Introduction to biological network analysis

Rui Benfeitas

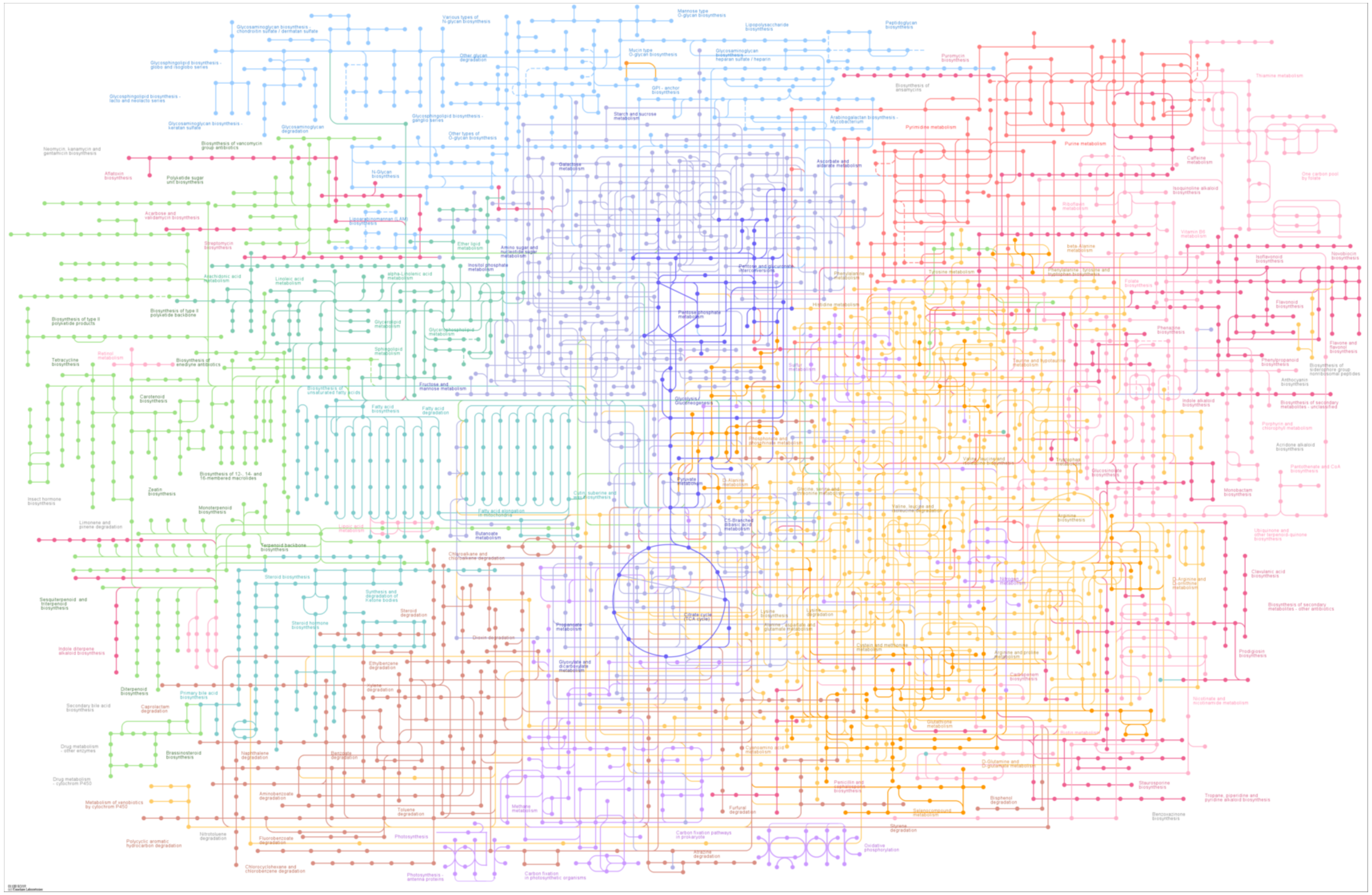
NBIS - National Bioinformatics Infrastructure Sweden
Science for Life Laboratory, Stockholm
Stockholm University

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.

How to tackle biological complexity?



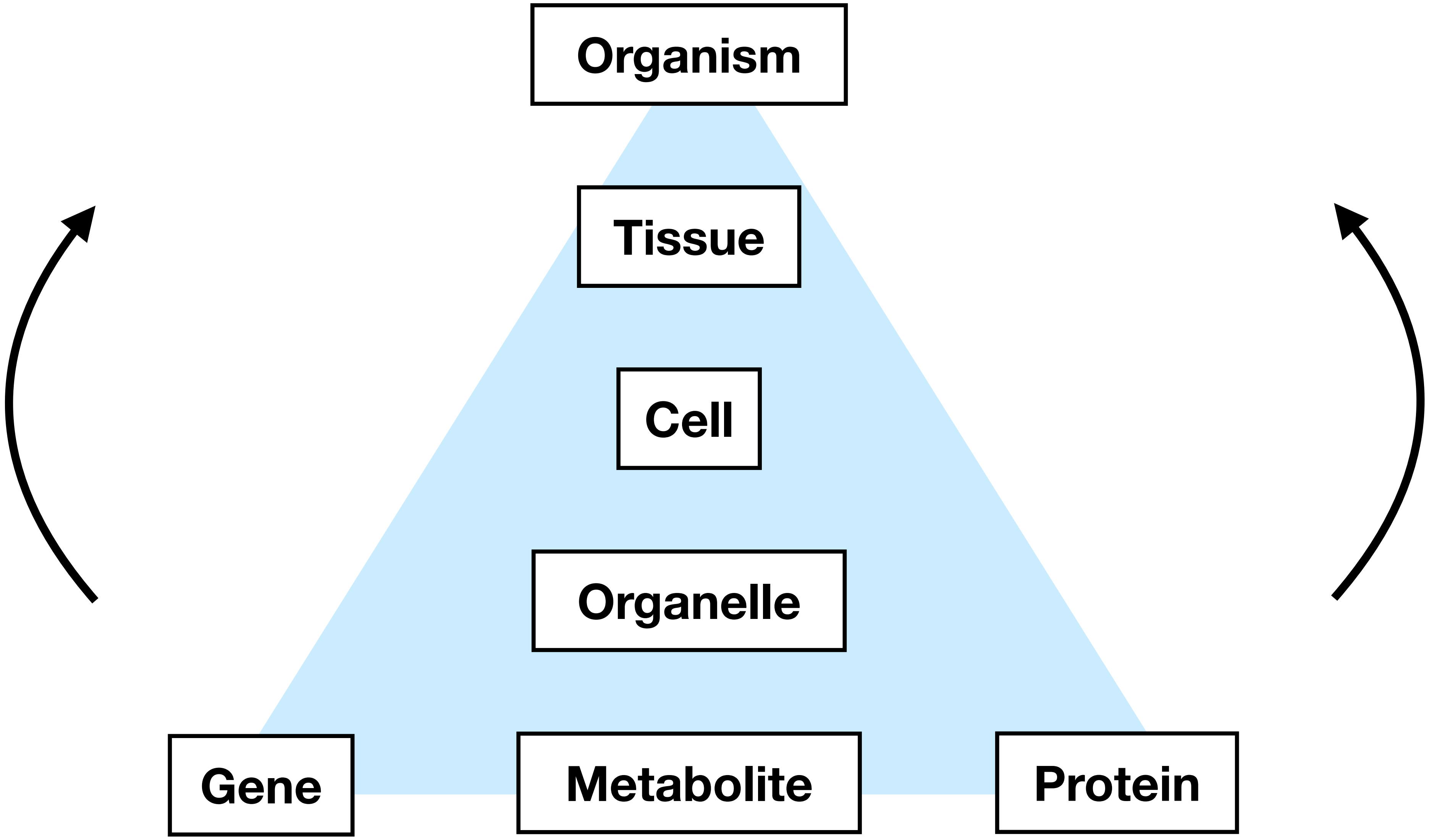
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate
Various types of N-glycan biosynthesis
Mannose type O-glycan biosynthesis
Lipopolysaccharide biosynthesis
Peptidoglycan biosynthesis
Glycosphingolipid biosynthesis - globo and isoglobo series
Glycosphingolipid biosynthesis - lacto and neolacto series
Mucin type O-glycan biosynthesis
Glycosaminoglycan biosynthesis - heparan sulfate / heparin
Glycosaminoglycan biosynthesis - keratan sulfate
Glycosaminoglycan degradation
Glycosphingolipid biosynthesis - ganglio series
Other types of O-glycan biosynthesis
Starch and sucrose metabolism
Pyrimidine metabolism
Purine metabolism
Thiamine metabolism
Caffeine metabolism
Galactose metabolism
Ascorbate and aldarate metabolism
Isoquinoline alkaloid biosynthesis
Pentose phosphate pathway
Glycolysis / Gluconeogenesis
Citrate cycle (TCA cycle)
Fatty acid biosynthesis
Fatty acid degradation
Steroid biosynthesis
Steroid degradation
Carotenoid biosynthesis
Tetracycline biosynthesis
Zeatin biosynthesis
Monoterpenoid biosynthesis
Diterpenoid biosynthesis
Flavonoid biosynthesis
Phenylalanine metabolism
Tyrosine metabolism
Histidine metabolism
Arginine biosynthesis
Photosynthesis
Methane metabolism
Oxidative phosphorylation
Carbon fixation in photosynthetic organisms
Nitrotoluene degradation
Toluene degradation
Naphthalene degradation
Benzoate degradation
Drug metabolism - cytochrome P450
Drug metabolism - other enzymes
Metabolism of xenobiotics by cytochrome P450
Polycyclic aromatic hydrocarbon degradation
Tropane, piperidine and pyridine alkaloid biosynthesis

Focus: feature-feature relationships

genome.jp/kegg/

How to tackle biological complexity?

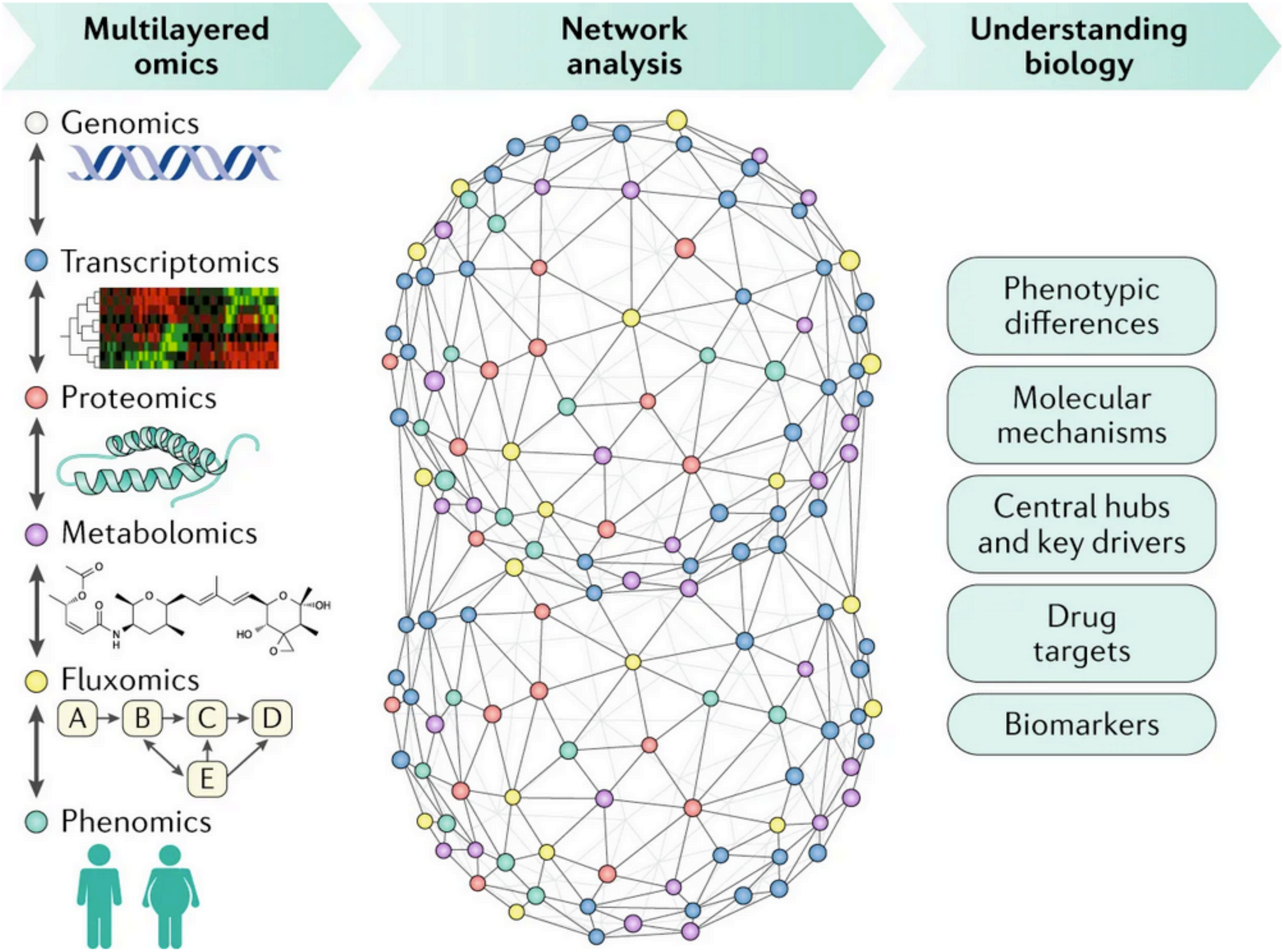
Moving from reductionist approaches towards global characterisations



How to tackle biological complexity?

Integrative approaches, and global patterns

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



Mardinoglu 2018

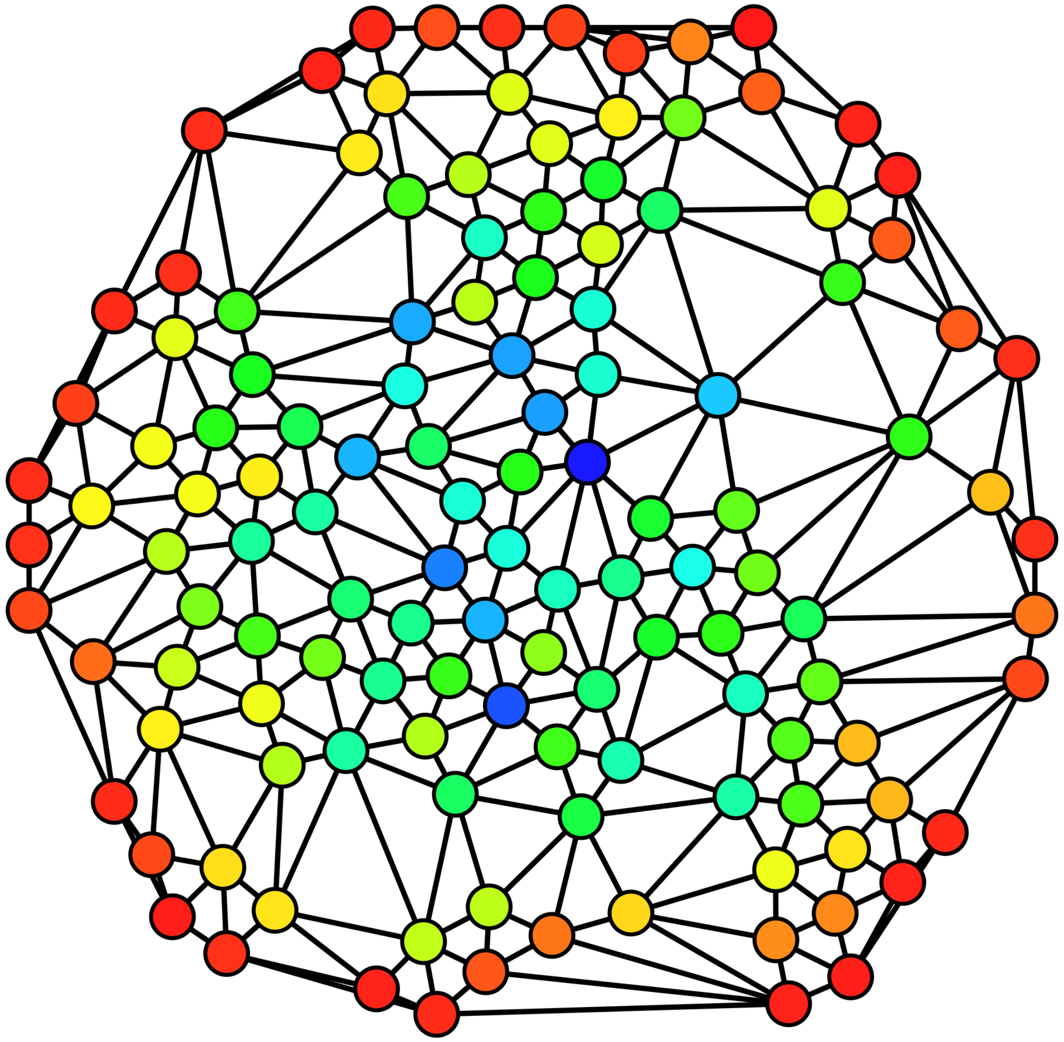
What are networks?

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



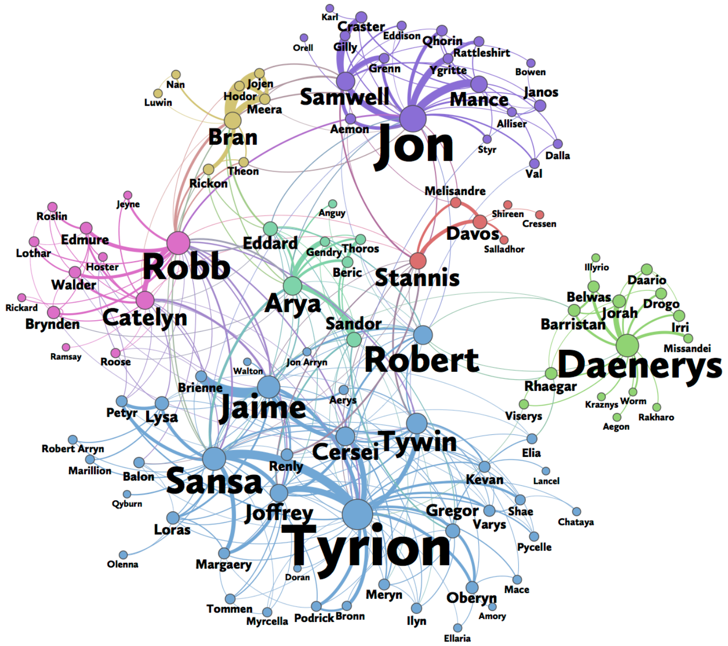
What are networks?

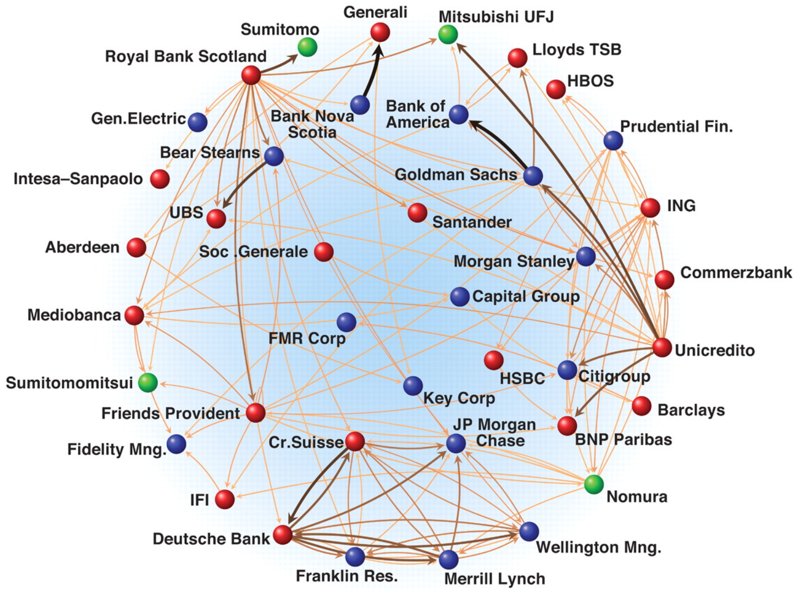
Social

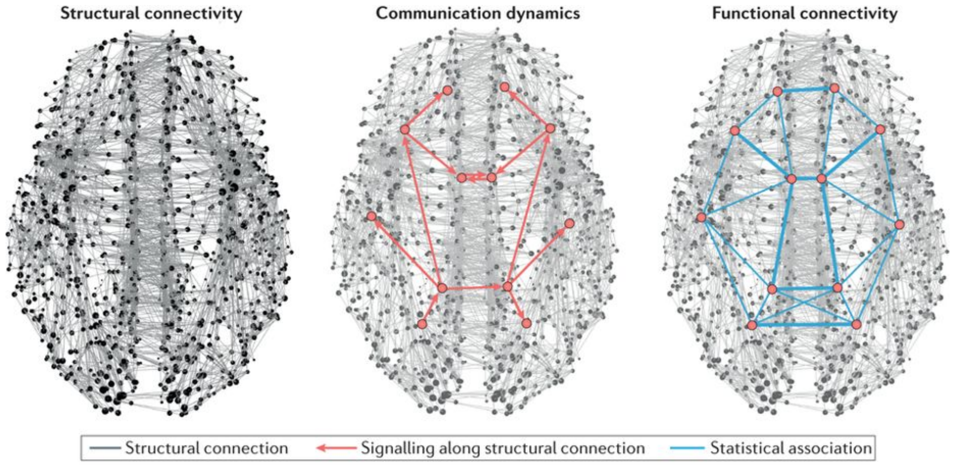
Economic

Communication

Neuronal







Nature Reviews | Neuroscience

Beveridge & Shan 2017
Helen Knight MIT News 2013
Schweitzer et al 2009
Avena-Koenigsberger et al 2018

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?

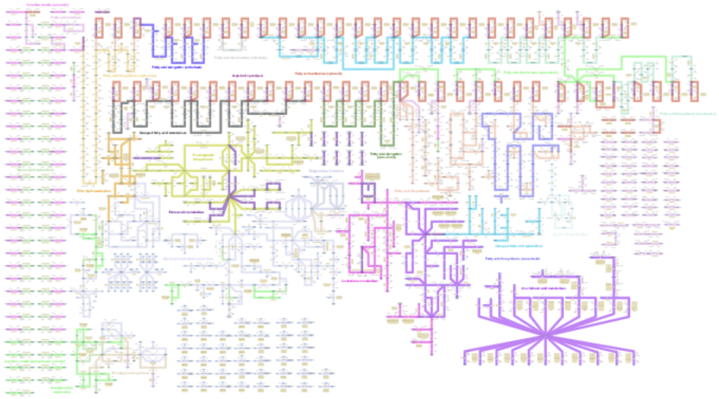
Metabolite - Enzyme - Signal - Genes (GEMs)

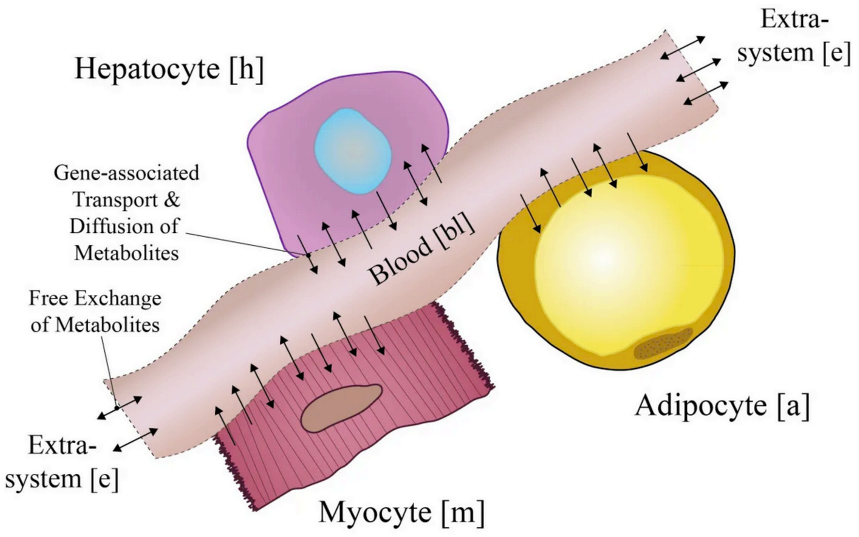
Multi-tissue networks

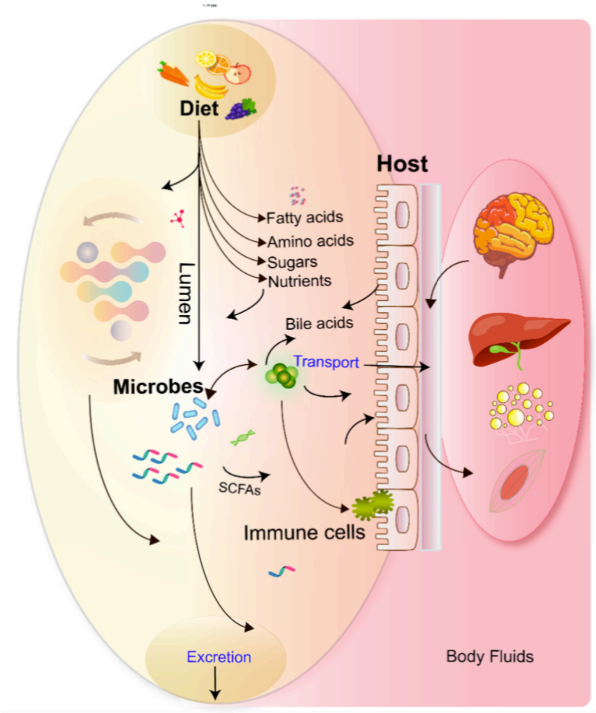
Multi-species networks

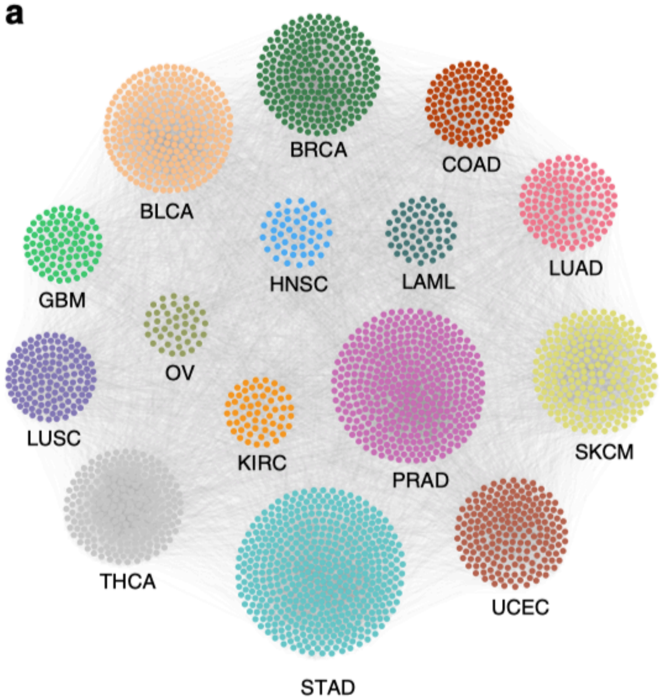
Disease networks

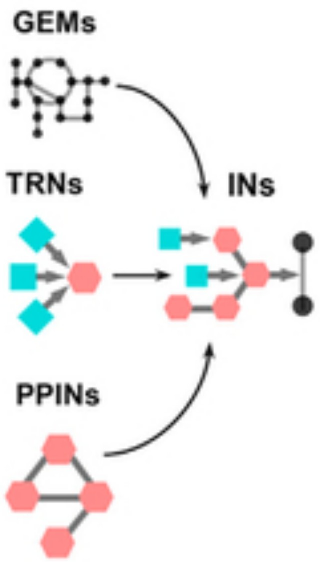
Integrated networks







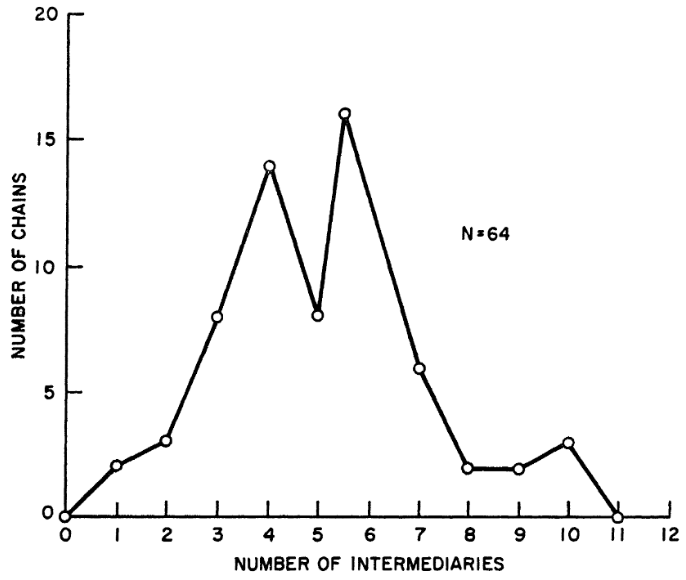




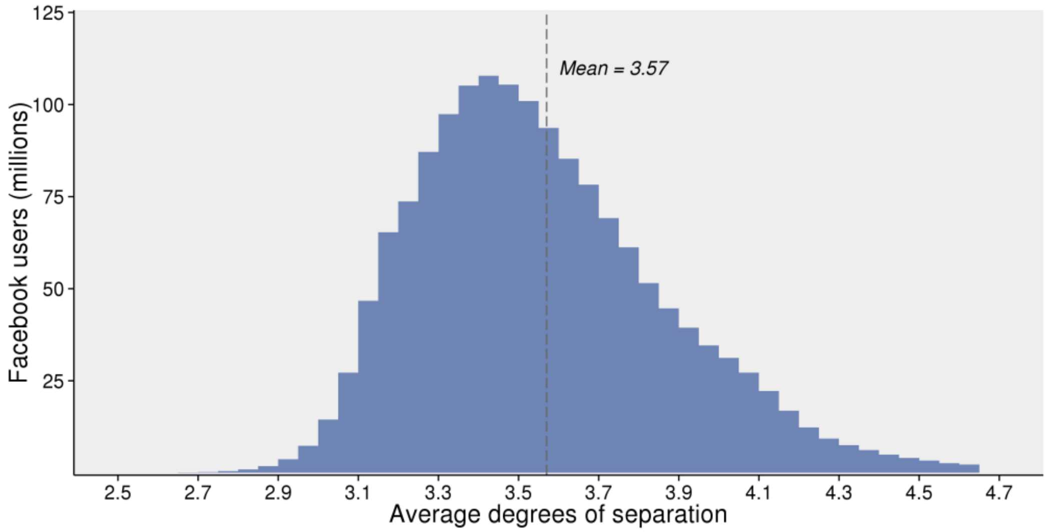
https://metabolicatlas.org/
Bordbar et al 2011
Sen & Oresic 2019
Cheng et al 2019
Lee et al 2016

