

Introduction to biological network analysis

Rui Benfeitas

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

rui.benfeitas@scilifelab.se



SciLifeLab

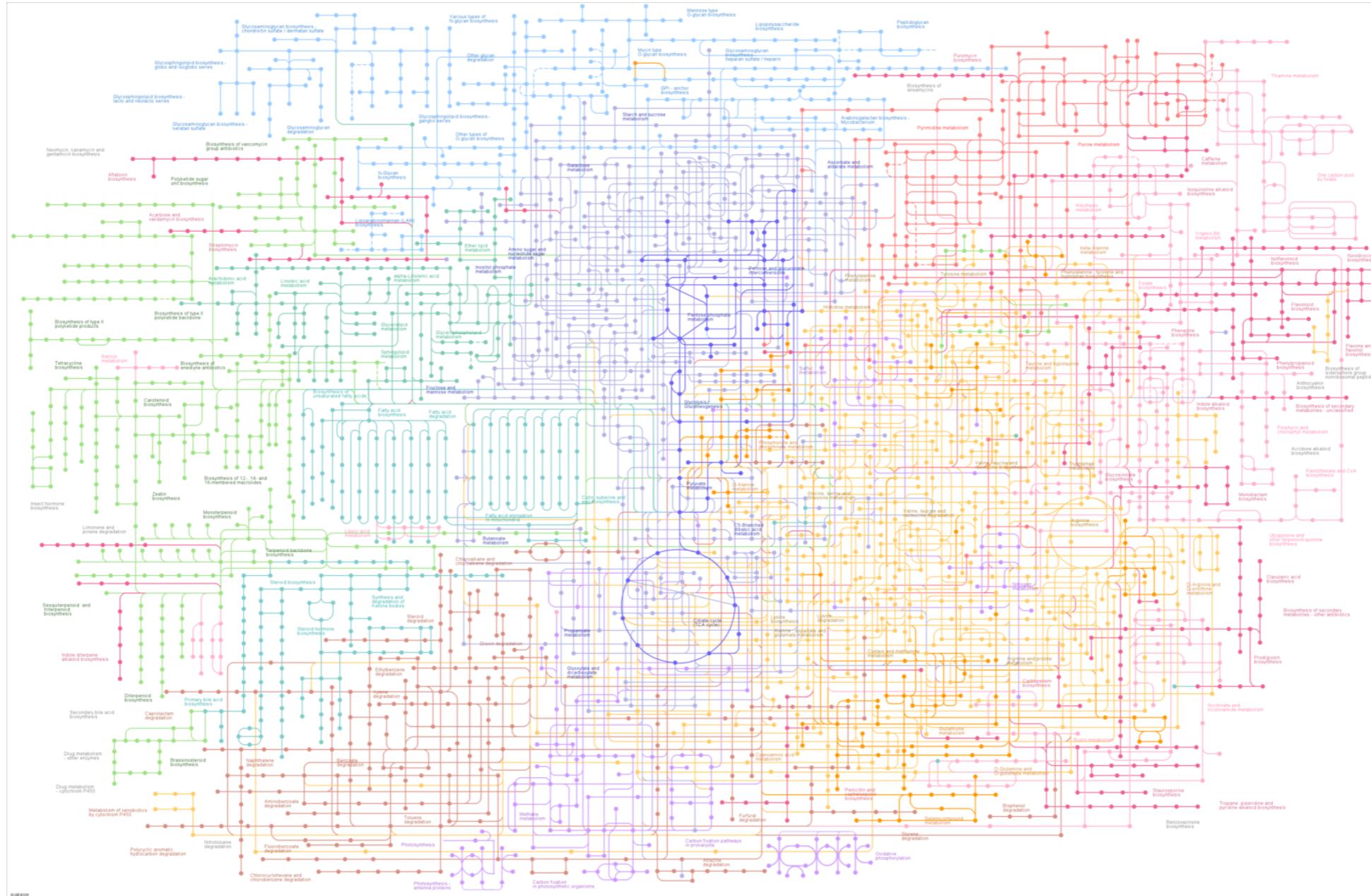


Overview

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

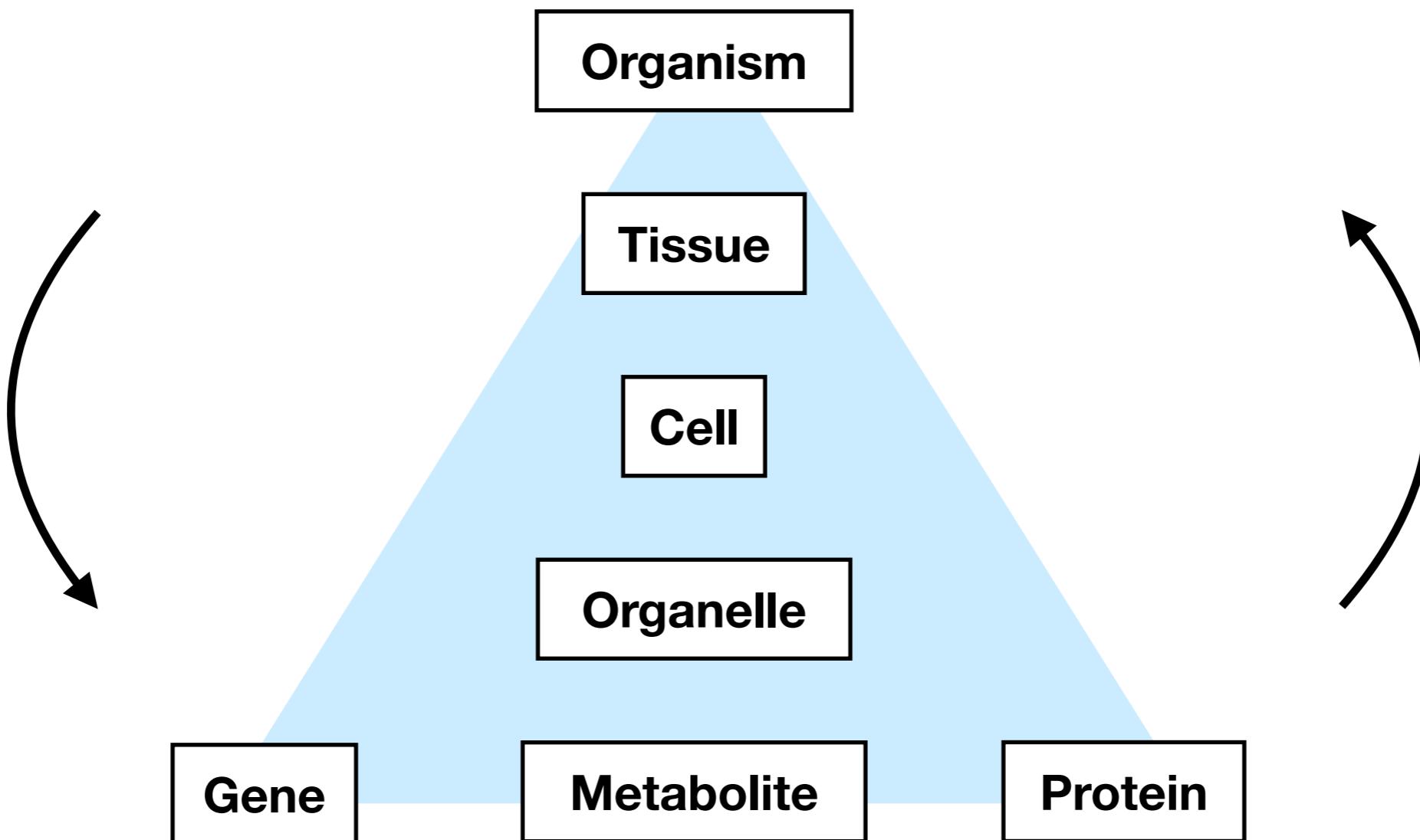
Original sources of images provided as reference and hyperlinks, where applicable.

How to tackle biological complexity?



How to tackle biological complexity?

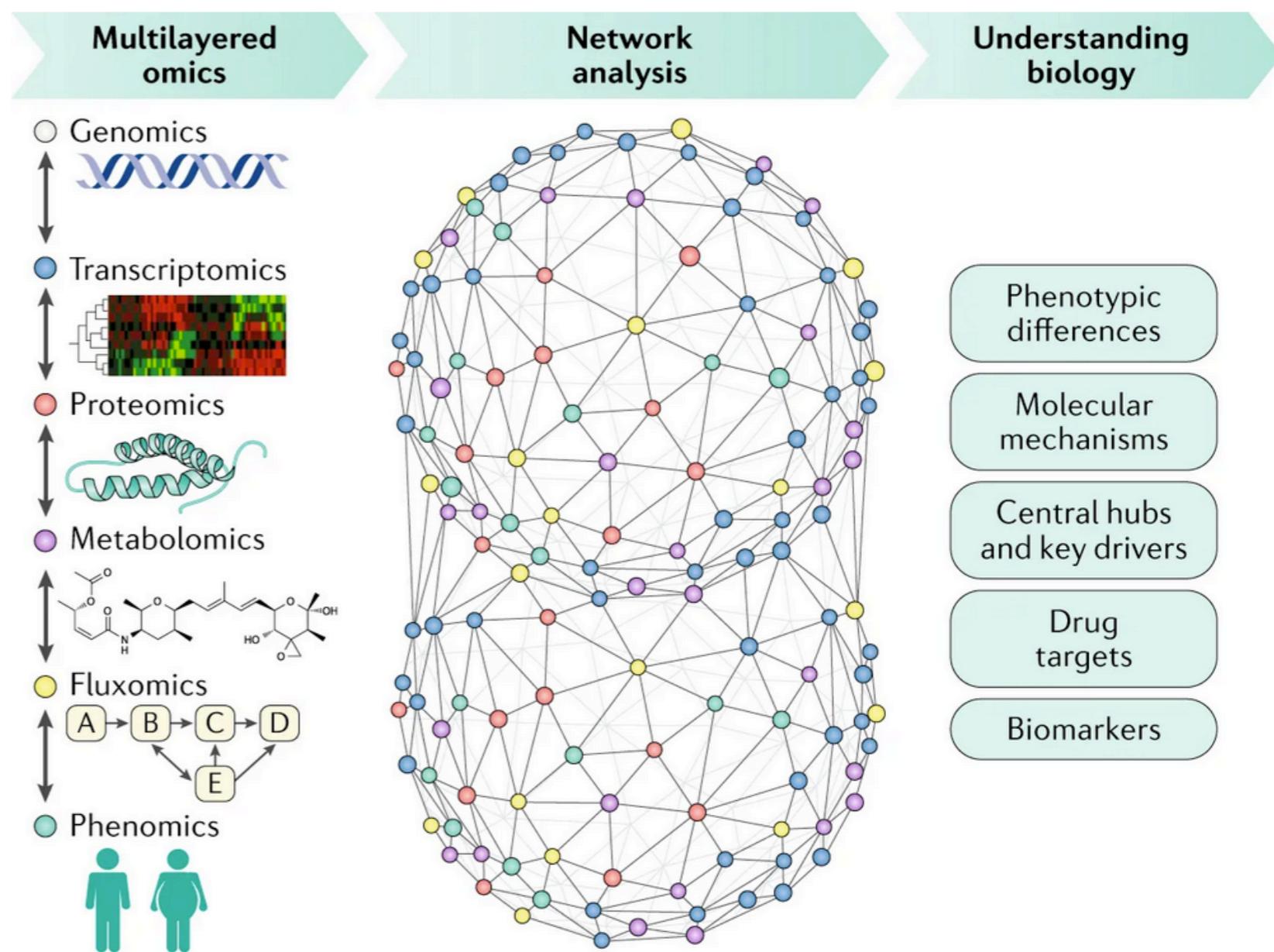
Moving from reductionist approaches towards global characterisations



How to tackle biological complexity?

Integrative approaches, and global patterns

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



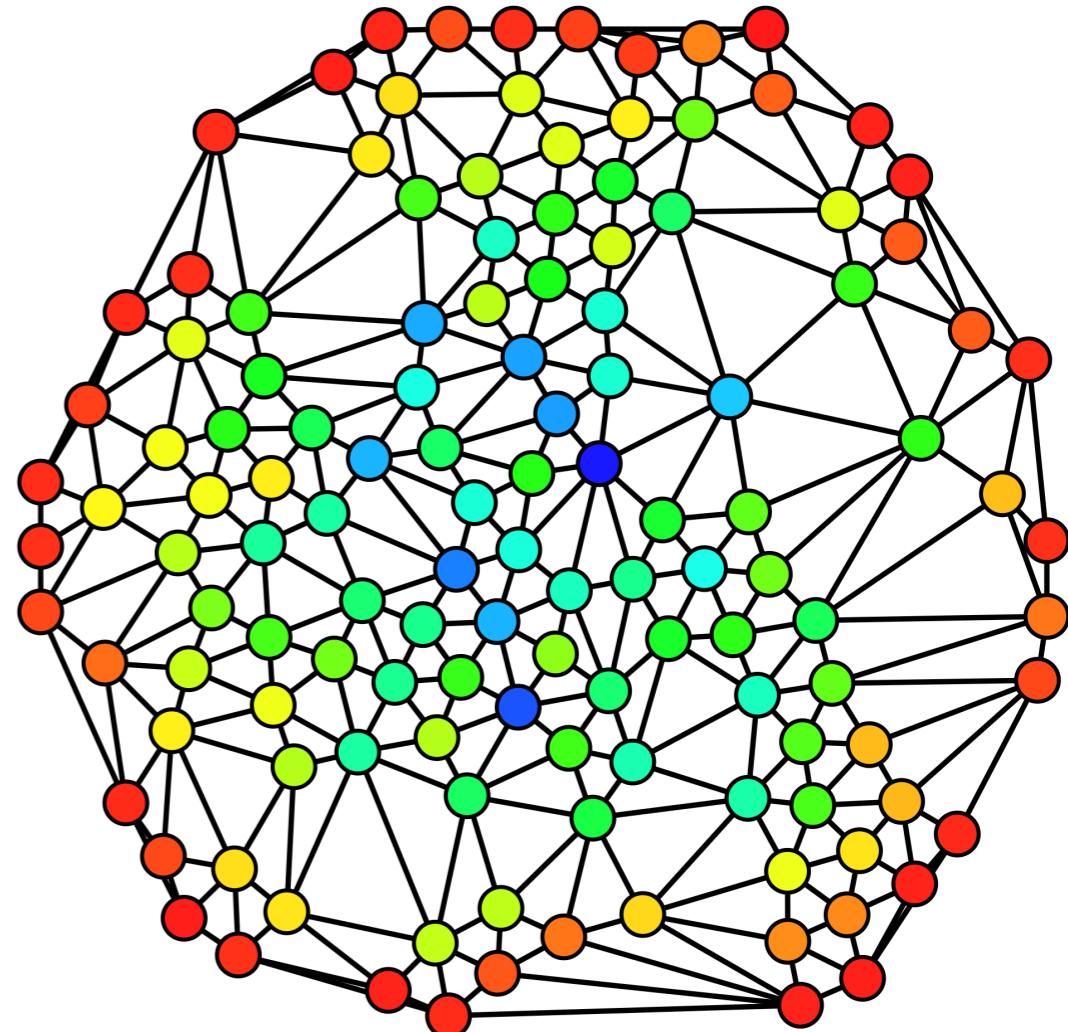
What are networks?

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



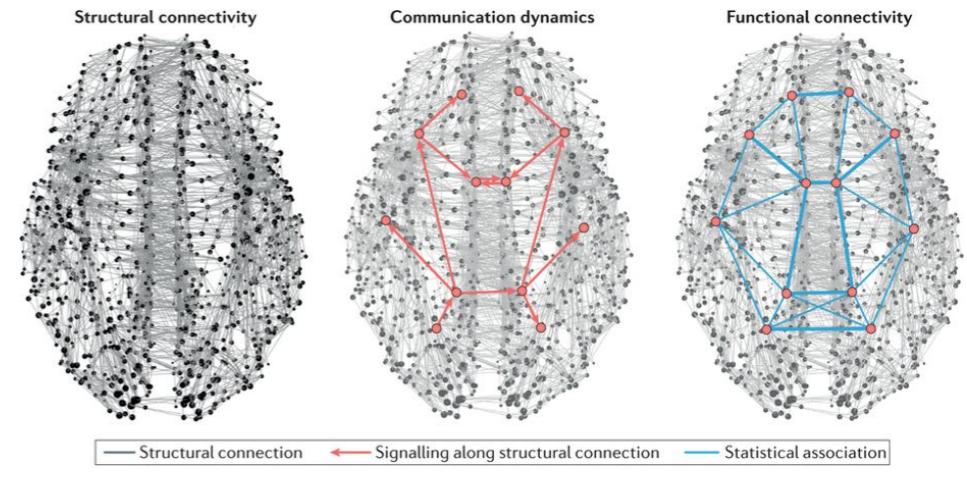
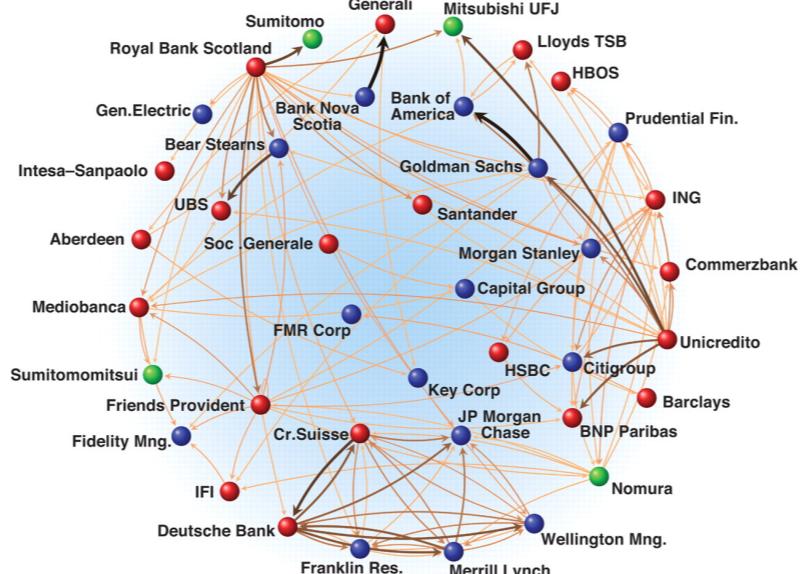
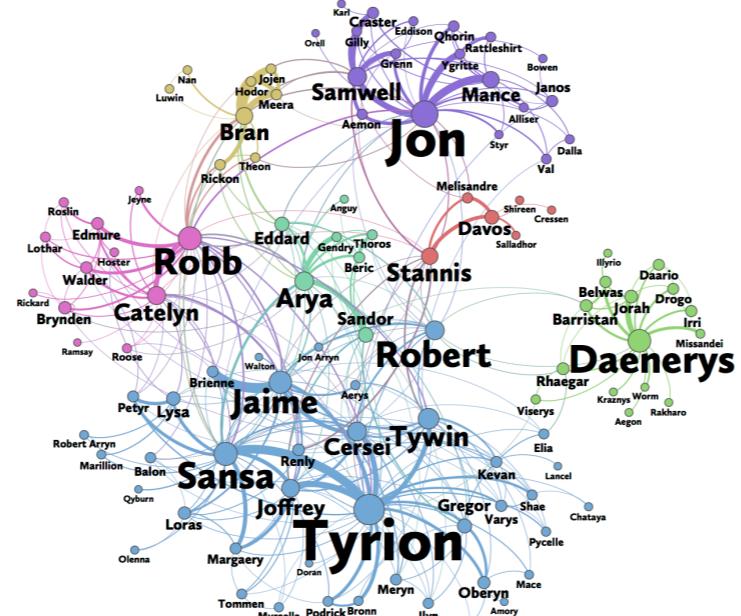
What are networks?

Social

Economic

Communication

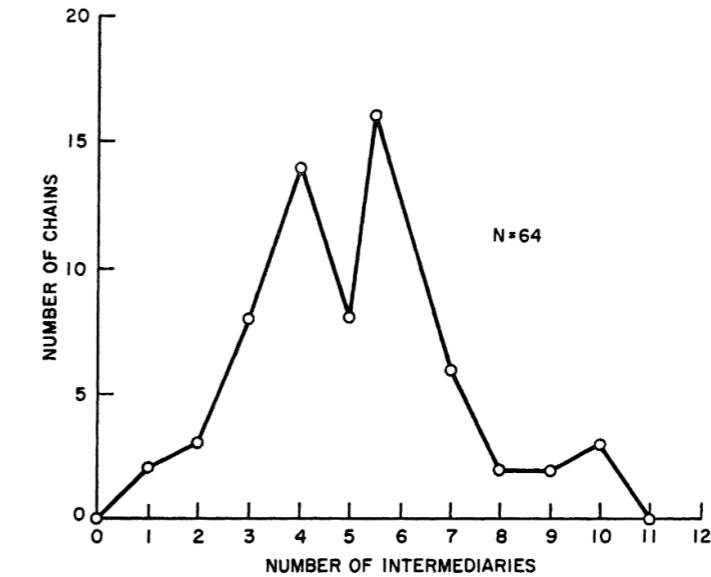
Neuronal



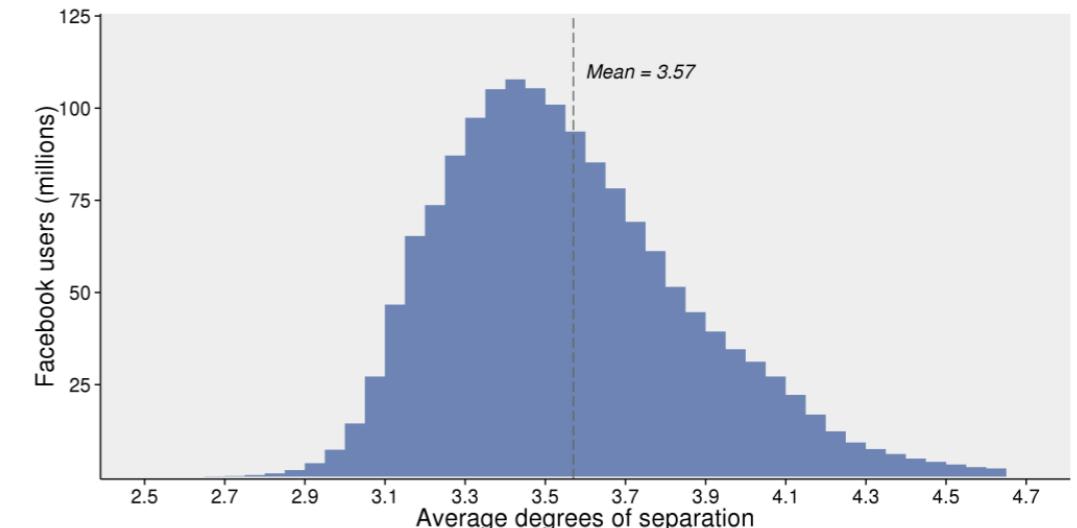
Nature Reviews | Neuroscience

Small world

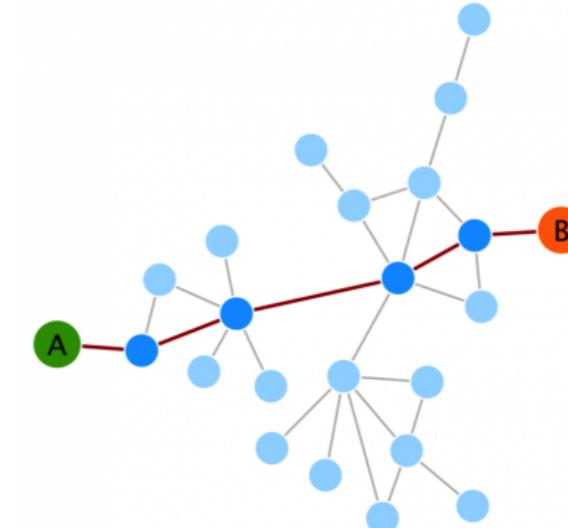
Stanley Milgram (1967) - 6 degrees
("6 degrees of Kevin Bacon")



Backstrom et al. (2016) - 3.6 degrees



Biological Networks



What are biological networks?

