École Thématique de Bioinformatique Intégrative

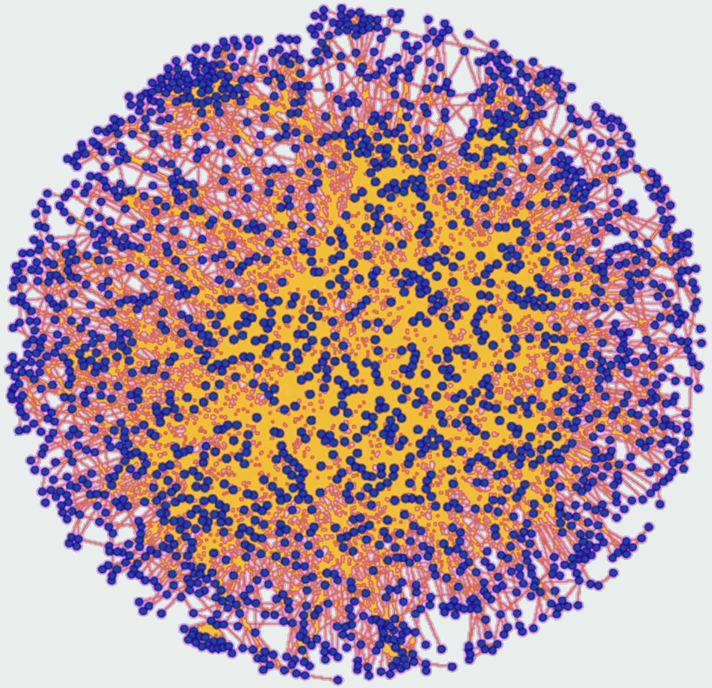
Fréjus, march 2026

Network theory

Galadriel Brière, Morgane Térézol, Bastien Chassagnol

Inspired by the slides of Anaïs Baudot

https://orcid.org/0000-0002-4090-2573 (Morgane Térézol)
https://orcid.org/0000-0003-0885-7933 (Anaïs Baudot)
https://orcid.org/0000-0003-1808-3759 (Galadriel Brière)
https://orcid.org/0000-0002-8955-2391 (Bastien Chassagnol)
https://orcid.org/0009-0005-0362-1835 (Benjamin Loire)

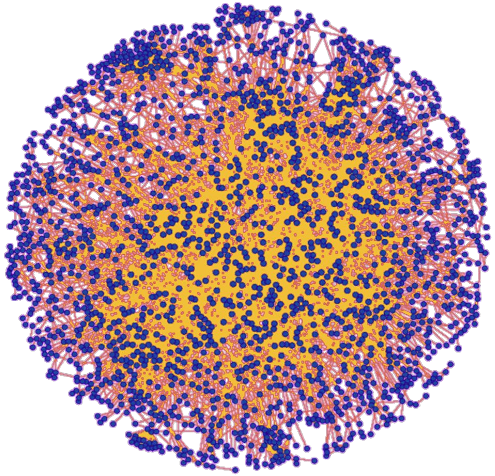


ÉCOLE THÉMATIQUE ETBII BIOINFORMATIQUE INTÉGRATIVE IFB

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

Outline



Source: https://www3.nd.edu/~tmilenko/research.html.

General Introduction, Key Concepts

Network Biology, Network Science, Complex Systems, …

Network Construction/Inference

Data sources, Distances/Similarities, Correlation, …

Network Analysis

Measures

Degree, Centrality, Distances, …

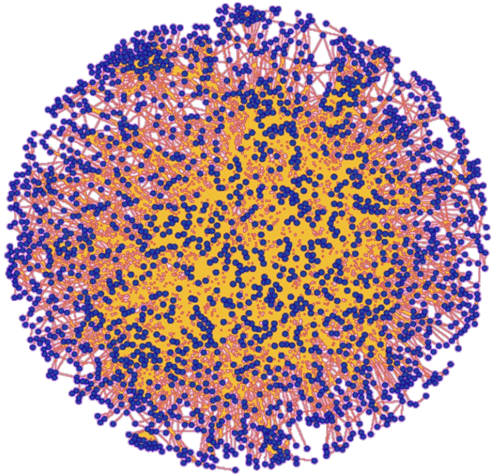
Algorithms

Diffusion, Clustering, Embedding, …

Networks and Integration

Multiplex Networks, Multilayer Networks, Knowledge Graphs, …

Outline



Source: https://www3.nd.edu/~tmilenko/research.html.