Small world

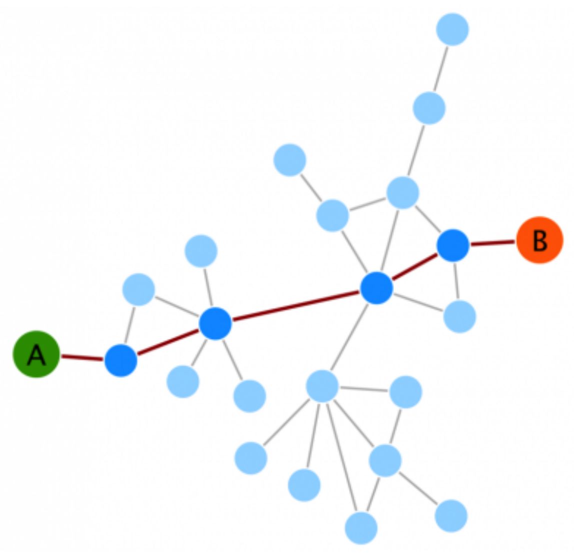
Stanley Milgram (1967) - 6 degrees
("6 degrees of Kevin Bacon")



Backstrom et al. (2016) - 3.6 degrees



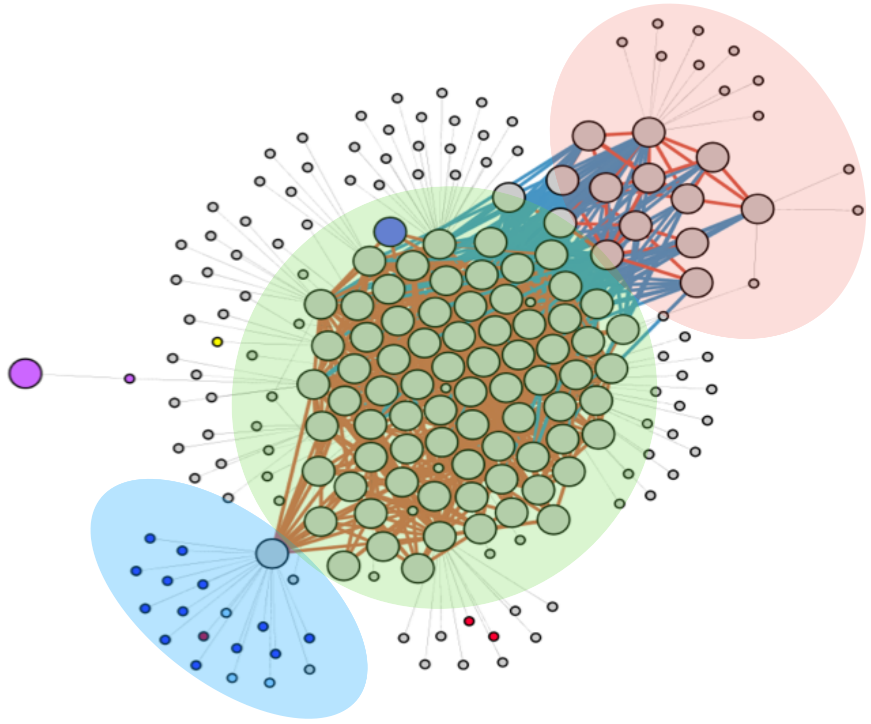
Biological Networks



Why look at network topology?

Use networked systems to:

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

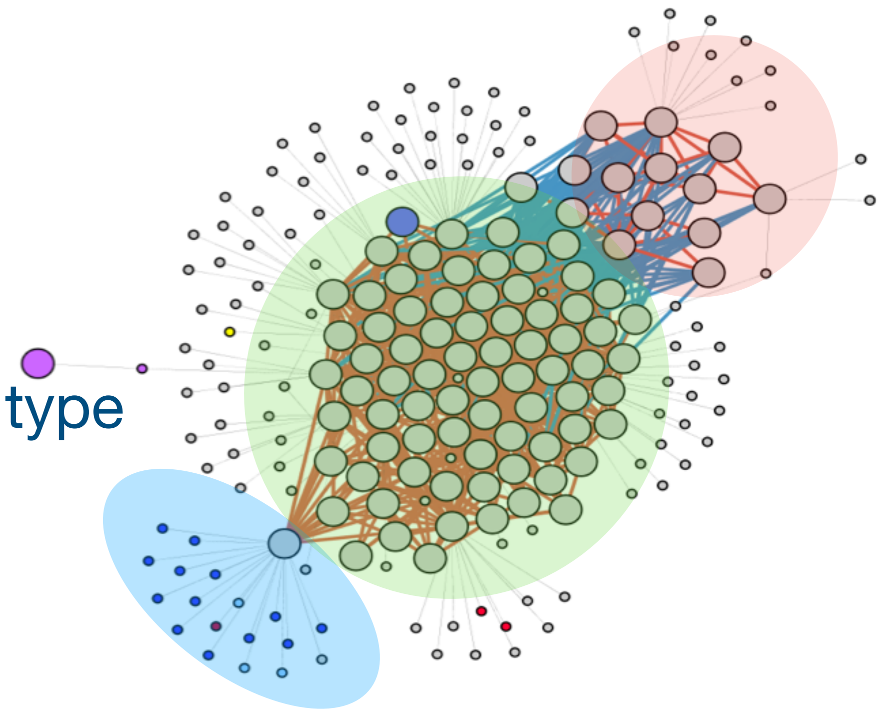


Examples:

- 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

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.

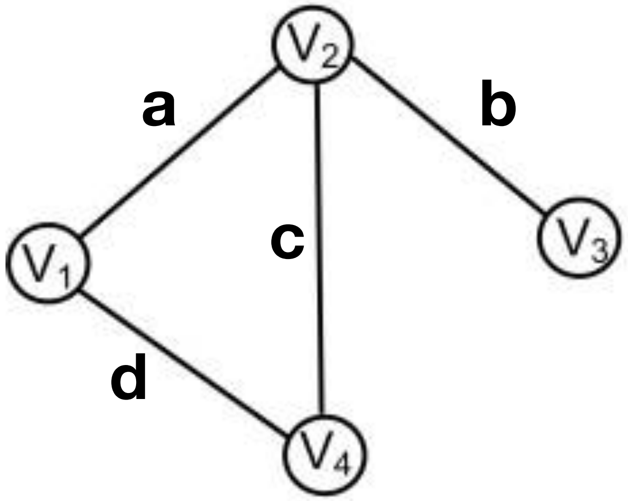
Graphs, nodes, edges

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

$G = (V, E)$

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

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



Nodes sometimes called **vertices**

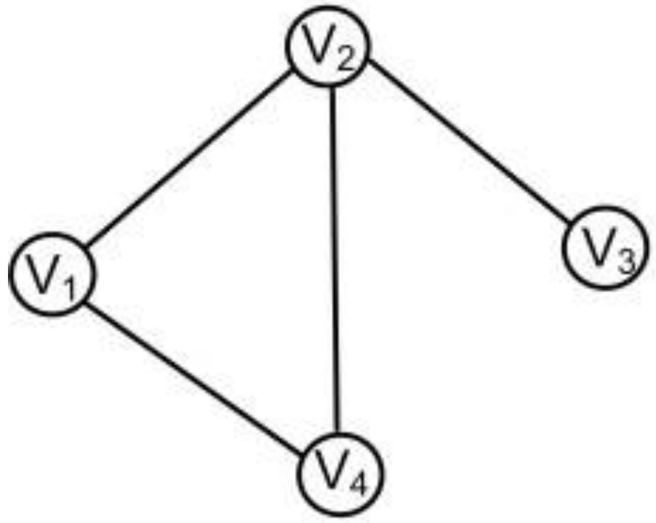
Two connected nodes are called **neighbours, adjacent,** or **end-nodes**

Simple vs multigraphs

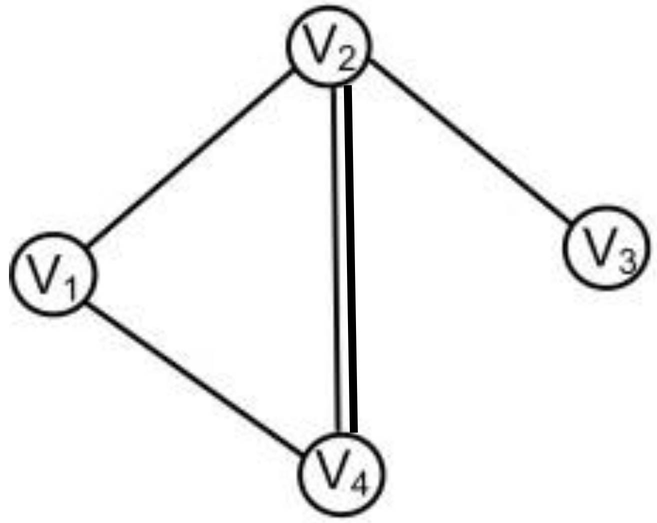
Multigraphs contain parallel edges

Multi-edged connections indicate different properties

Simple

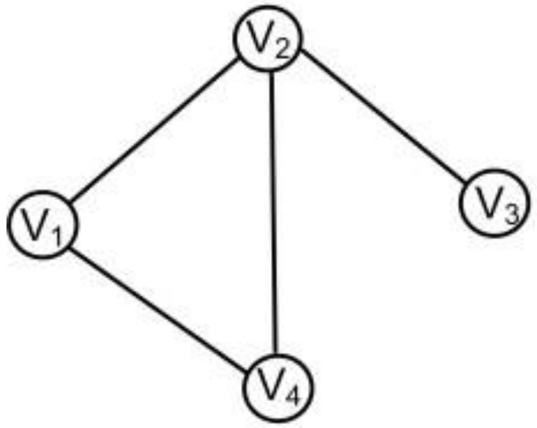


Multigraphs



Directed vs undirected graphs

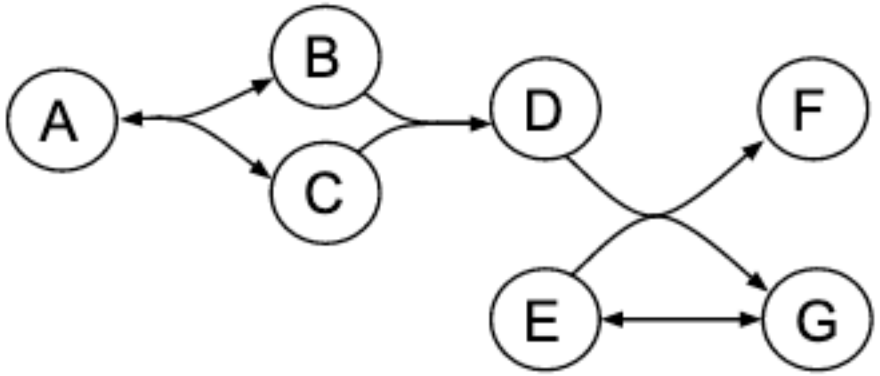
Undirected graphs:
co-expression networks



Directed graphs:
metabolic networks

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: $G \rightarrow E$

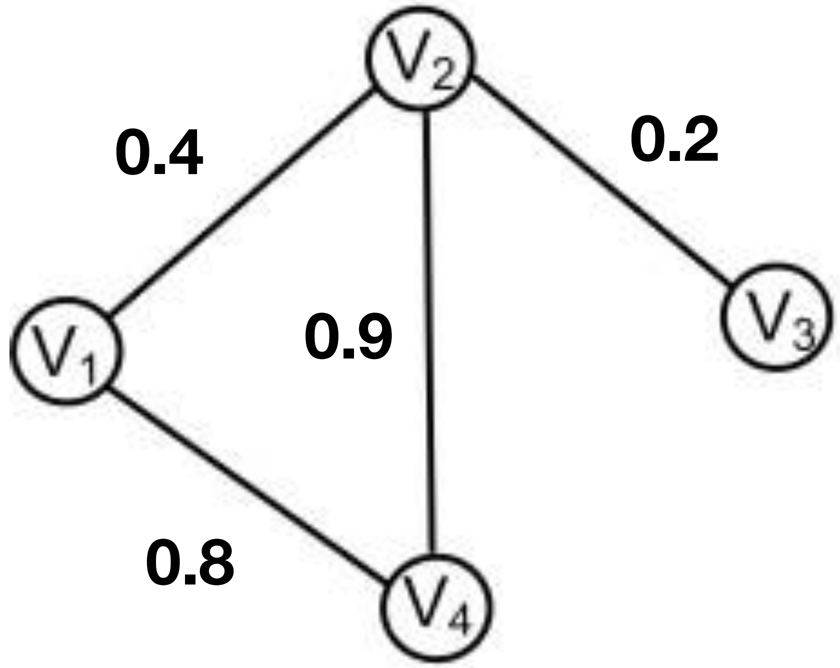
(a) Reaction network



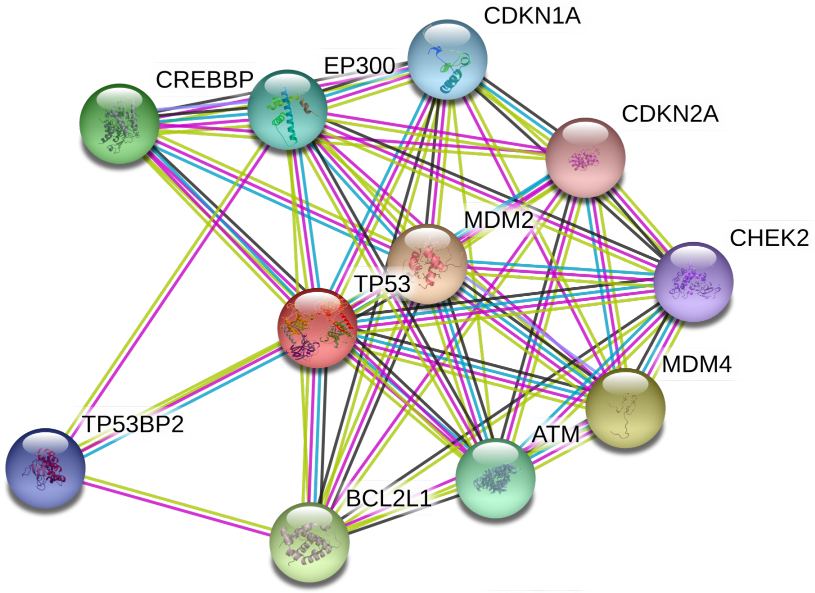
Weighted vs unweighted graphs

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

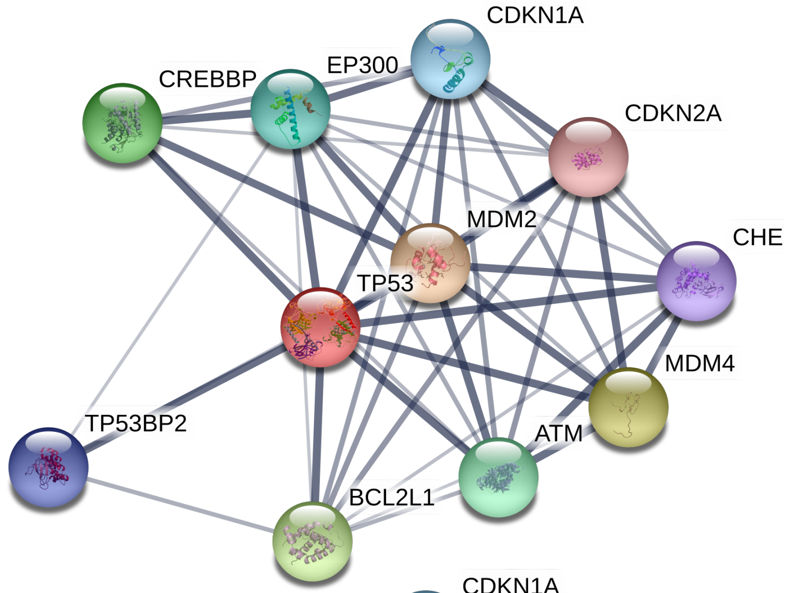
E.g. weighted co-expression networks



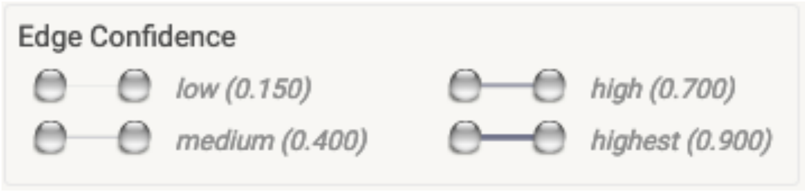
STRING-db.org: TP53



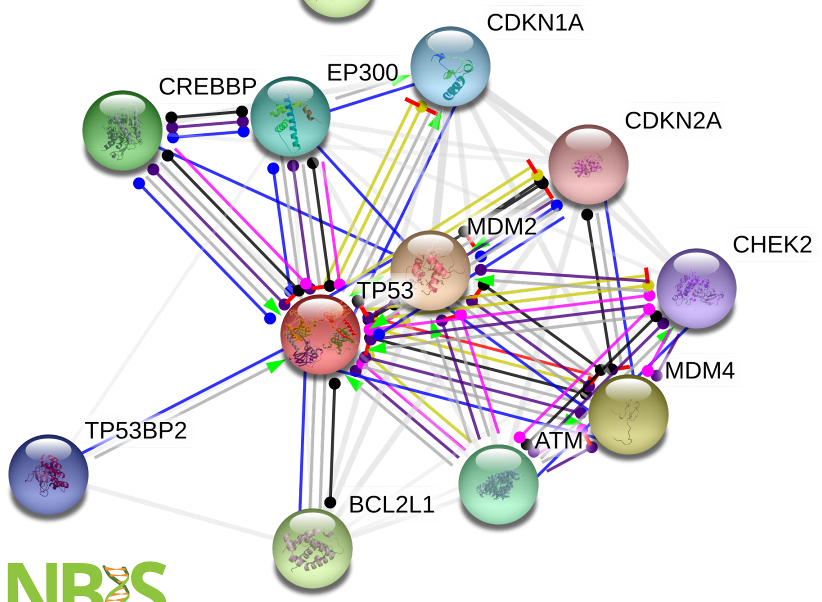
Multi-edged



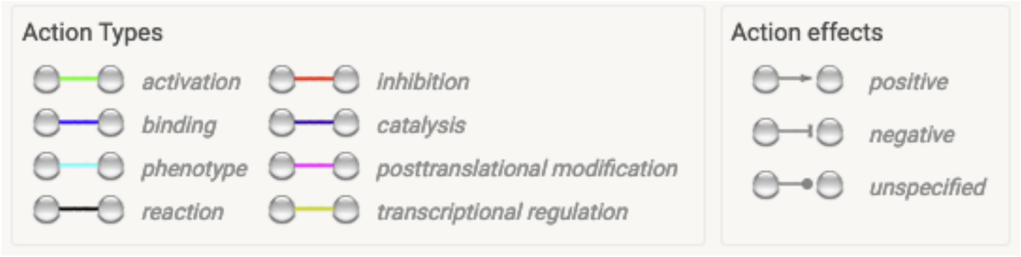
Weighted multi-edged


Edge Confidence

low (0.150)
medium (0.400)
high (0.700)
highest (0.900)



Multi-edged directed


Action Types

activation
binding
phenotype
reaction
inhibition
catalysis
posttranslational modification
transcriptional regulation

Action effects

positive
negative
unspecified

Bipartite graphs

A graph

$$G=(V,E)$$

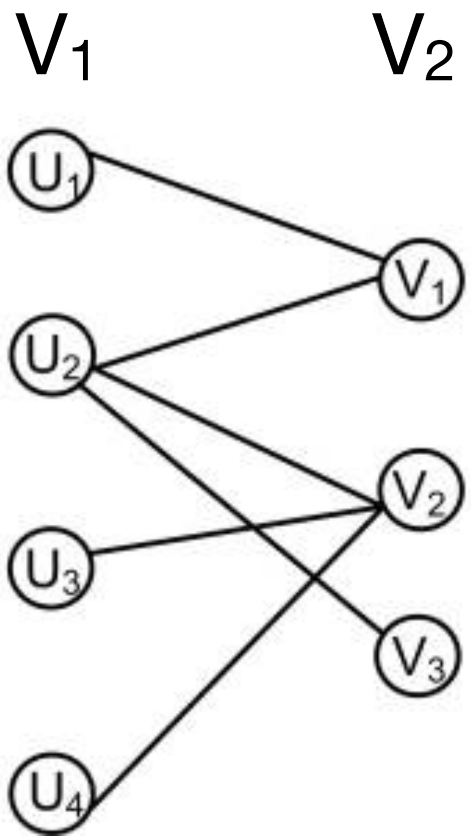
may be partitioned into two sets of nodes (V_1, V_2) such that

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

All e_i has end-nodes in V_1, V_2

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

$$G_1 = (V_1, E_1)$$



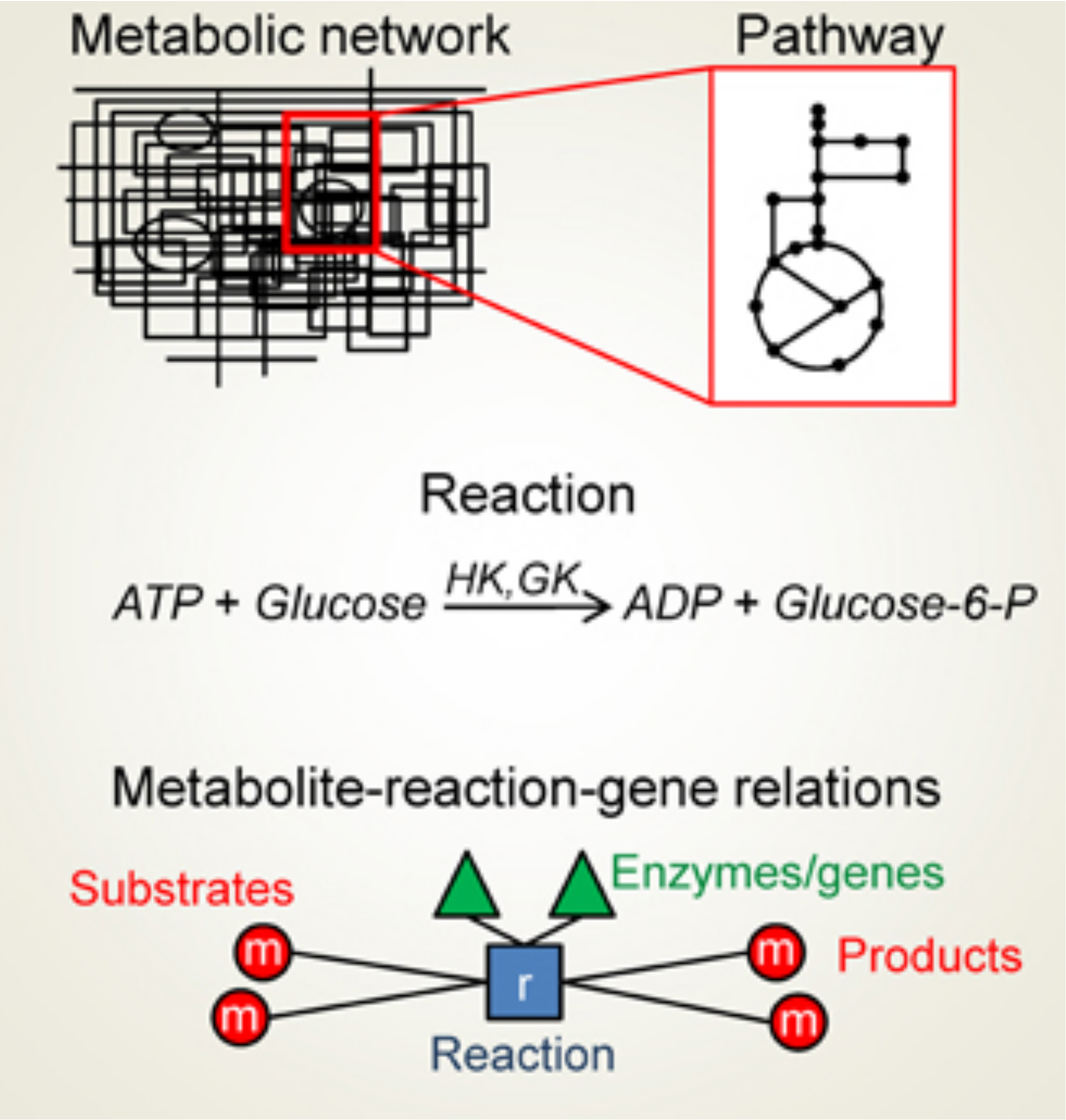
Bipartite and *k*-partite graphs

Example of k-partite graph:

Enzyme - Reaction

Metabolite - reaction - enzyme

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


Metabolic network
Pathway
Reaction
ATP + Glucose $\xrightarrow{HK,GK}$ ADP + Glucose-6-P
Metabolite-reaction-gene relations
Substrates
Enzymes/genes
Products
m
r
Reaction