Protein - Protein interaction (PPI) networks

Transcription-factor regulatory networks

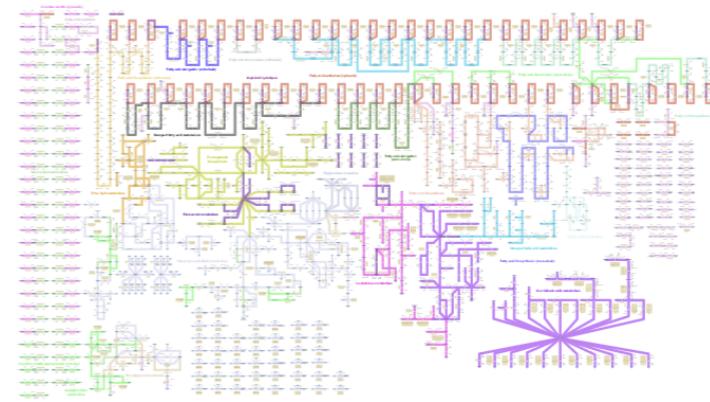
Gene - gene co-expression networks

Signal transduction networks

Drug-disease association networks

What are biological networks?

Metabolite - Enzyme - Signal - Genes (GEMs)

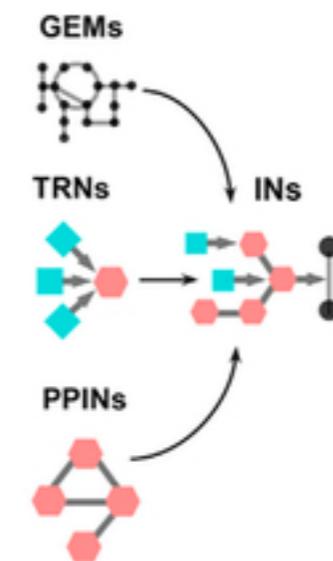
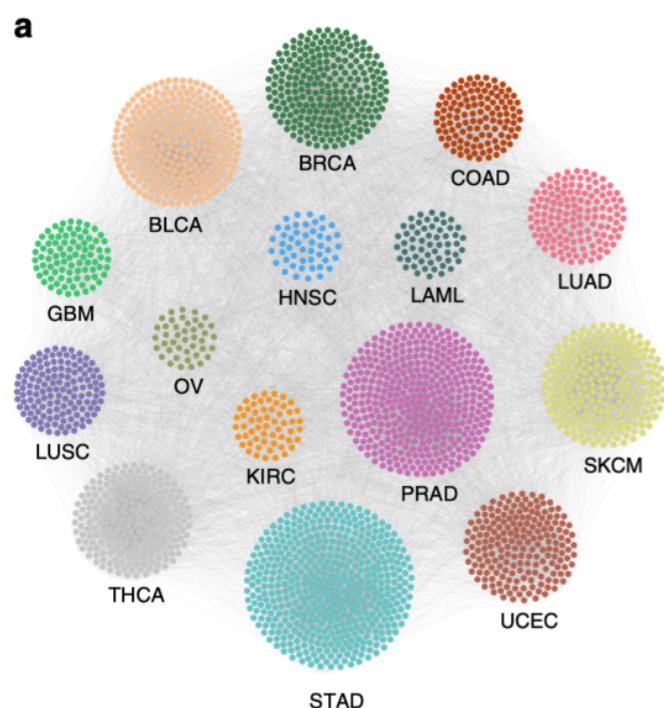
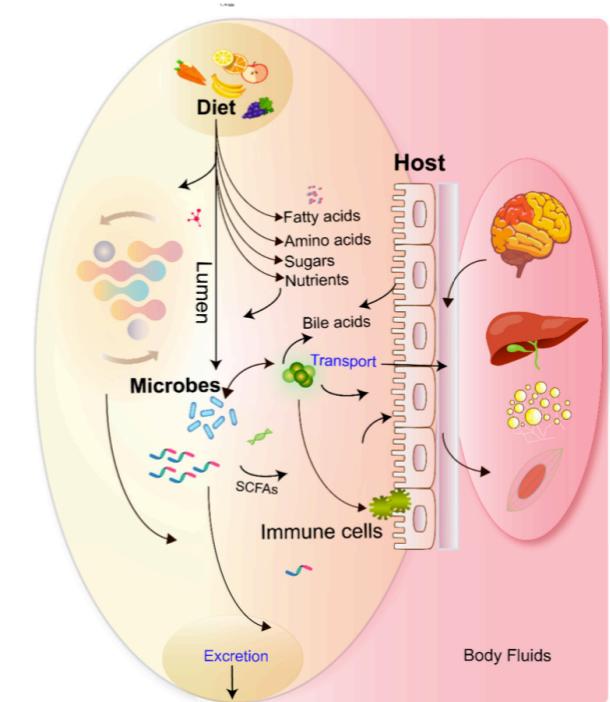
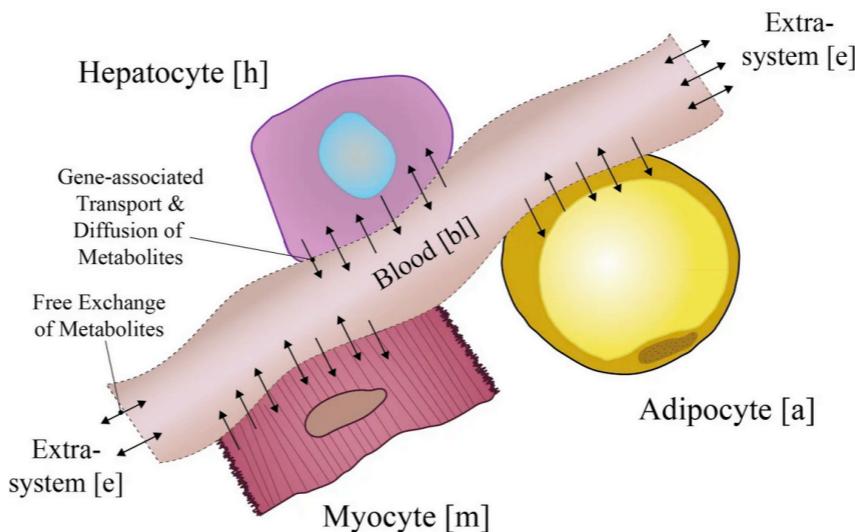


Multi-tissue networks

Multi-species networks

Disease networks

Integrated networks



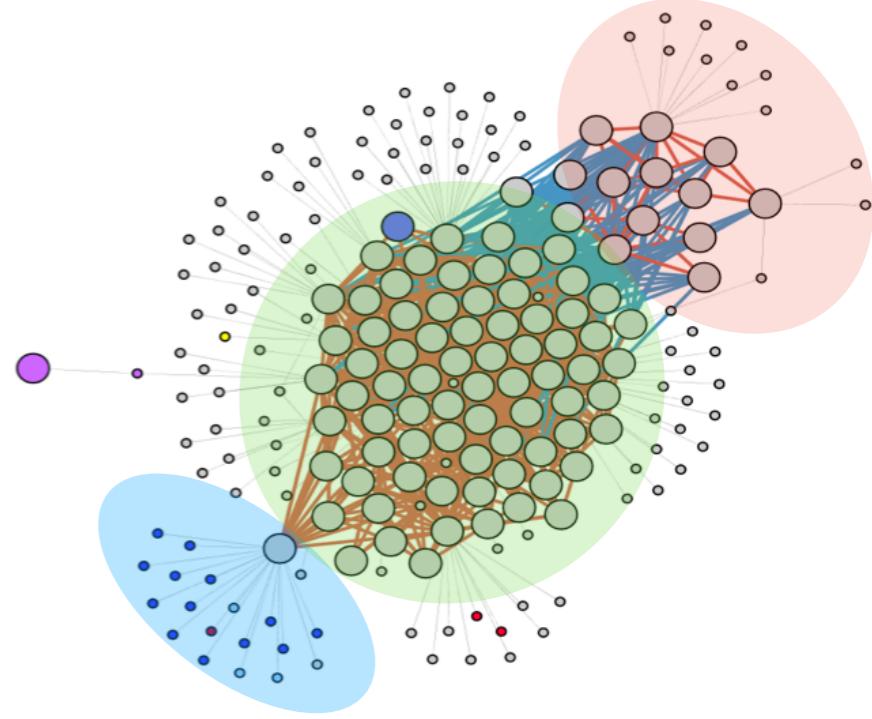
Why look at network topology?

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?

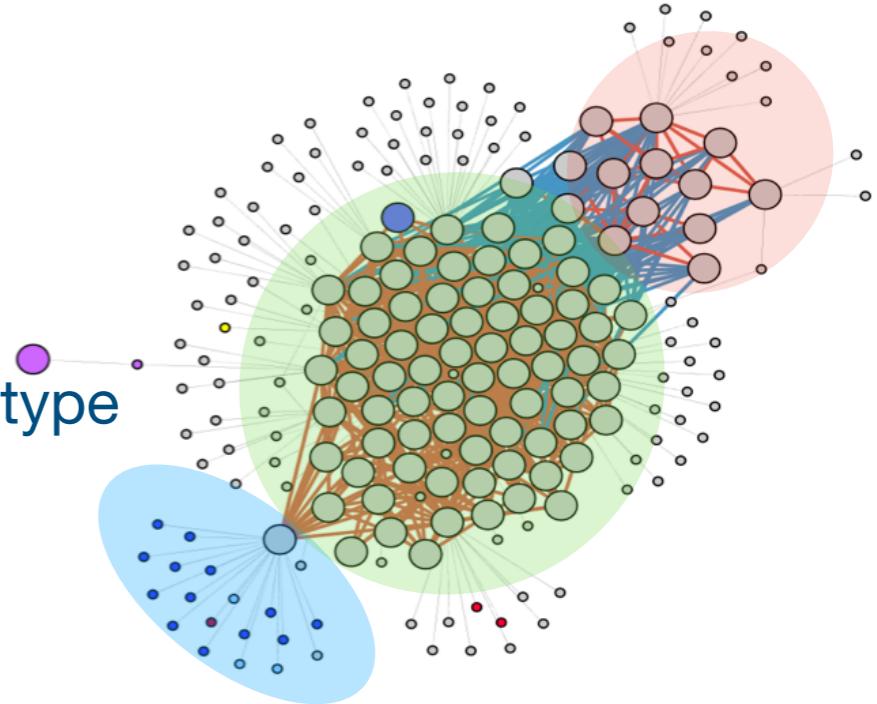


What is my biological network?

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)

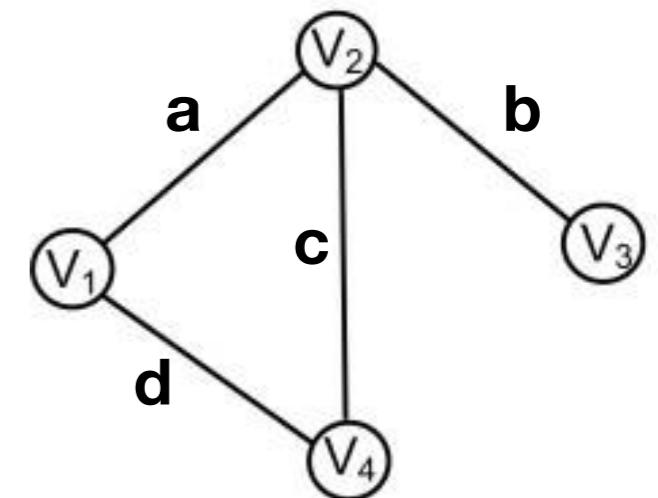
$$G = (V, E)$$

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

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

Nodes sometimes called **vertices**

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

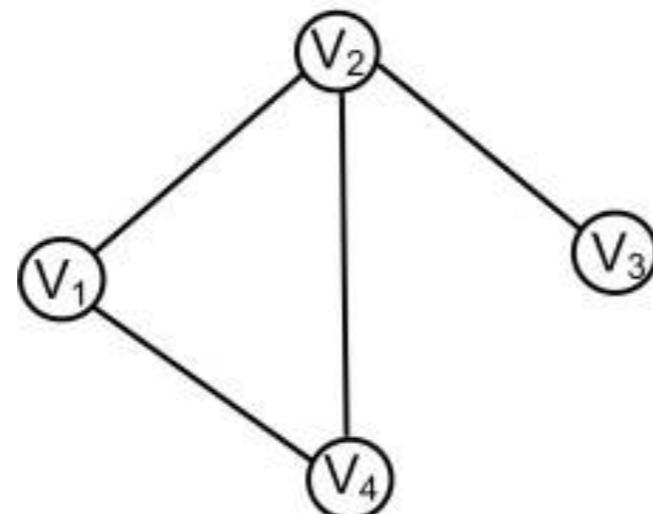


Simple vs multigraphs

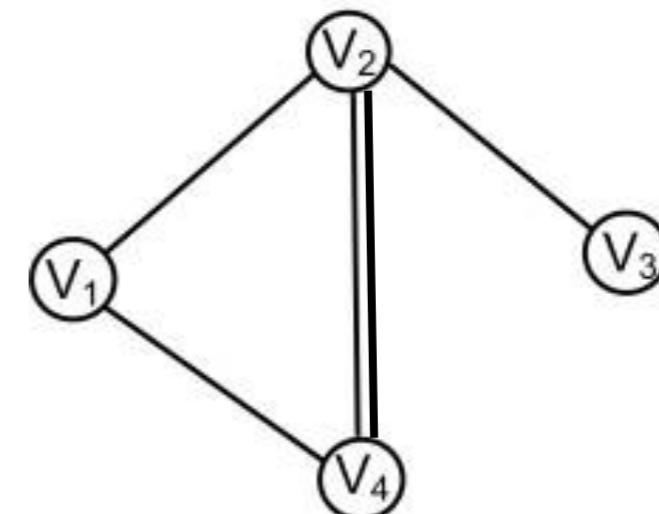
Multigraphs contain parallel edges

Multi-edged connections indicate different properties

Simple

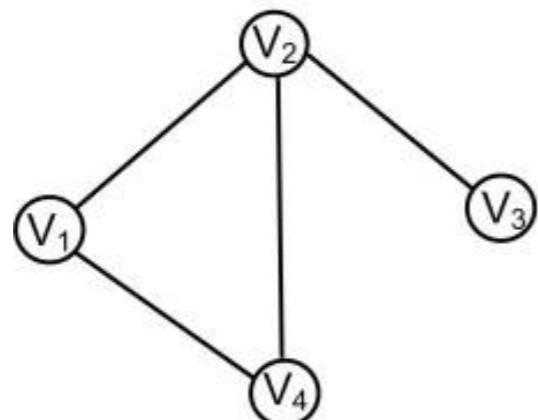


Multigraphs



Directed vs undirected graphs

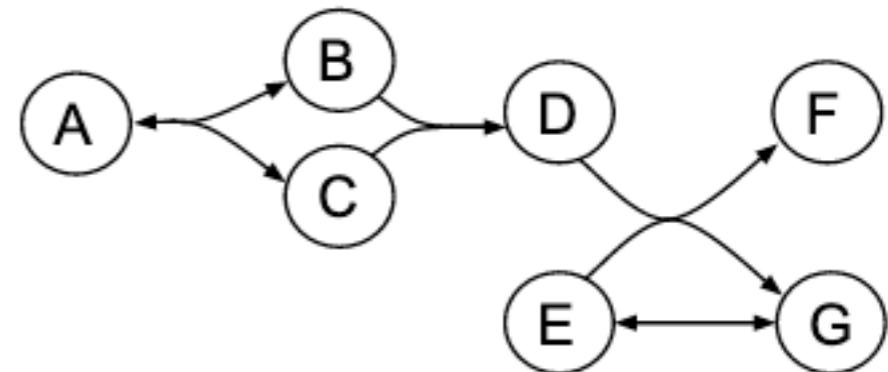
Undirected graphs:
co-expression networks



Directed graphs:
metabolic networks

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

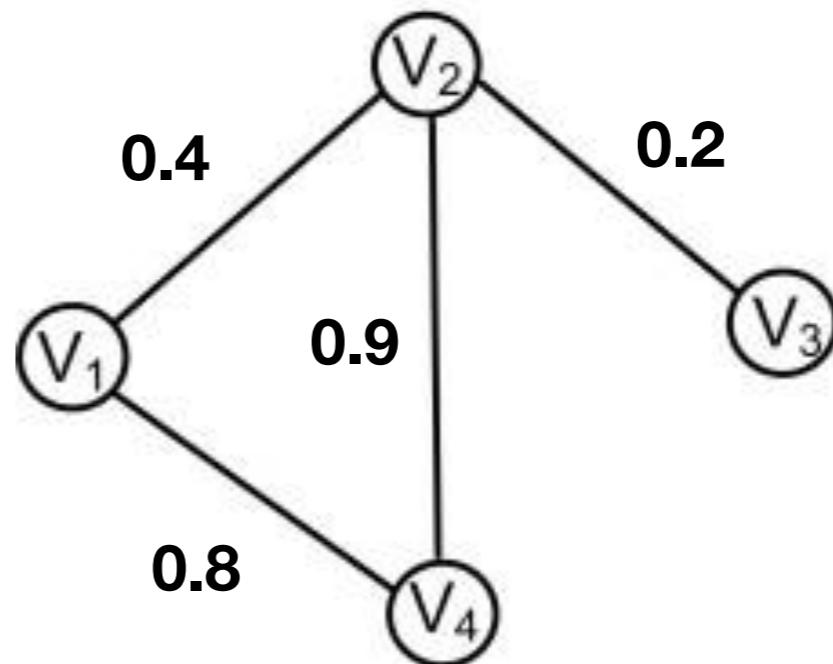
(a) Reaction network



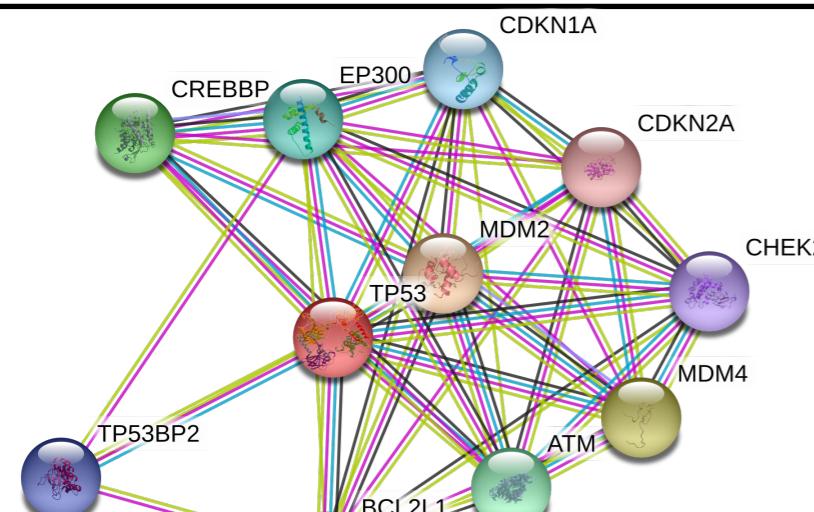
Weighted vs unweighted graphs

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

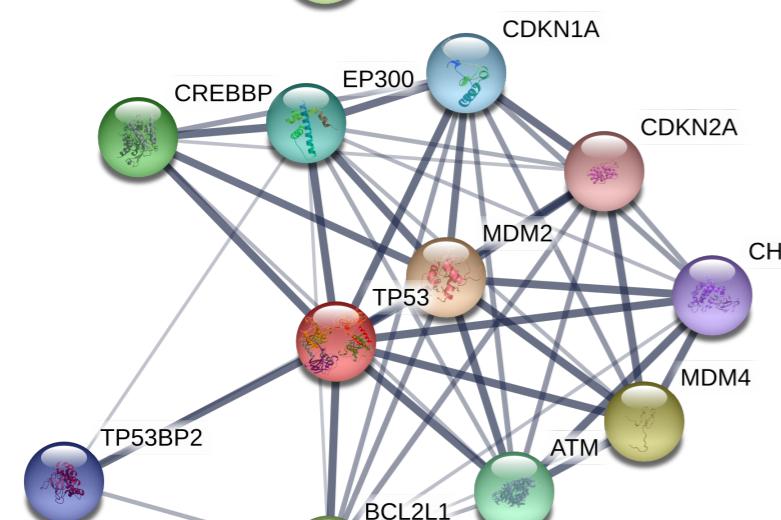
E.g. weighted co-expression networks



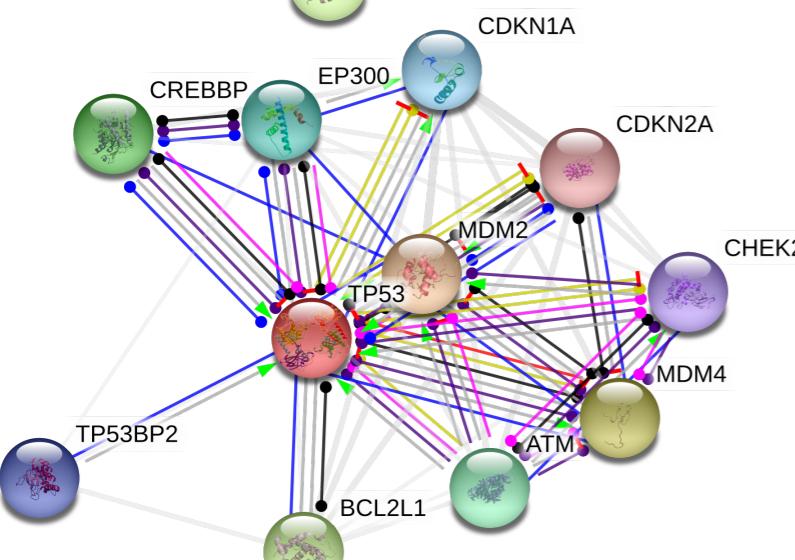
STRING-db.org: TP53



Multi-edged



Weighted multi-edged



Multi-edged directed



Bipartite graphs

A graph

$$G=(V,E)$$

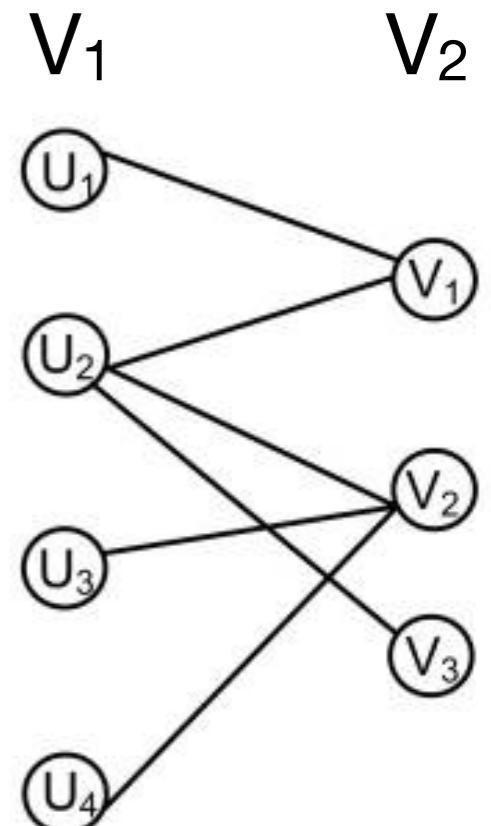
may be partitioned into two sets of nodes (V_1 , V_2)
such that

