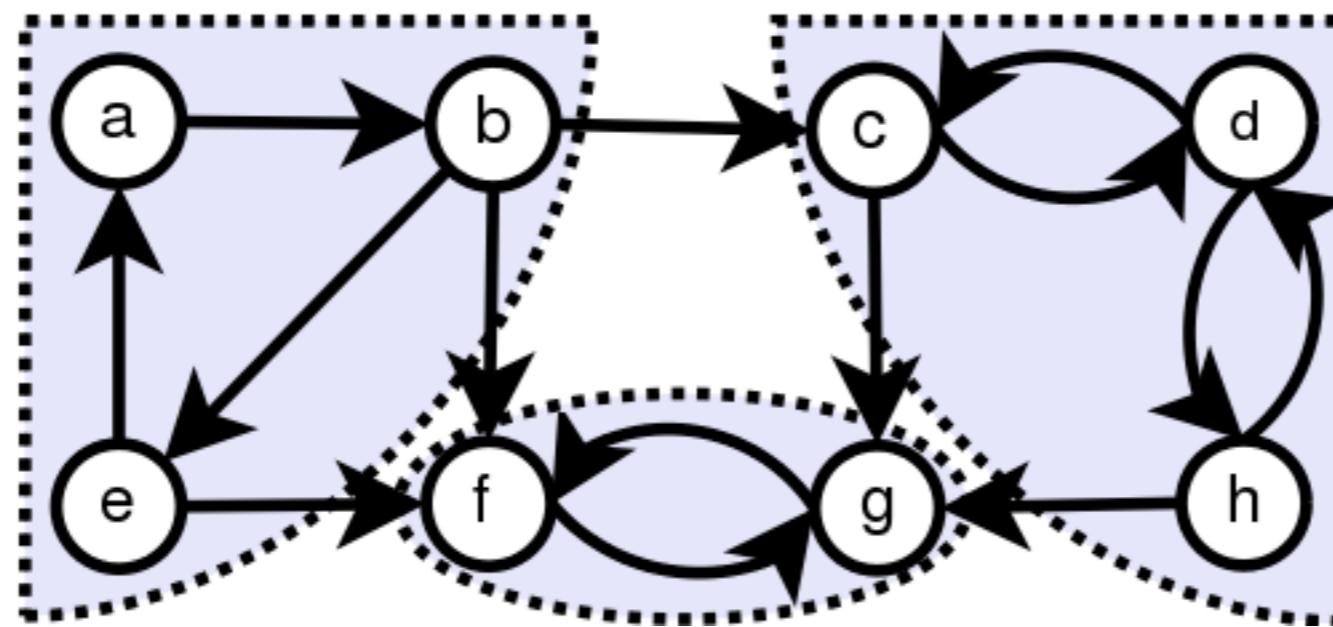


# Connected components

---

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

## Weak vs strong components



# Additional reading

---

- [Network Science](#) - Textbook on graph theory and network analysis.
- [Communication dynamics in complex brain networks](#) - Network topology applied to study the brain
- [A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models](#) - 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.