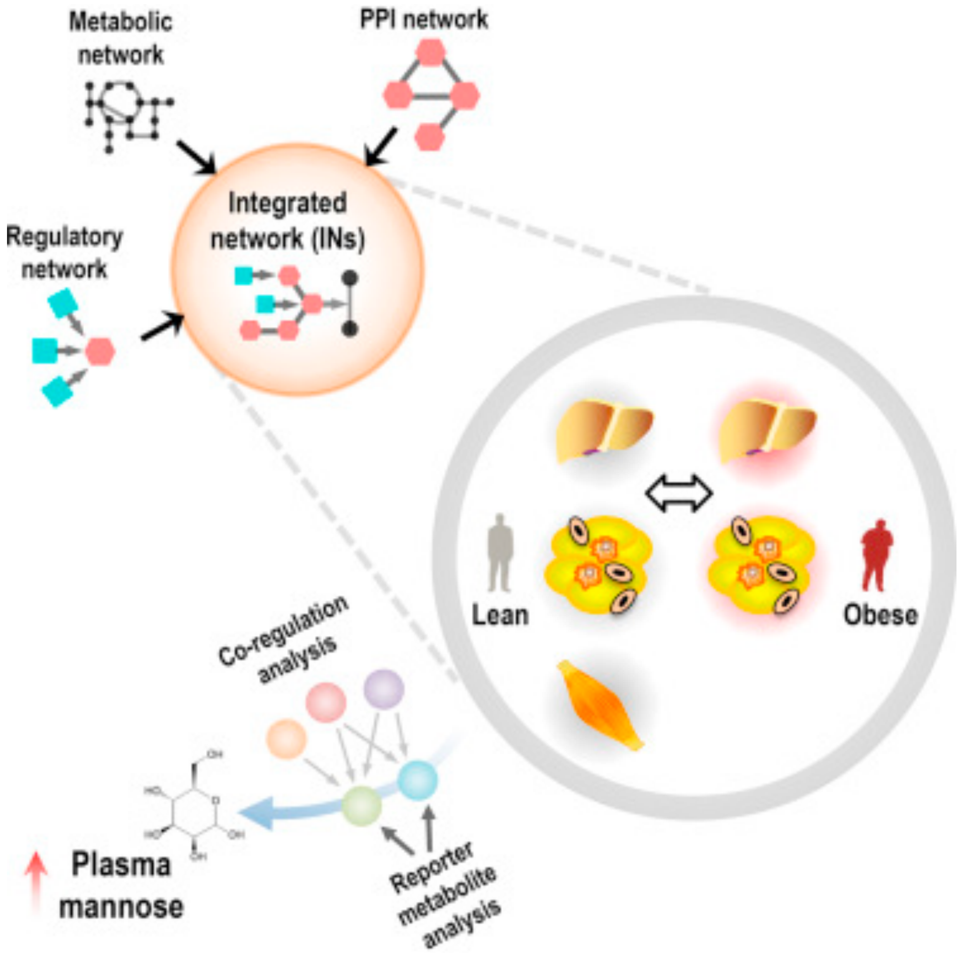
Väremo 2013

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



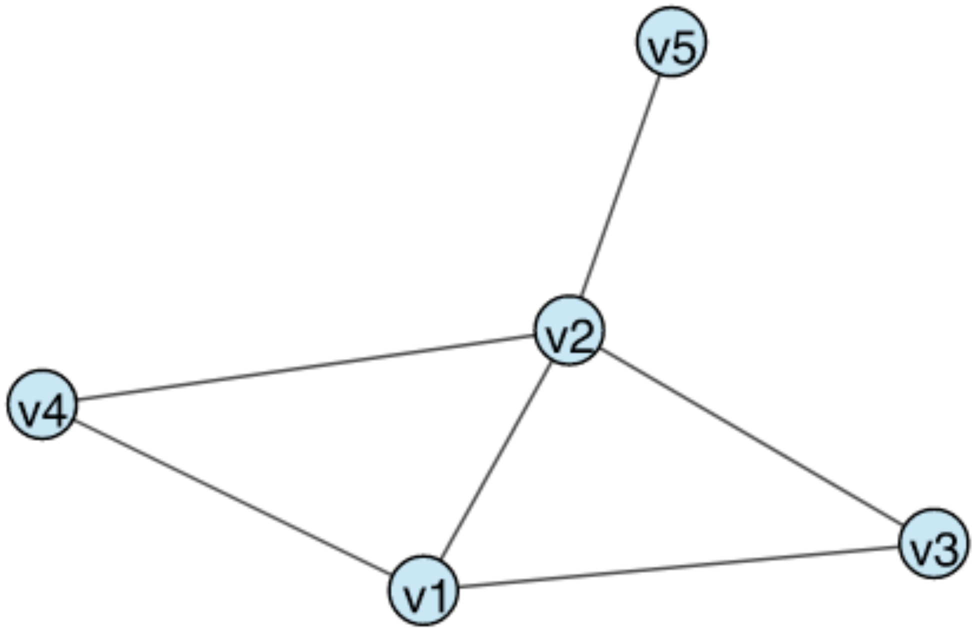
Lee 2016

Adjacency matrix (undirected graphs)

Vertex association

(undirected network)

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



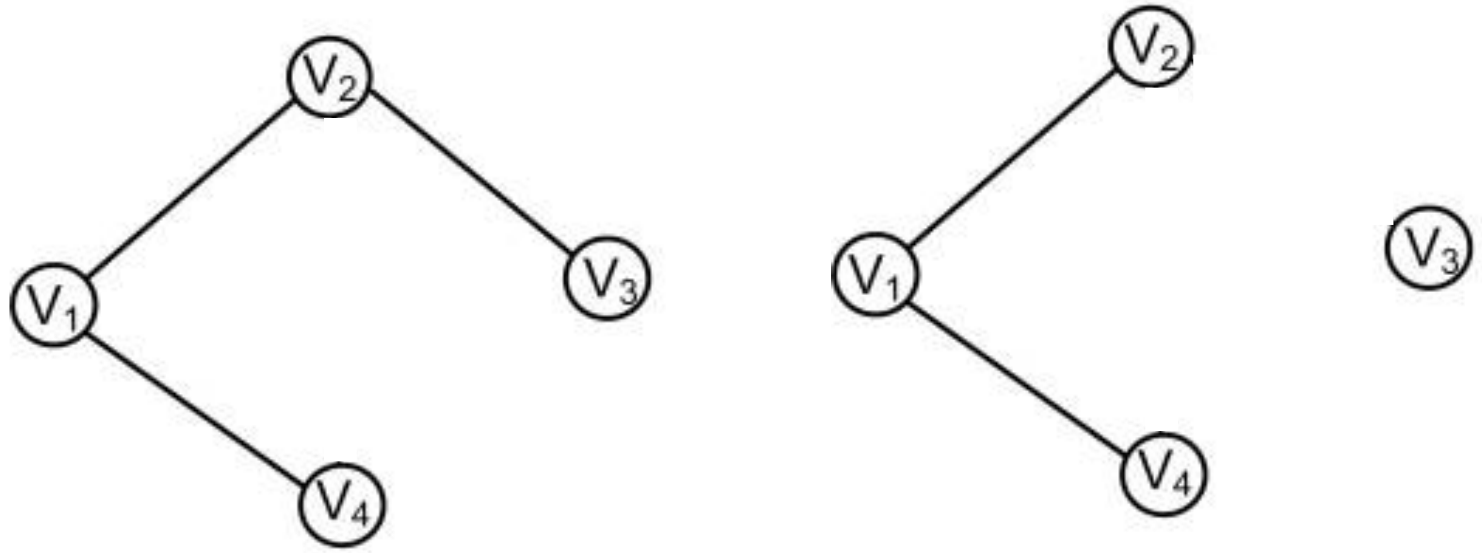
Adjacency matrix is symmetric

	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

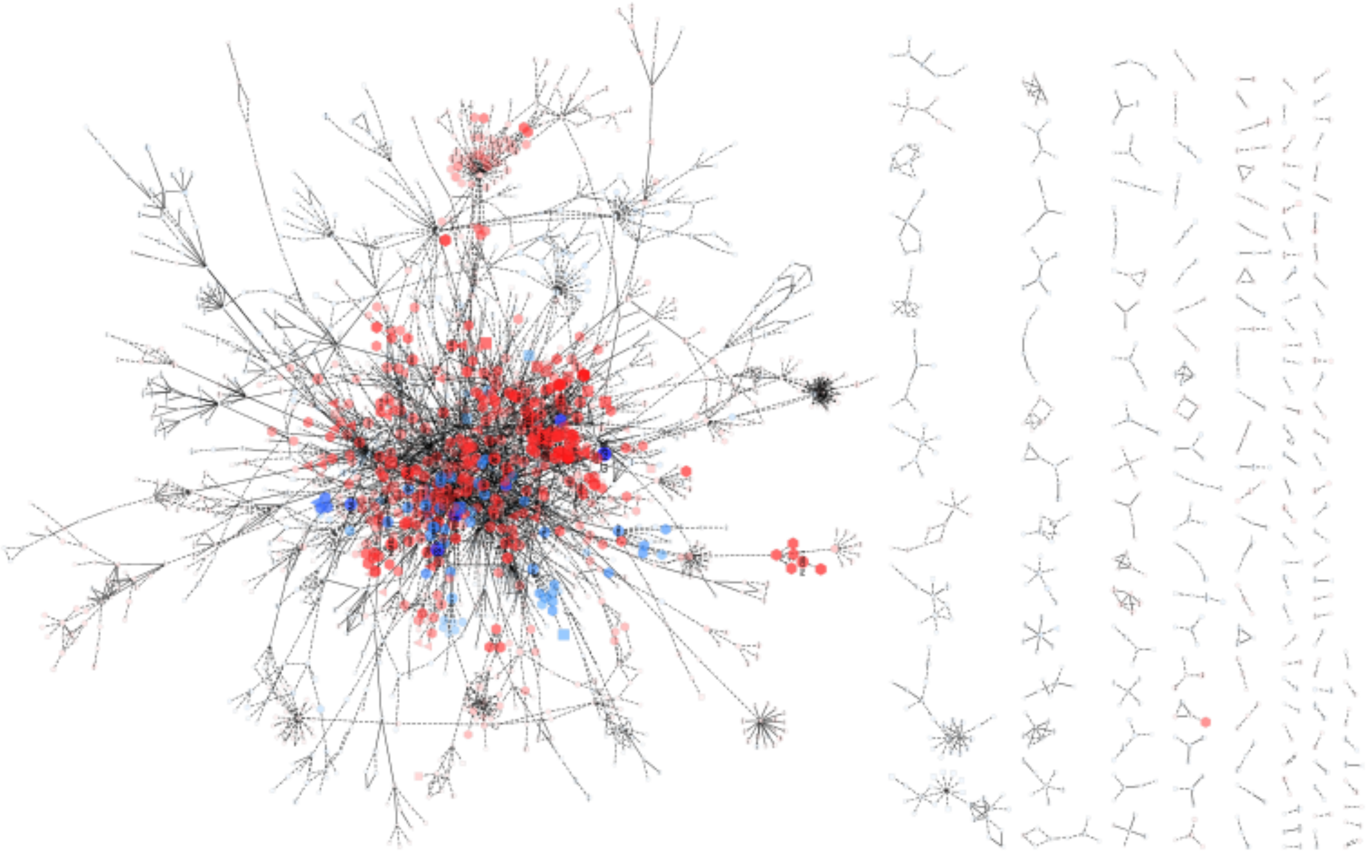
Disconnected network: some of the nodes are unreachable



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)**

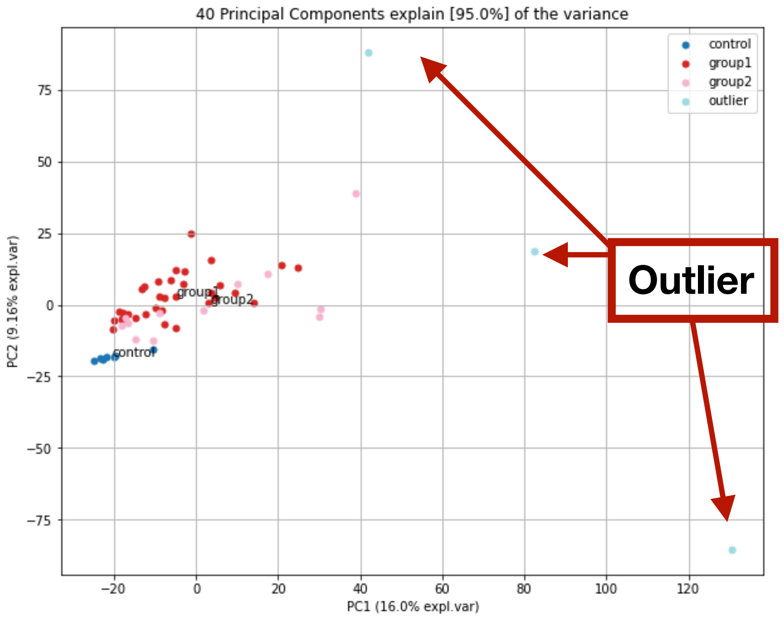


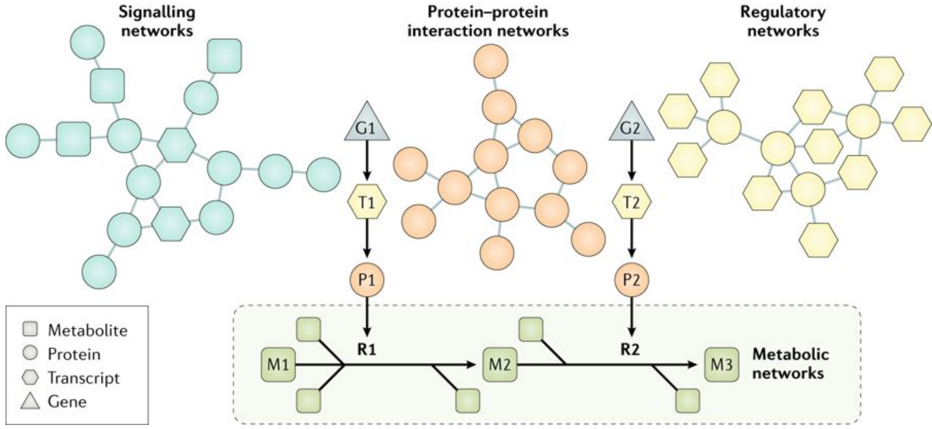
Overview

1. Introduction to network analysis
2. Terminology
3. **Network inference**
4. Key network properties
5. Community analysis

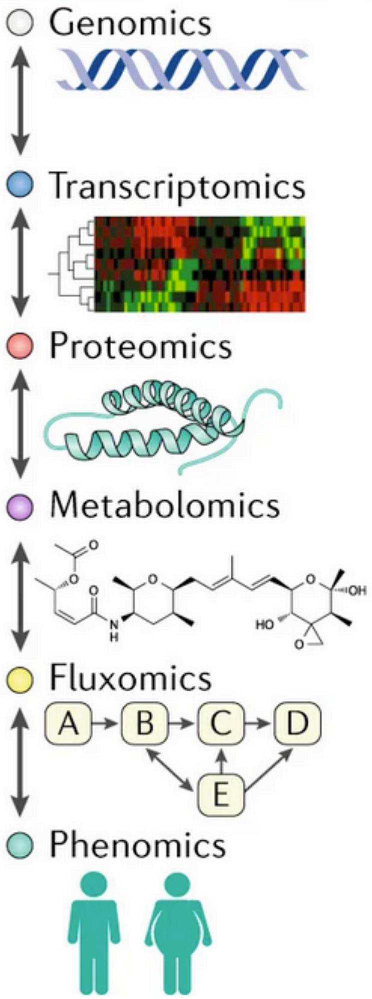
Original sources of images provided as reference and hyperlinks, where applicable.

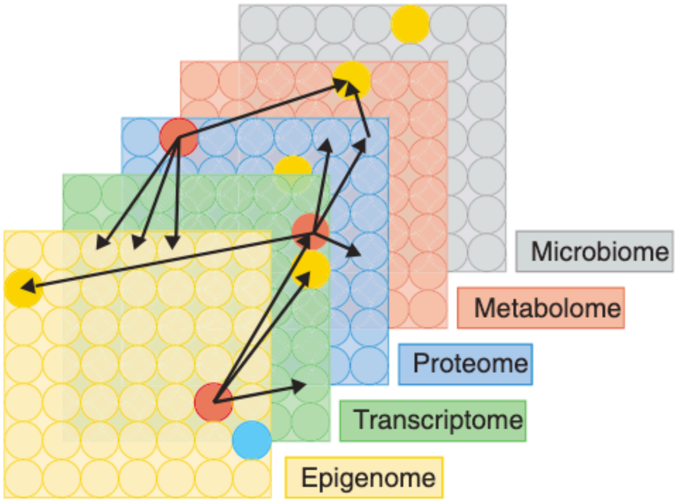
Building networks





Raw → Pre-processing → Distance calculation → Graph analysis





Hasin 2017

Piening 2018

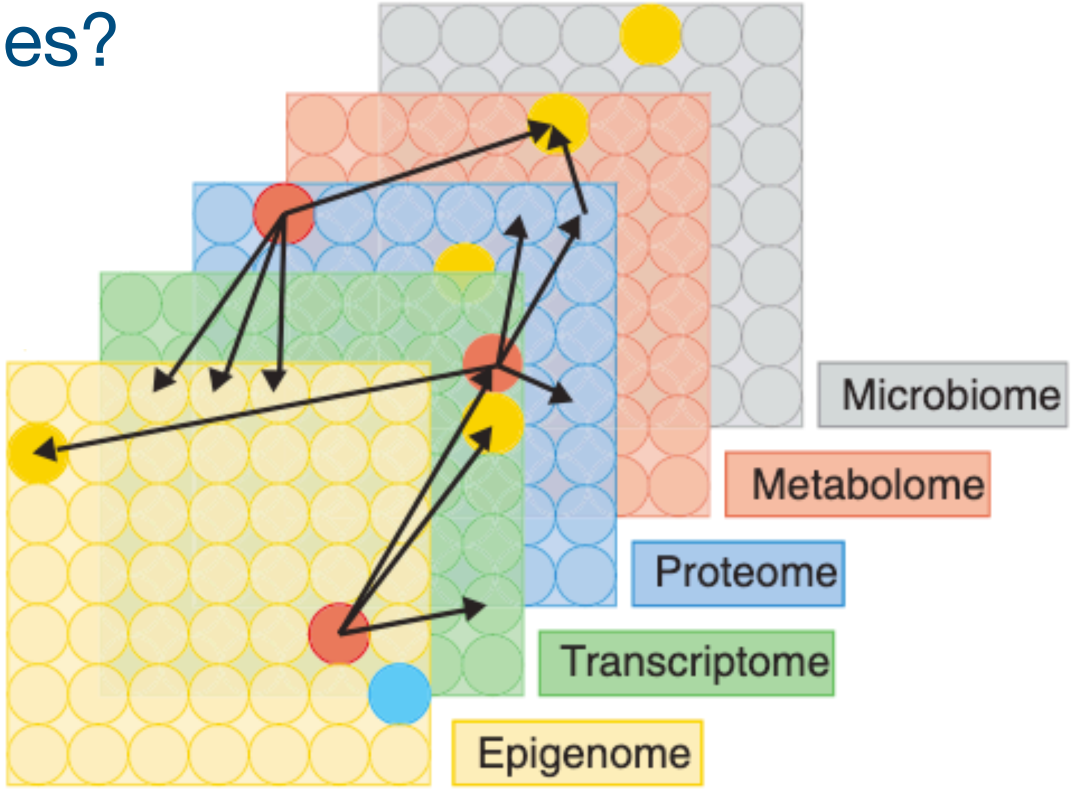
Mardinoglu 2018

Interomic vs Intraomic networks

Networks may be build for individual omics or for their integration

What is my biological question?

- Do I want to analyse vertical relationships between features?
- Biological motivation for integrating omics with different coverage (e.g. transcriptomic and proteomic)
- Do I want to extract functional properties?



Hasin 2017

Different approaches for network inference

1. Feature association
2. K-nearest neighbour graph (k-NNG) construction
3. Pathway-based
4. Genome-scale metabolic models
5. Network deconvolution

No prior graph structure

Based on available information

Filter indirect effects

1. Association analysis

Balanced dataset for group sizes

GroupA (80 samples) vs GroupB (20 samples)

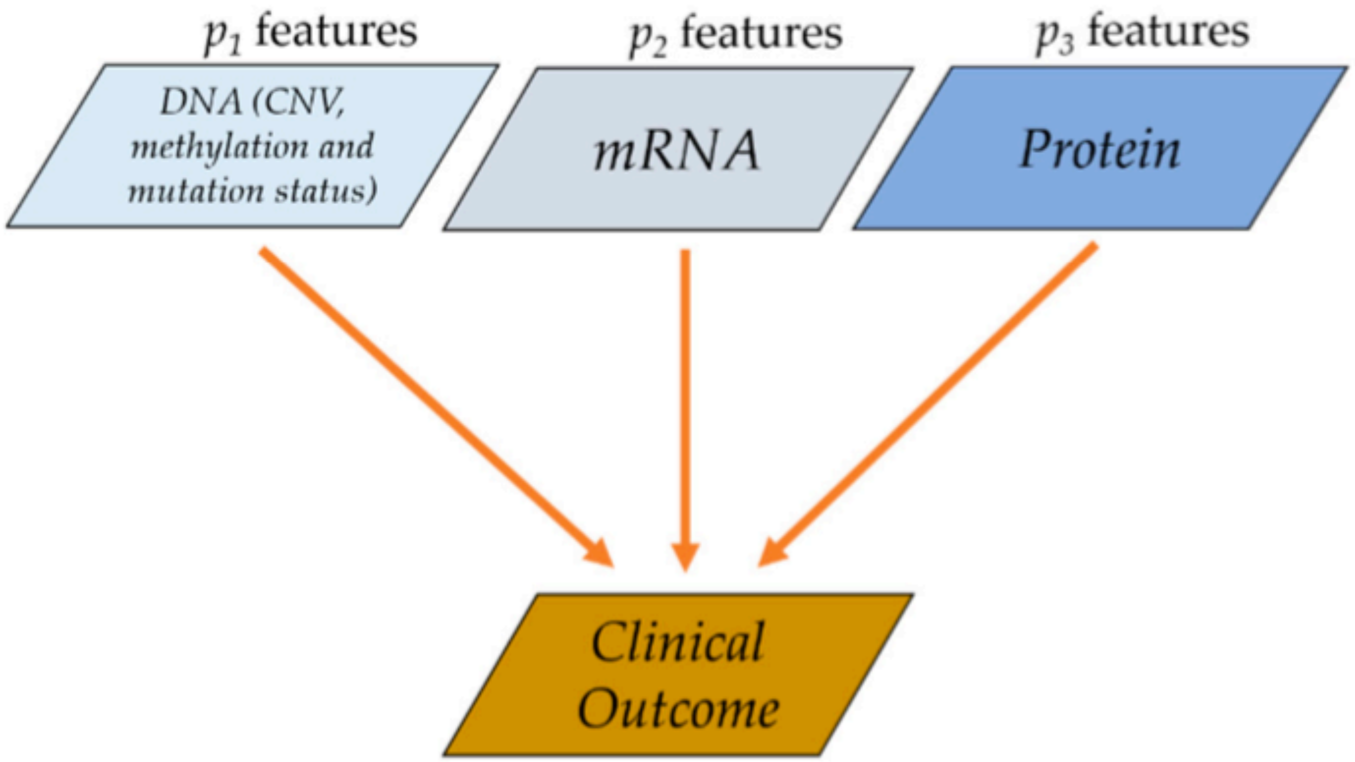
GroupA (50 samples) vs GroupB (50 samples)