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

All e_i has end-nodes in V_1 , V_2

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

$$G_1 = (V_1, E_1)$$



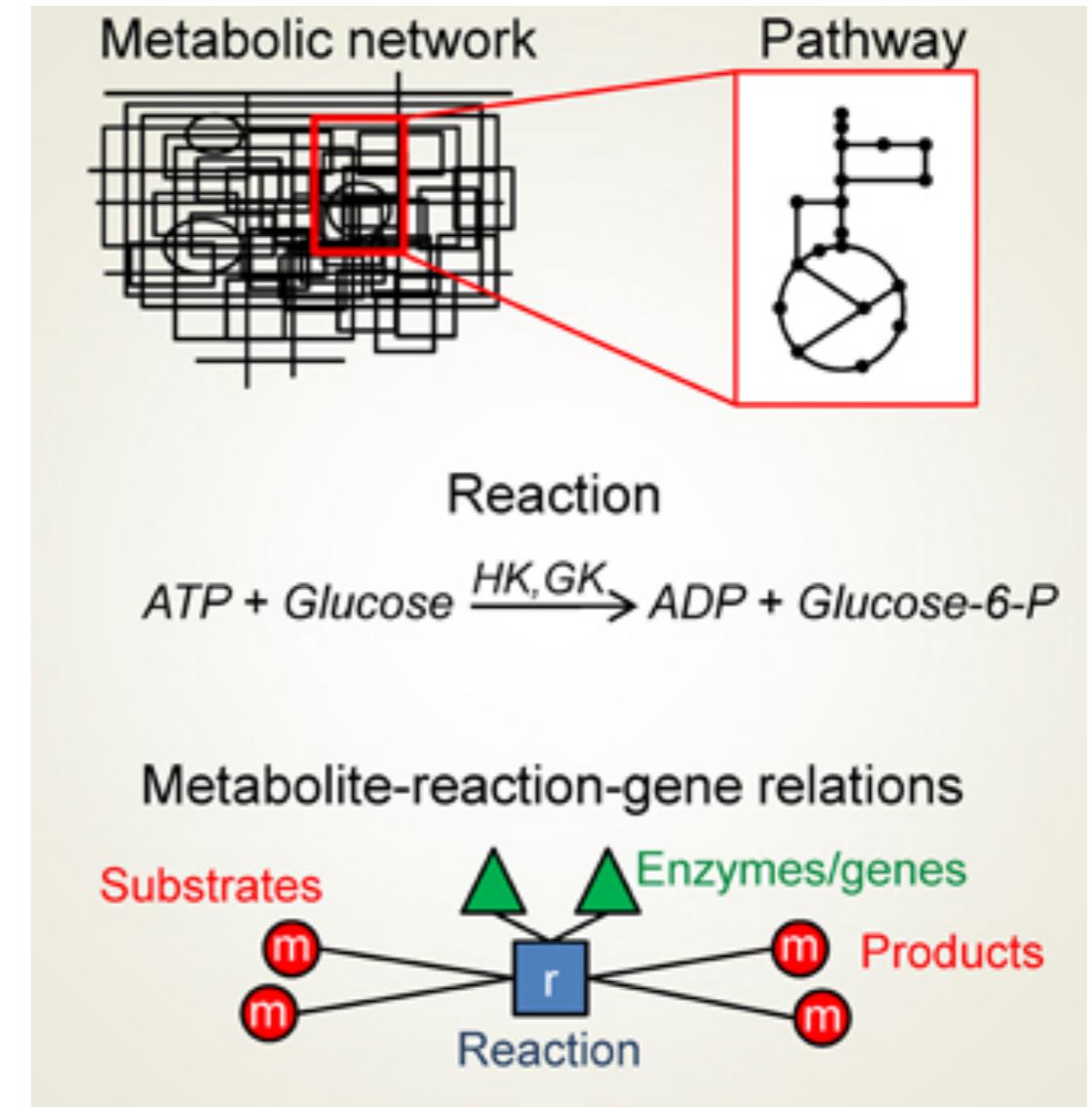
Bipartite and k -partite graphs

Example of bipartite graph:

Enzyme - Reaction

Metabolite - reaction - enzyme

k -partite graphs display k -types of nodes

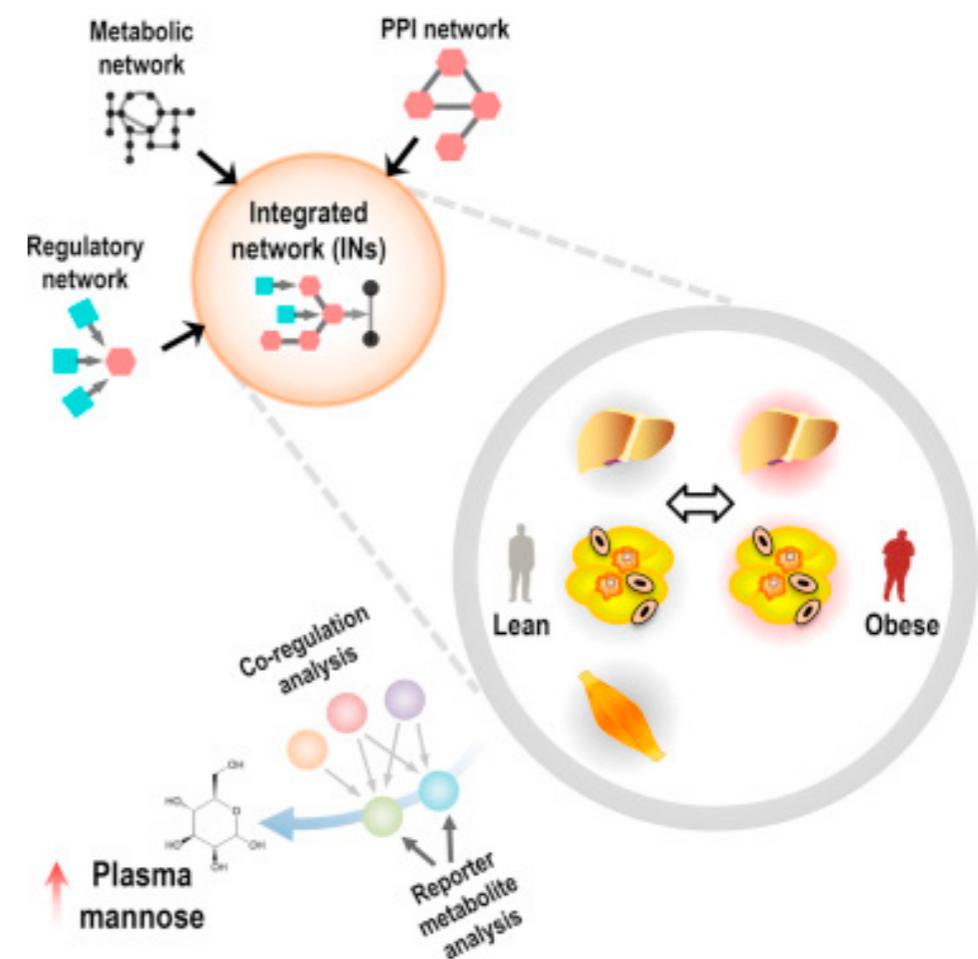


k-partite graphs

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)

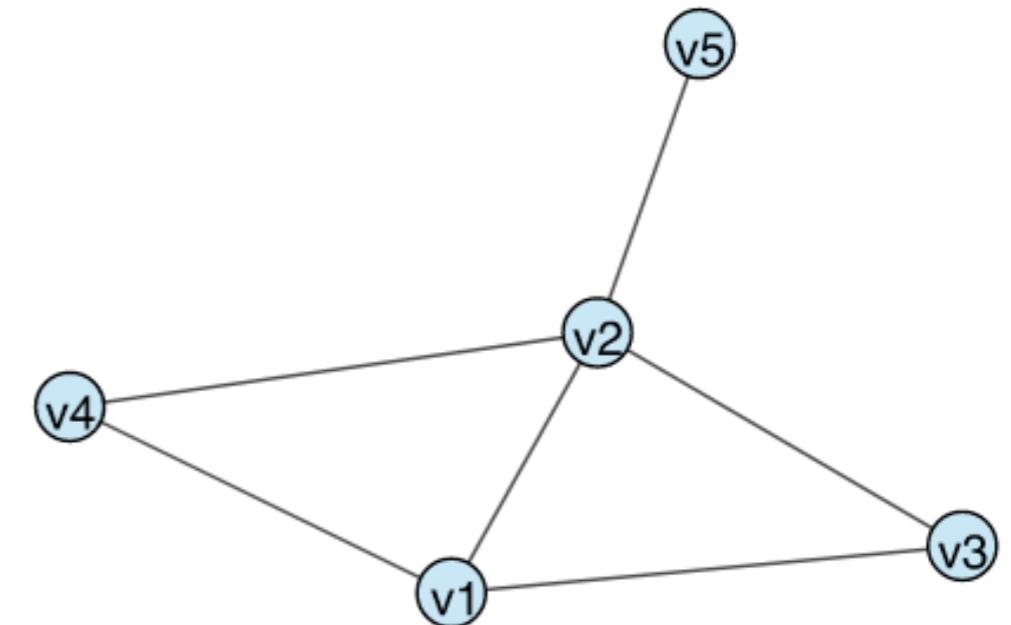
Integrated Networks



Adjacency matrix (undirected graphs)

**Vertex association
(undirected network)**

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

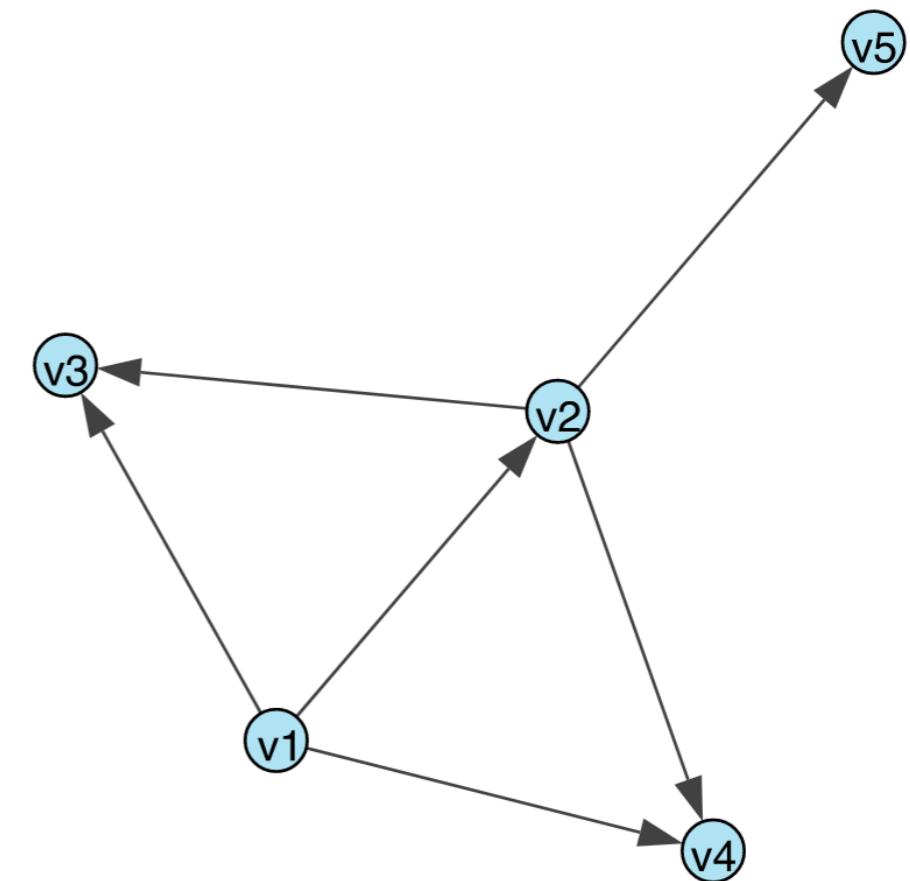


Adjacency matrix is symmetric

	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

Adjacency matrix (directed graphs)

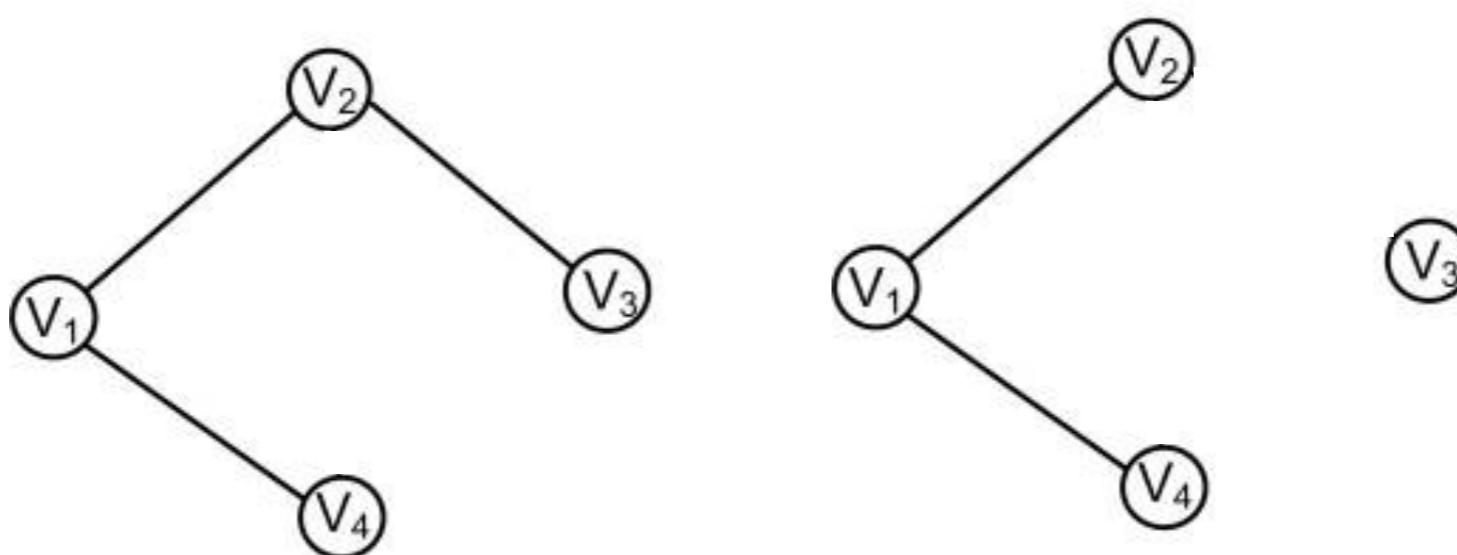
	Target				
	v1	v2	v3	v4	v5
v1	0	1	1	1	0
v2	0	0	1	1	1
v3	0	0	0	0	0
v4	0	0	0	0	0
v5	0	0	0	0	0



Connected vs disconnected networks

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

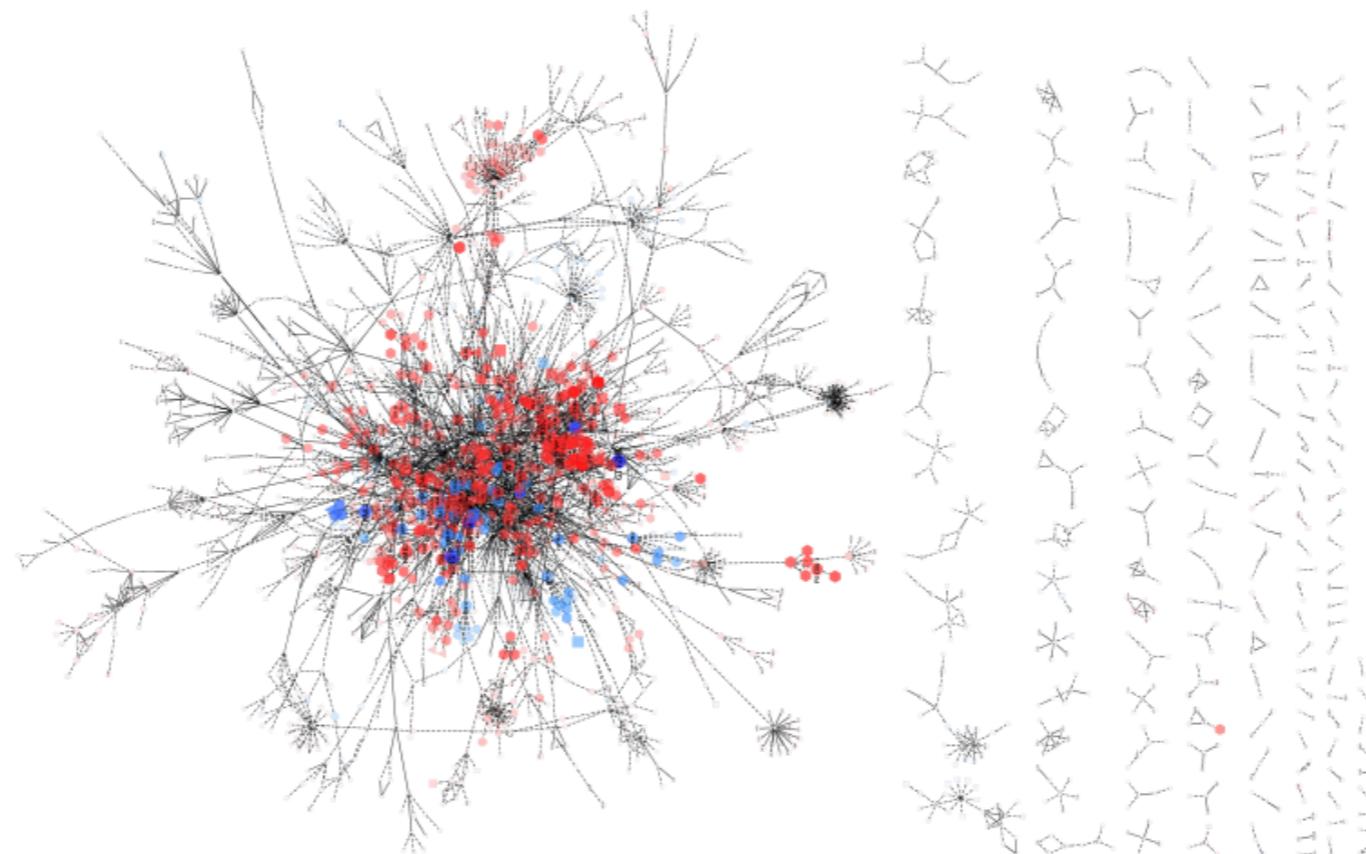
Disconnected network: some of the nodes are unreachable



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)**

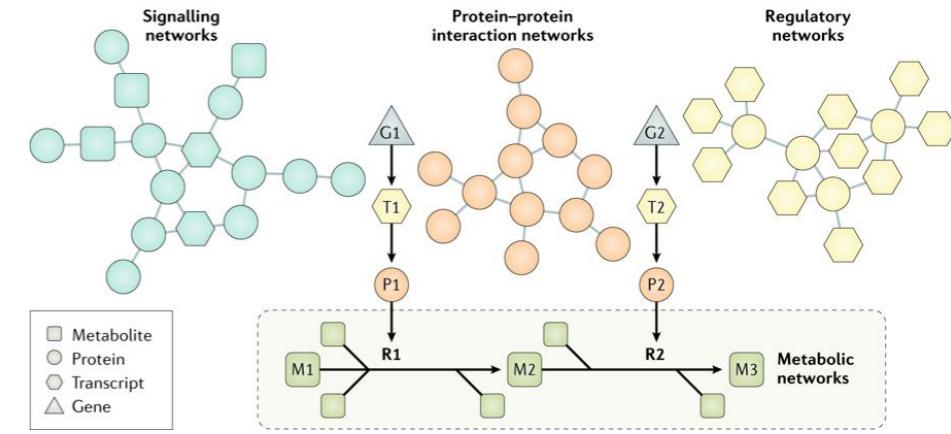
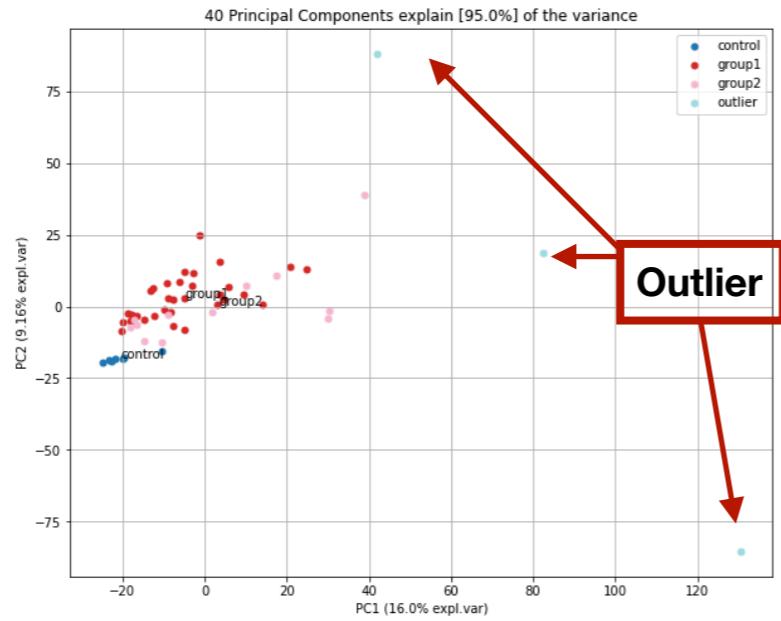


Overview

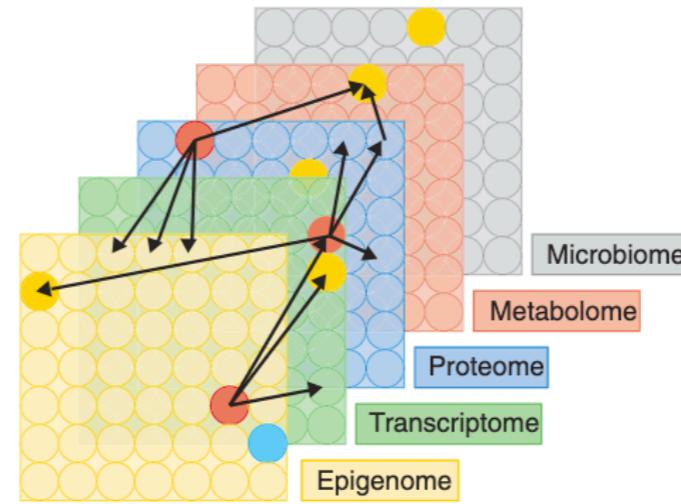
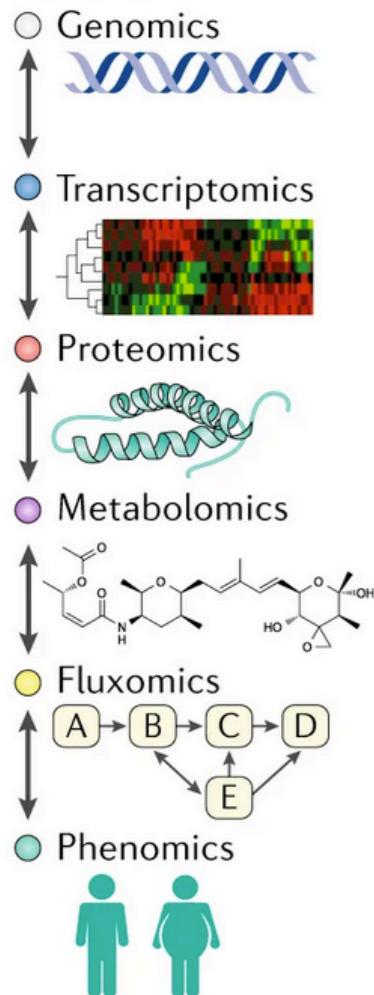
1. Introduction to network analysis
2. Terminology
- 3. Network inference**
4. Key network properties
5. Community analysis

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

Building networks



Raw → Pre-processing → Distance calculation → Graph analysis



Hasin 2017

Piening 2018

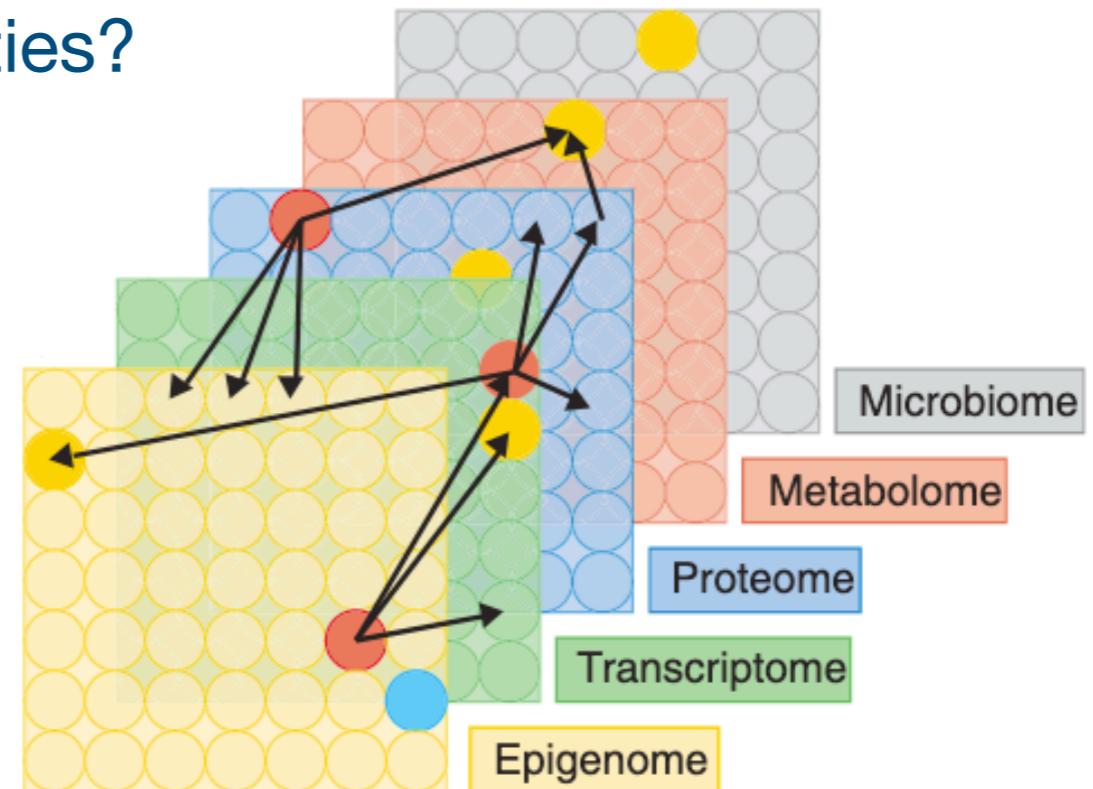
Mardinoglu 2018

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?



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

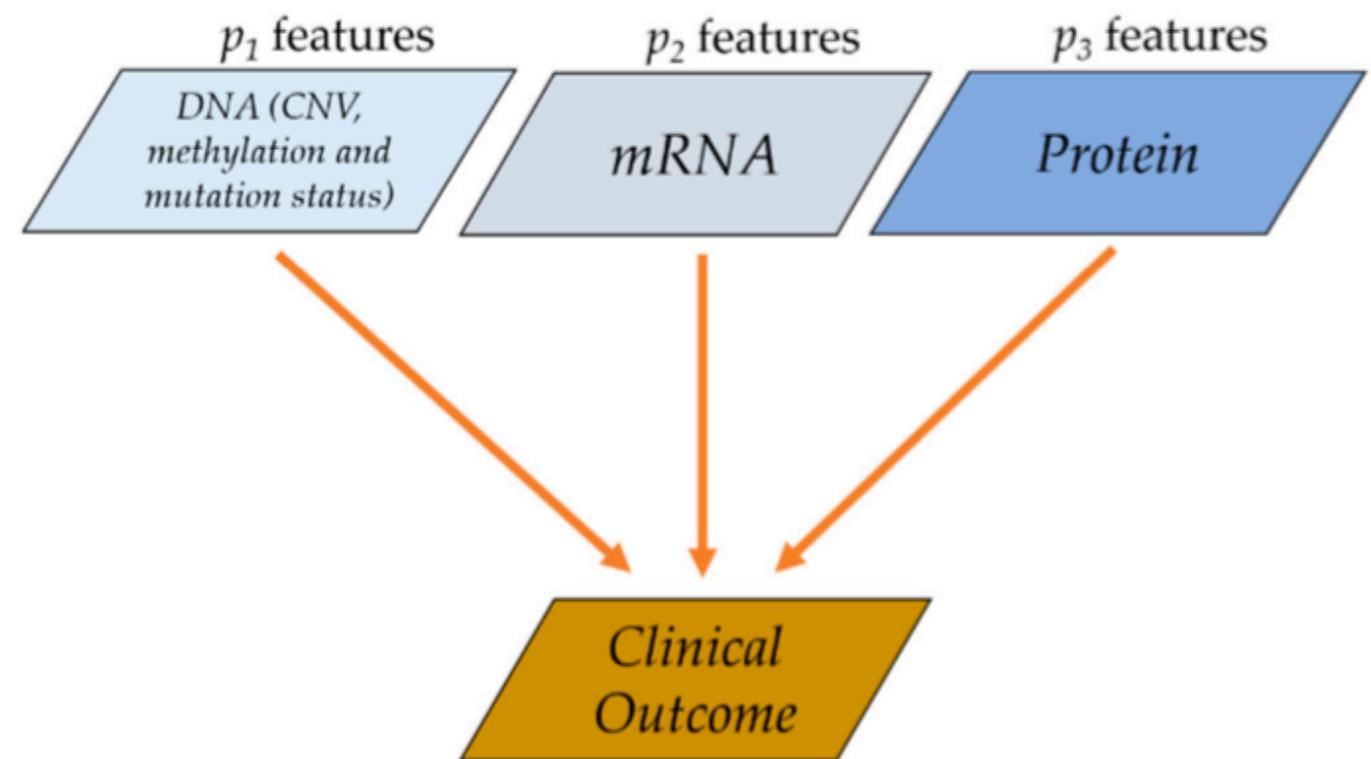
Balanced dataset for group sizes

Normalization may be needed to make omic datasets comparable (e.g. standardization)

Common approach: compute correlations between different features

- Spearman
- Pearson

Extend known associations



1. Association analysis

Easy to interpret

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

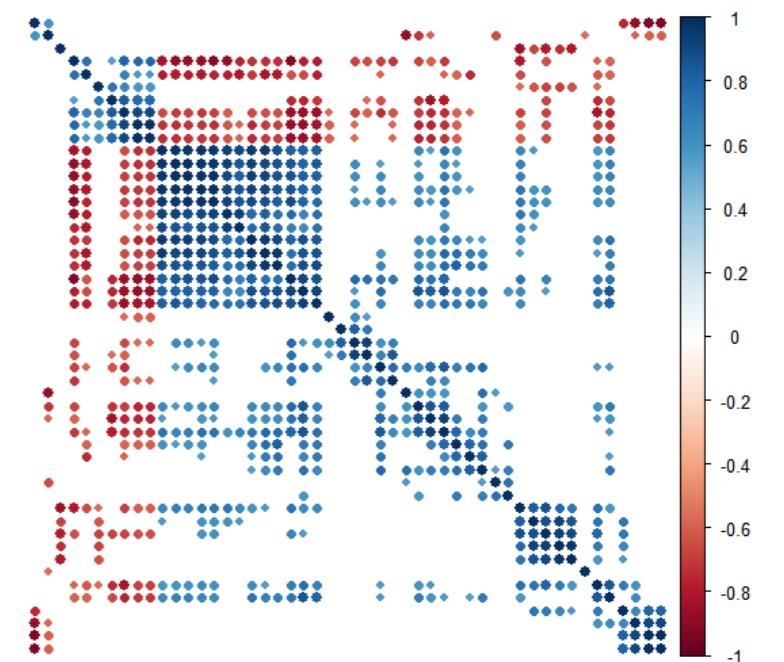
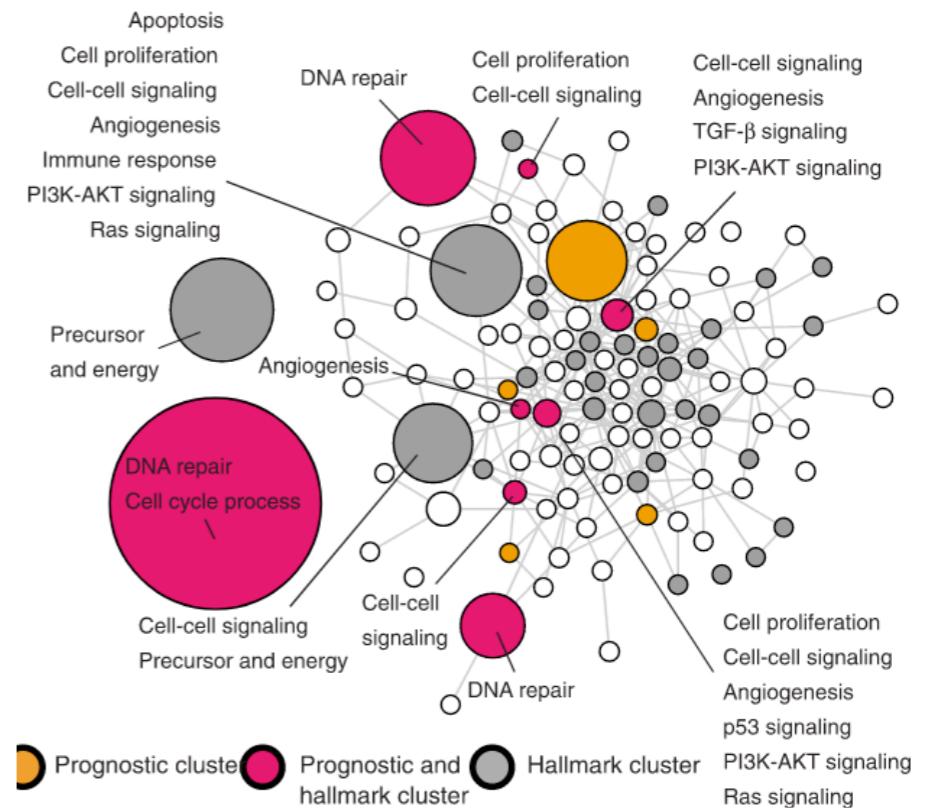
Unbalanced networks

Prone to type I errors

Filtering

- FDR vs Bonferroni
- Effect size cut-off

Need adjustment to possible confounding factors

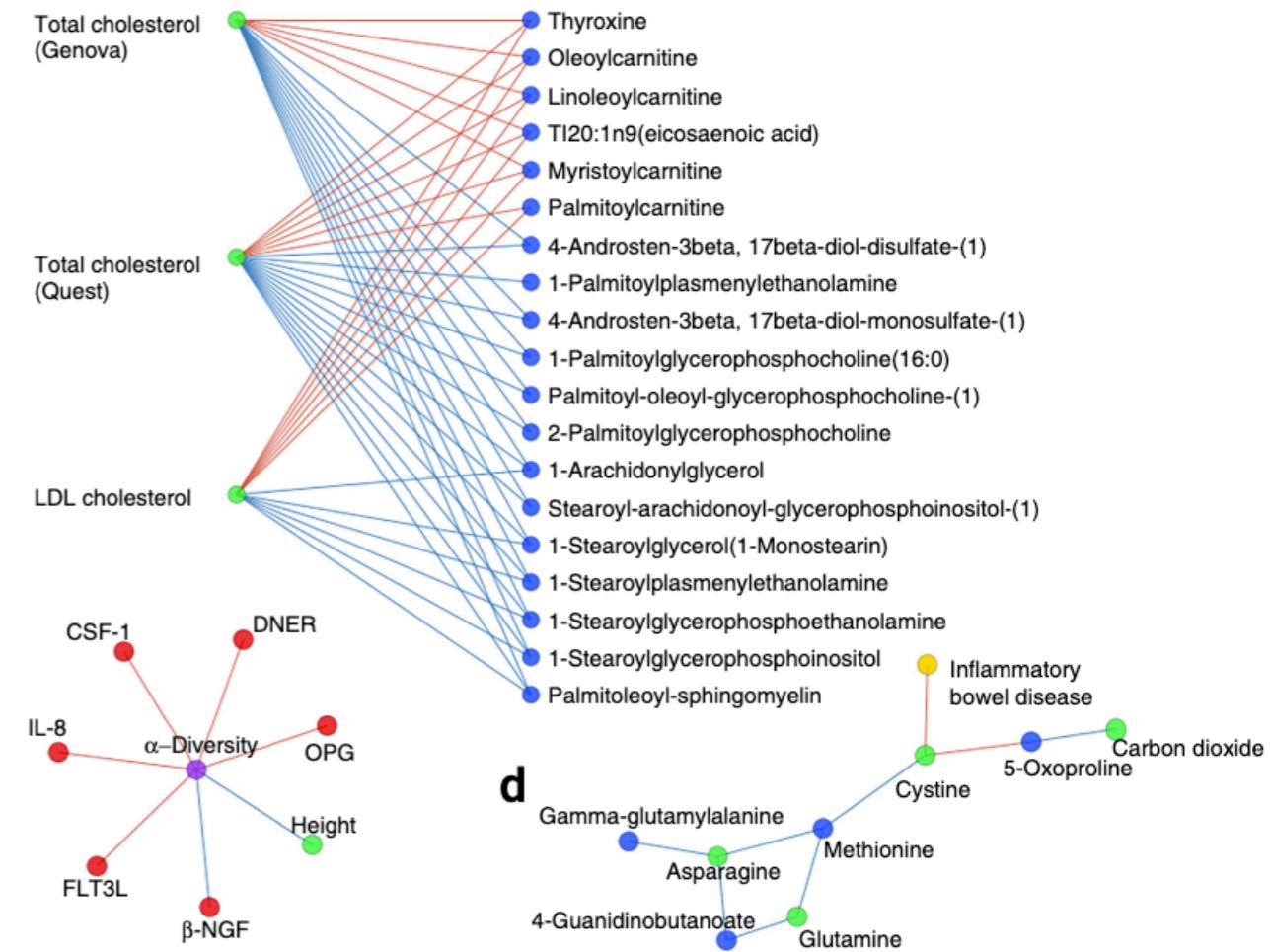
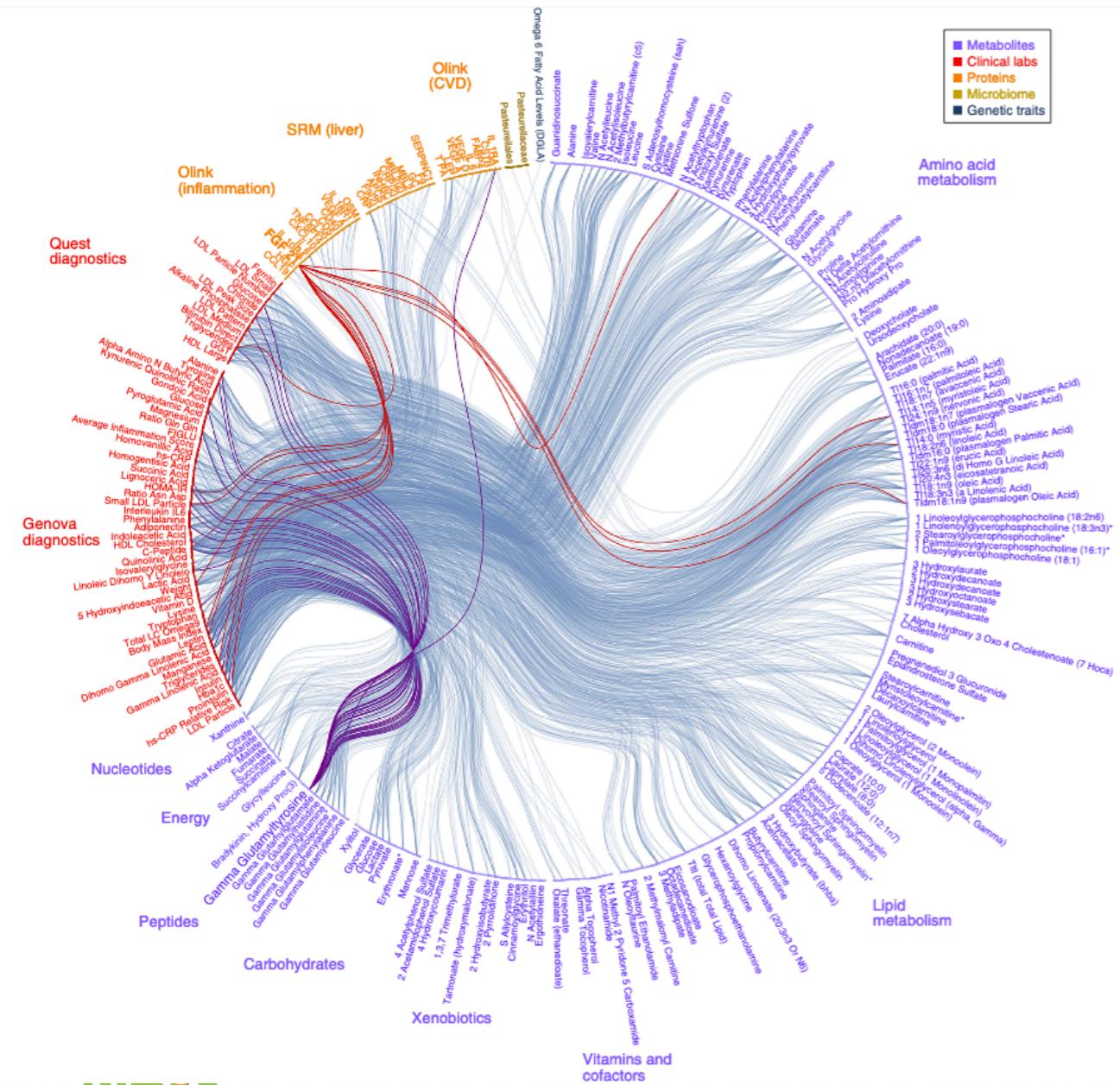


1. Association analysis

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



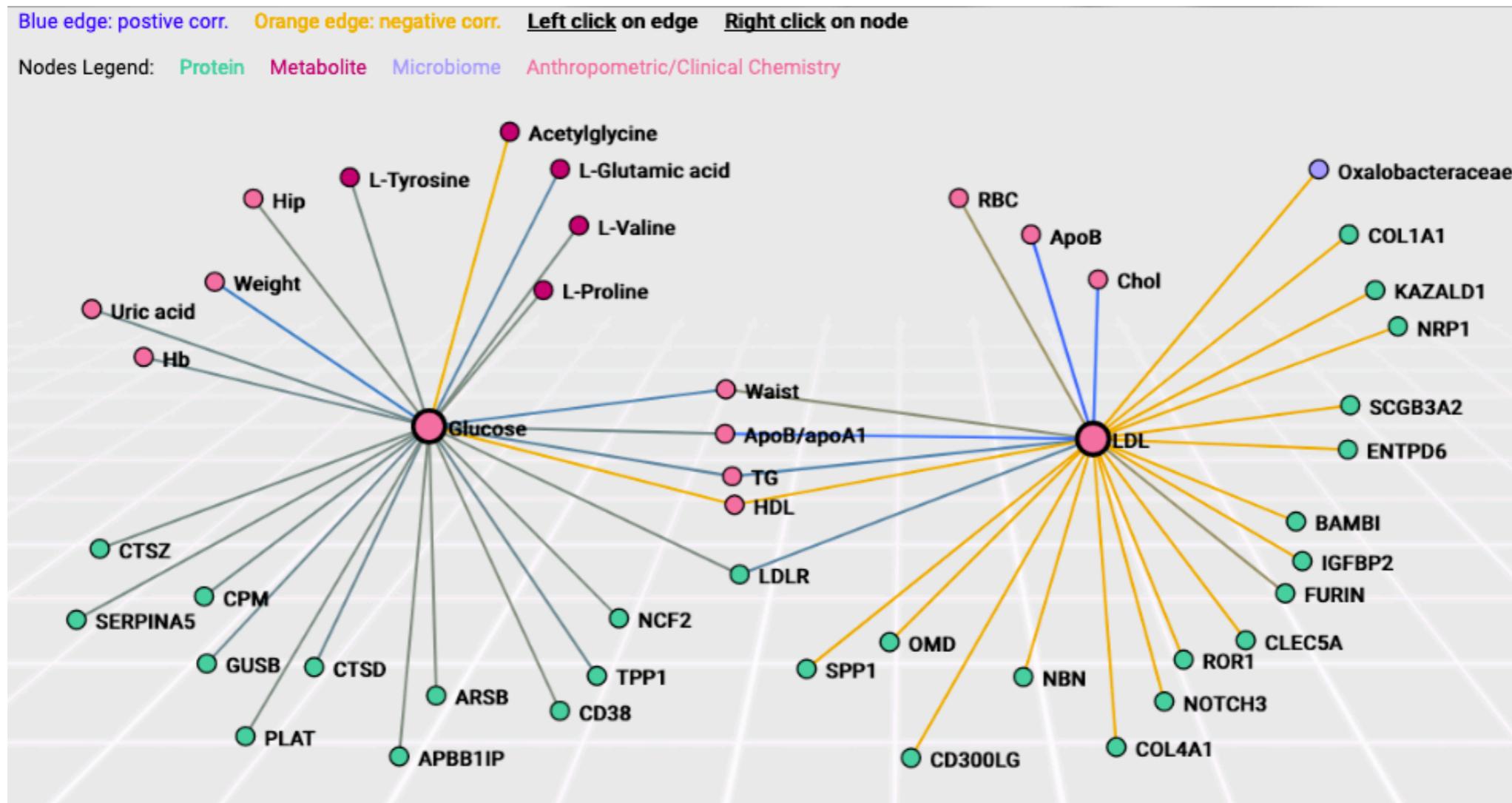
Overview

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?



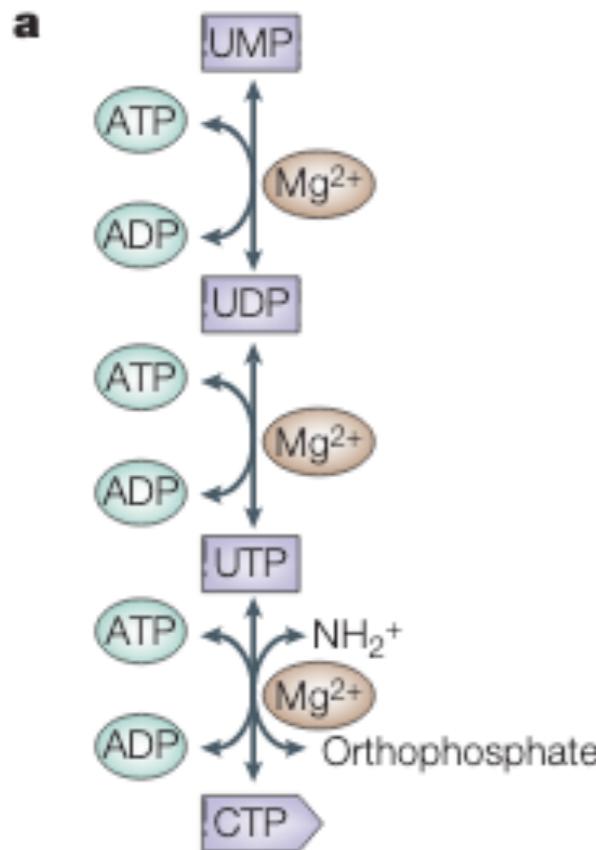
Key network properties to discuss

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

1. Network representations

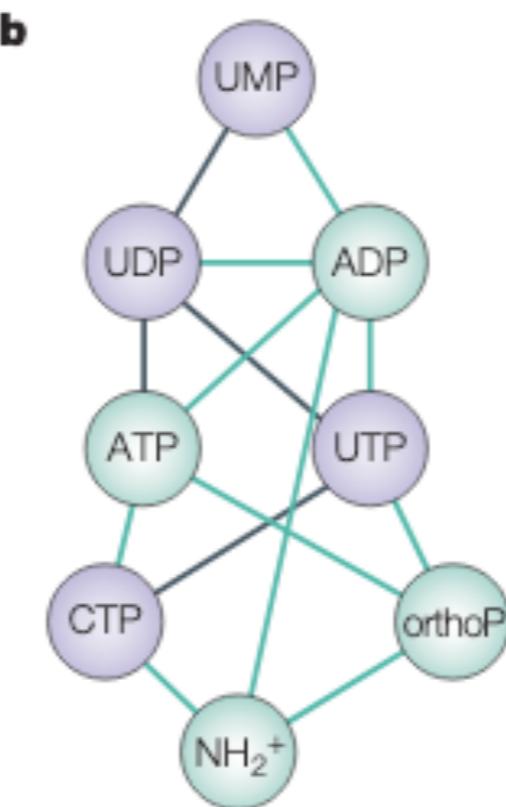
Representations of a metabolic network: pyrimidine metabolism

Metabolism



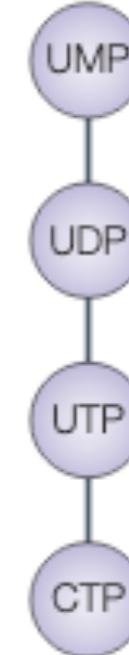
(directed graph)

Graph representation:
metabolites and co-factors



(undirected graph)

metabolite-metabolite
association



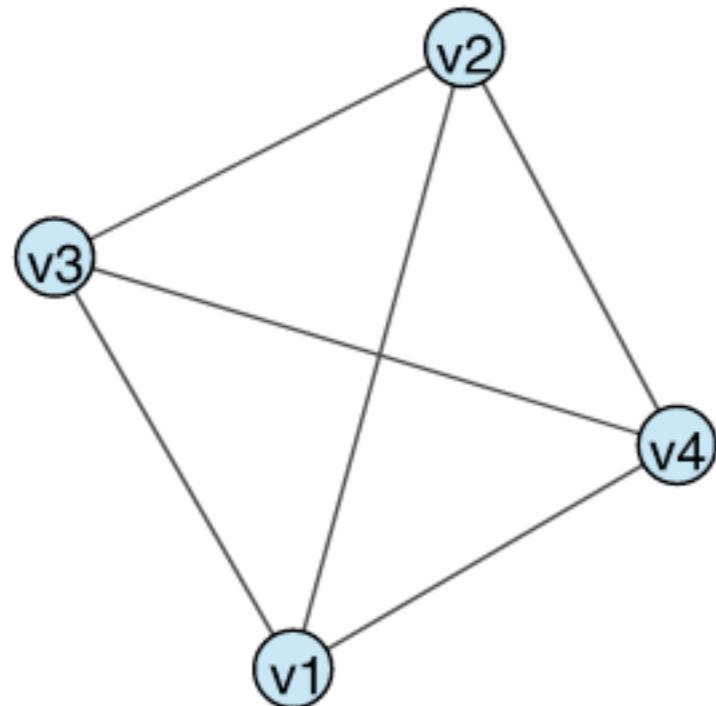
(undirected graph)

Other representations: Protein-Protein, Protein-Metabolite

2. Network density

For a graph with **V nodes**, the total number of possible **edges** is given by

Undirected graphs:
$$\frac{|V| \cdot (|V| - 1)}{2}$$



2. Network density

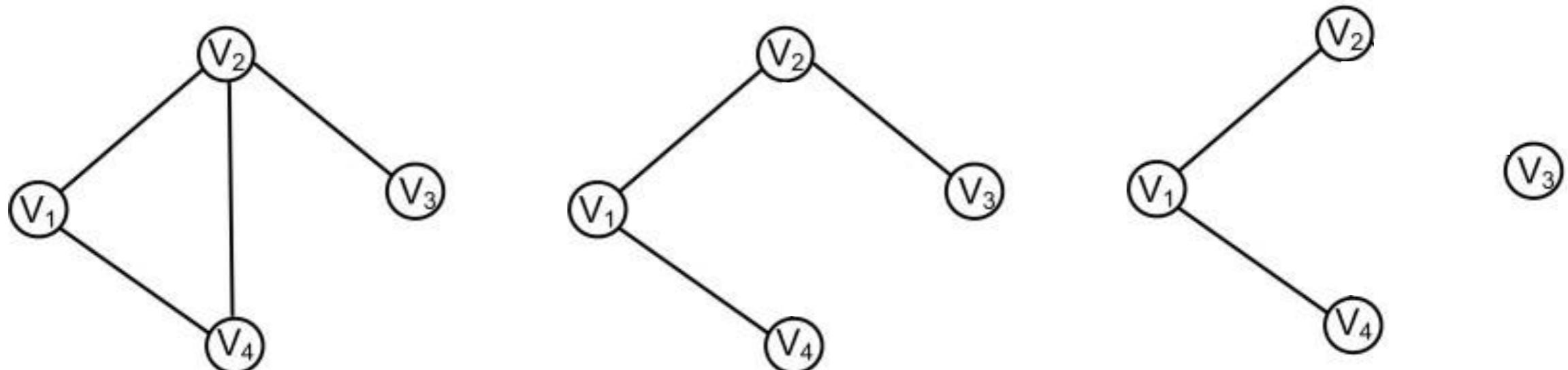
A **dense graph** is a graph where the number of edges approximates the maximum possible number of edges for the given node number.