Common approach: compute correlations between different features

- Spearman

- Pearson

Extend known associations



Wu 2019

1. Association analysis

Easy to interpret

Unweighted vs weighted $(-1 \leq \rho \leq 1)$

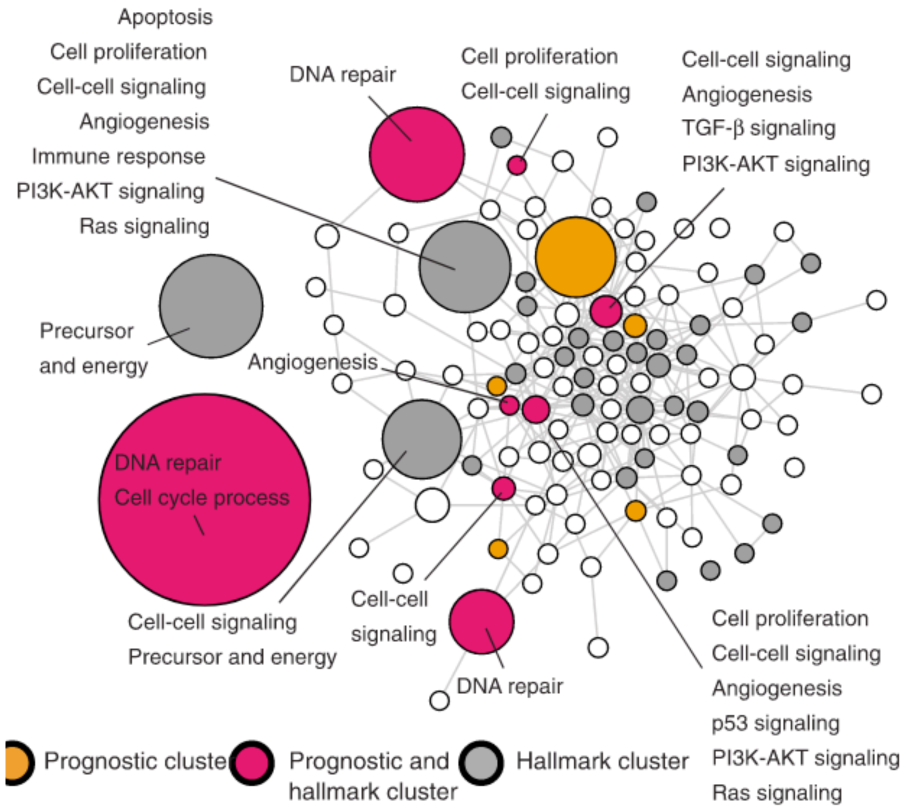
Unbalanced networks

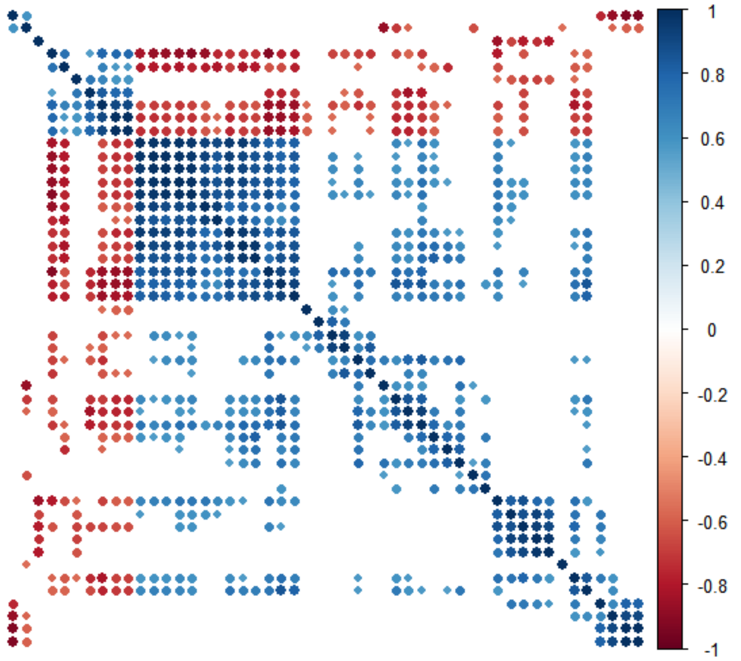
Prone to type I errors

Filtering

- FDR vs Bonferroni
- Effect size cut-off

Need adjustment to possible confounding factors





Wu 2019
Uhlen 2017

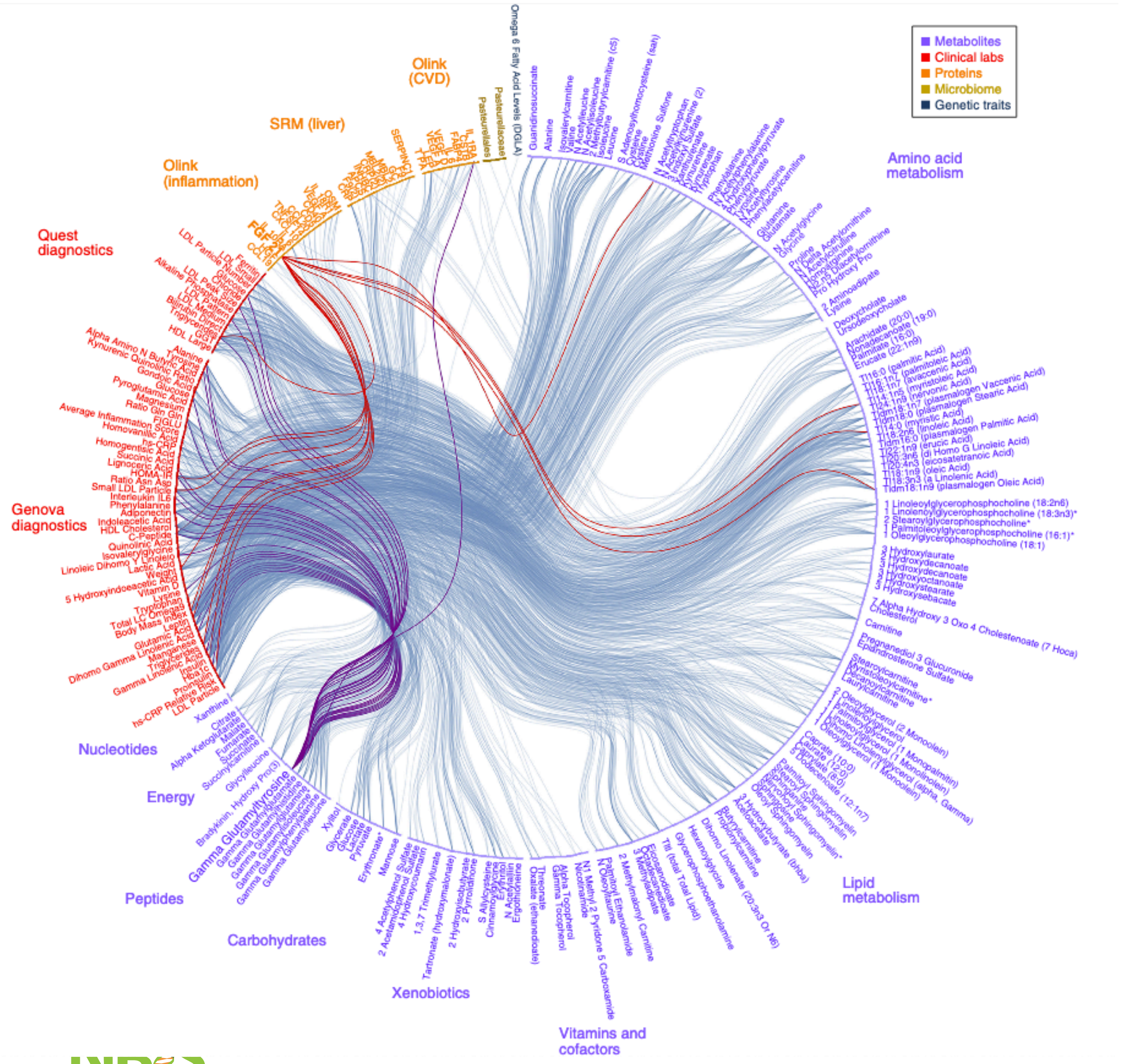
1. Association analysis

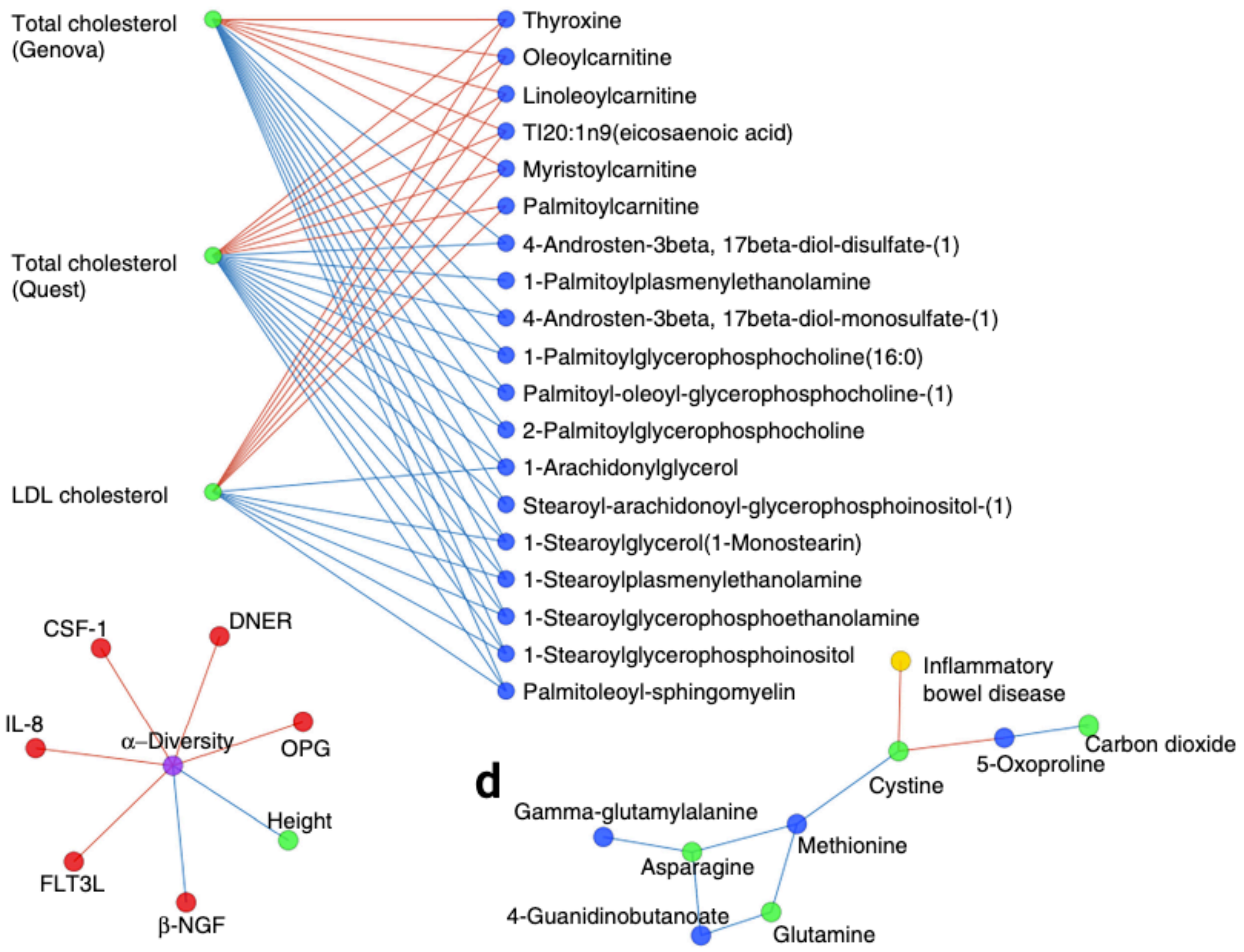
Adjusting for confounding factors: partial correlation analysis

Below:

- gender and age are known confounding factors

- feature regression on confounding factors, followed by correlation on the residuals of each model





Total cholesterol (Genova)
Total cholesterol (Quest)
LDL cholesterol
Thyroxine
Oleoylcarnitine
Linoleoylcarnitine
TI20:1n9(eicosaenoic acid)
Myristoylcarnitine
Palmitoylcarnitine
4-Androsten-3beta, 17beta-diol-disulfate-(1)
1-Palmitoylplasmenylethanolamine
4-Androsten-3beta, 17beta-diol-monosulfate-(1)
1-Palmitoylglycerophosphocholine(16:0)
Palmitoyl-oleoyl-glycerophosphocholine-(1)
2-Palmitoylglycerophosphocholine
1-Arachidonylglycerol
Stearoyl-arachidonoyl-glycerophosphoinositol-(1)
1-Stearoylglycerol(1-Monostearin)
1-Stearoylplasmenylethanolamine
1-Stearoylglycerophosphoethanolamine
1-Stearoylglycerophosphoinositol
Palmitoleoyl-sphingomyelin
CSF-1
DNER
IL-8
α–Diversity
OPG
Height
FLT3L
β-NGF
Inflammatory bowel disease
Carbon dioxide
5-Oxoproline
Cystine
d
Gamma-glutamylalanine
Methionine
Asparagine
4-Guanidinobutanoate
Glutamine

Price 2016

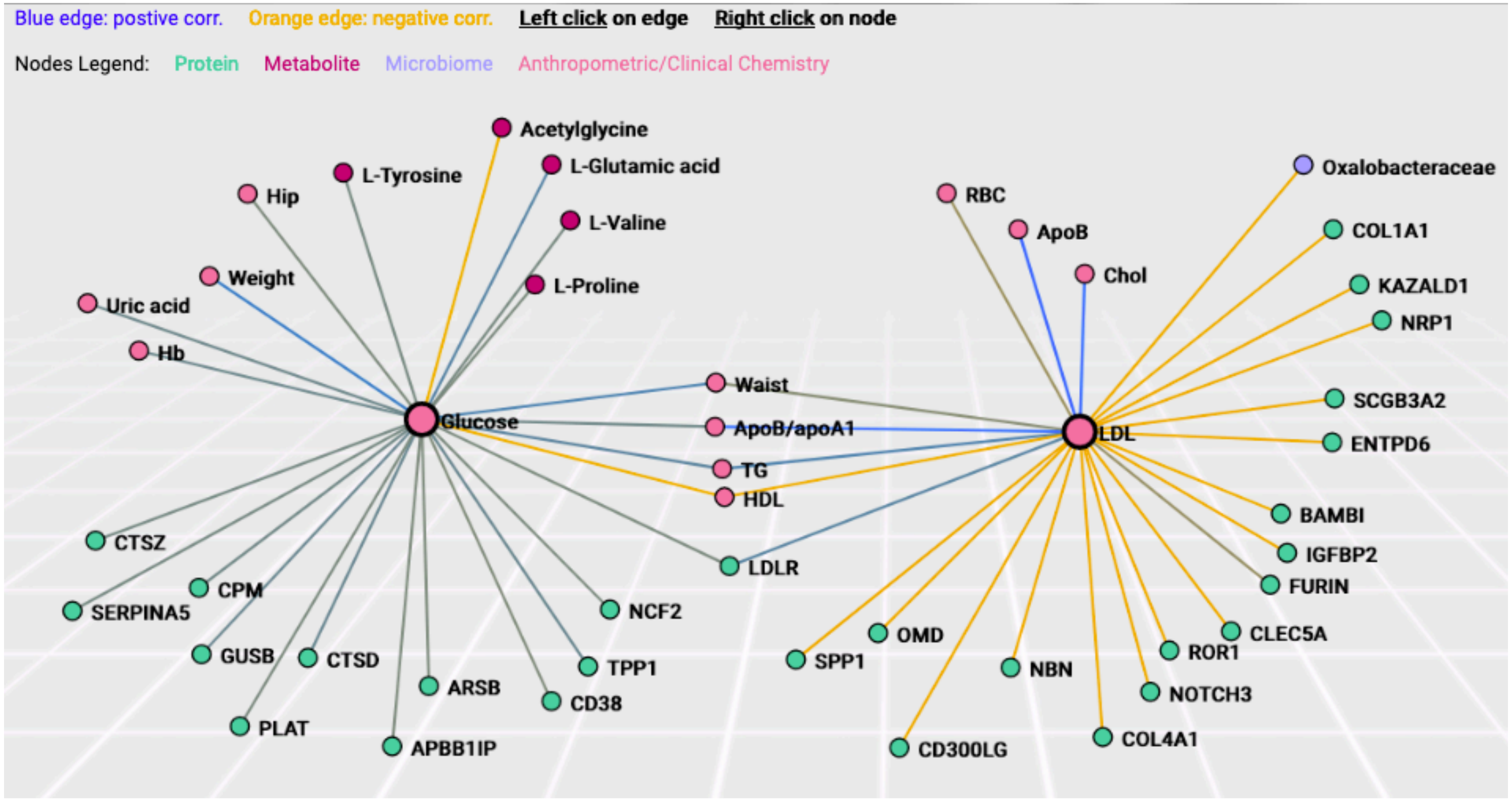
Overview

1. Introduction to network analysis
2. Terminology
3. Network inference
4. **Key network properties**
5. Community analysis

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

Motivation

You have built an association network (e.g. PPI, multi-omic). How to identify pivotal features, their organization, and biological characteristics?



multiomics.inetmodels.com/

Key network properties

1. **Network representations**
2. **Network density**
3. **Paths**
4. **Centrality**
5. **Clustering coefficient**
6. **Degree and connectivity distributions**

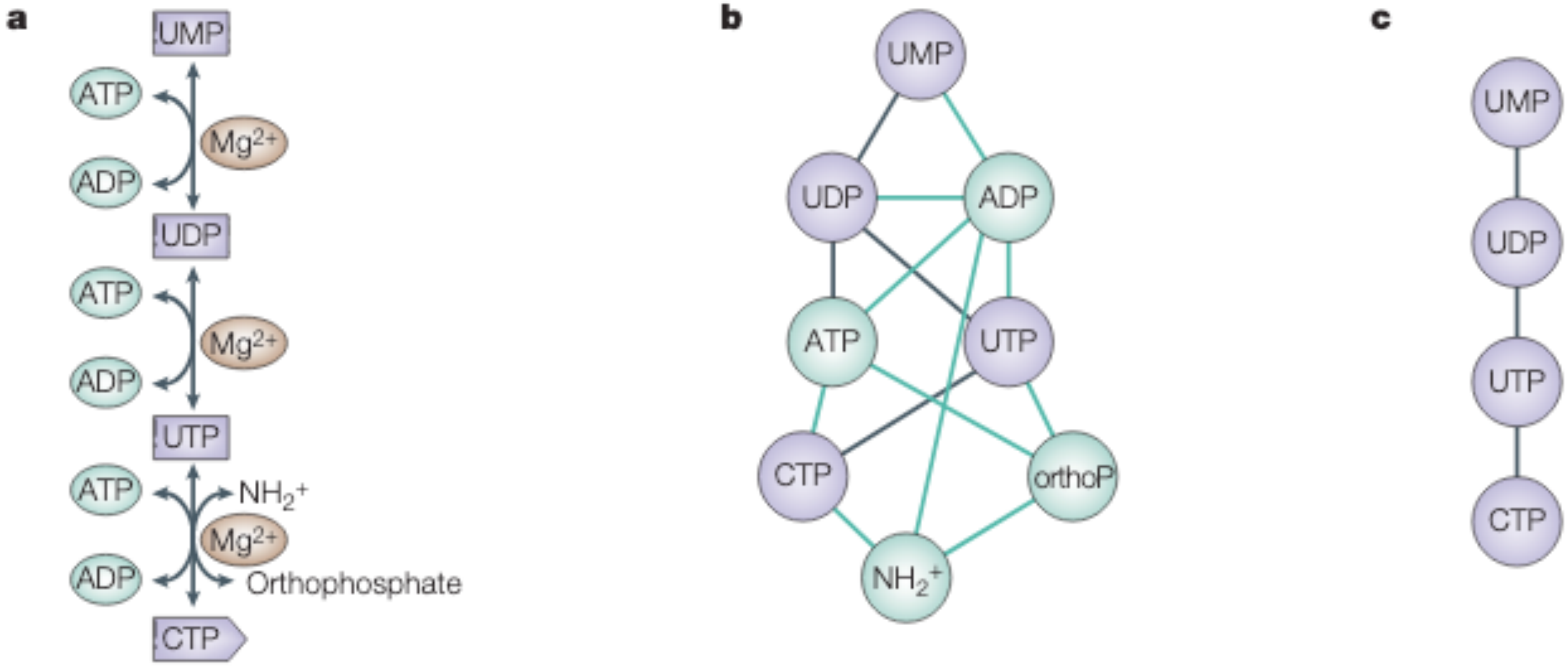
1. Network representations

Representations of a metabolic network: pyrimidine metabolism

Metabolism

Graph representation: metabolites and co-factors

metabolite-metabolite association



(directed graph)

(undirected graph)

(undirected graph)

Other representations: Protein-Protein, Protein-Metabolite

2. Network density

A **dense graph** is a graph where the number of edges approximates the maximum possible number of edges for the given node number.

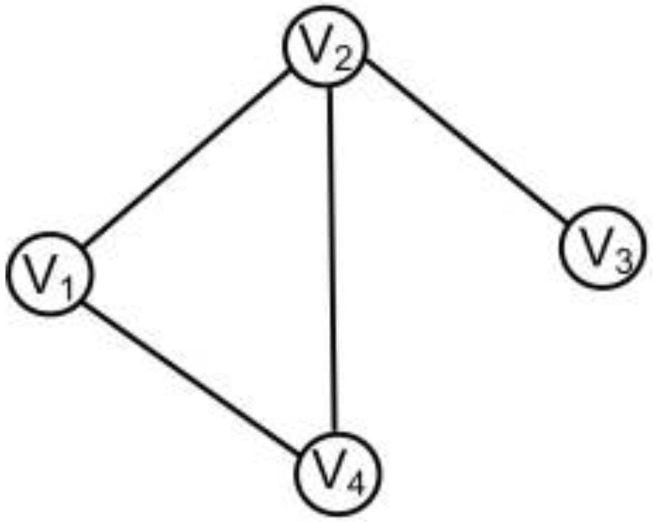
We can thus compute the network **density** (or **global connectivity**) as

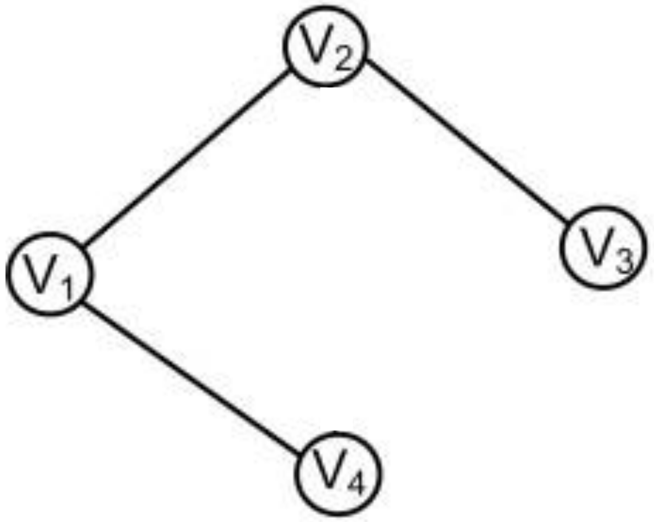
$$\textbf{Undirected graphs: } D = \frac{2 * E}{V \cdot (V - 1)}$$

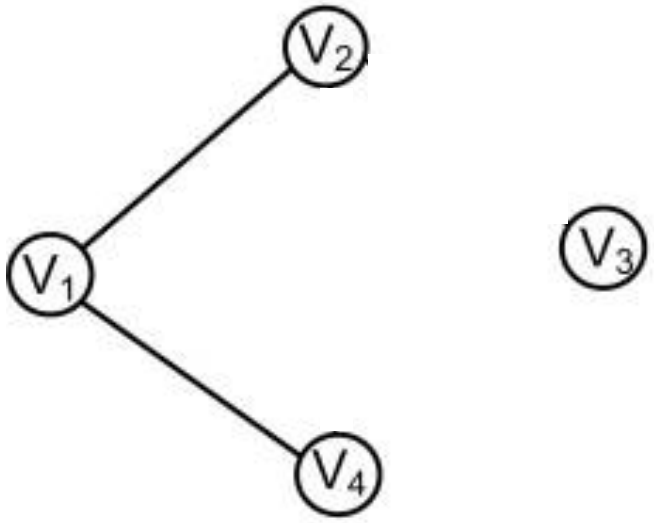
E : number of edges

V : number of vertices

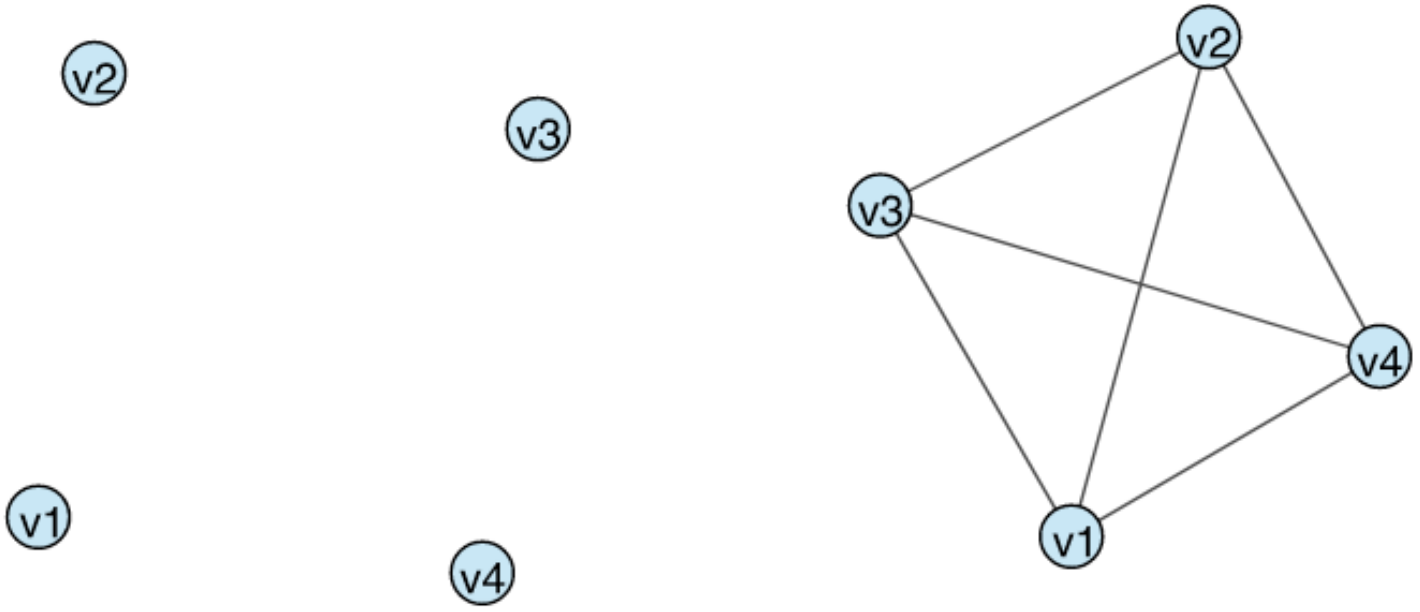
$$\text{Possible edges} = \frac{V \cdot (V - 1)}{2}$$





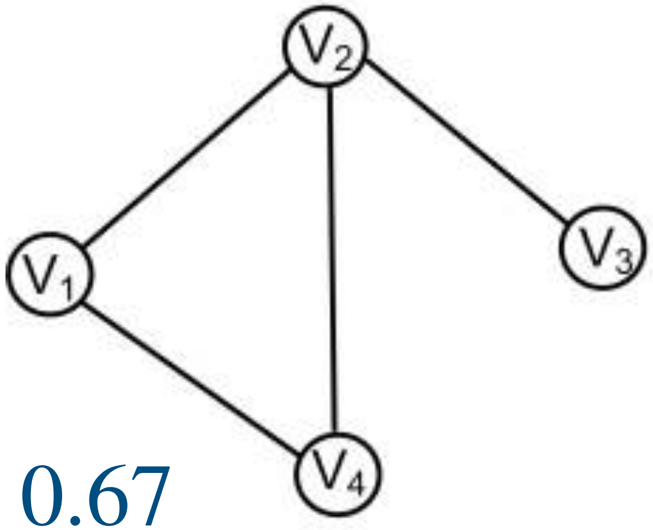


2. Network density

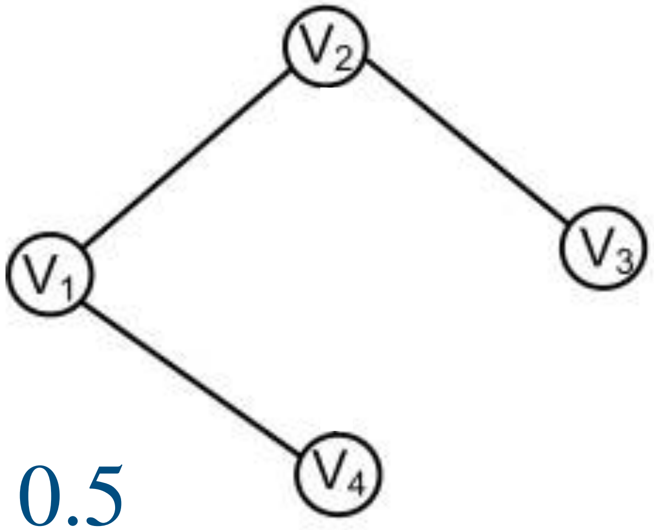


$$0 \leq D \leq 1$$

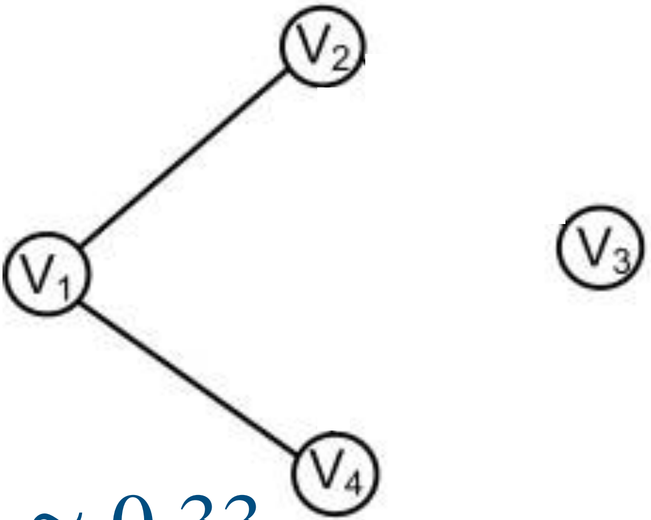
Higher density indicates higher associations in the network, which implies lower resilience to changes.