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

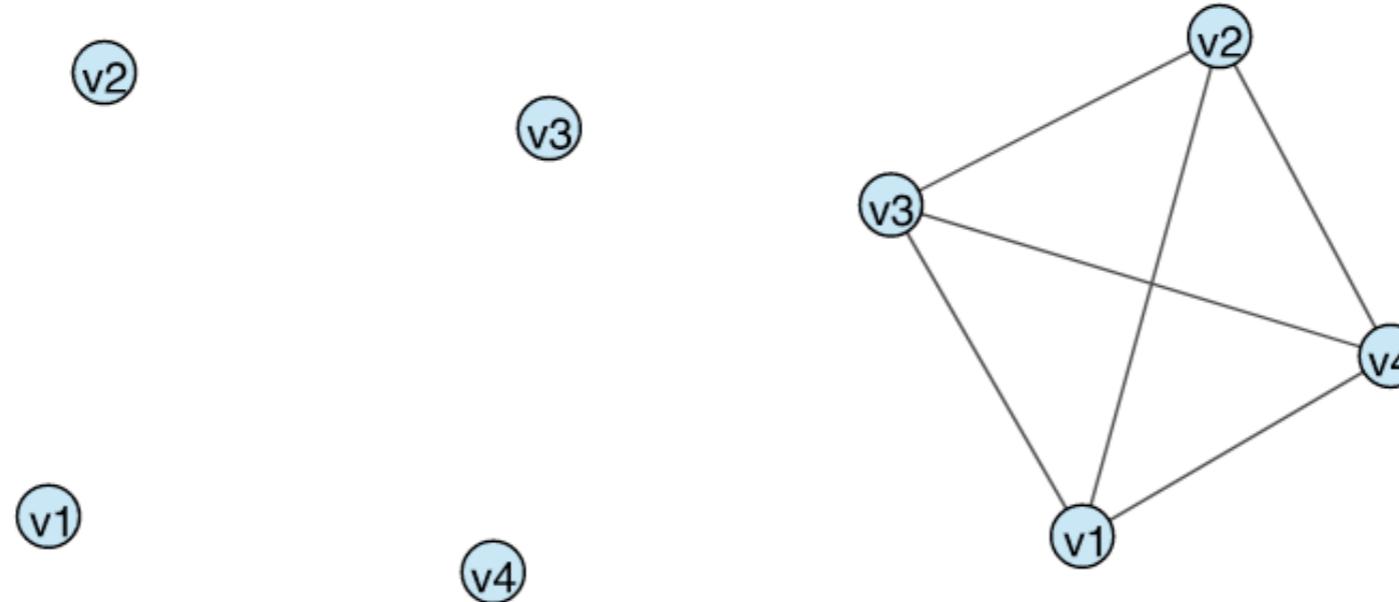
$$\text{Undirected graphs: } D = \frac{E}{V \cdot (V - 1)}$$

E : number of edges

V : number of vertices

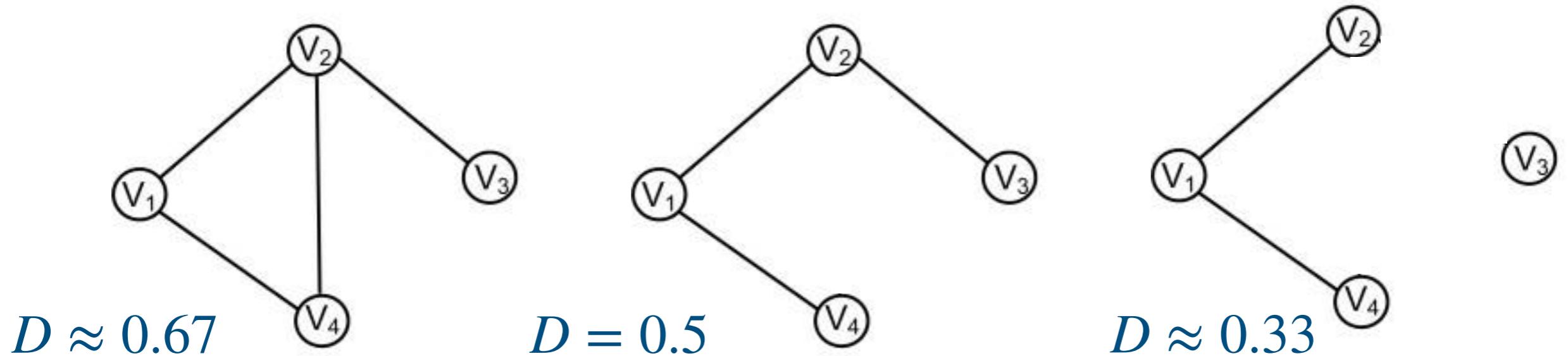


2. Network density



$$0 \leq D \leq 1$$

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



2. Biological network density

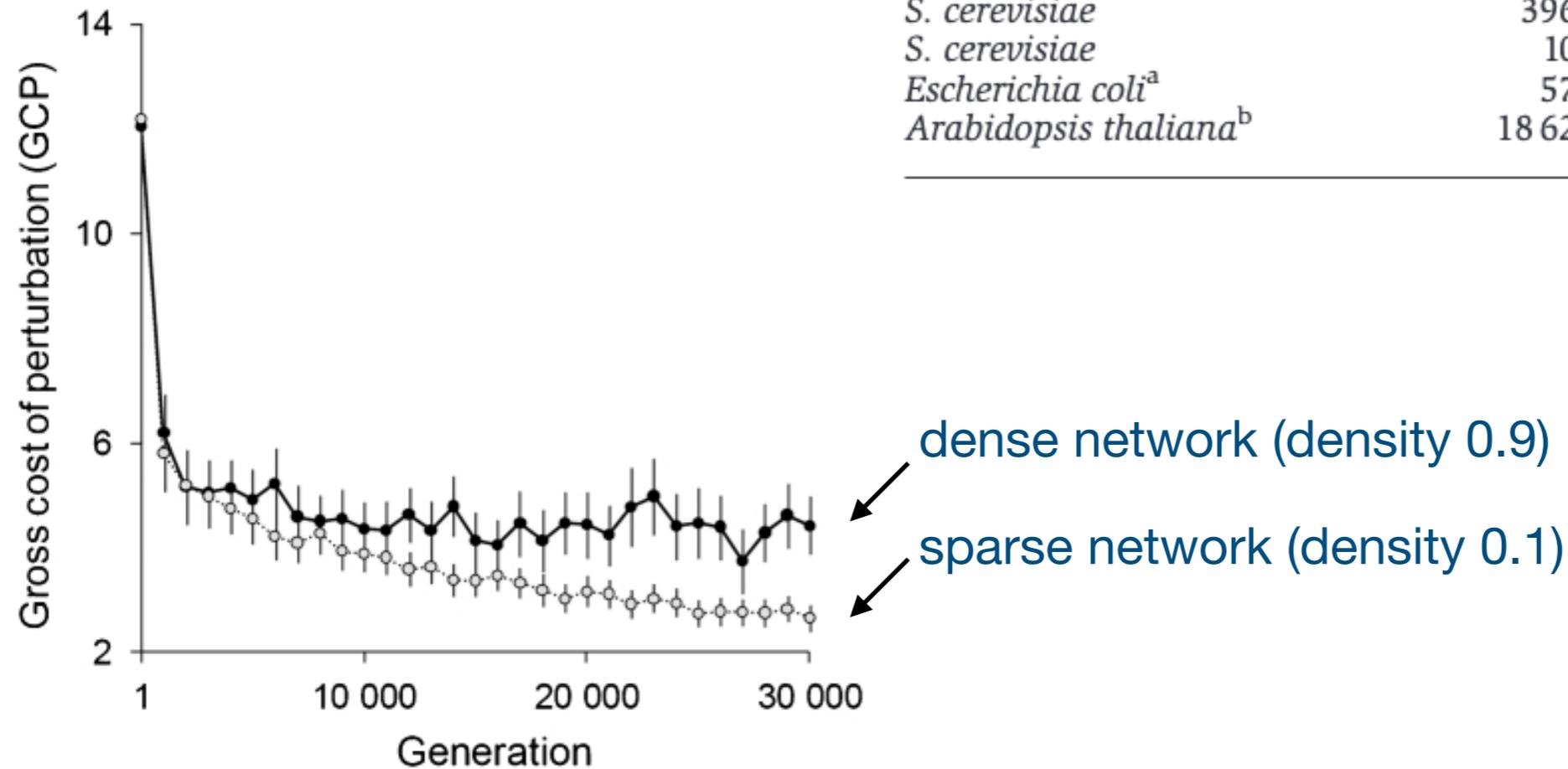
Evolutionary analysis of biological networks indicates general sparsity

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

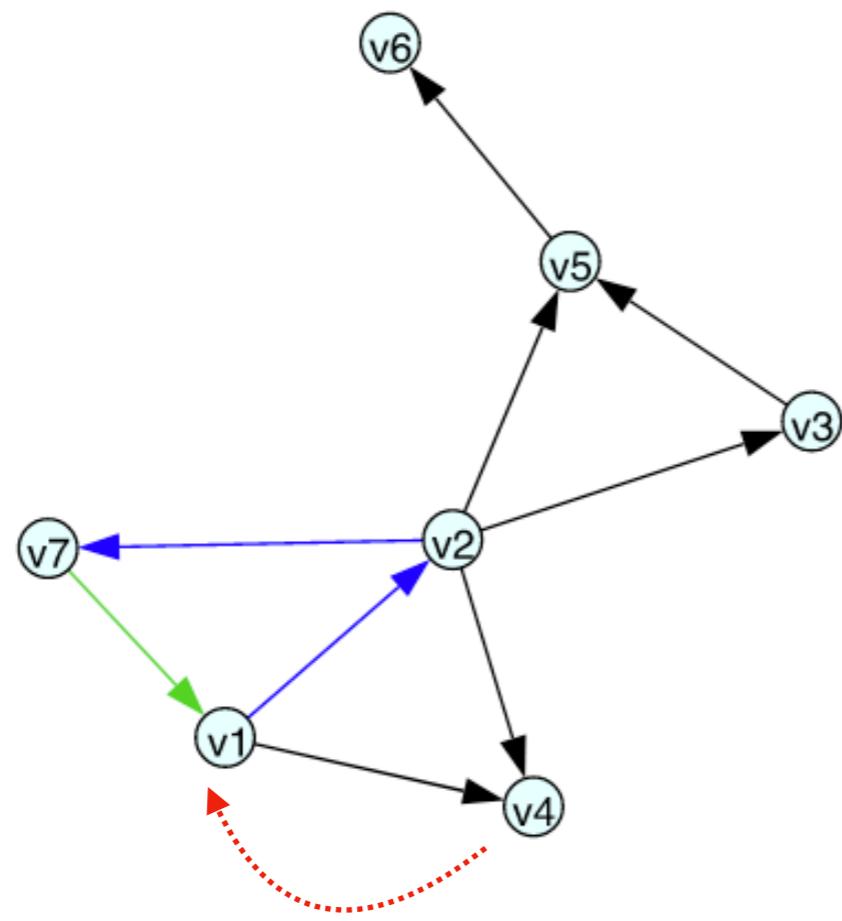
Organism	Interactions	Genes	D
<i>Drosophila melanogaster</i>	29	14	0.148
<i>D. melanogaster</i>	45	25	0.072
Sea urchin	82	44	0.0065
<i>Saccharomyces cerevisiae</i>	1052	678	0.0023
<i>S. cerevisiae</i>	3969	2341	0.0007
<i>S. cerevisiae</i>	106	56	0.0338
<i>Escherichia coli</i> ^a	578	423	0.0032
<i>Arabidopsis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



3. Paths

Distance between nodes is measured in path length

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

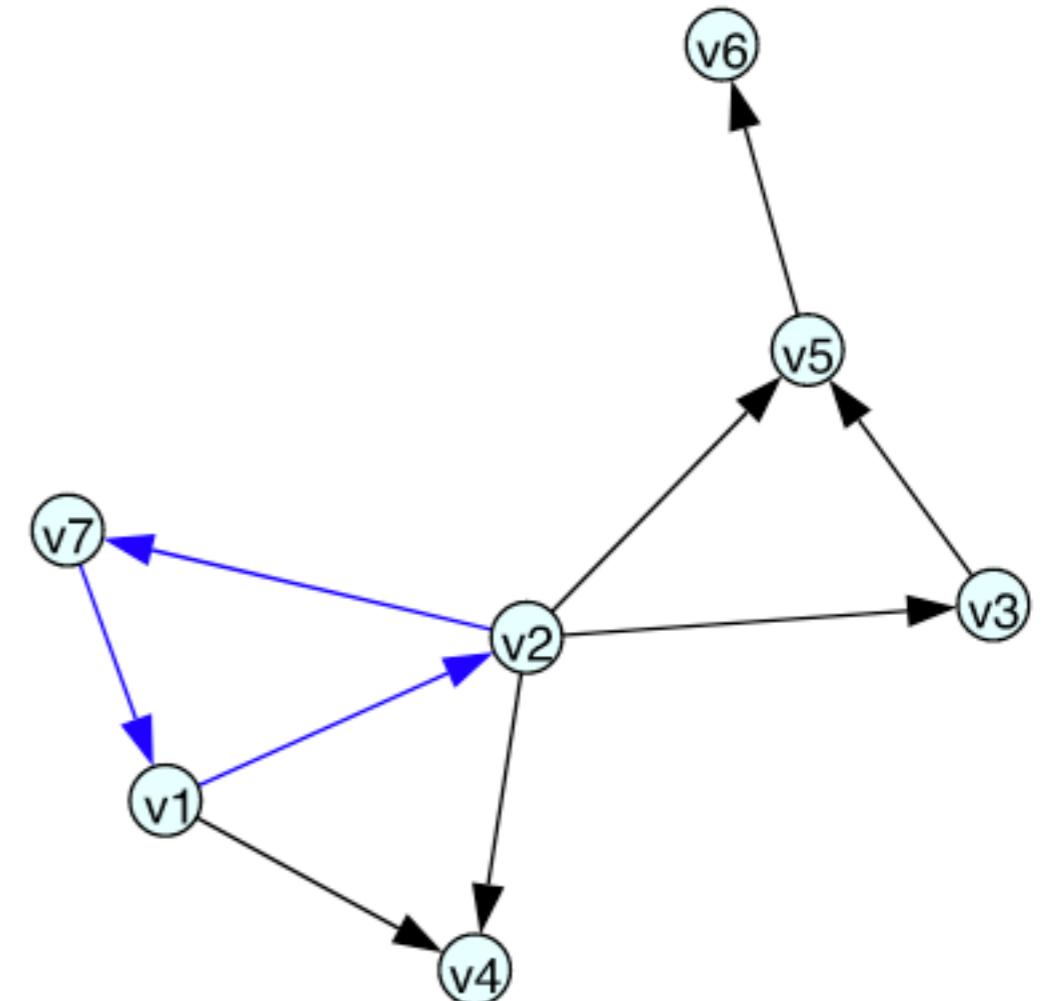


	v1	v2	v4	v3	v5	v7	v6
v1	0.0	1.0	1.0	2.0	2.0	2.0	3.0
v2	2.0	0.0	1.0	1.0	1.0	1.0	2.0
v4	inf	inf	0.0	inf	inf	inf	inf
v3	inf	inf	inf	0.0	1.0	inf	2.0
v5	inf	inf	inf	inf	0.0	inf	1.0
v7	1.0	2.0	2.0	3.0	3.0	0.0	4.0
v6	inf	inf	inf	inf	inf	inf	0.0

3. Paths

Cycles and acyclic graphs

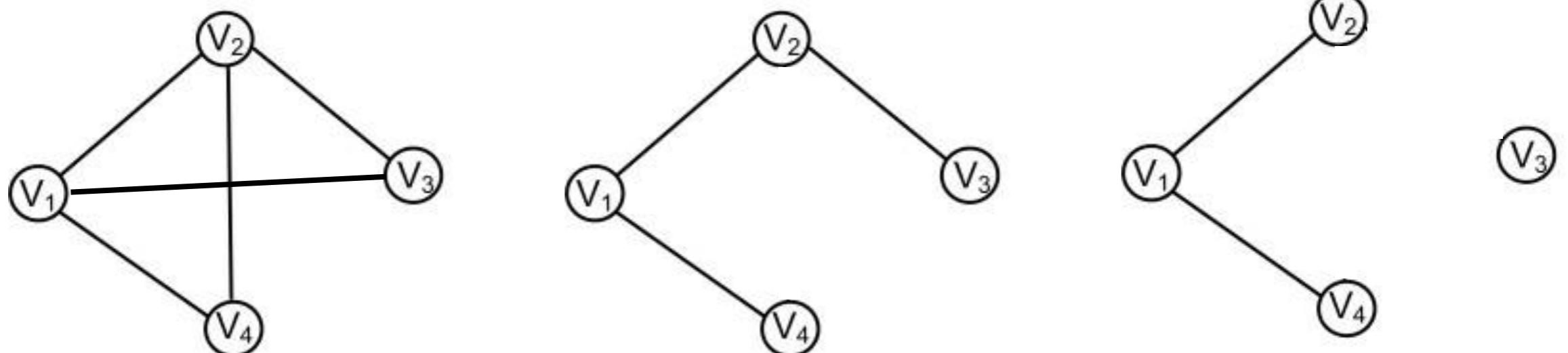
The **average path** gives a measure of network navigability
(~feature relationships)



4. Connectivity

Node connectivity $\kappa(G)$: minimum number of **nodes** whose removal renders the network disconnected

Edge connectivity $\lambda(G)$: minimum number of **edges** whose removal renders the network disconnected

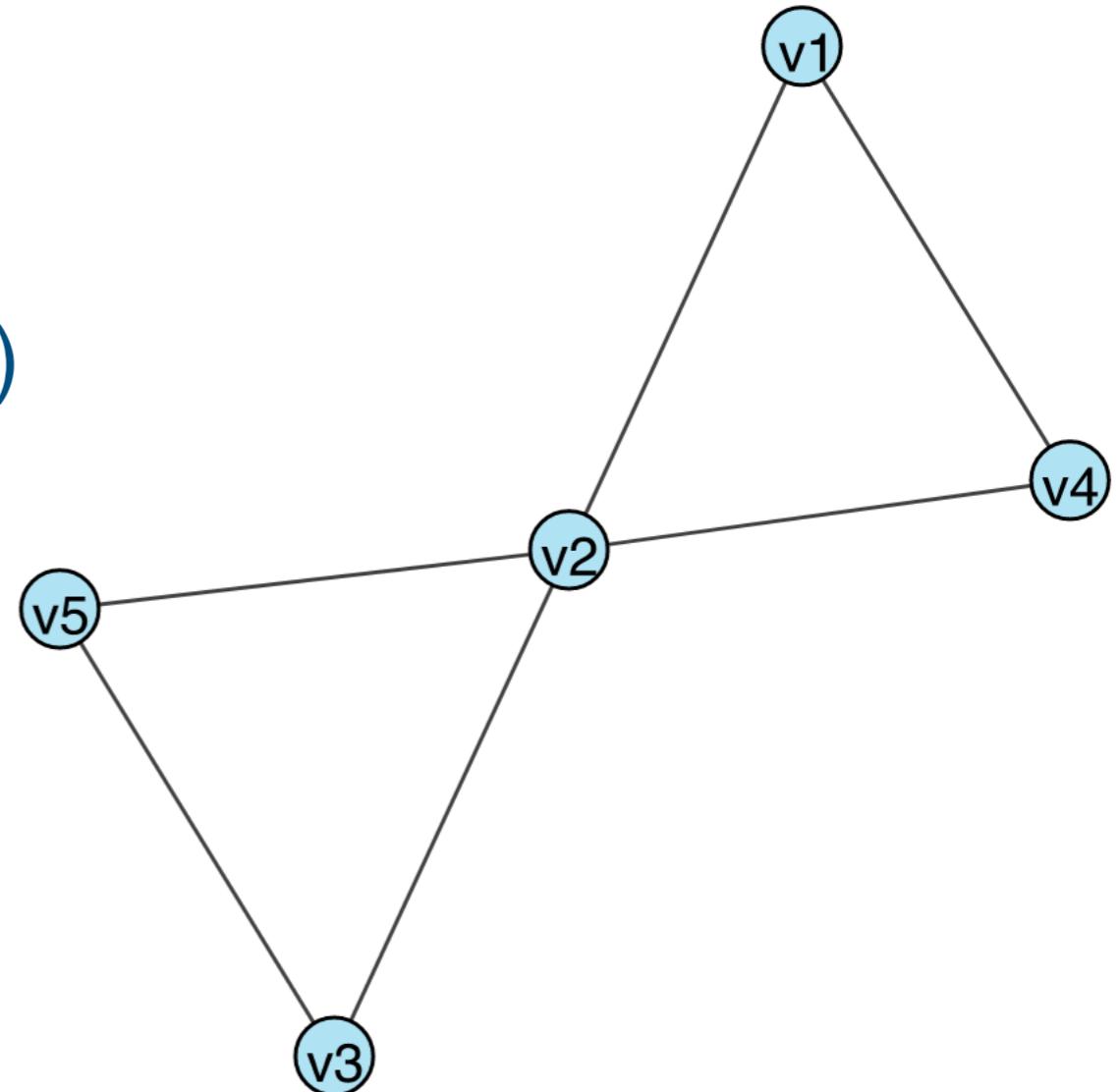


4. Connectivity

$\kappa(G) = 1$; **cut:** v_2

$\lambda(G) = 2$; **bridge:** ((v_2, v_1) & (v_2, v_4))

Local connectivity may also be computed for any given pair of vertices
(e.g. v_3, v_1 : v_2 and associated edges)



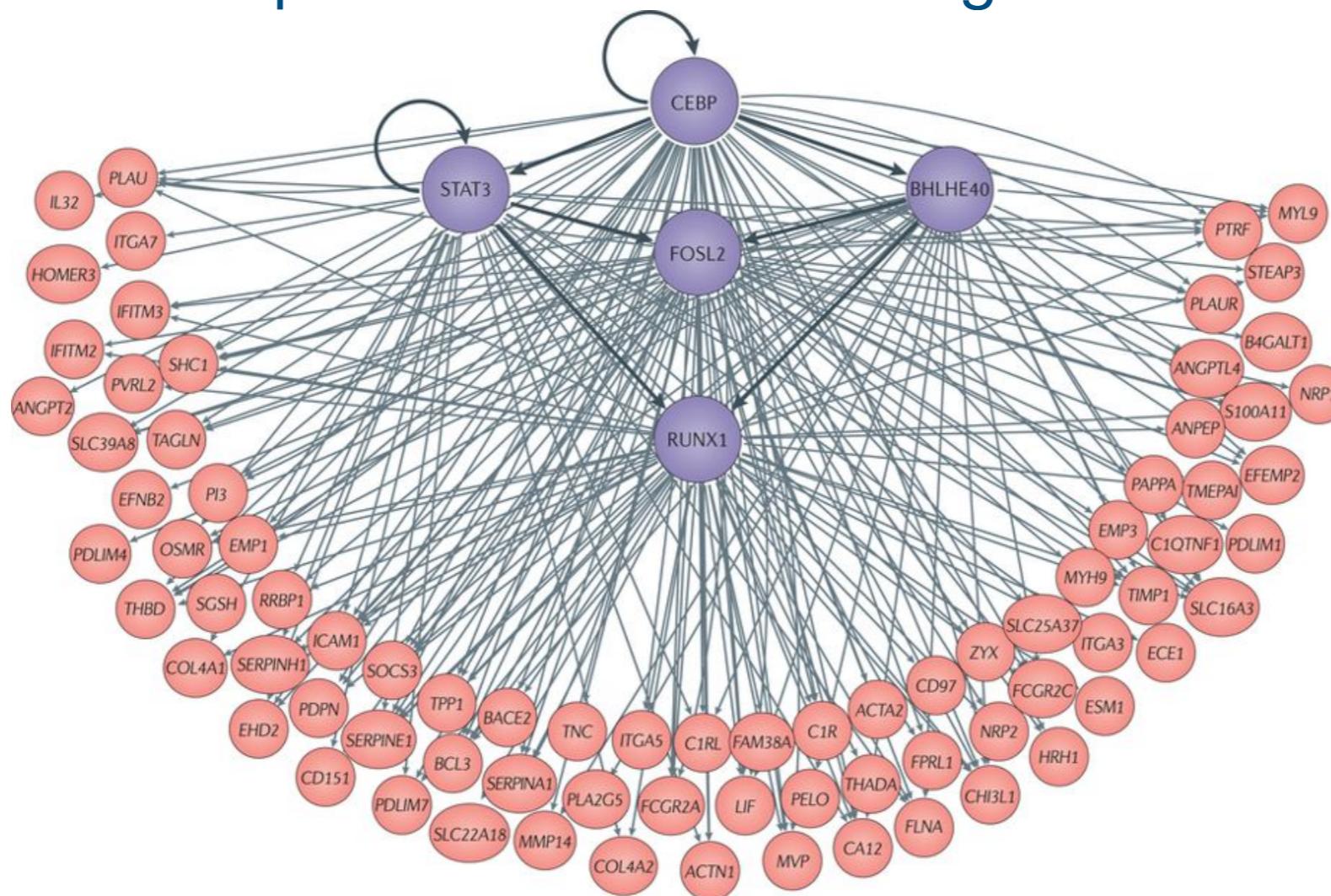
5. Centrality

Indicate the most central nodes in a network

Why look at the central nodes?

Hubs

Example: Transcription Factor Master Regulators



5. Centrality

Indicate the most central nodes in a network

Central nodes **possibly** most important in the network

There are many different measures of centrality:

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

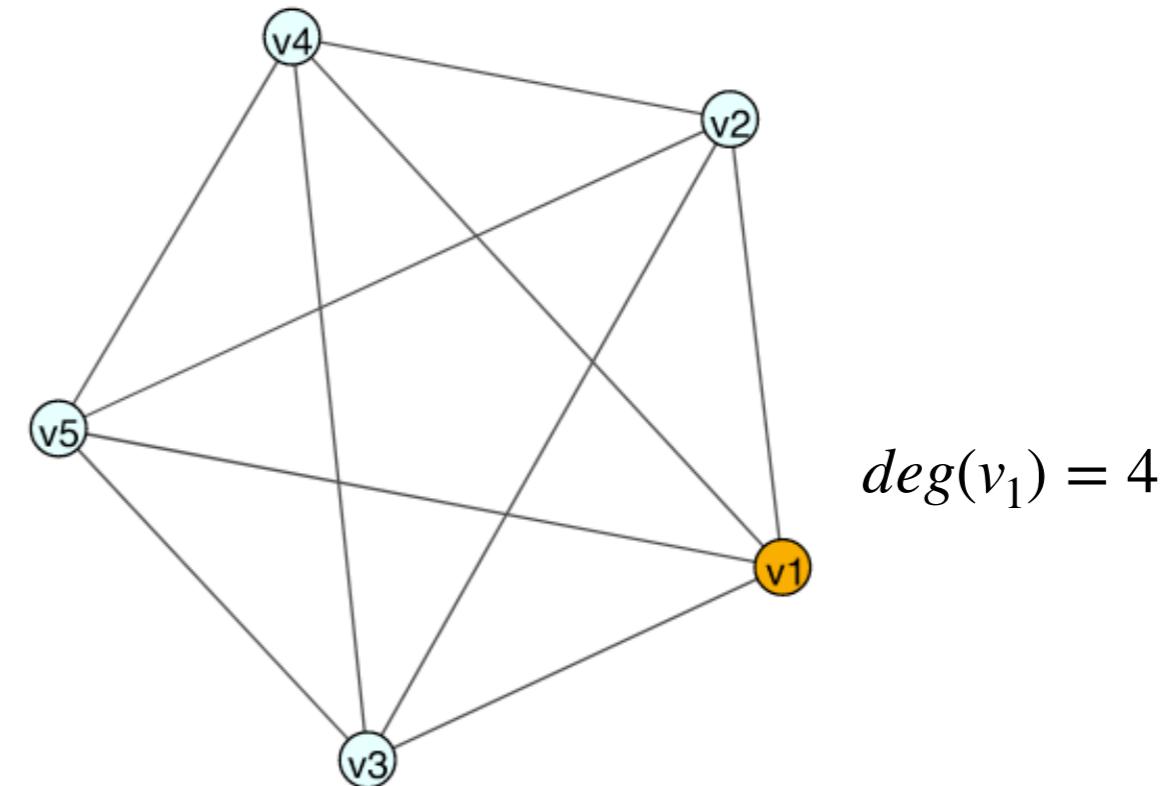
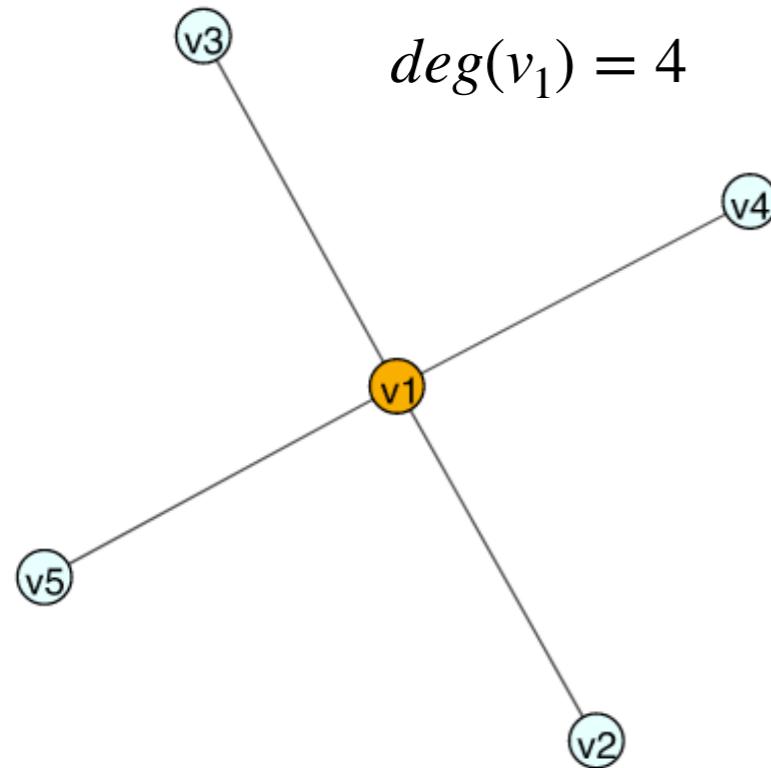
...

5. 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.



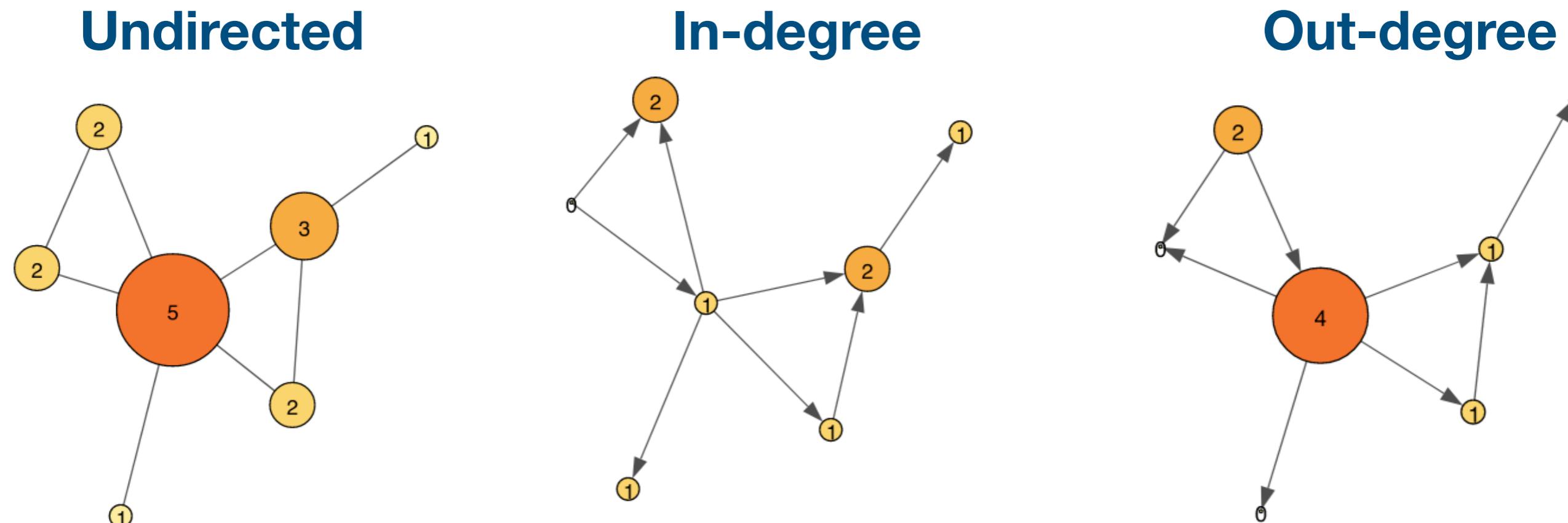
5. Centrality: degree centrality

Undirected networks vs directed networks

In-degree vs Out-degree

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

Numbers indicate degree:



5. Centrality: degree centrality

Degree centrality

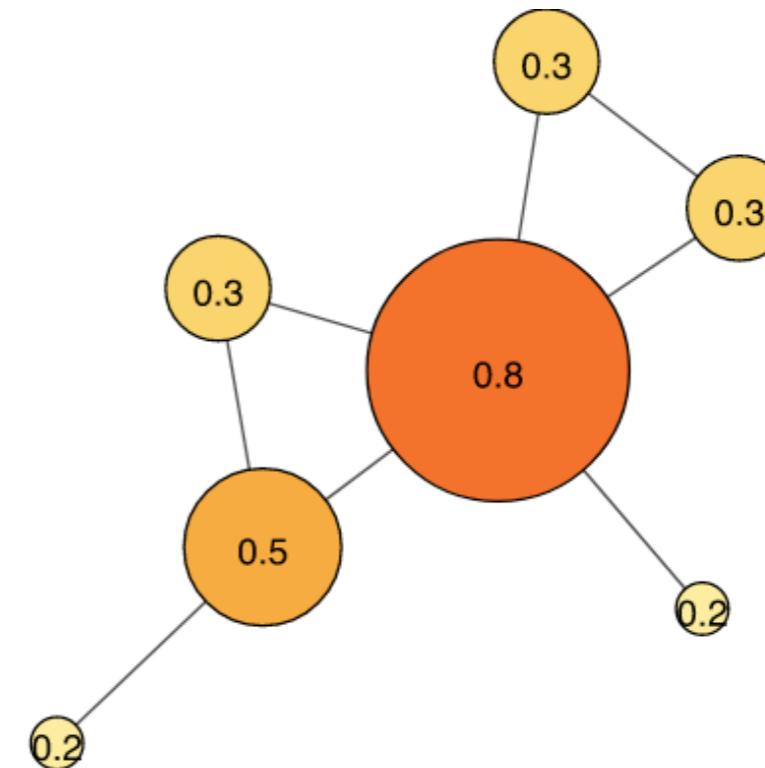
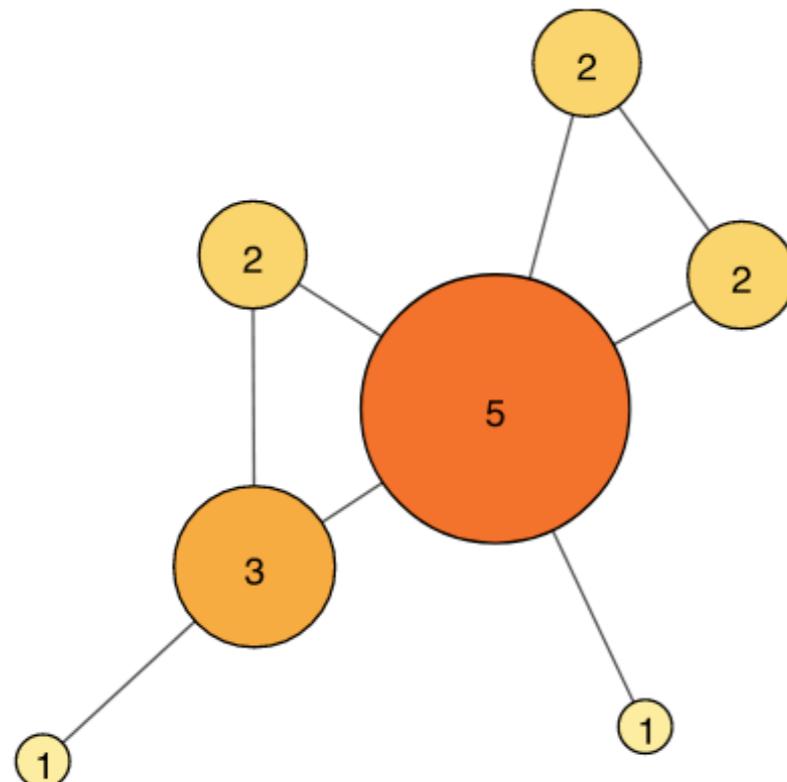
$$C_D(v_i) = \sum_{j=1}^N e_{ij}$$

Normalized
degree centrality

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

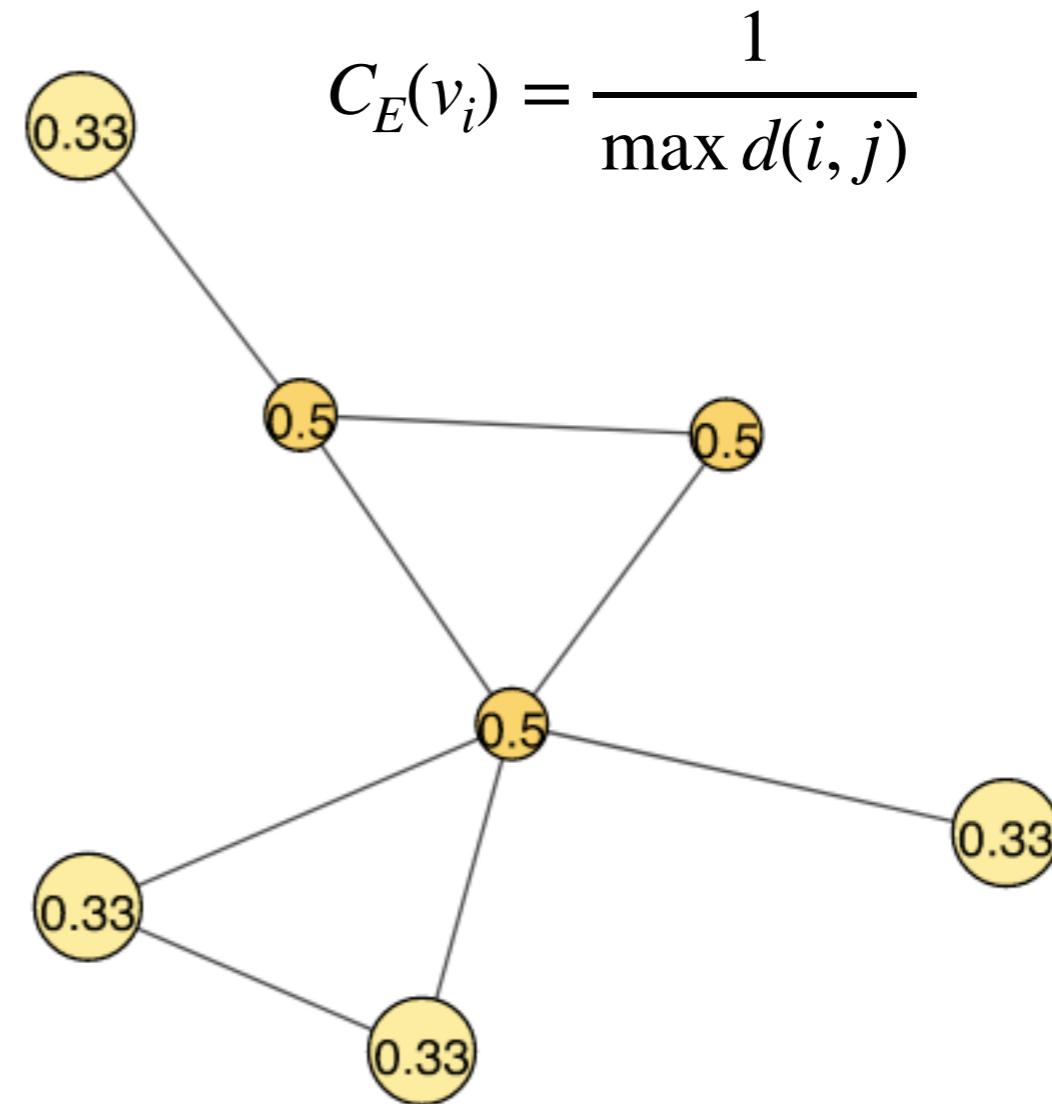
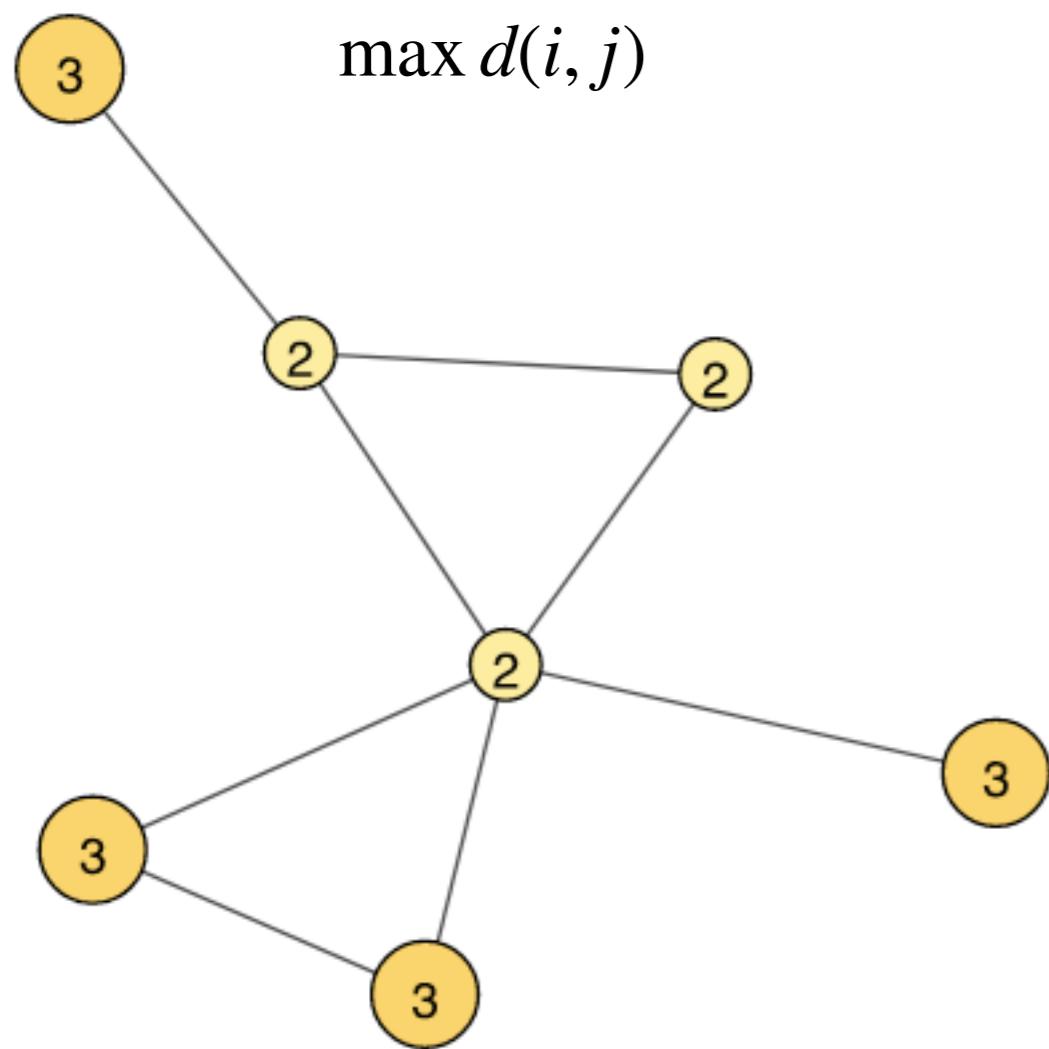
Normalized degree centrality accounts for the total possible number of connections

Centrality normalization allows for comparison between networks of different sizes



5. Centrality: eccentricity centrality

Eccentricity considers a node's max path to all other nodes



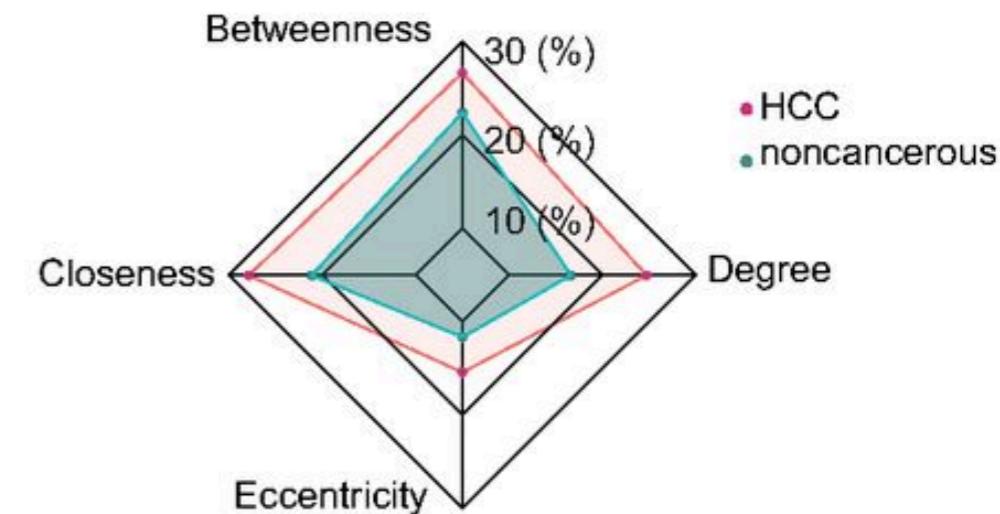
5. Centrality: limitations & influence

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**, modifications of centrality
 - **Accessibility**
 - **Dynamic influence**
 - **Impact**
 - **Expected force**

Measure **information transmission**



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



SciLifeLab



Key network properties to discuss

1. Network representations
2. Network density
3. Paths
4. Connectivity
5. Centrality
- 6. Clustering coefficient**
- 7. 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.