$D \approx 0.67$



$D = 0.5$



$D \approx 0.33$

2. Biological network density

Evolutionary analysis of biological networks indicates general sparsity

Network structure must balance robustness to mutation, stochasticity and environmental queues

Sparse networks show higher robustness when accounting for costs and benefits of complexity

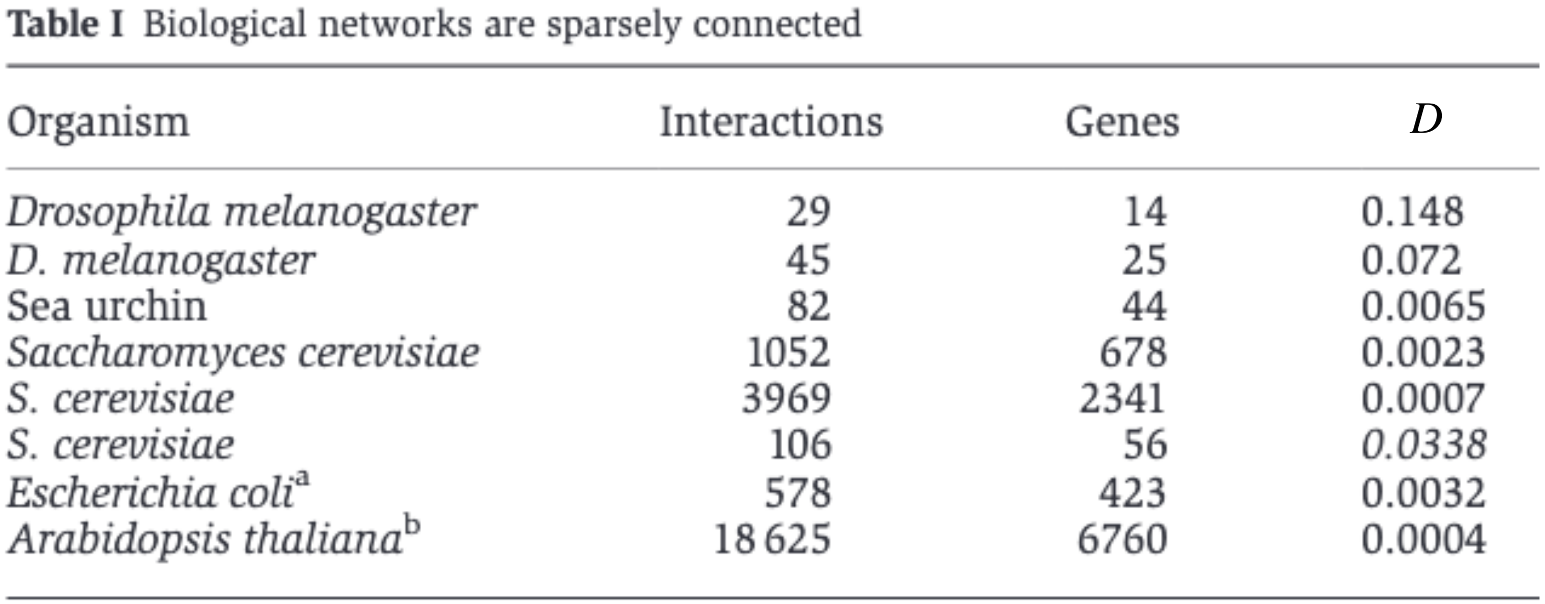
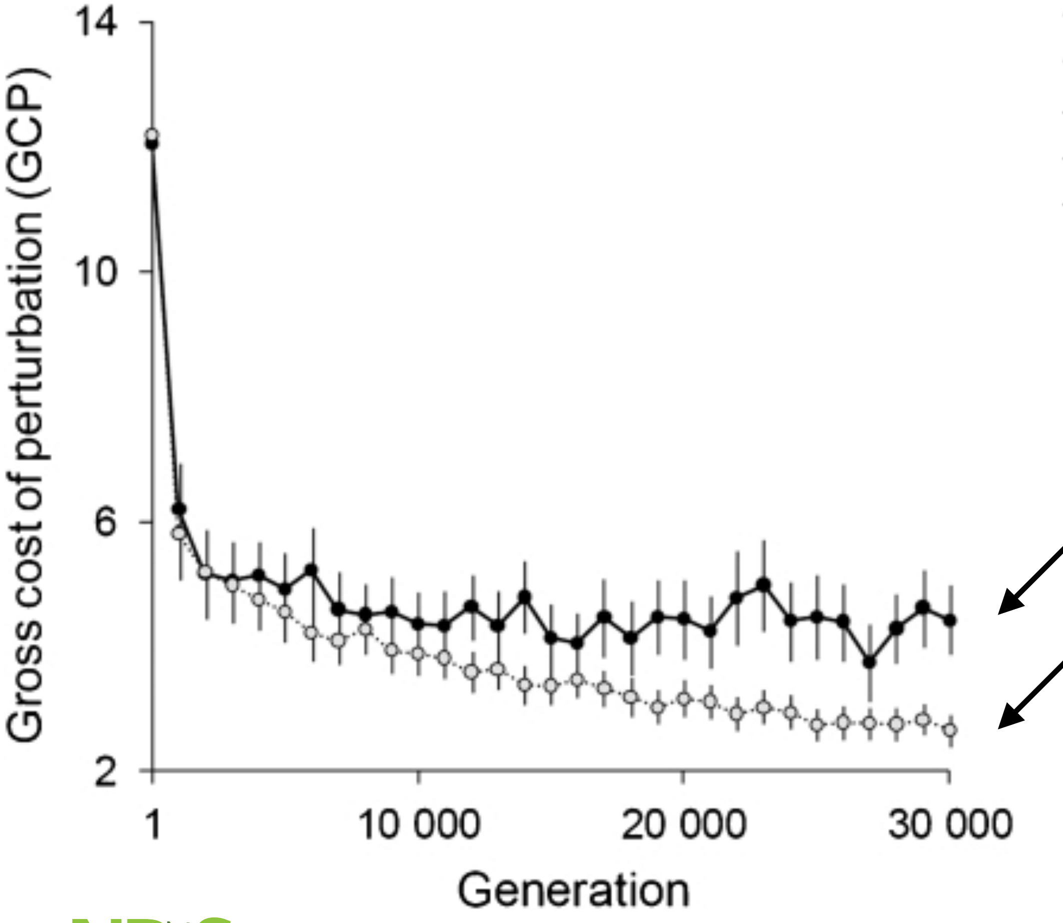


Table I Biological networks are sparsely connected

Organism	Interactions	Genes	*D*
Drosophila melanogaster	29	14	0.148
D. melanogaster	45	25	0.072
Sea urchin	82	44	0.0065
Saccharomyces cerevisiae	1052	678	0.0023
S. cerevisiae	3969	2341	0.0007
S. cerevisiae	106	56	*0.0338*
Escherichia coli[a]	578	423	0.0032
Arabidopsis thaliana[b]	18 625	6760	0.0004



dense network (density 0.9)

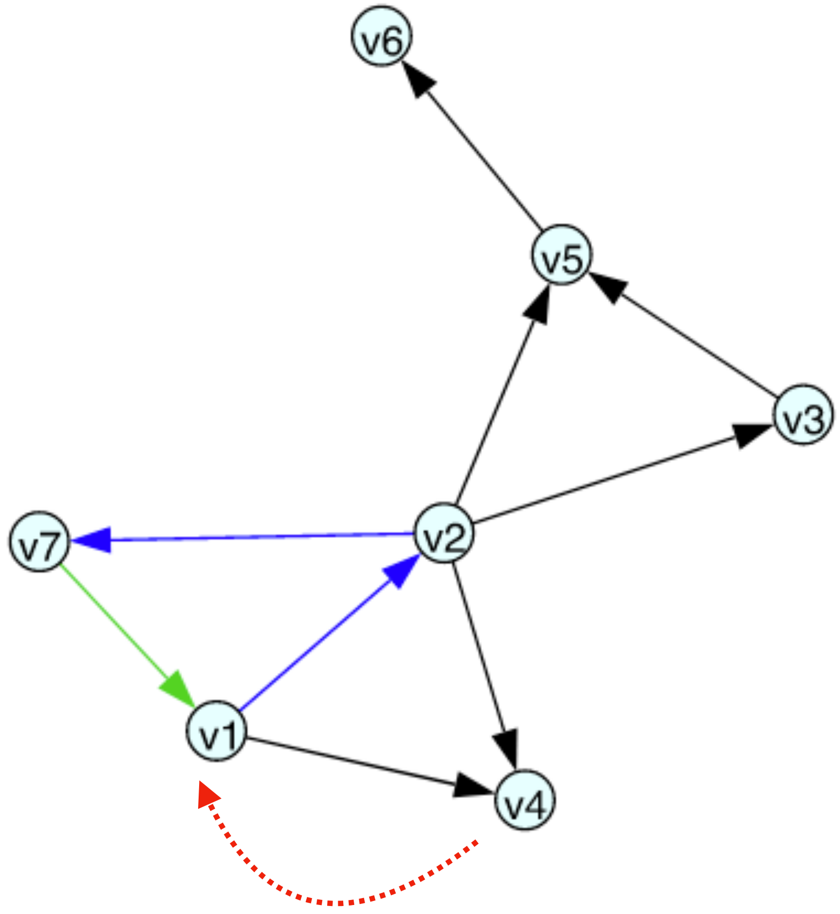
sparse network (density 0.1)

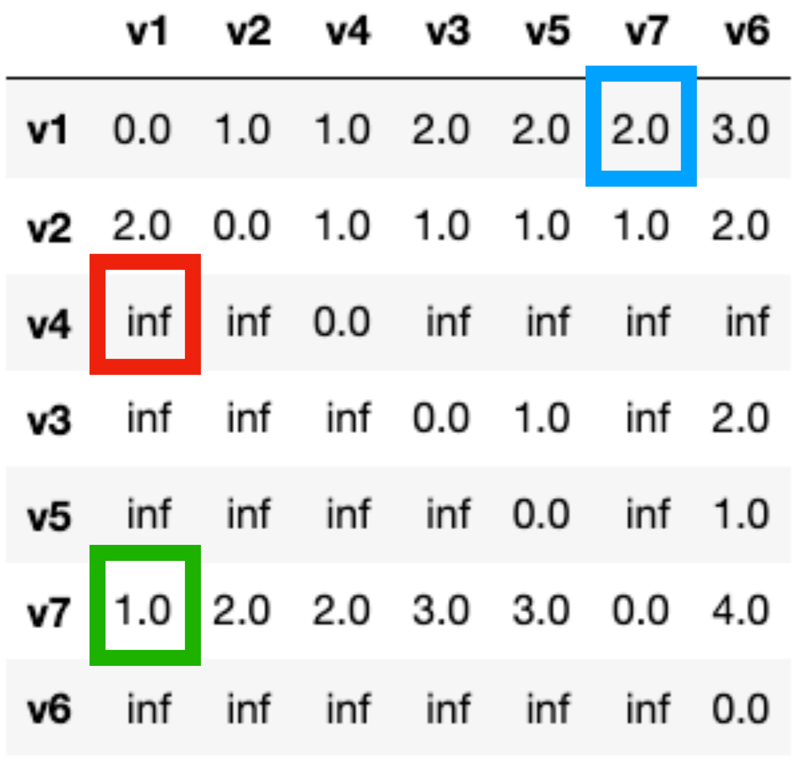
Leclerc 2008

3. Paths

Distance between nodes is measured in path length

In directed graphs, the shortest path between $(a, b) \neq (b, a)$



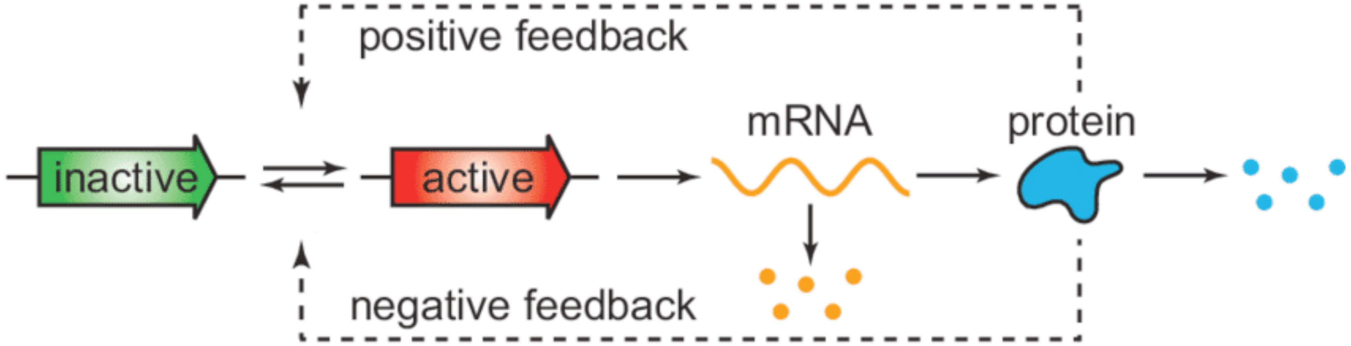


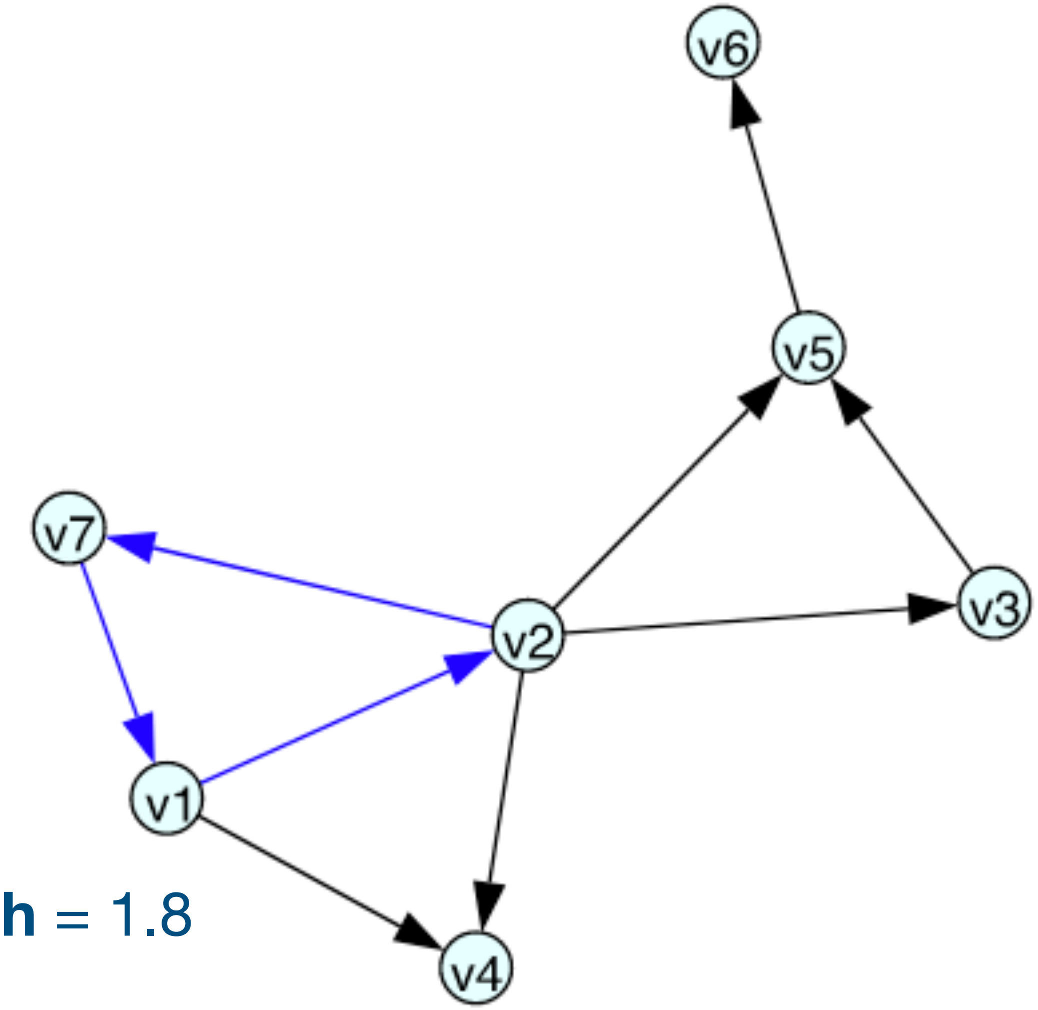
	v1	v2	v4	v3	v5	v7	v6
v1	0.0	1.0	1.0	2.0	2.0	2.0	3.0
v2	2.0	0.0	1.0	1.0	1.0	1.0	2.0
v4	inf	inf	0.0	inf	inf	inf	inf
v3	inf	inf	inf	0.0	1.0	inf	2.0
v5	inf	inf	inf	inf	0.0	inf	1.0
v7	1.0	2.0	2.0	3.0	3.0	0.0	4.0
v6	inf	inf	inf	inf	inf	inf	0.0

3. Paths

Cycles and acyclic graphs

The **average path** gives a measure of network navigability (~feature relationships)





Average path length = 1.8

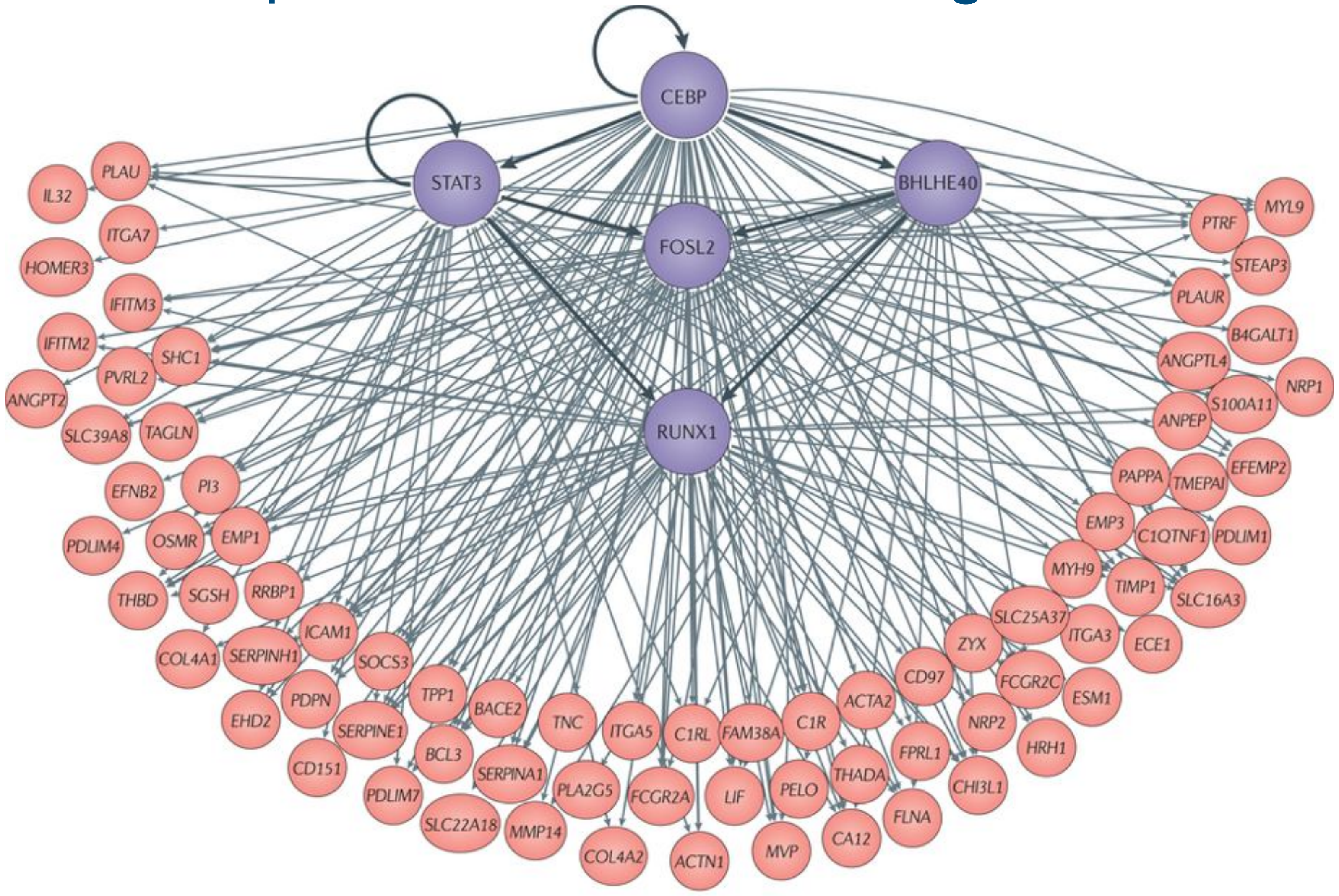
4. Centrality

Indicate the most central nodes in a network

Why look at the central nodes?

Hubs

Example: Transcription Factor Master Regulators



Nature Reviews | Cancer

Tyagarajan 2014

4. Centrality

Indicate the most central nodes in a network

Central nodes **possibly** important in the network

There are many different measures of centrality:

- **Degree**
- **Eccentricity**
- *Closeness*
- *Betweenness*
- *Eigenvector*
- Katz
- PageRank
- Percolation
- Cross-clique

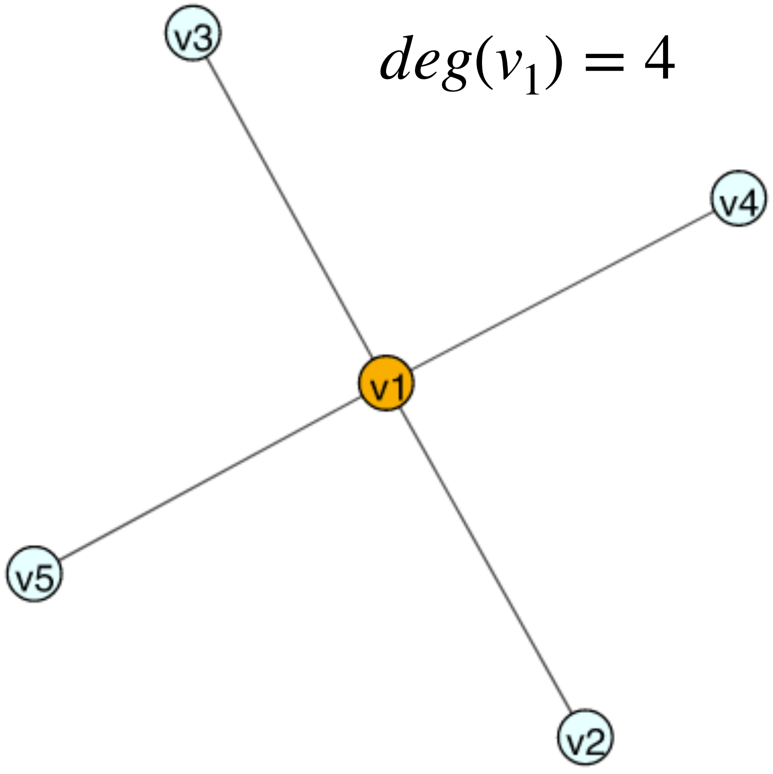
…

4. Centrality: degree centrality

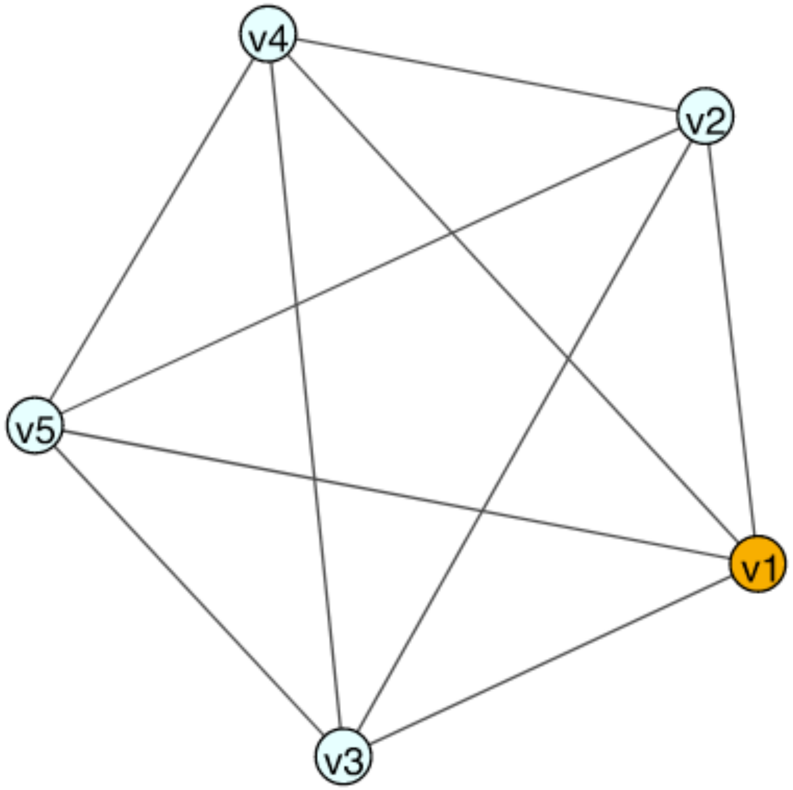
Degree indicates the number of connections with a node

$$d(v) = |N(i)|$$

where *N(i)* is the number of 1st neighbours of a node.



$deg(v_1) = 4$



$deg(v_1) = 4$

4. Centrality: degree centrality

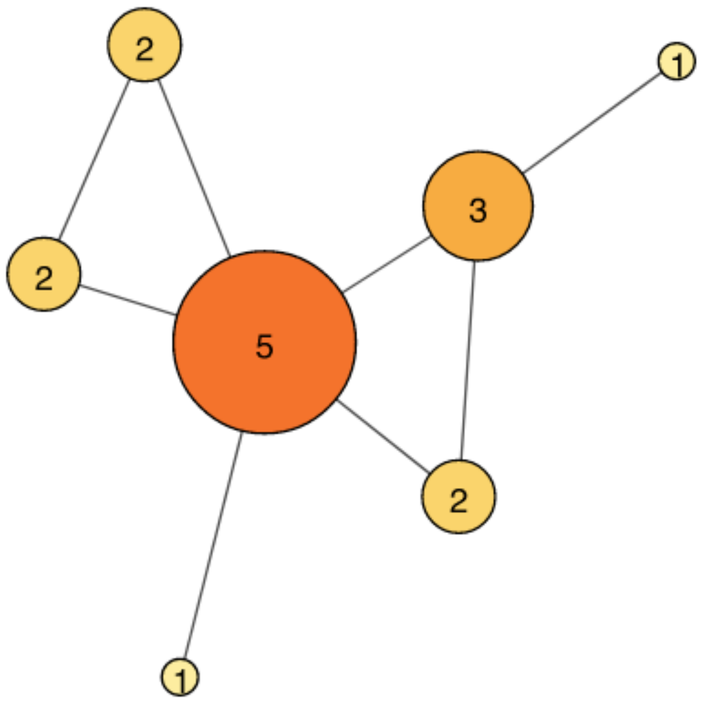
Undirected networks vs directed networks

In-degree vs **Out-degree**

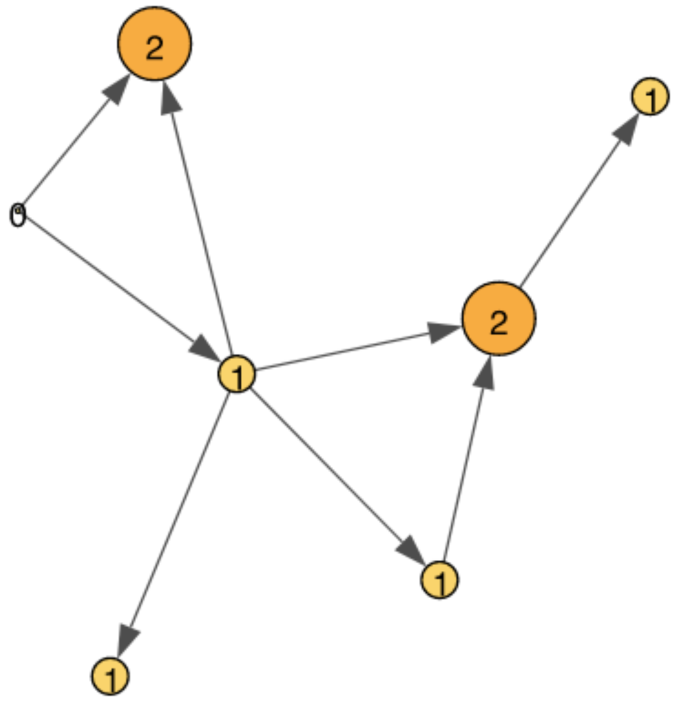
$$C_D(v_i) = \sum_{j=1}^{N} e_{ij}$$

Numbers indicate degree:

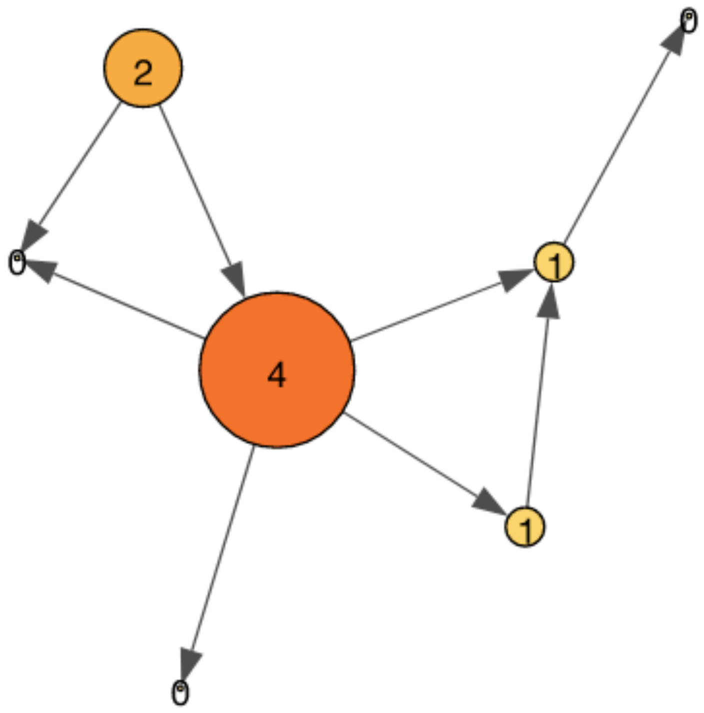
Undirected



In-degree



Out-degree



4. Centrality: degree centrality

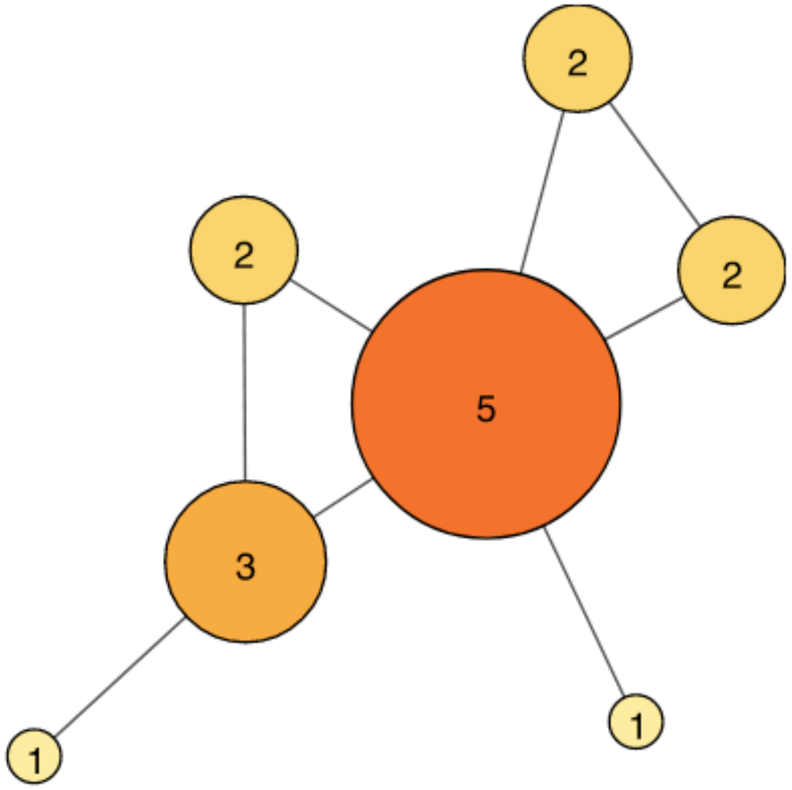
Degree centrality

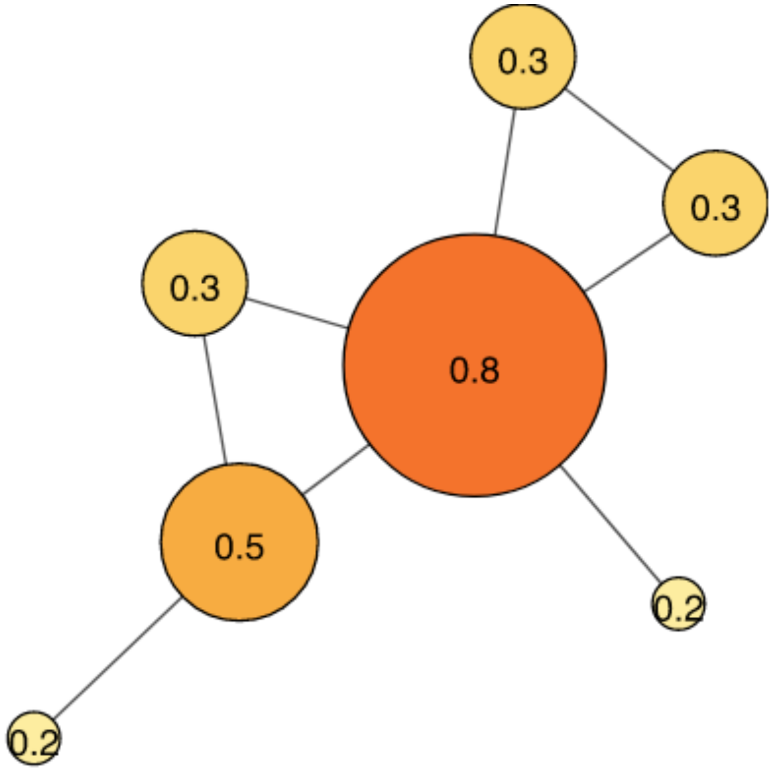
$$C_D(v_i) = \sum_{j=1}^{N} e_{ij}$$

Normalized degree centrality

$$C_D(v_i) = \frac{\sum_{j=1}^{N} e_{ij}}{N-1}$$

Centrality normalization allows for comparison between networks of different sizes

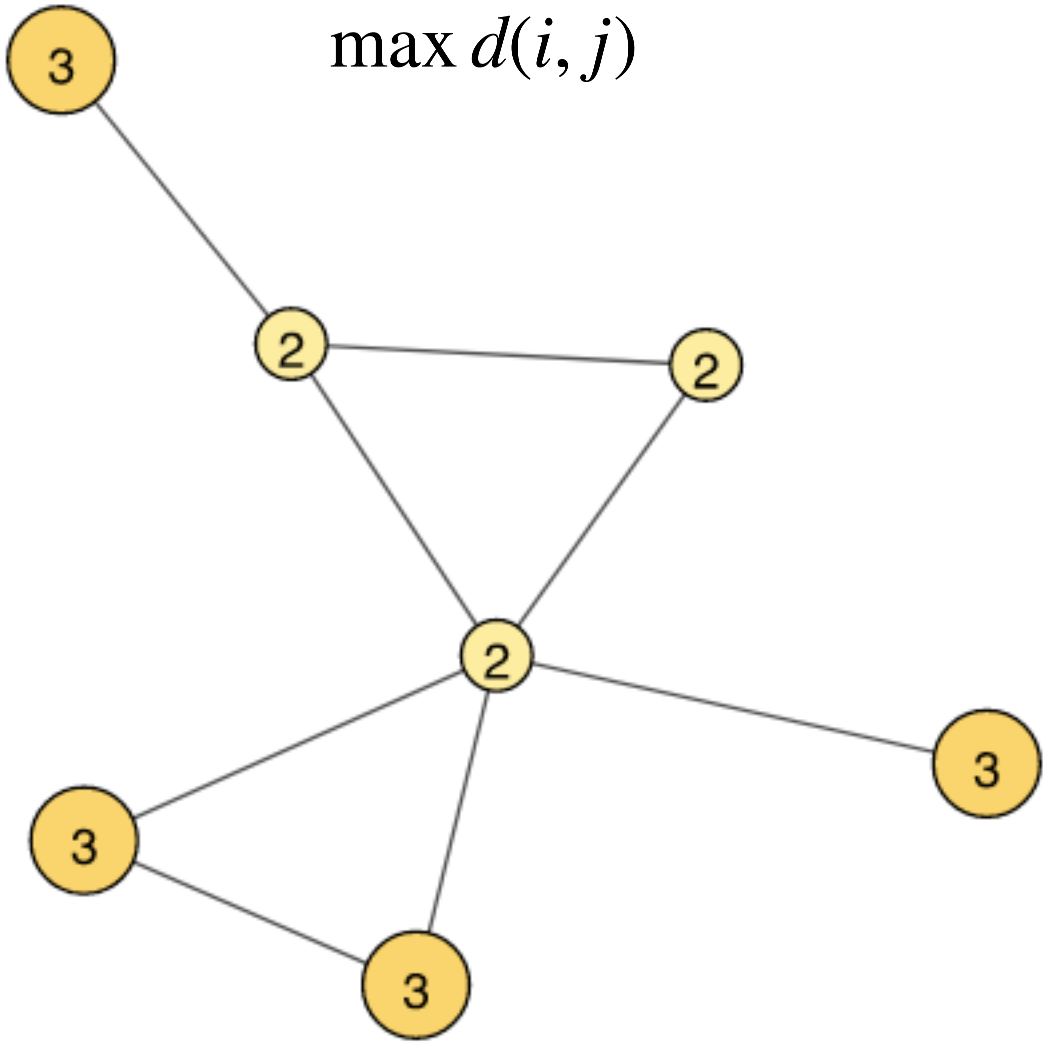




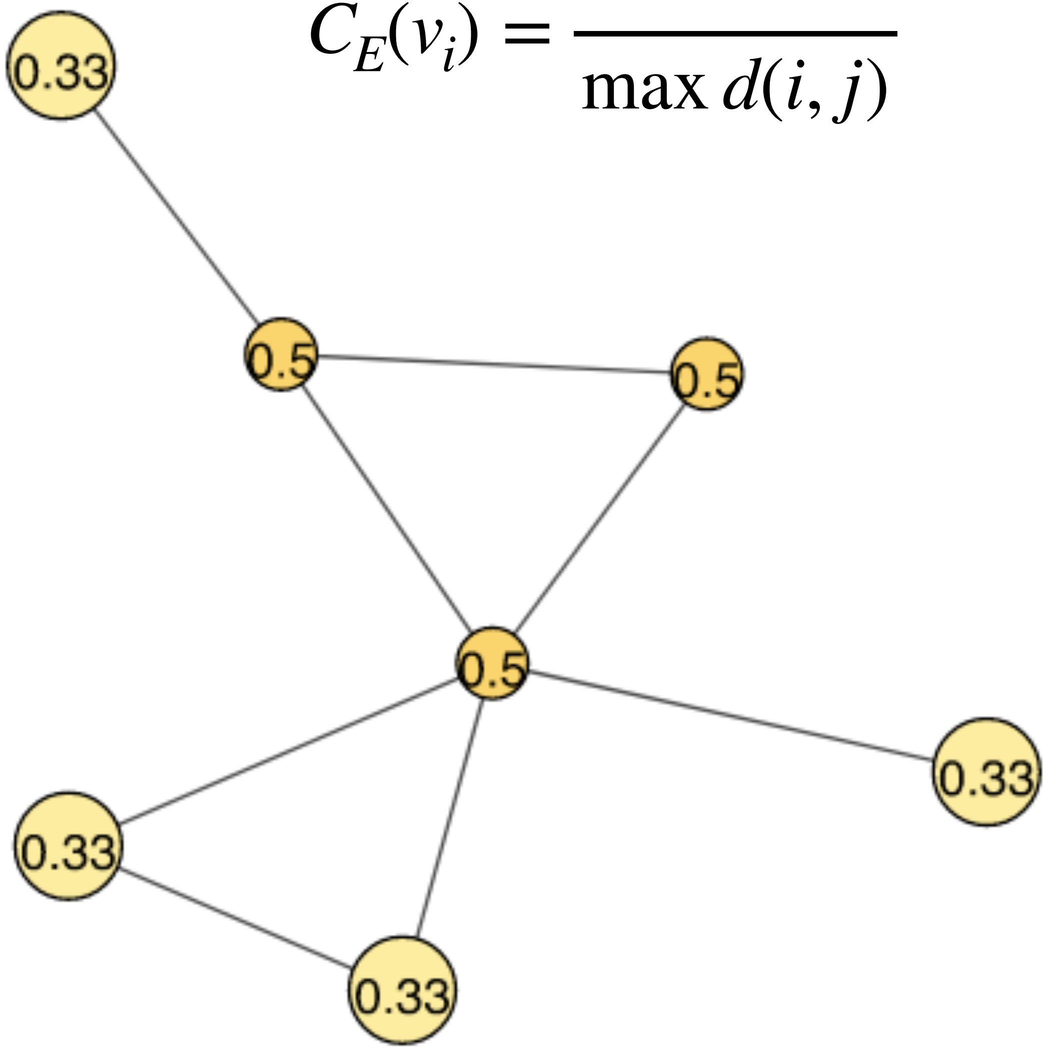
4. Centrality: eccentricity centrality

Eccentricity considers a node’s maximum shortest path to all other nodes

$$\max d(i,j)$$



$$C_E(v_i) = \frac{1}{\max d(i,j)}$$



4. Centrality: limitations & influence

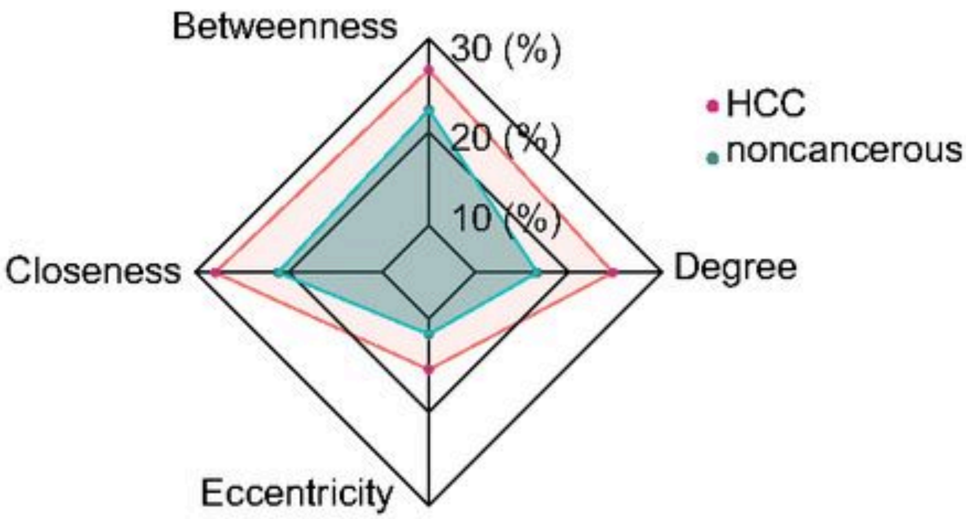
Node centrality does not necessarily imply **importance**

How to tackle this?

1. Complement with experimental observations
2. Compute multiple metrics and summarise joint observations
3. Compute node **influence**

- **Accessibility**
- **Dynamic influence**
- **Impact**
- **Expected force**

Measure **information transmission**



Bidkhori 2018

Break

Introduction to biological network analysis - part 2

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

rui.benfeitas@scilifelab.se

Key network properties to discuss

1. Network representations
2. Network density
3. Paths
4. **Centrality**
5. **Clustering coefficient**
6. **Degree and connectivity distributions**

6. Clustering coefficient

How likely is it that two connected nodes are part of a highly connected group of nodes?

If node v_1 is connected with v_2 and v_3, it is very likely that v_2 and v_3 are also connected.

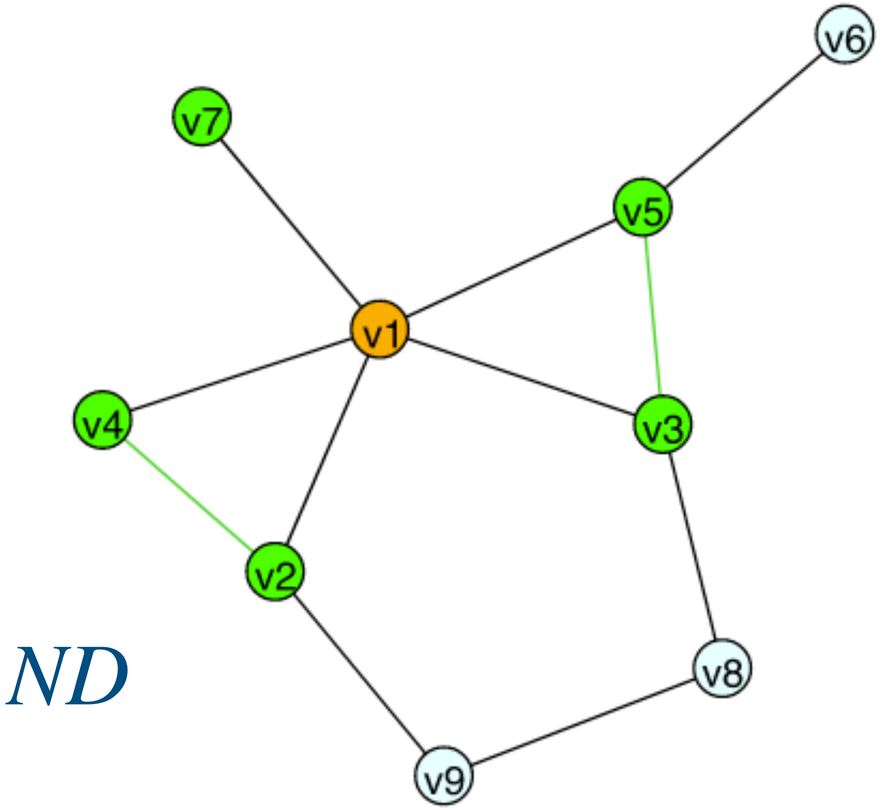
Takes into account degree of a node and the degree of its 1st neighbours

For node v_1

- $deg(v_1) = k = 5$
- n connections between 1st neighbours of $v1 = 2$

$$C_i = \frac{2 \cdot n}{k_i \cdot (k_i - 1)}$$

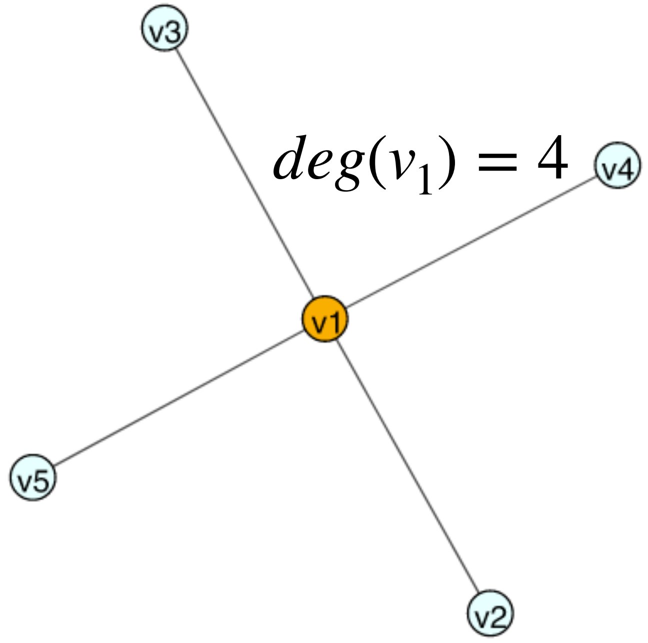
$$C(v_1) = \frac{2 \cdot 2}{5 \cdot 4} = 0.2 \qquad C(v_7) = \frac{2 \cdot 0}{1 \cdot 0} = 0 \text{ or } ND$$



6. Clustering coefficient

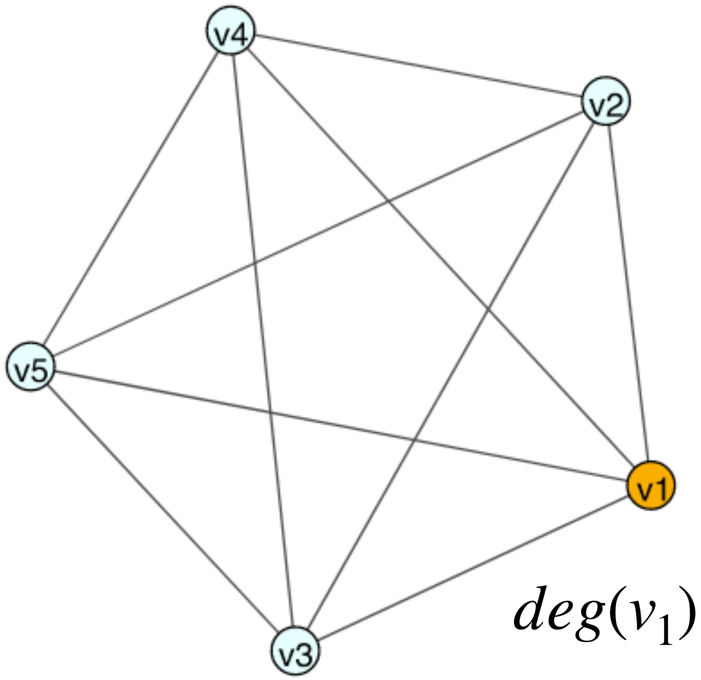
$C_i = \dfrac{2 \cdot n}{k_i \cdot (k_i - 1)}$ gives the **fraction of possible interconnections** for neighbours of node i

where $\dfrac{k_i \cdot (k_i - 1)}{2}$ is the maximum number of triangles through a node



$deg(v_1) = 4$

$$0 \leq C_i \leq 1$$



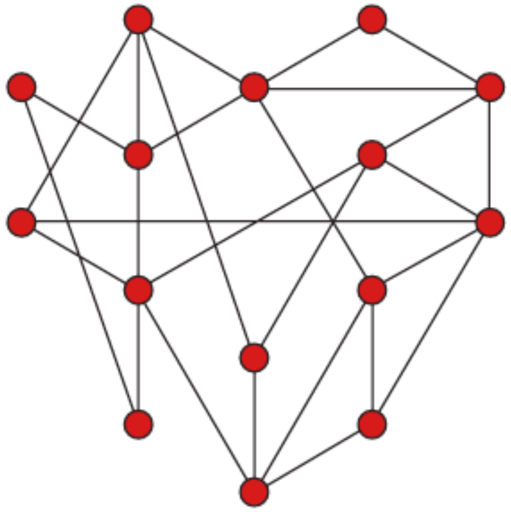
$deg(v_1) = 4$

The global clustering coefficient $C(G)$ is simply the average of its clustering coefficients

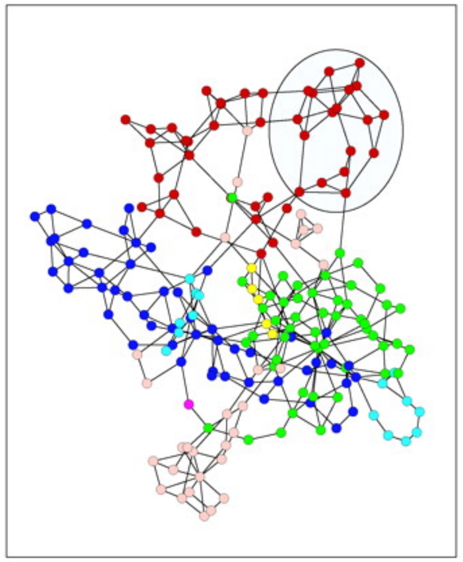
What distinguishes biological networks from random?

Do metabolic networks display different network properties from random networks?

Random network



Metabolic network

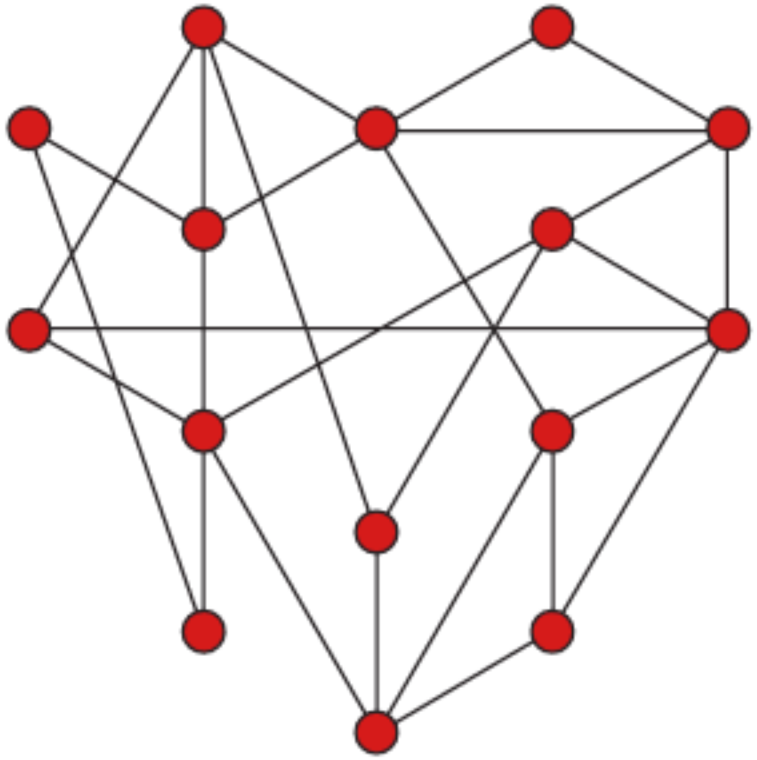


Barabasi 2004
Jeong 2000
Ravasz 2002

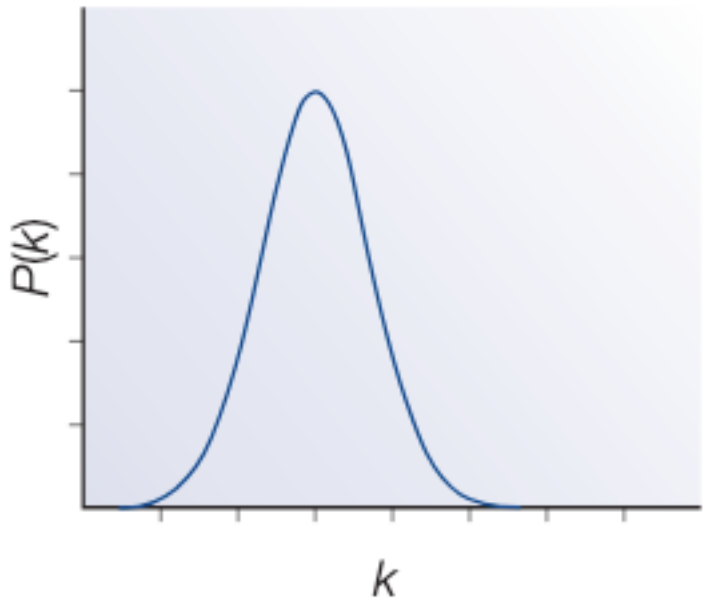
7. Degree and clustering coefficient distribution

Degree distributions allow us to compare network organization

Random network
(e.g. Erdös-Rényi model)