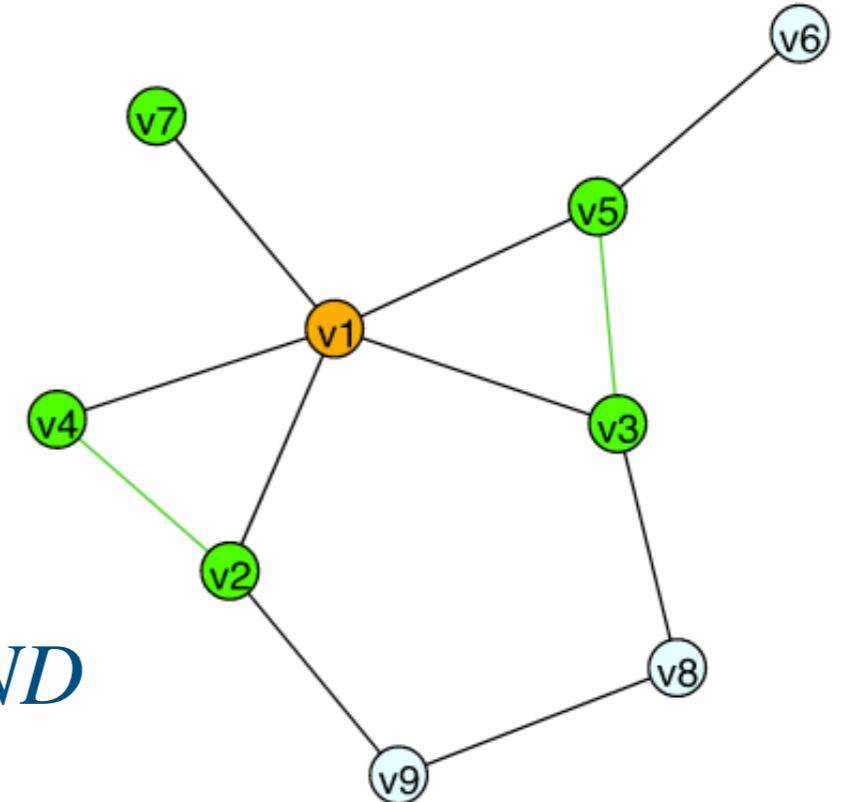
Takes into account degree of a node and the degree of its 1st neighbours

For node v_1

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

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

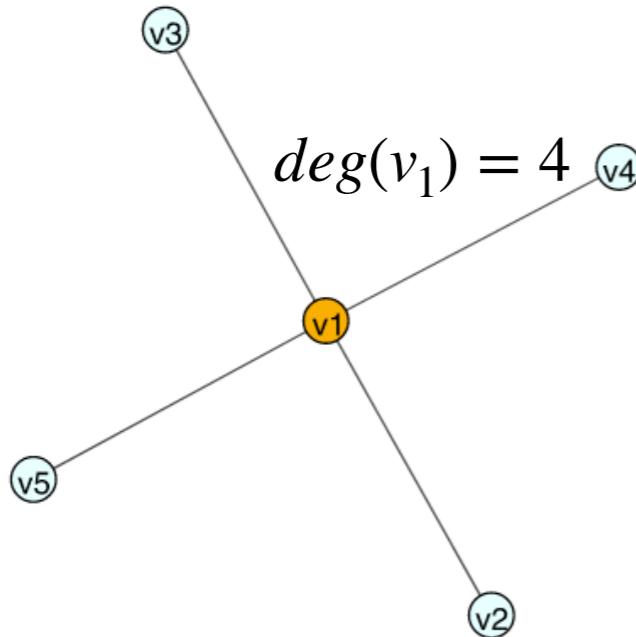
$$C(v_1) = \frac{2 \cdot 2}{5 \cdot 4} = 0.2 \quad C(v_7) = \frac{2 \cdot 0}{1 \cdot 0} = 0 \text{ or } ND$$



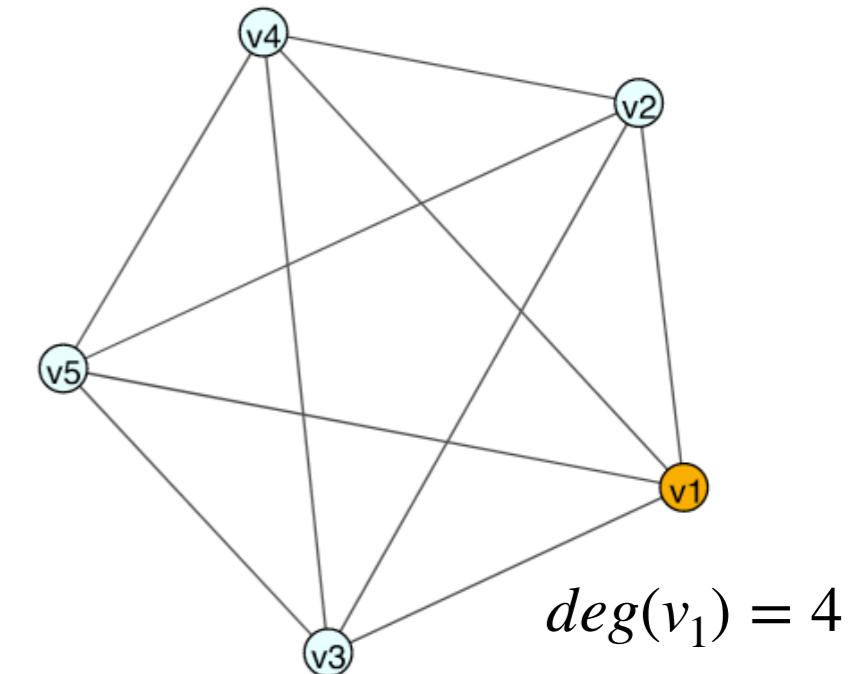
6. Clustering coefficient

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

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



$$0 \leq C_i \leq 1$$

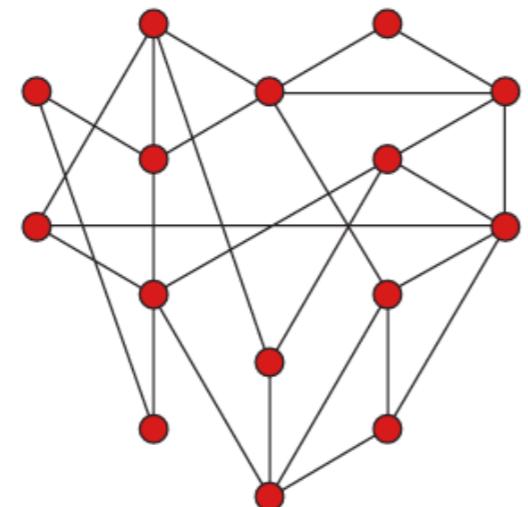


The global clustering coefficient $C(G)$ is simply the average of its clustering coefficients

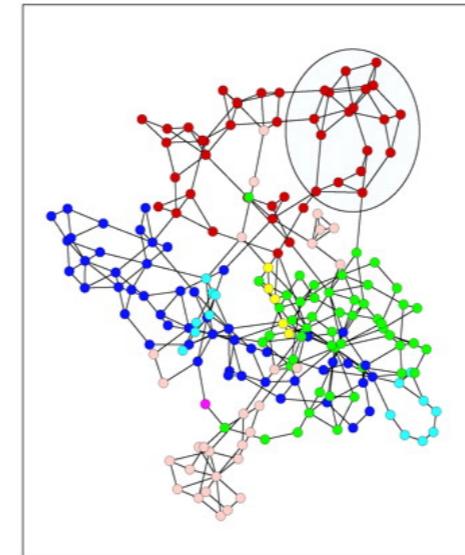
What distinguishes biological networks from random?

Do metabolic networks display different network properties from random networks?

Random network



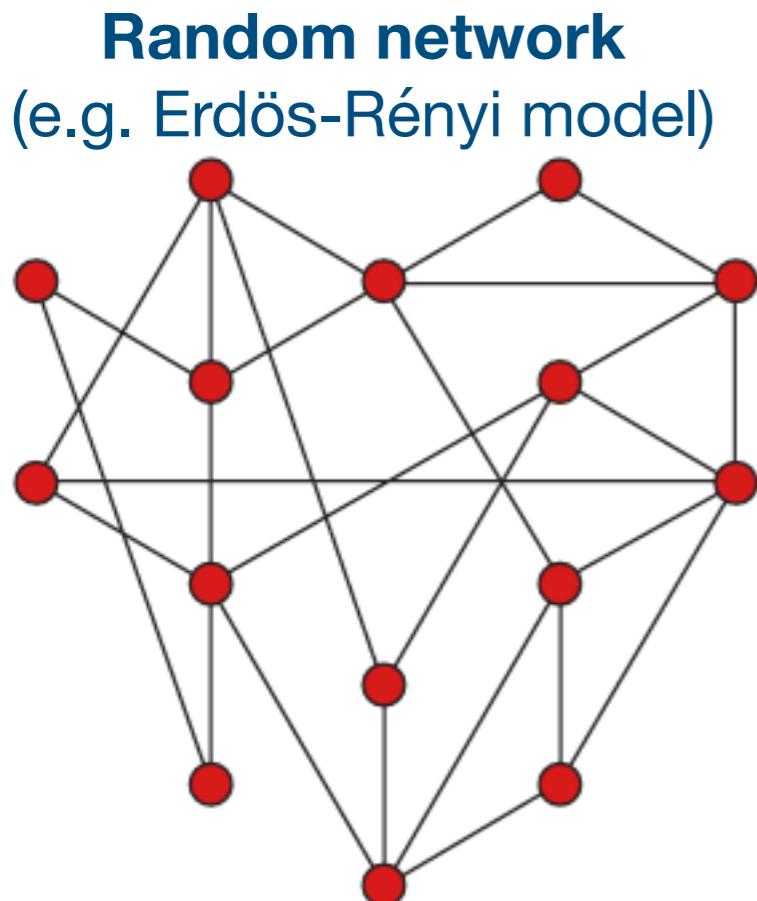
Metabolic network



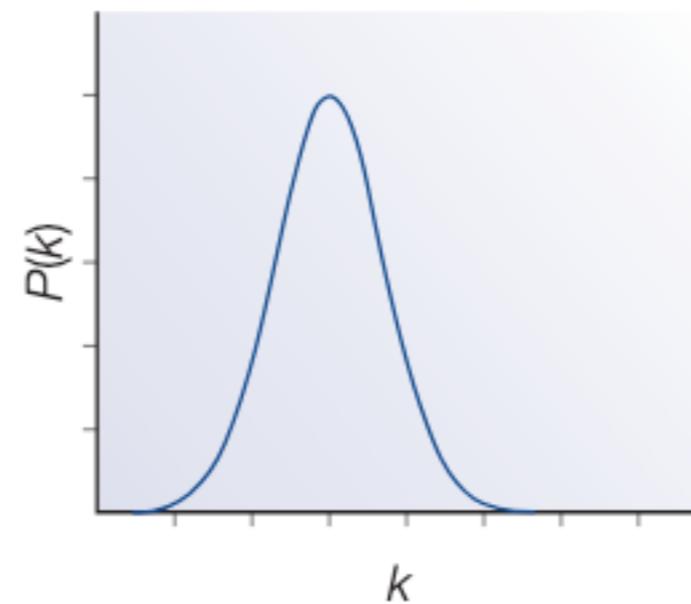
Barabasi 2004
Jeong 2000
Ravasz 2002

7. Degree and clustering coefficient distribution

Degree distributions allow us to compare network organization



Poisson degree distribution
shows no highly connected nodes

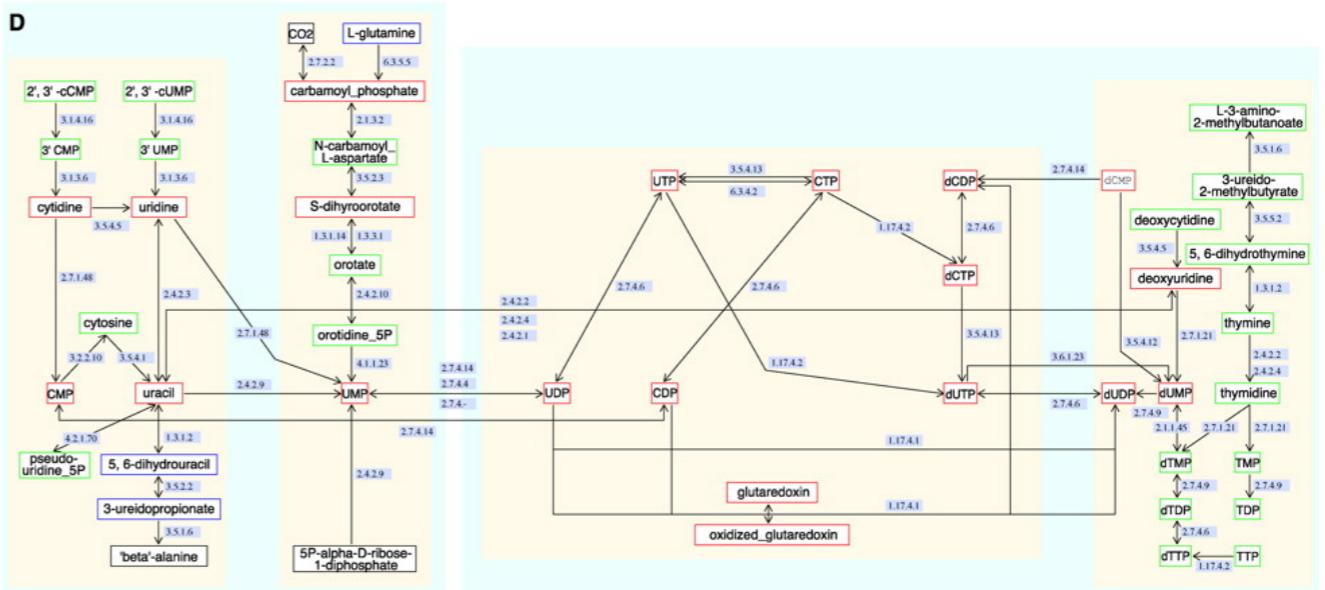
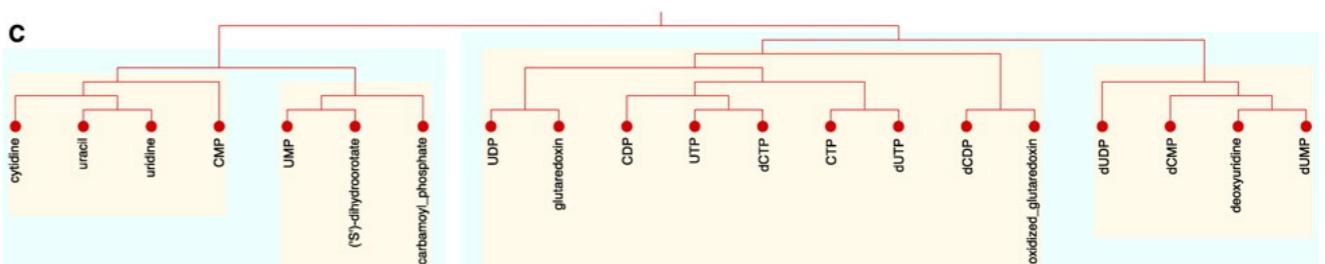
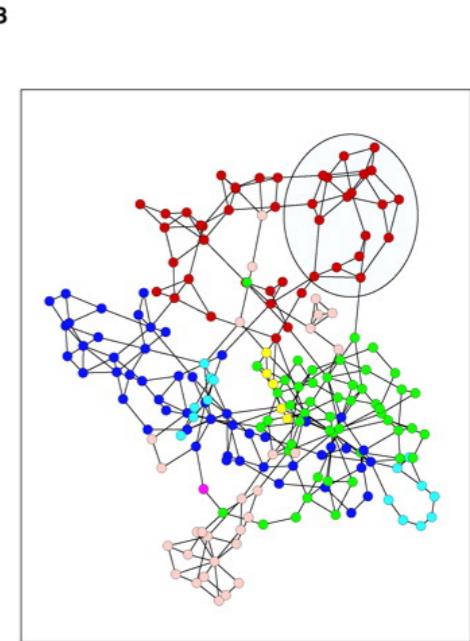


Most nodes have near $\langle k \rangle$

Metabolic networks show hierarchical topology

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

Their combination shows a hierarchical structure



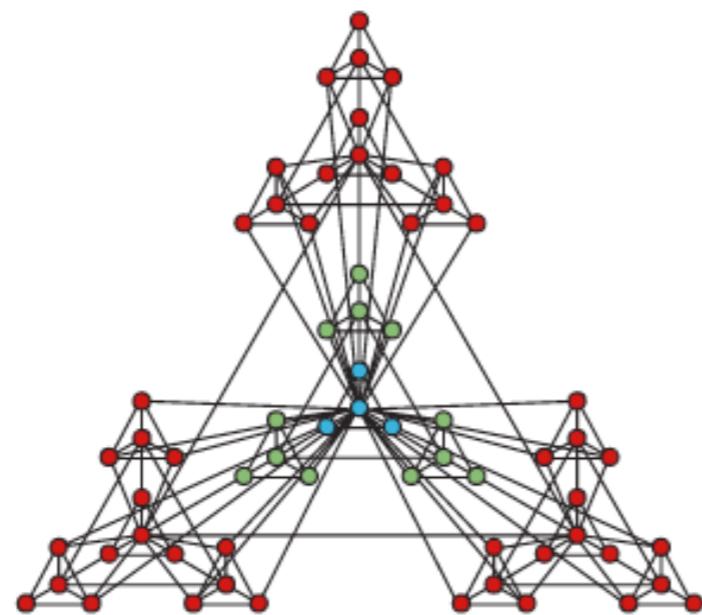
7. Degree distribution

Biological networks do not follow topology features of random networks.

Analysis of metabolic networks of 43 organisms shows common patterns

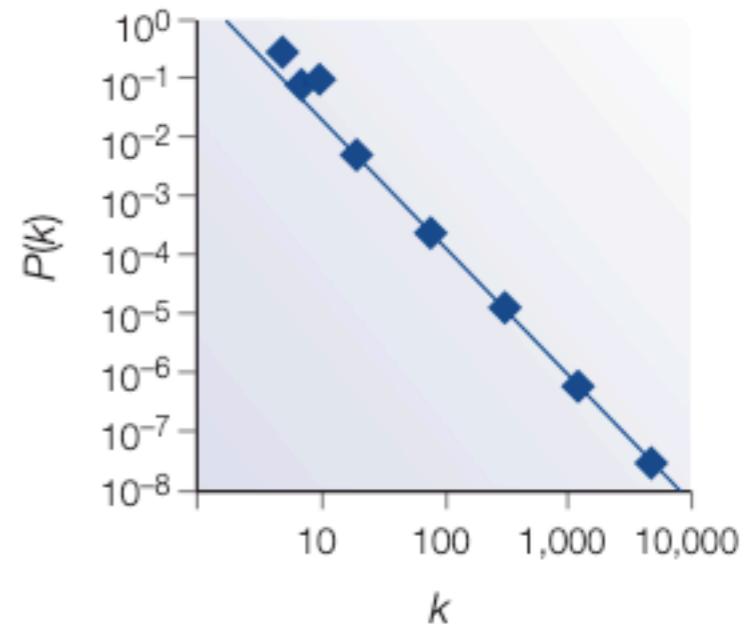
Biological networks tend to display high robustness to node failure:
removal of <80% nodes still retains paths between any two nodes

Hierarchical network



Degree distribution

shows many with low degrees
a few highly connected nodes



7. Degree and clustering coefficient distribution

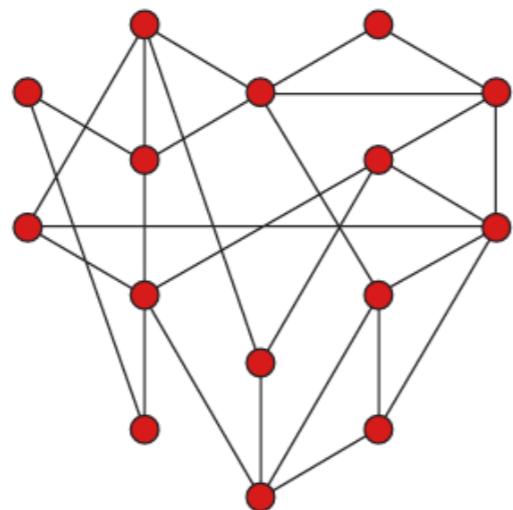
$C(k)$ shows no relationship with k in random networks: no modular organisation

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

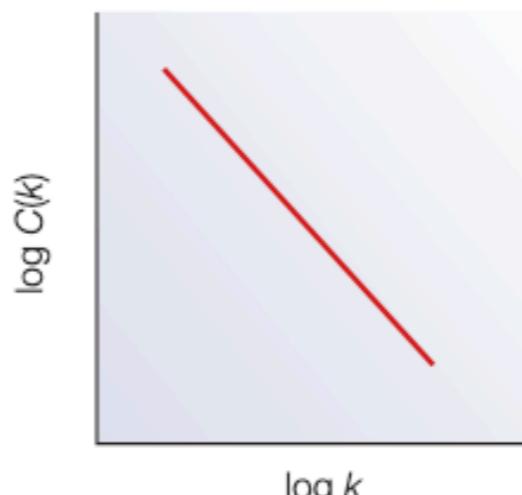
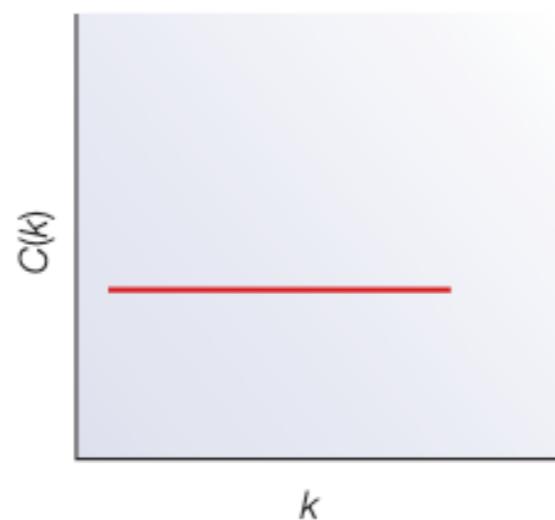
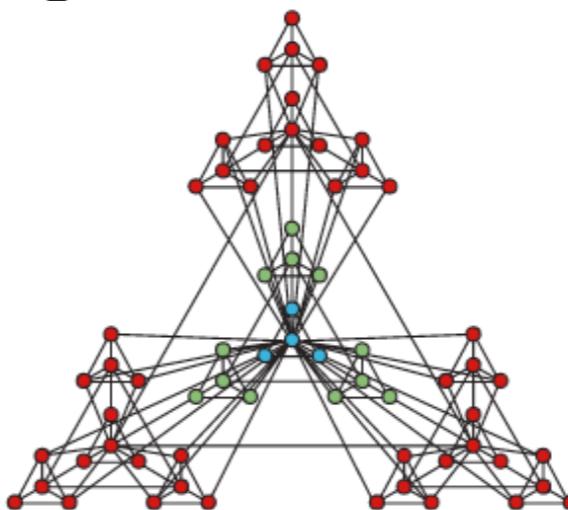
Sparingly connected nodes are part of highly modular areas

Communication between highly clustered neighbourhoods maintained by a few hubs

Random network



Hierarchical network



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

$$l(G) \approx \log N \text{ for a network of size } N$$

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

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

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

(social networks: **assortative**)