Poisson degree distribution
shows no highly connected nodes



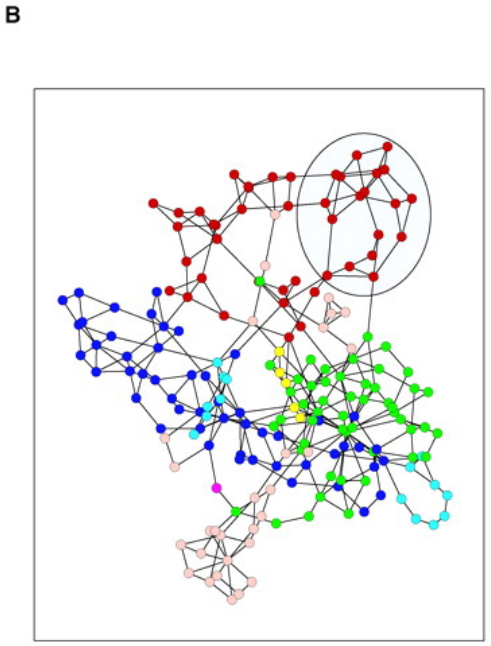
Most nodes have near <*k*>

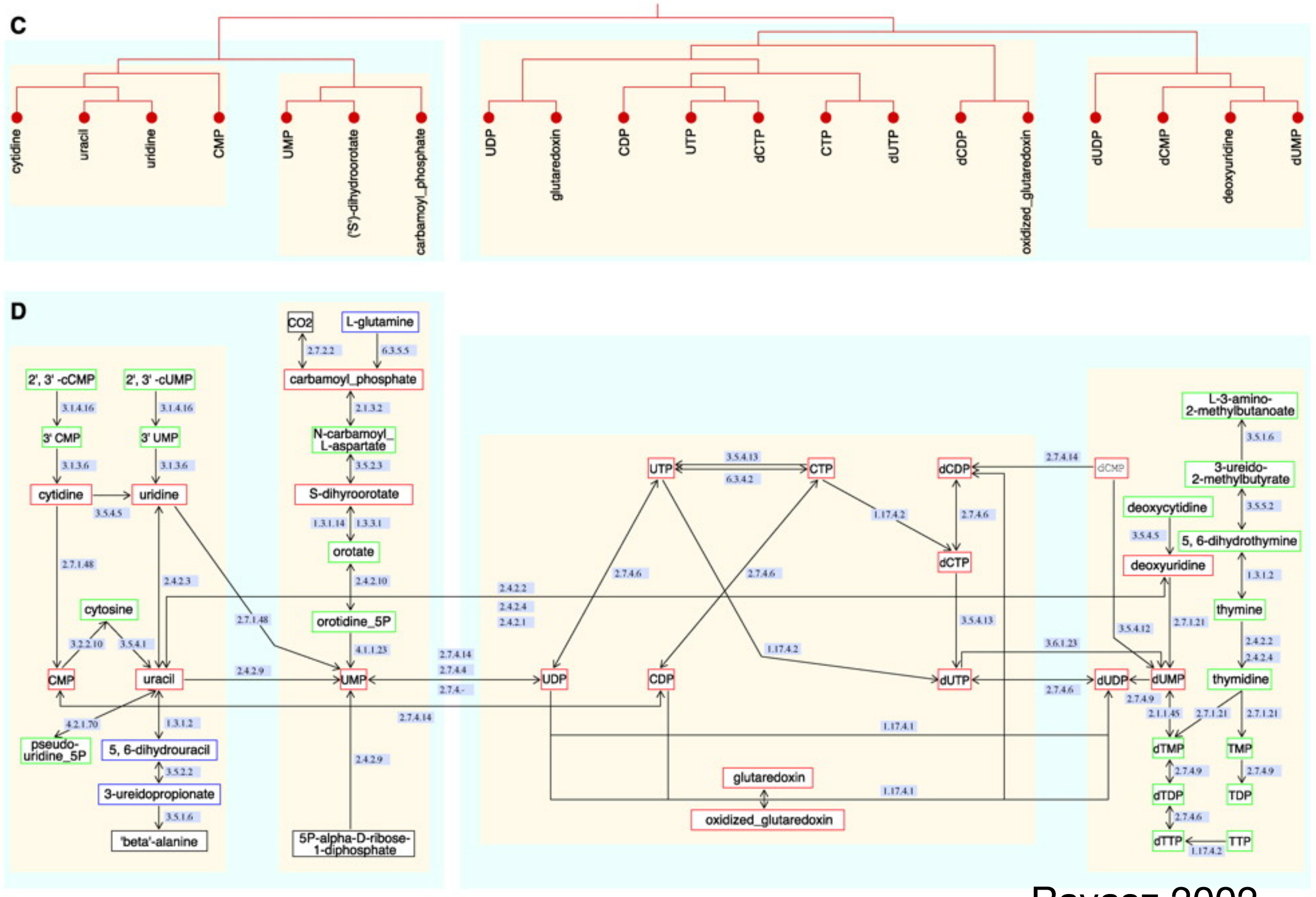
Barabasi 2004

Metabolic networks show hierarchical topology

Metabolic networks of 43 organisms are organised into **small, tightly connected modules**

Their combination shows a hierarchical structure





Ravasz 2002

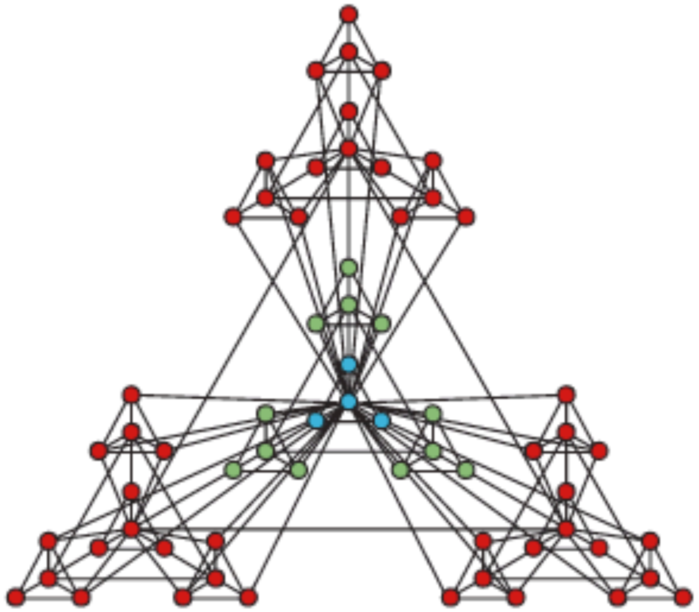
7. Degree distribution

Biological networks do not follow topology features of random networks.

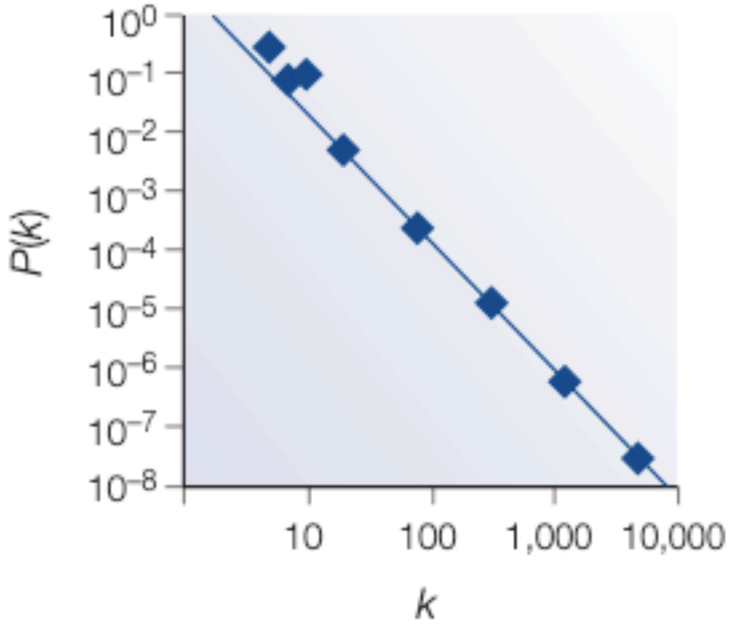
Analysis of metabolic networks of 43 organisms shows common patterns

Biological networks tend to display high robustness to node failure: removal of <80% nodes still retains paths between any two nodes

Hierarchical network



Degree distribution
shows many with low degrees
a few highly connected nodes



Barabasi 2004
Jeong 2000

7. Degree and clustering coefficient distribution

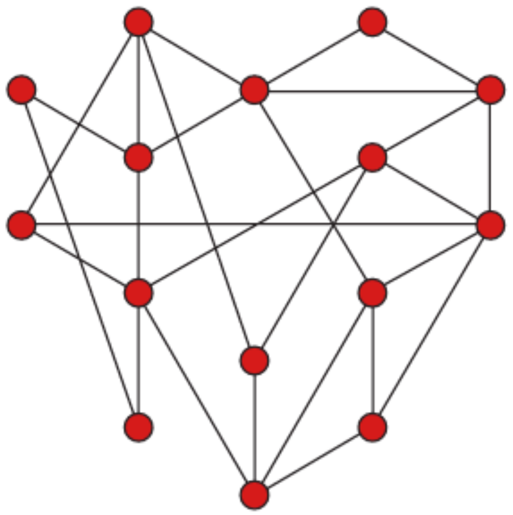
$C(k)$ shows no relationship with *k* in random networks: no modular organisation

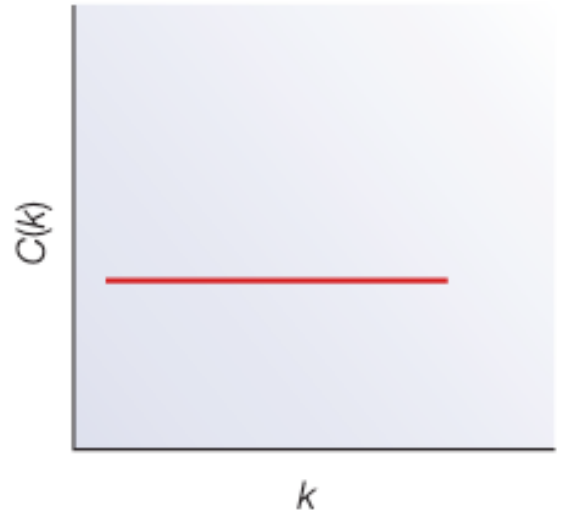
$C(k) = k^{-1}$ in hierarchical networks

Sparsely connected nodes are part of highly modular areas

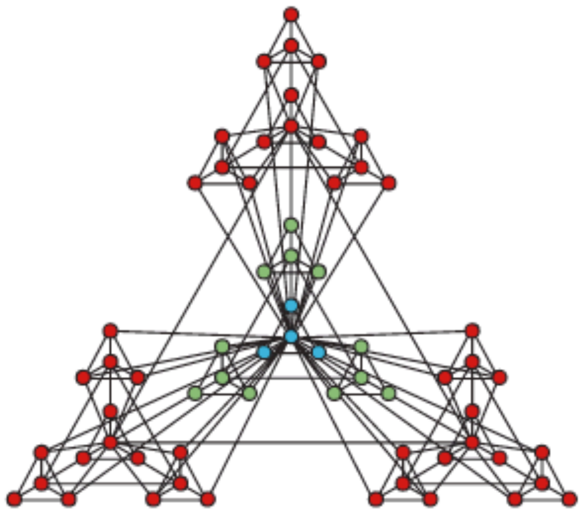
Communication between highly clustered neighbourhoods maintained by a few hubs

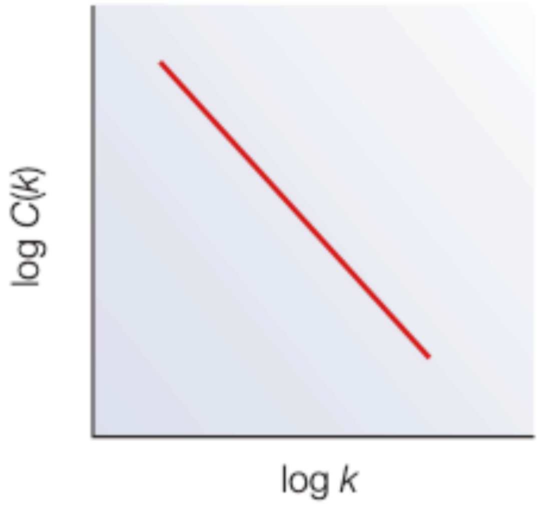
Random network





Hierarchical network





Barabasi 2004
Jeong 2000

7. Small world

Any two nodes can be connected in a small number of steps.

This is a property seen in **random networks** where the mean path length

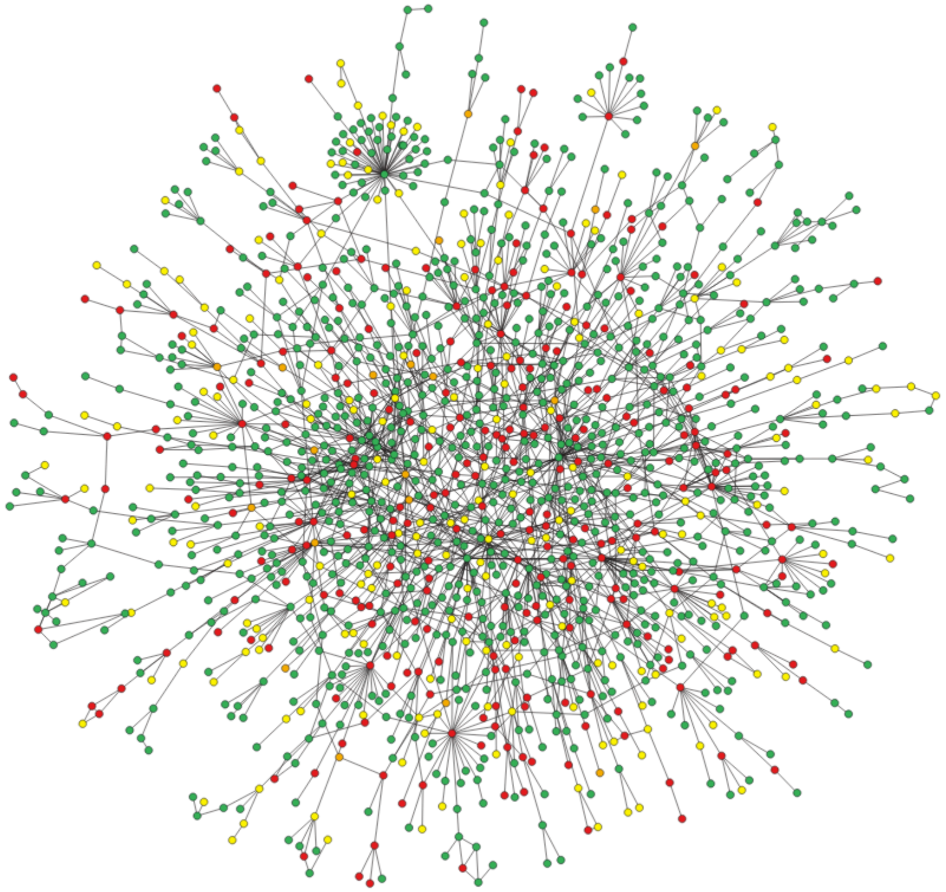
$l(G) \approx logN$ for a network of size *N*

Scale-free networks show **ultra-small world**:

$$l(G) \approx log(logN)$$

Highly central hubs tend **not** to be connected in biological networks: they are **disassortative**

(social networks: **assortative**)



Barabasi 2004

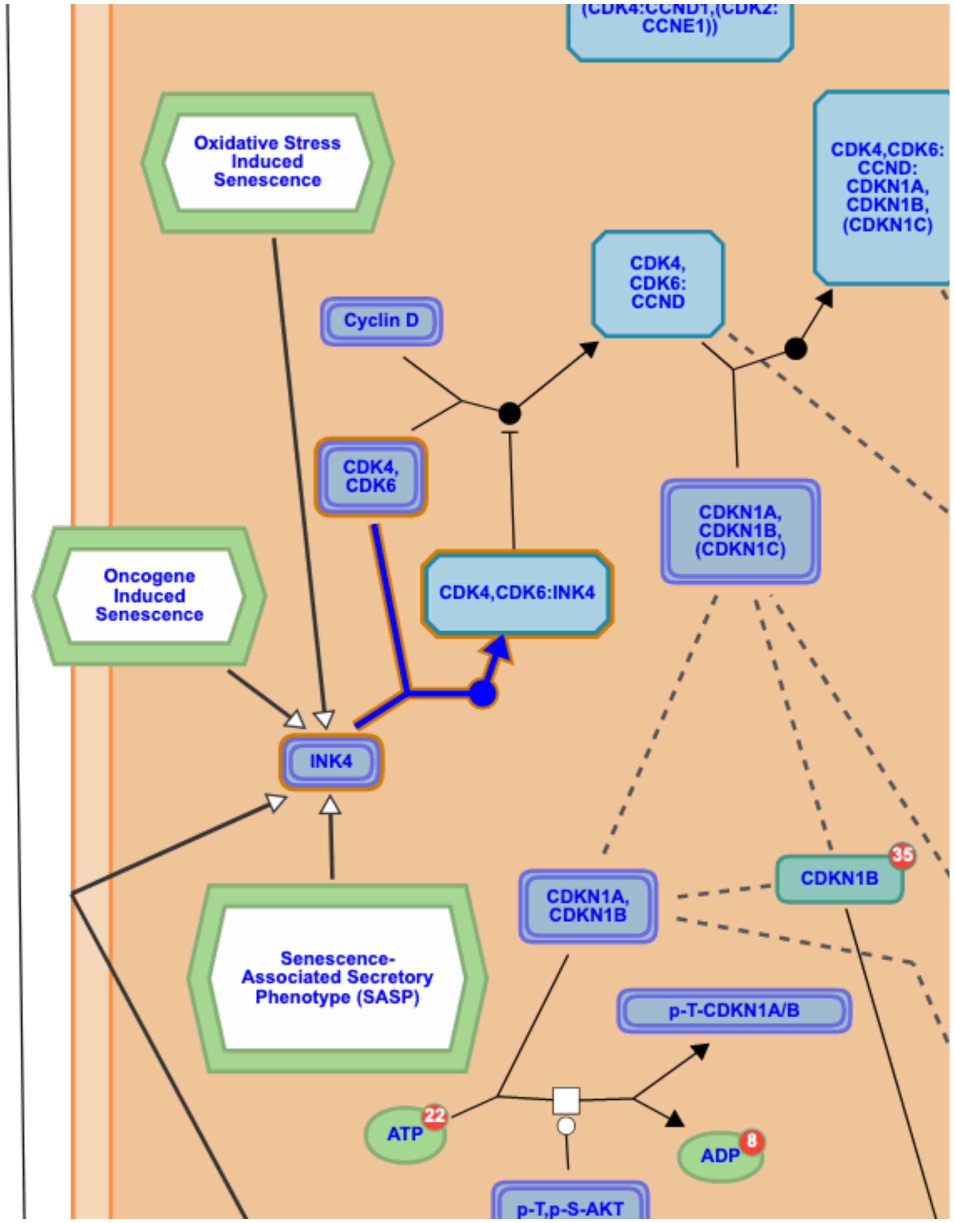
Overview

1. Introduction to network analysis
2. Terminology
3. Network inference
4. Key network properties
5. **Community analysis**

What are modules?

Modules are physically or functionally associated nodes that work together to achieve a distinct function

Protein complexes are physical modules

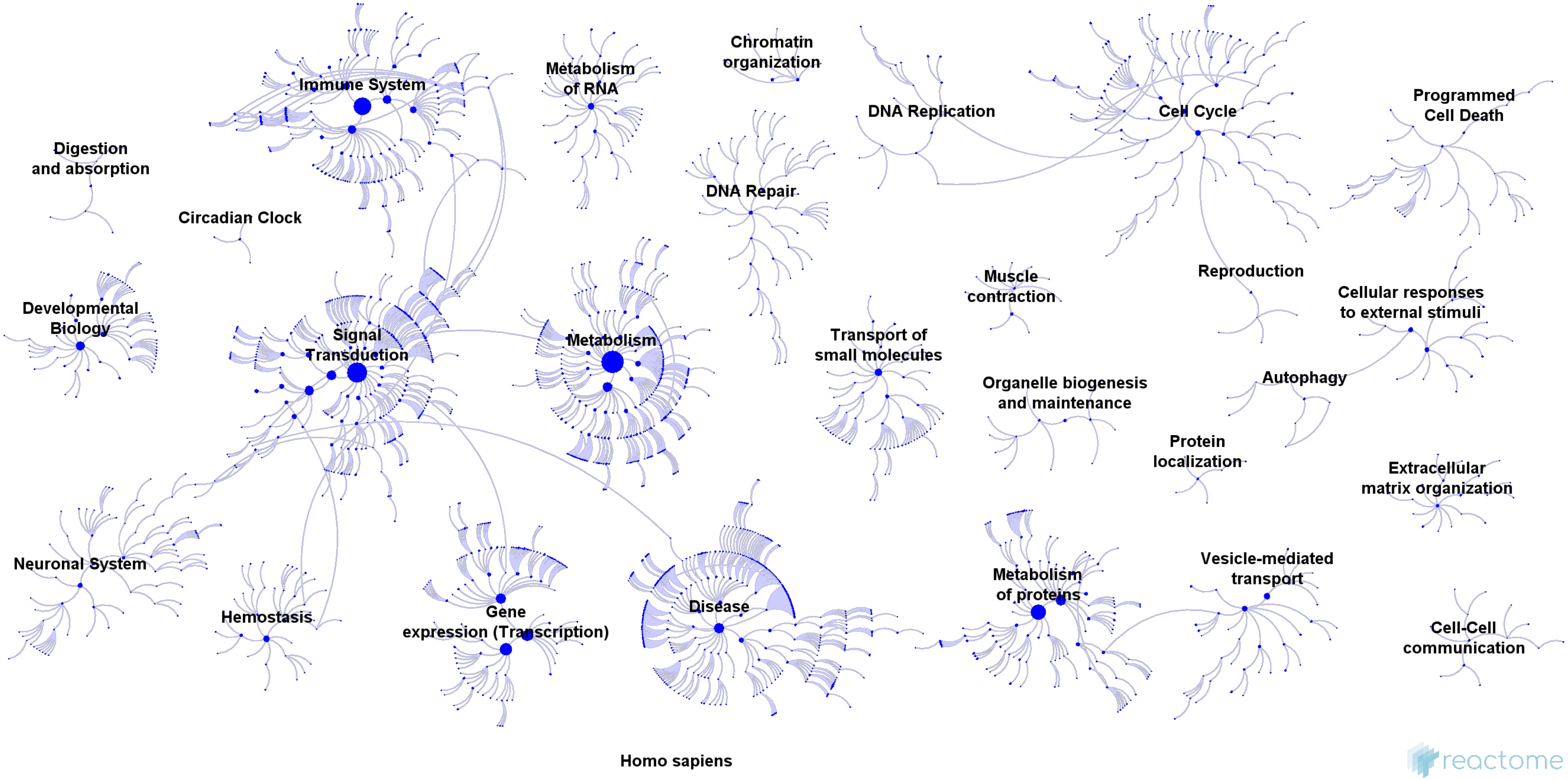


reactome.org

What are modules?

Pathway-associated proteins ***may*** represent functional modules

Gene Ontology



reactome.org

What are modules?

In addition to physical or functional modules, one may identify other types of modules

Topological: derived from their high within-module degree

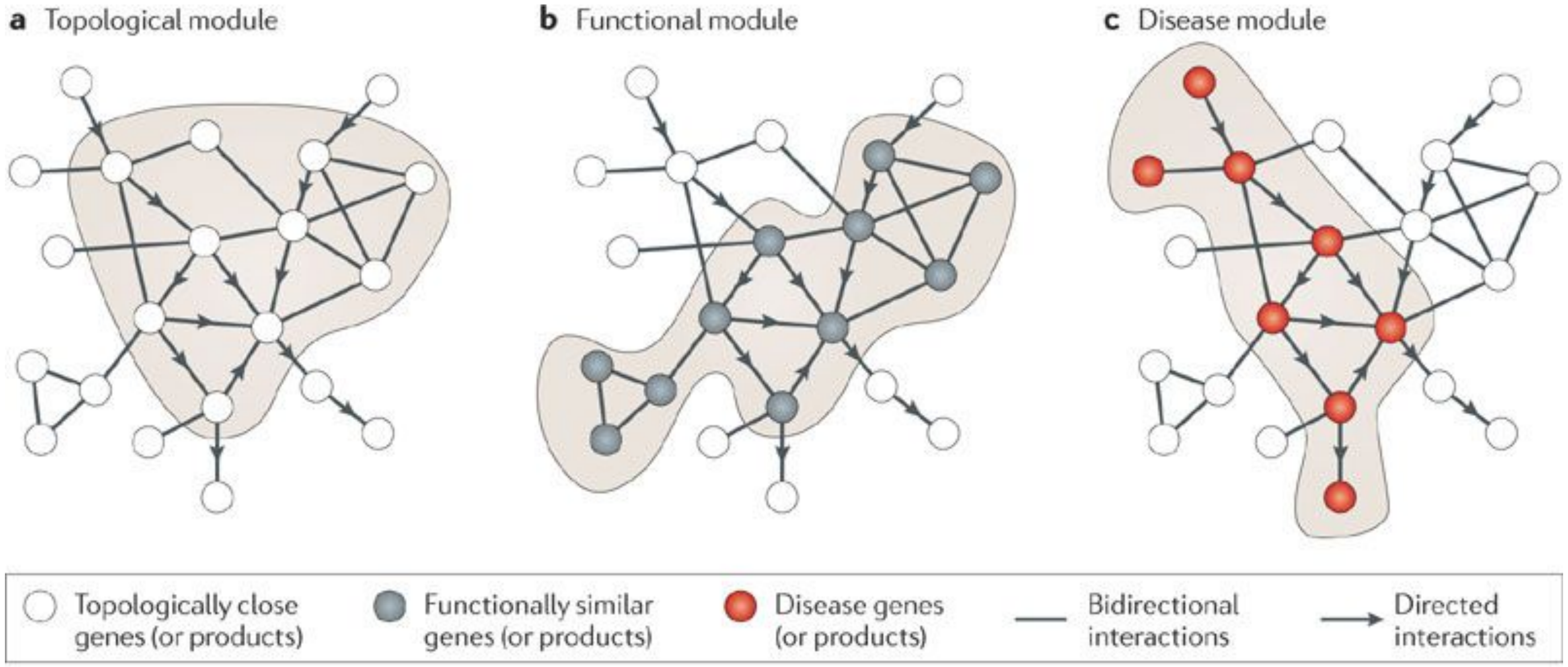
Disease: highly interconnected nodes associated with a disease response

Drug: highly interconnected nodes associated with a drug response

Subgroup: highly interconnected nodes associated with a sample subgroup (e.g. cancer subtype)

Tissue-, cell-type-specific: highly interconnected nodes associated with a specific tissue or cell type

Highly interlinked local regions of a network


a Topological module
b Functional module
c Disease module
Topologically close genes (or products)
Functionally similar genes (or products)
Disease genes (or products)
Bidirectional interactions
Directed interactions

Barabasi 2011

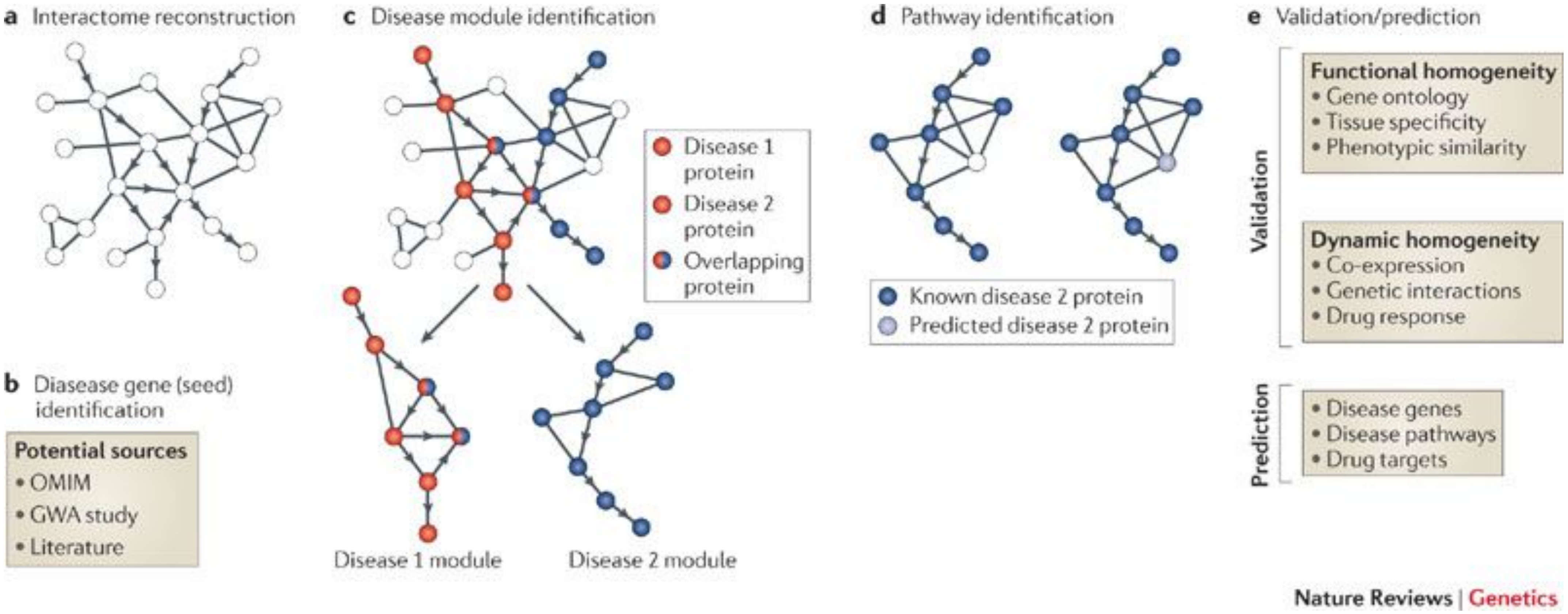
The challenge: identify and characterise modules

Moving from full network to modular characterisation

Hypothesis: common functional properties (diseases, biological processes, etc.) are associated with the same module

Prediction: *in silico*, relies on available knowledge

Validation: experimental responses



Barabasi 2011

Modularity

Modularity is a property of the network

Modularity (Q) measures the tendency of a graph to be organised into modules

Modules computed by comparing probability that an edge is in a module vs what would be expected in a random network

For a given partitioning of the network into individual groups s, compute

$$Q \propto \sum_{s\in S} [(e_s) - (\text{expected } e_s)]$$

# edges in group s

Random network with
same number of nodes, edges and
degree per node

$$-1 < Q < 1$$

Q = 1: much higher number of edges than expected by chance

Q = -1: lower number of edges than expected by chance

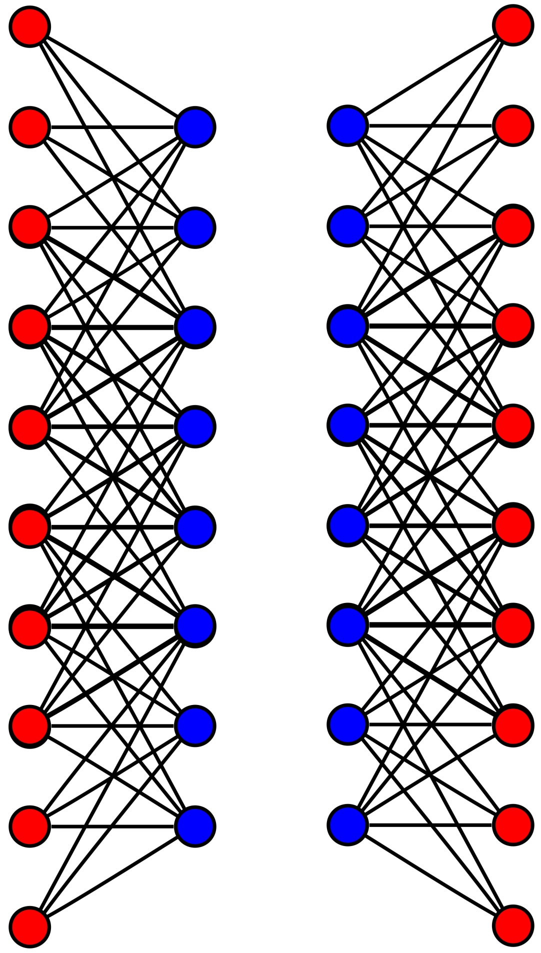
Q > 0.3 - 0.7 means significant community structure

Modularity

Modularity is different than **clustering coefficient**:

Graph composed of two bipartite complete subgraphs:

high Q but low connectivity (C)



Modules

A **module** (or **community**) is a set of nodes with a lot of **internal connections**, but **fewer external connections.**

How to identify modules? Maximise Q

$$Q \propto \sum_{s \in S} [(e_s) - (\text{expected } e_s)]$$

Brute-force approach:

1. Start with 1 node/module
2. Compute distances between nodes
3. Join closest node
4. Re-compute distances between a 2n module and each 1n module
5. Join them if Q increases

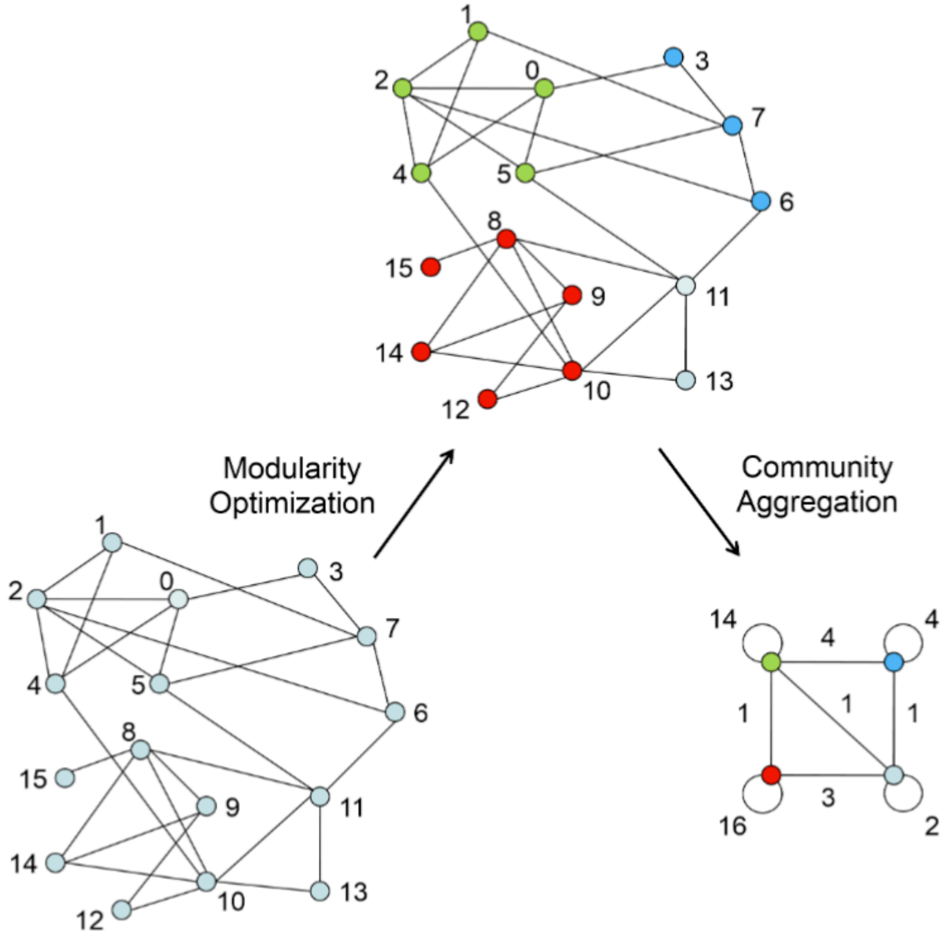
Module detection: Louvain algorithm

Phase 1: greedy modularity optimisation

1. Start with 1n/community
2. Compute Q by moving *i* to the community of *j*
3. If ΔQ>1, node is placed in community
4. Repeat 1-3 until no improvement is found. Ties solved arbitrarily

Phase 2: coarse grained community aggregation

5. Link nodes in a community into single node.
6. Self loops show intra-community associations
7. Inter-community weights kept
8. Repeat phase 1 on new network



Campigotto 2014
Traag 2019

Community characterisation

Clustering coefficient and degree distribution

Enrichment analysis

Hypothesis: community-associated features show coordinated changes associated with common biological processes

Can significantly enriched biological processes serve as “validation”?

- Mutual feature associations may reinforce data characterisations not evident by individual features
- …or need of further network curation based on top biological terms

Mooney 2015

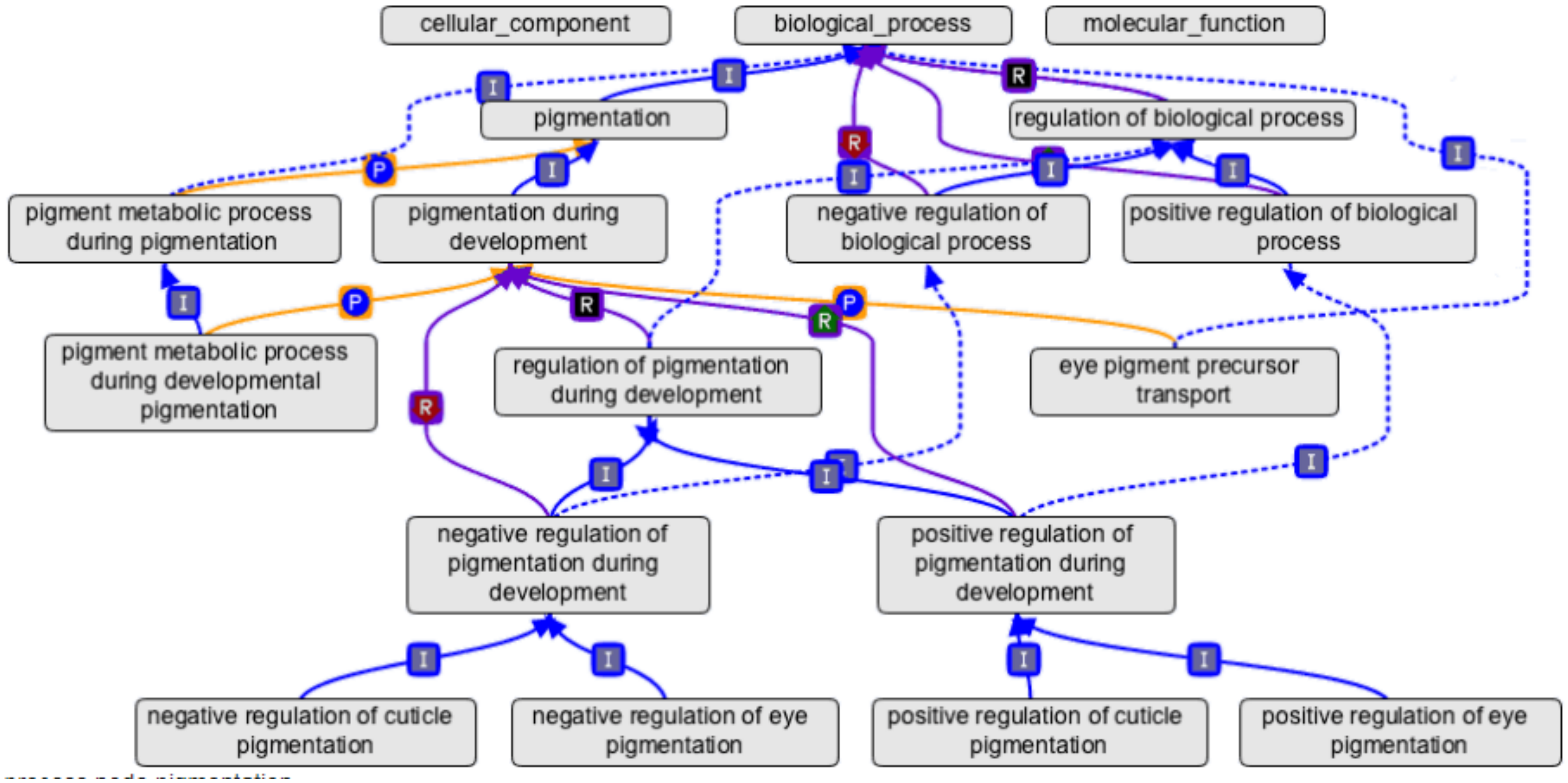
Enrichment analysis

GO-terms, pathways, subcellular location, TF-targets, disease, drugs

Tests for significant overlap between groups

All considerations from standard enrichment analyses apply

Some biological processes may have no biological meaning in your analysis


cellular_component
biological_process
molecular_function
pigmentation
regulation of biological process
pigment metabolic process during pigmentation
pigmentation during development
negative regulation of biological process
positive regulation of biological process
pigment metabolic process during developmental pigmentation
regulation of pigmentation during development
eye pigment precursor transport
negative regulation of pigmentation during development
positive regulation of pigmentation during development
negative regulation of cuticle pigmentation
negative regulation of eye pigmentation
positive regulation of cuticle pigmentation
positive regulation of eye pigmentation
I
P
R

://geneontology.org/

Enrichment analysis

Important databases with gene-sets:

- MSigDB (gene)
- Enrichr (gene)
- KEGG (metabolite, gene)
- DIANA (miRNA)
- MetaboAnalyst (metabolite)
- DAVID (web)
- Reactome (web)

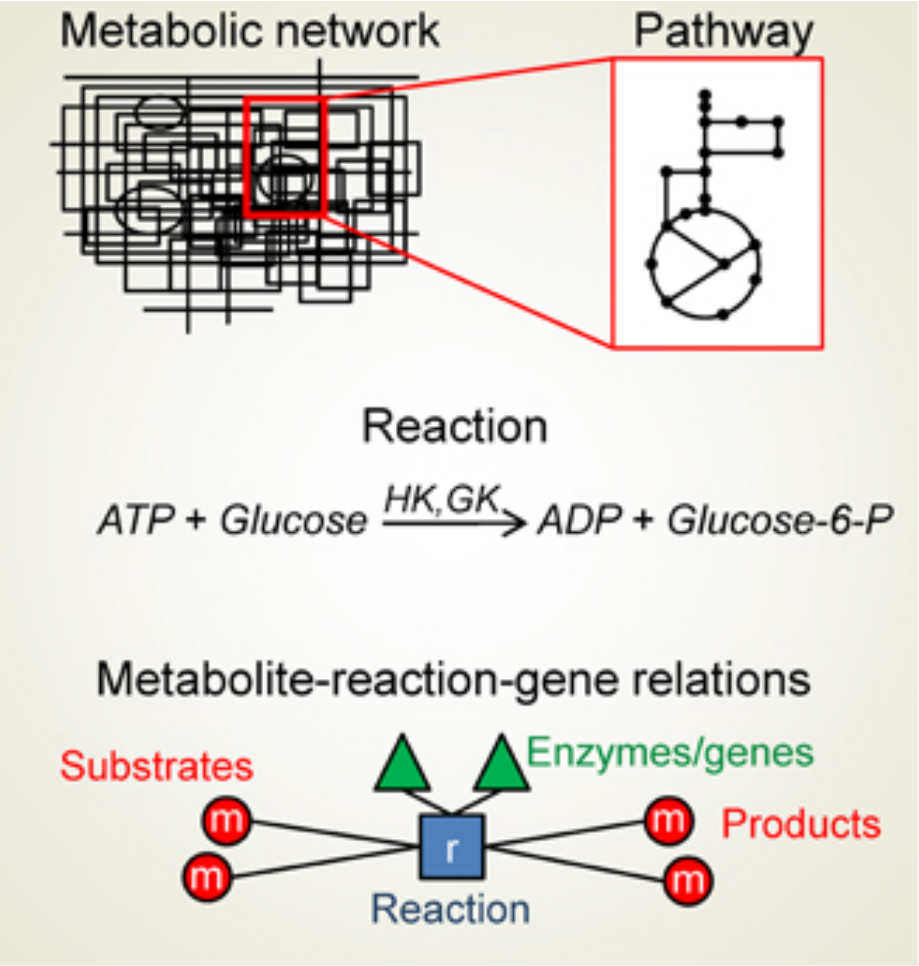
Creating custom sets and joint sets

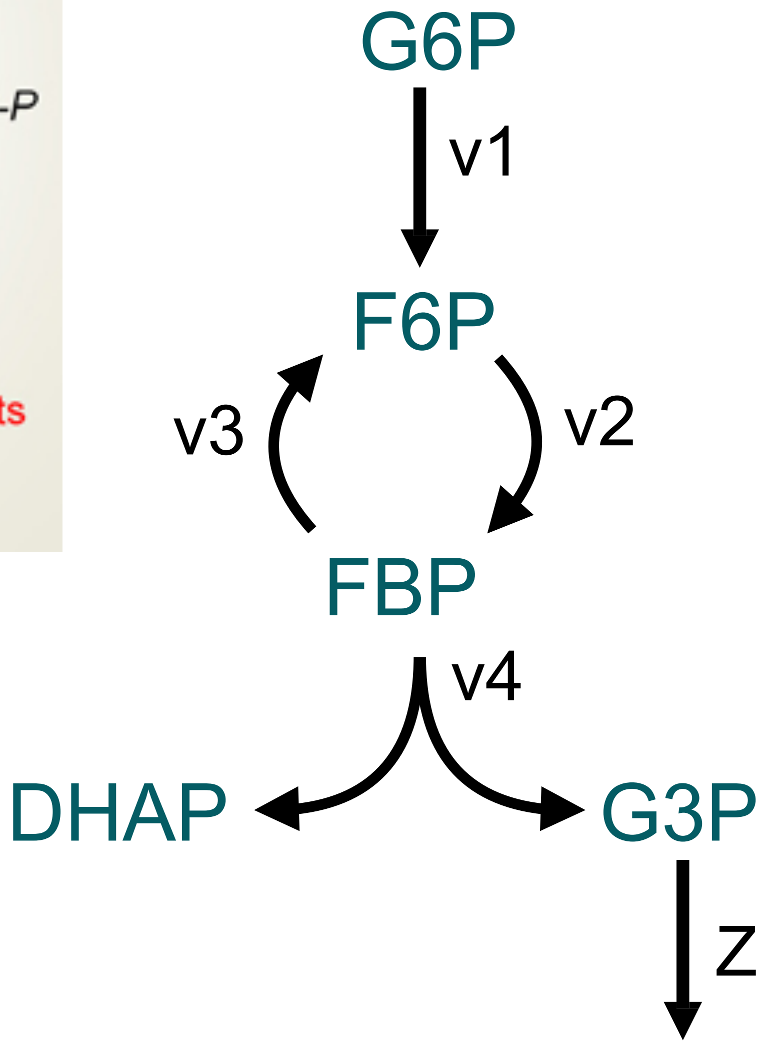
Mapping your data to common IDs

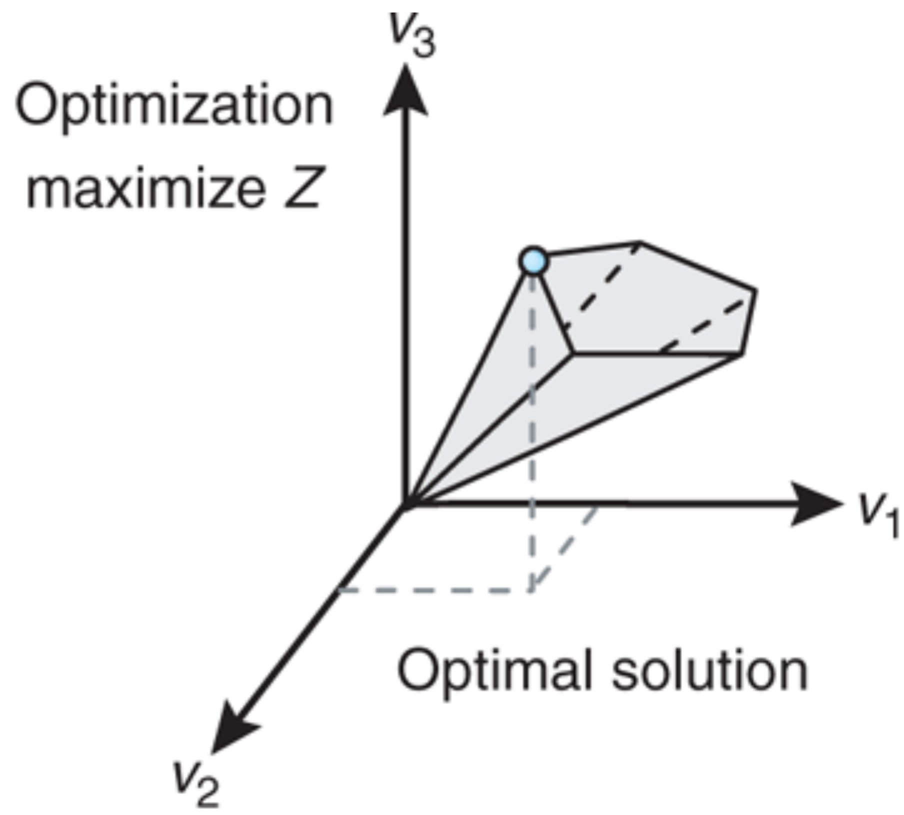
- Easy for genes and proteins: use DAVID, Biomart, or MyGene (in Python or R)
- Hard for other data types

References as hyperlinks

Genome-scale metabolic models as integrative networks

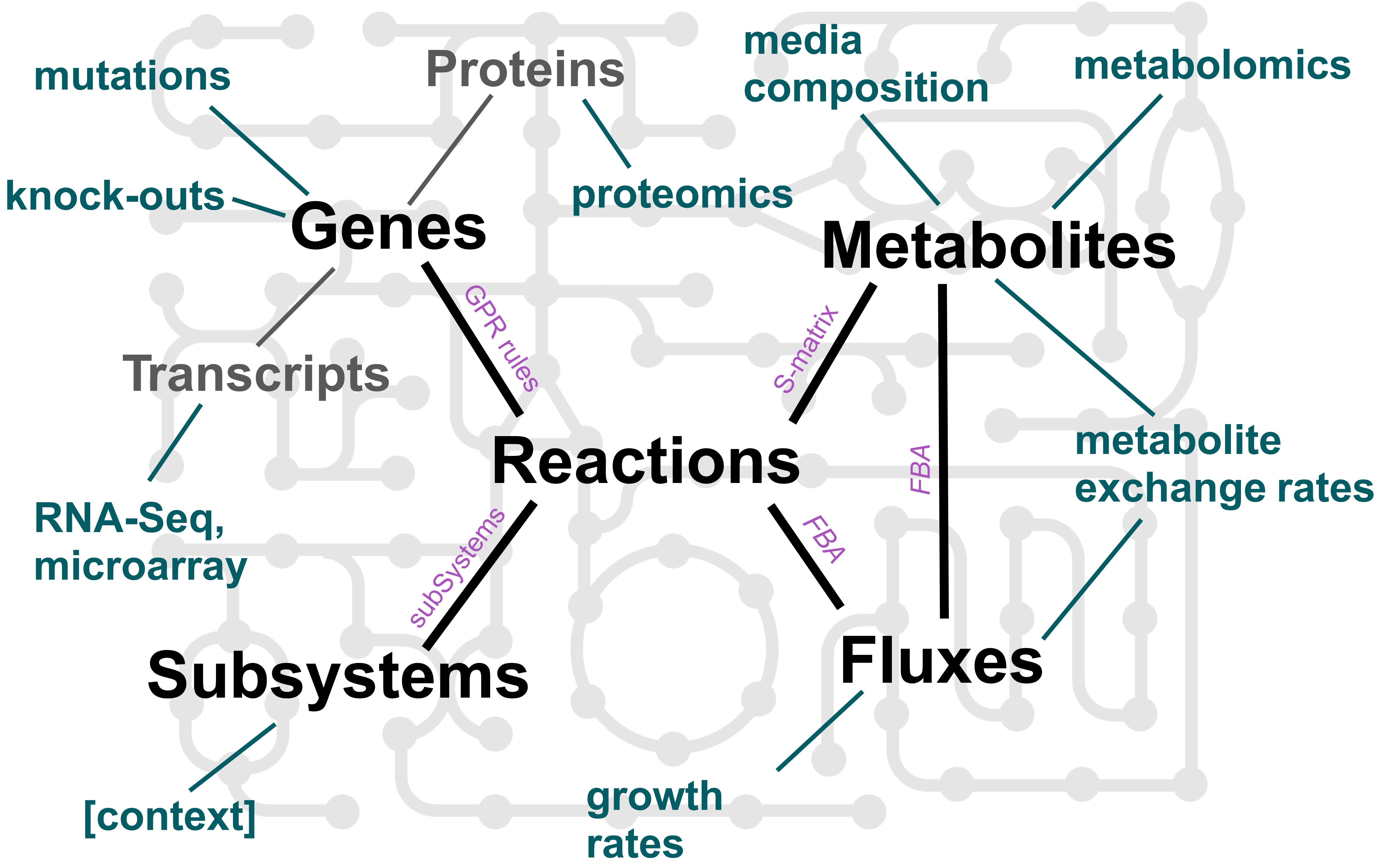

Metabolic network
Pathway
Reaction
ATP + Glucose $\xrightarrow{HK,GK}$ ADP + Glucose-6-P
Metabolite-reaction-gene relations
Substrates
Enzymes/genes
Products
m
r
Reaction





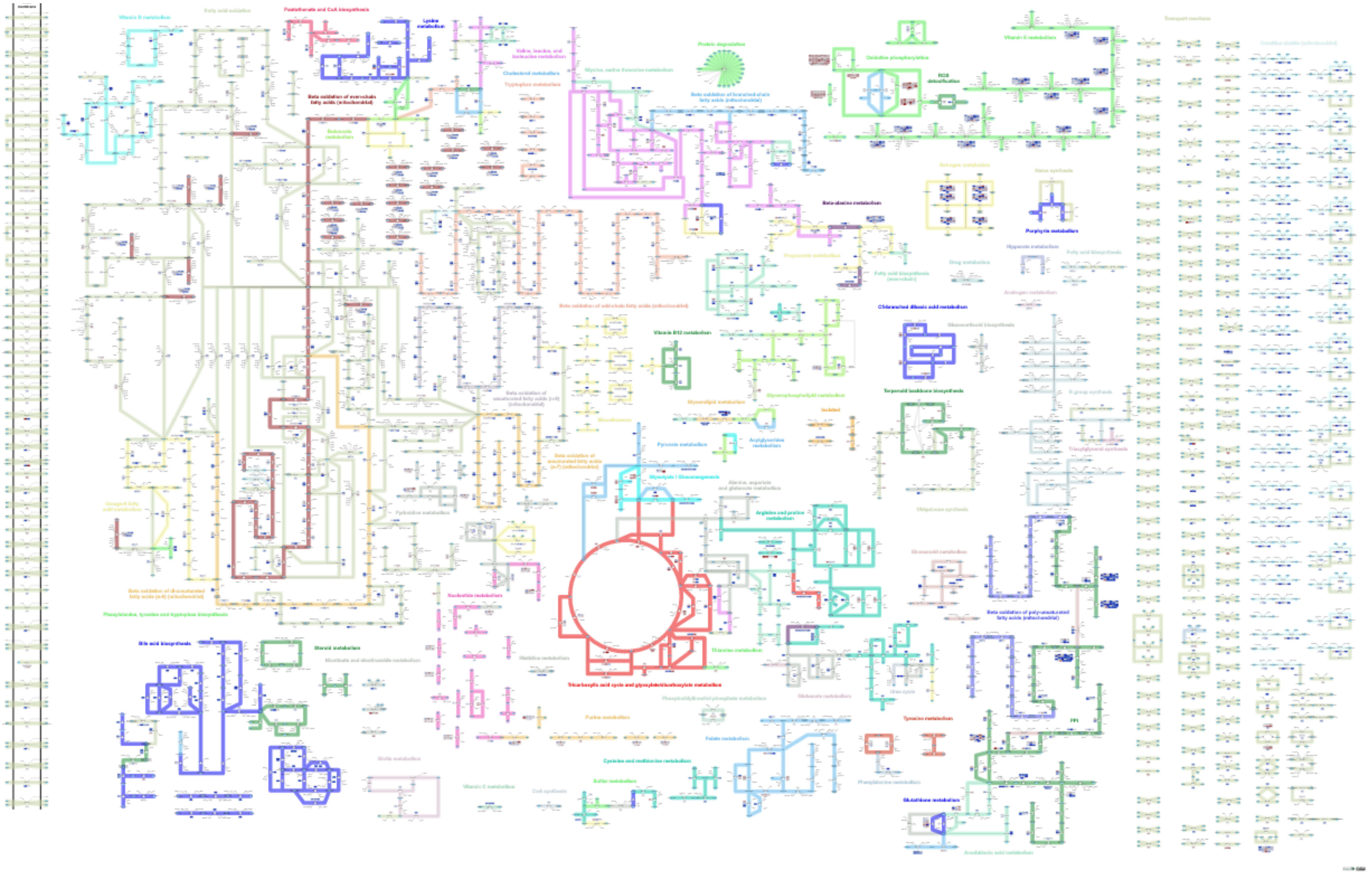
Väremo 2013
Orth et al Nat Biotechnol (2010).

Genome-scale metabolic models as integrative networks



Genome-scale metabolic models as integrative networks

metabolic ATLAS



Simulate flux distributions

Dysregulated pathways

Reporter metabolites

Essential genes

Targetable enzymes

May be combined with standard graph analysis

https://metabolicatlas.org/

Additional reading

- Network Science - Textbook on graph theory and network analysis.
- Communication dynamics in complex brain networks - 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.

Additional reading

- Analysis of Biological Networks - General introduction into biological networks, network notation, and analysis, including graph theory.
- Using graph theory to analyze biological networks - overview of the usage of graph theory in biological network analysis
- Survival of the sparsest: robust gene networks are parsimonious - analysis of network complexity and robustness.
- Network biology: understanding the cell's functional organization - Overview of key concepts in biological network structure
- Graph Theory and Networks in Biology - extended perspective on how graph analysis is applied in biology
- Scale free networks are rare
- Modularity and community structure in networks

Additional references displayed as hyperlinks in each figure.