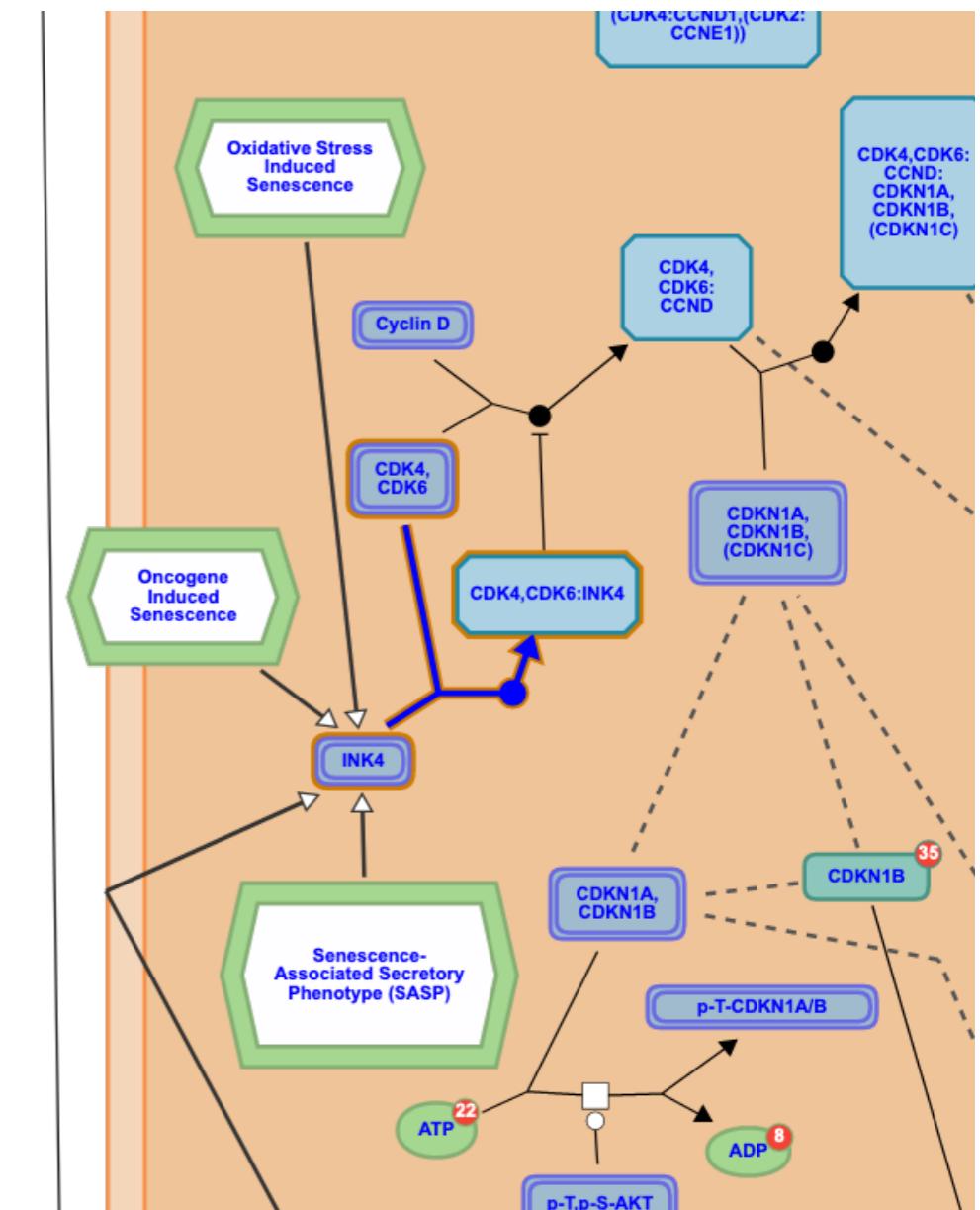
Overview

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

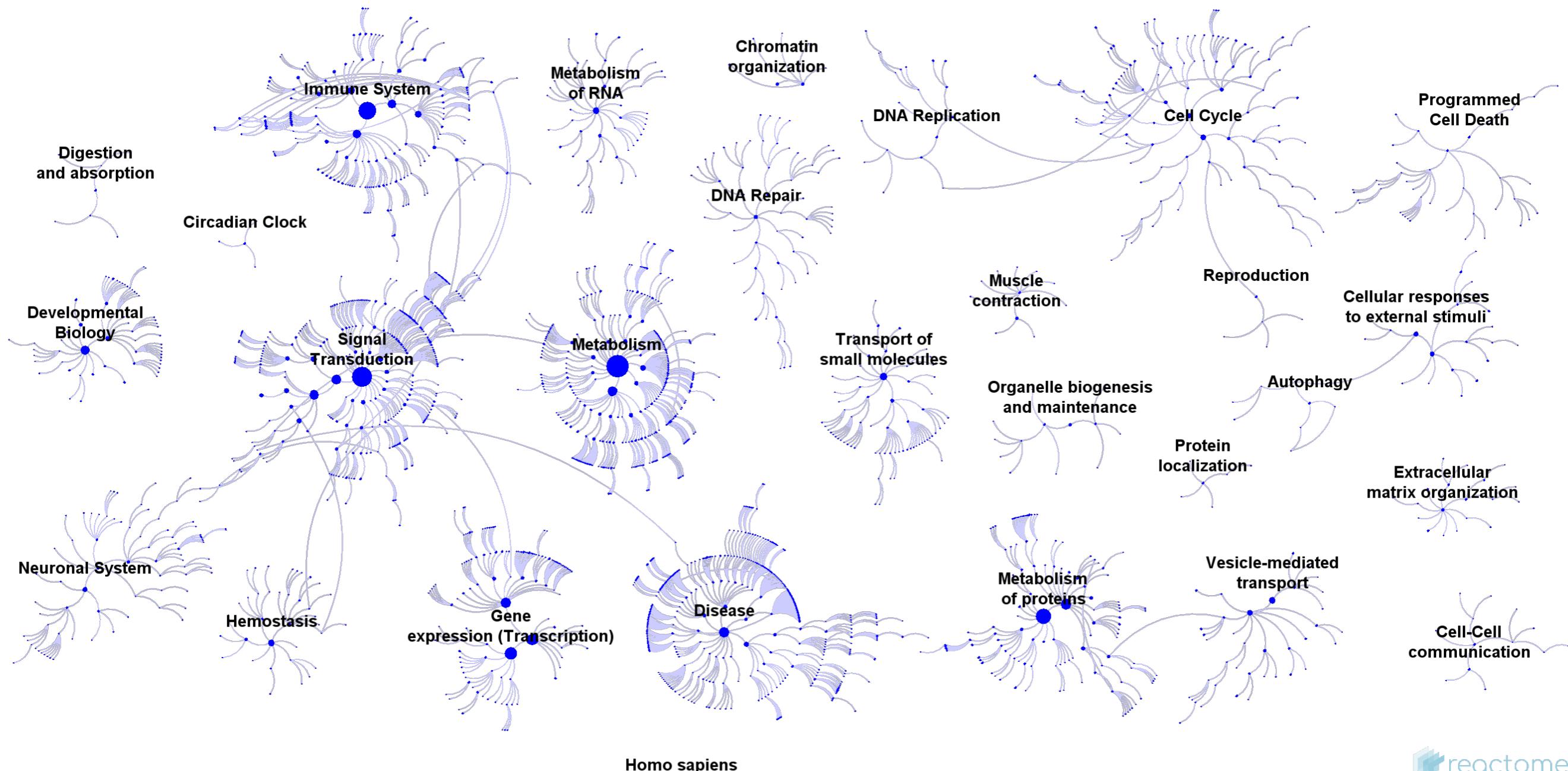
Protein complexes are physical modules



What are modules?

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

Gene Ontology



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

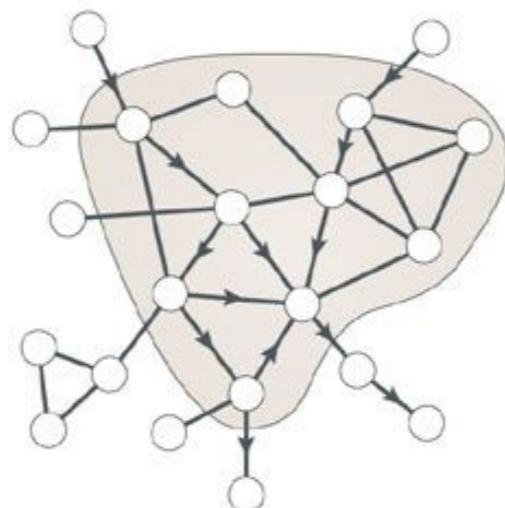
Drug: highly interconnected nodes associated with a drug response

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

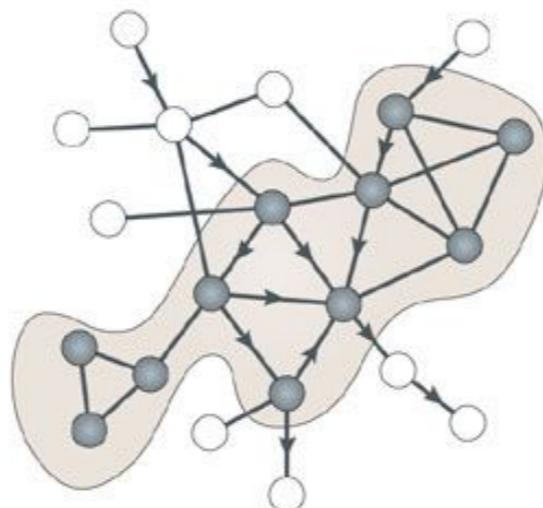
Tissue-, cell-type-specific: highly interconnected nodes associated with a specific tissue or cell type

Highly interlinked local regions of a network

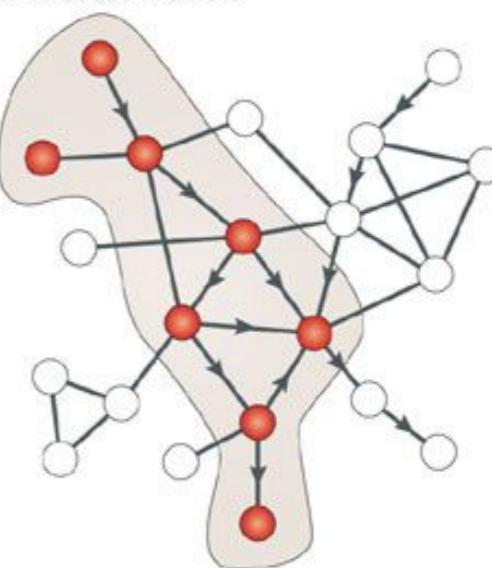
a Topological module



b Functional module



c Disease module



○ Topologically close genes (or products)

● Functionally similar genes (or products)

● Disease genes (or products)

— Bidirectional interactions

→ Directed interactions

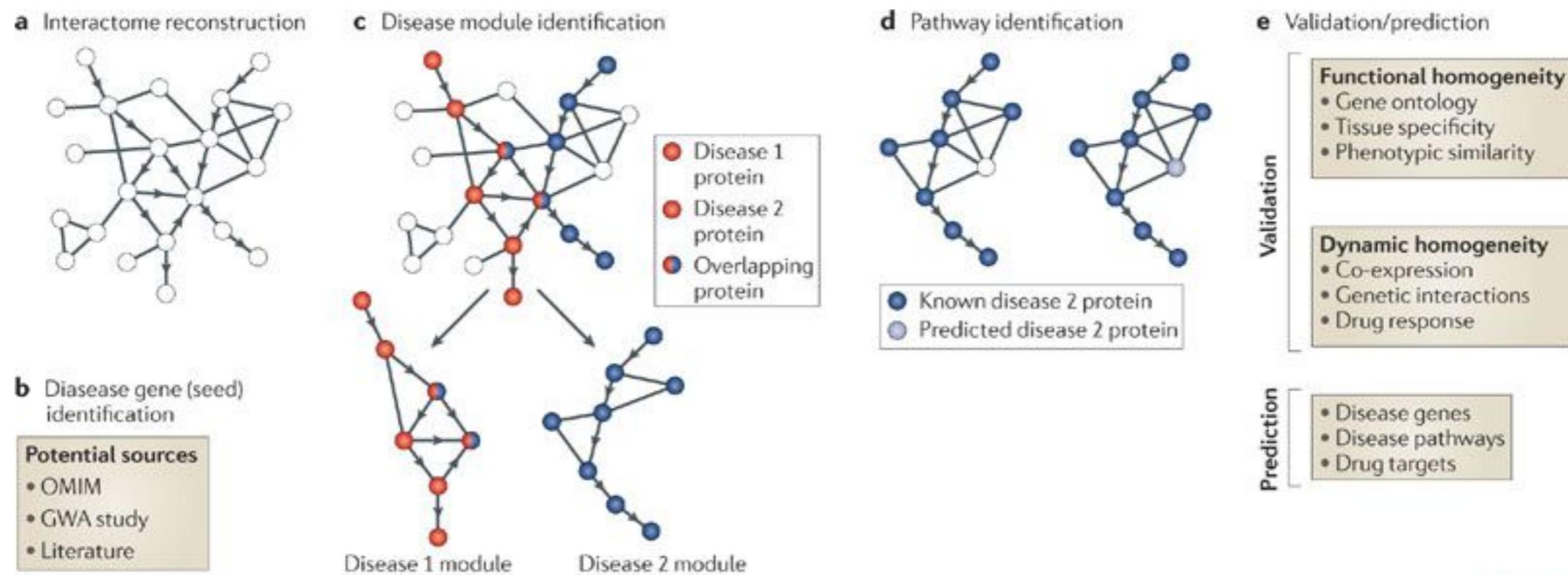
The challenge: identify and characterise modules

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



Nature Reviews | Genetics

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

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

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

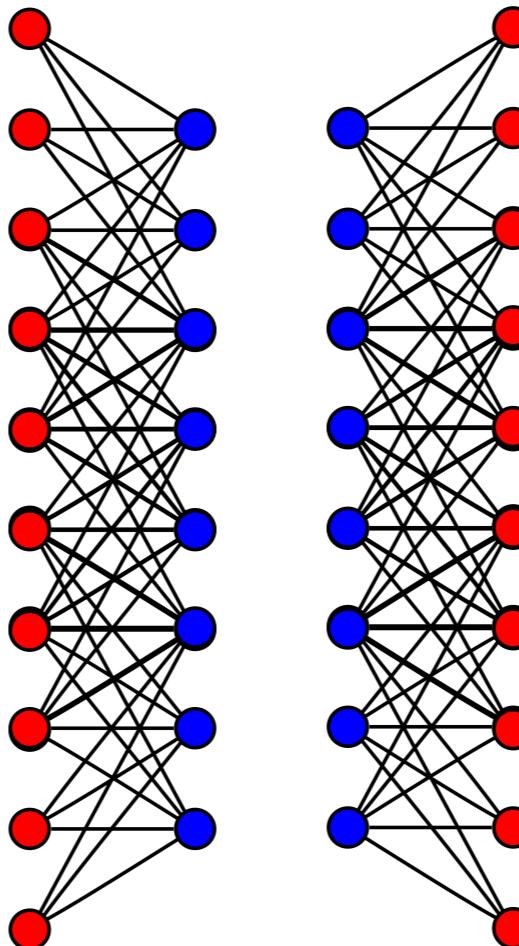
$Q > 0.3 - 0.7$ means significant community structure

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

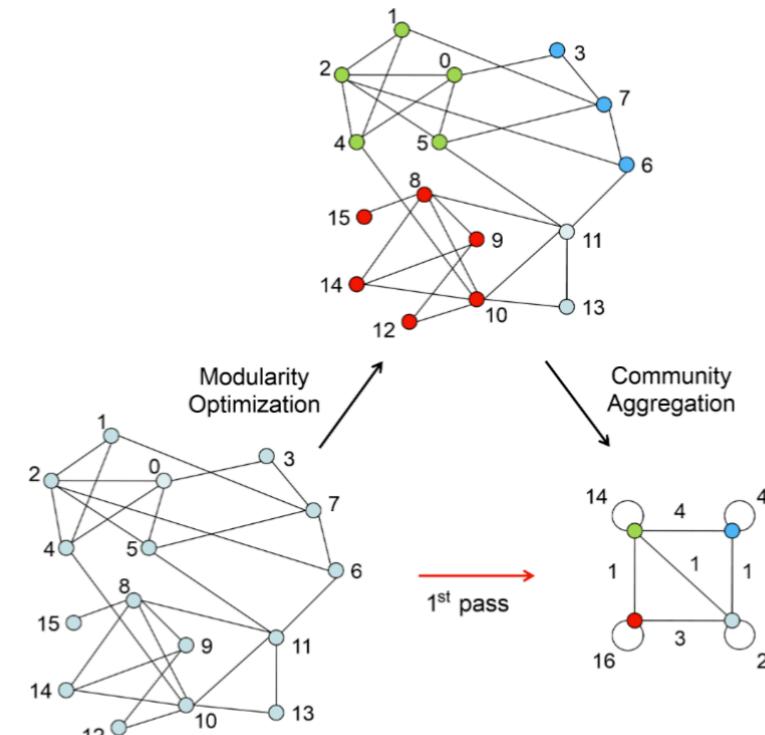
Module detection: Louvain algorithm

Phase 1: greedy modularity optimisation

1. Start with 1n/community
2. Compute Q by moving i to the community of j
3. If $\Delta Q > 1$, 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



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”?

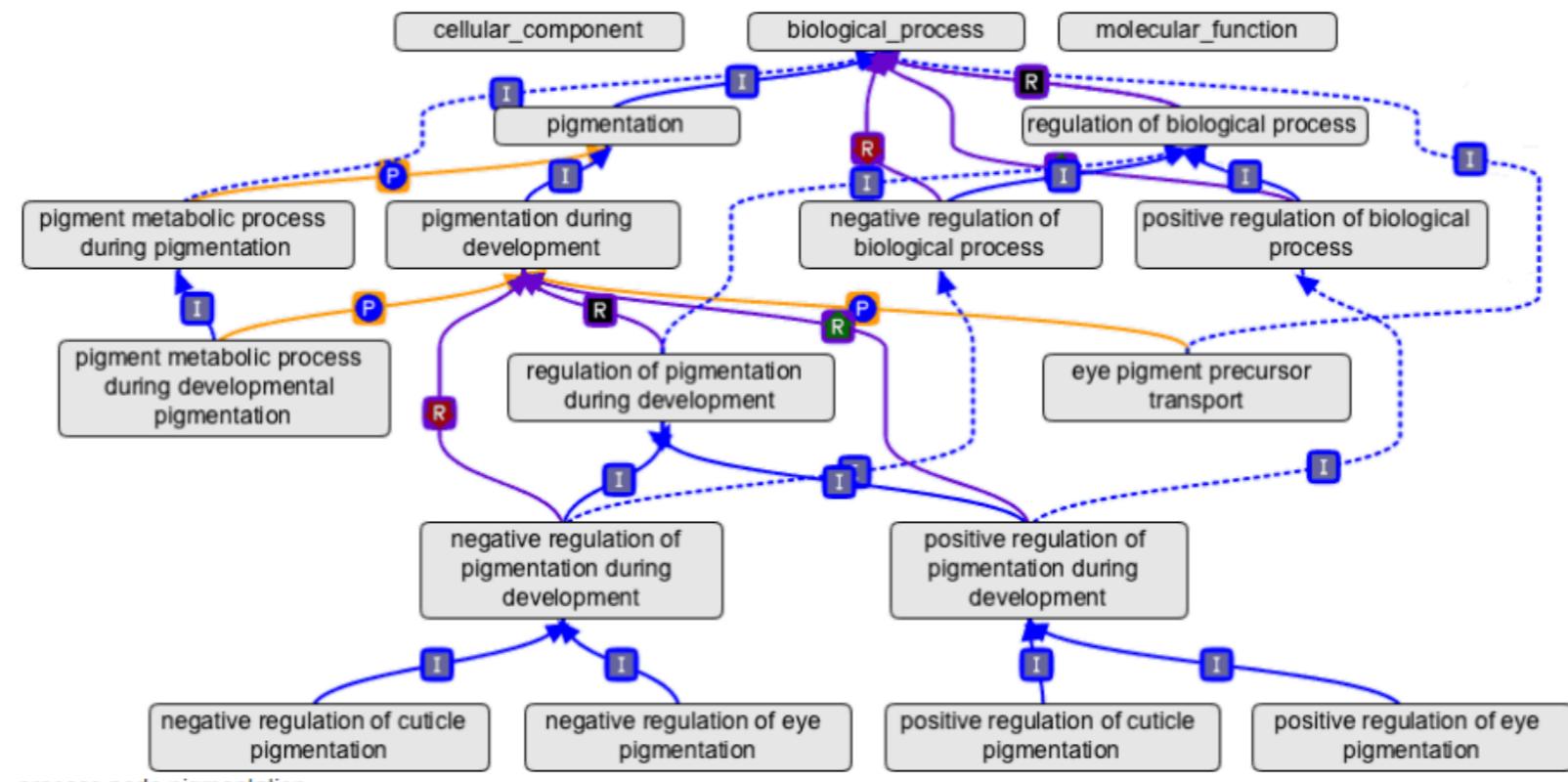
- Mutual feature associations may reinforce data characterisations not evident by individual features
- ...or need of further network curation based on top biological terms

Enrichment analysis

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

Tests for significant overlap between groups

Some biological processes may have no biological meaning in your analysis



Enrichment analysis

MSigDB



GSEA
Gene Set Enrichment Analysis

GSEA Home Downloads Molecular Signatures Database Documentation Contact

Overview

Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether *a priori* defined set of genes shows statistically significant, concordant differences between two biological states (e.g. phenotypes).

From this web site, you can:

- ▶ [Download](#) the GSEA software and additional resources to analyze, annotate and interpret enrichment results.
- ▶ [Explore the Molecular Signatures Database \(MSigDB\)](#), a collection of annotated gene sets for use with GSEA software.
- ▶ [View documentation](#) describing GSEA and MSigDB.

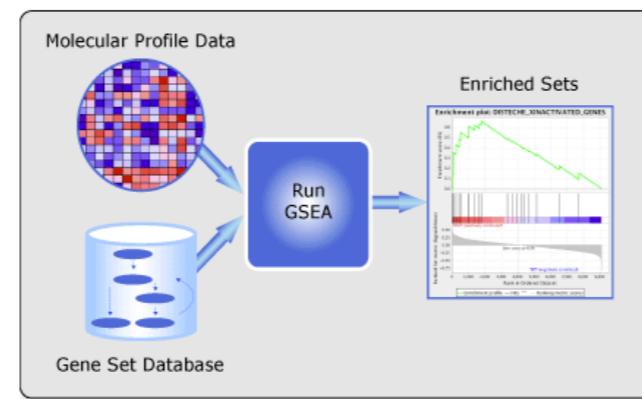
What's New

20-Aug-2019: MSigDB 7.0 released. This is a major release that includes a complete overhaul of gene symbol annotations, Reactome and GO gene sets, and corrections to miscellaneous errors. See the [release notes](#) for more information.

20-Aug-2019: GSEA 4.0.0 released. This release includes support for MSigDB 7.0, plus major internal updates for Java 11 support and performance improvements. See the [release notes](#) for more information.

16-Jul-2018: MSigDB 6.2 released. This is a minor release that includes updates to gene set annotations, corrections to miscellaneous errors, and a handful of new gene sets. See the [release notes](#) for more information.

[Follow @GSEA_MSigDB](#)



License Terms

GSEA and MSigDB are available for use under [these license terms](#).

Please [register](#) to download the GSEA software, access our web tools, and view the MSigDB gene sets. After registering, you can log in at any time using your email address. Registration is free. Its only purpose is to help us track usage for reports to our funding agencies.

Contributors

GSEA and MSigDB are maintained by the [GSEA team](#). Our thanks to our many contributors. Funded by: National Cancer Institute, National Institutes of Health, National Institute of General Medical Sciences.



Citing GSEA

To cite your use of the GSEA software, please reference Subramanian, Tamayo, et al. (2005, PNAS 102, 15545-15550) and Mootha, Lindgren, et al. (2003, Nat Genet 34, 267-273).

Enrichr



Login | Register

21,153,478 lists analyzed

307,486 terms

154 libraries

Analyze What's New? Libraries Find a Gene About Help

Gene-set Library	Terms	Gene Coverage	Genes per Term
Genes_Associated_with_NIH_Grants	32876	15886	9.0 
Cancer_Cell_Line_Encyclopedia	967	15797	176.0 
Achilles_fitness_decrease	216	4271	128.0 
Achilles_fitness_increase	216	4320	129.0 
Aging_Perturbations_from_GEO_down	286	16129	292.0 
Aging_Perturbations_from_GEO_up	286	15309	308.0 
Allen_Brain_Atlas_down	2192	13877	304.0 
Allen_Brain_Atlas_up	2192	13121	305.0 
ARCHS4_Cell-lines	125	23601	2395.0 
ARCHS4_IDG_Coexp	352	20883	299.0 
ARCHS4_Kinases_Coexp	498	19612	299.0 
ARCHS4_TFs_Coexp	1724	25983	299.0 
ARCHS4_Tissues	108	21809	2316.0 
BioCarta_2013	249	1295	18.0 
BioCarta_2015	239	1678	21.0 
BioCarta_2016	237	1348	19.0 
BioPlex_2017	3915	10271	22.0 
ChEA_2013	353	47172	1370.0 
ChEA_2015	395	48230	1429.0 
ChEA_2016	645	49238	1550.0 
Chromosome_Location	386	32740	85.0 
Chromosome_Location_hg19	36	27360	802.0 
CORUM	1658	2741	5.0 
Data_Acquisition_Method_Most_Popular_Genes	12	1073	100.0 
dbGaP	345	5613	36.0 
DepMap_WG_CRISPR_Screens_Broad_CellLines_2019	558	7744	363.0 
DepMap_WG_CRISPR_Screens_Sanger_CellLines_2019	325	6204	387.0 
Disease_Perturbations_from_GEO_down	839	23939	293.0 
Disease_Perturbations_from_GEO_up	839	23561	307.0 
Disease_Signatures_from_GEO_down_2014	142	15406	300.0 

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

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.