General Introduction, Key Concepts

Network Biology, Network Science, Complex Systems, …

Network Construction/Inference

Data sources, Distances/Similarities, Correlation, …

Network Analysis

Measures

Degree, Centrality, Distances, …

Algorithms

Diffusion, Clustering, Embedding, …

Networks and Integration

Multiplex Networks, Multilayer Networks, Knowledge Graphs, …

Key Concept 1: **Systems Biology**

- Computational and mathematical analysis and modeling of **complex biological systems.**

  - **Complex Systems**

Systems composed of many components
These components may interact with each others
Properties emerge from these interactions

The whole is greater than the sum of its parts

Key Concept 1: **Systems Biology**

- Computational and mathematical analysis and modeling of complex biological systems.

  ○ **Complex Systems**

Systems composed of many components
These components may interact with each others
Properties emerge from these interactions

The whole is greater than the sum of its parts

Source: CatalyzeX. “DNN-Buddies: A Deep Neural Network-Based Estimation Metric for the Jigsaw Puzzle Problem: Paper and Code.” CatalyzeX. Accessed September 1, 2023. https://www.catalyzex.com/.

Key Concept 1: **Systems Biology**

➢ Computational and mathematical analysis and modeling of complex biological systems.

○ Complex Systems

Systems composed of many components
These components may interact with each others
Properties emerge from these interactions

The whole is greater than the sum of its parts

○ Systems Biology

Systems components: genes/proteins, …

Key Concept 1: **Systems Biology**

➢ Computational and mathematical analysis and modeling of **complex biological systems.**

○ **Complex Systems**

Systems composed of many components
These components may interact with each others
Properties emerge from these interactions

The whole is greater than the sum of its parts

○ **Systems Biology**

Systems components: genes/proteins, …
Interactions: PPI, co-expression, …

Key Concept 1: **Systems Biology**

➢ Computational and mathematical analysis and modeling of **complex biological systems.**

○ **Complex Systems**

Systems composed of many components
These components may interact with each others
Properties emerge from these interactions

The whole is greater than the sum of its parts

○ **Systems Biology**

Systems components: genes/proteins, …
Interactions: PPI, co-expression, …
Emerging properties: Phenotypes

Key Concept 1: **Systems Biology**

- Computational and mathematical analysis and modeling of **complex biological systems.**

○ **Complex Systems**

Systems composed of many components
These components may interact with each others
Properties emerge from these interactions

The whole is greater than the sum of its parts

○ **Systems Biology**

Systems components: genes/proteins, …
Interactions: PPI, co-expression, …
Emerging properties: Phenotypes

Phenotype does not emerge from isolated biological molecules but from their interactions

Key Concept 1: **Systems Biology**

➢ Computational and mathematical analysis and modeling of complex biological systems.

○ **Complex Systems**

Systems composed of many components
These components may interact with each others
Properties emerge from these interactions

The whole is greater than the sum of its parts

○ **Systems Biology**

Systems components: genes/proteins, …
Interactions: PPI, co-expression, …
Emerging properties: Phenotypes

Phenotype does not emerge from isolated biological molecules but from their interactions

○ **Ecological Systems**

Systems components: organisms, …
Interactions: prey, symbiosis, competition, …
Emerging properties: resilience, stability, …

○ **Nervous System**

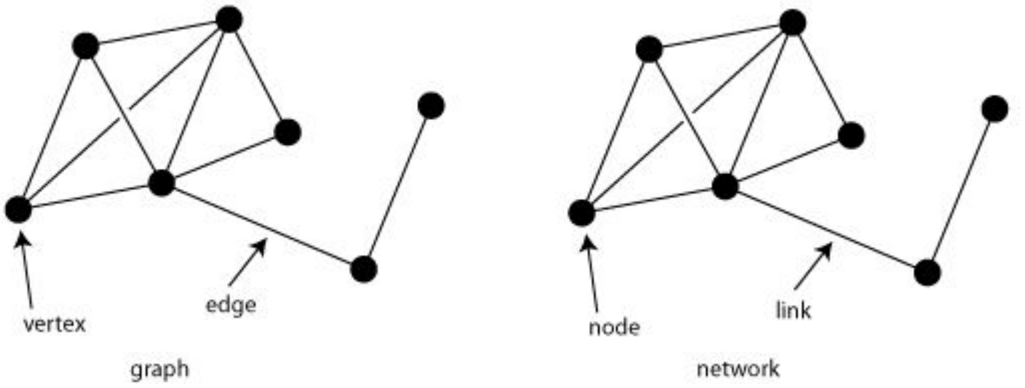
Systems components: neurons, axons, dendrites, …
Interactions: synaptic transmission, …
Emerging properties: memory, cognition, …

○ **etc….**

Key Concept 2: **Graph Theory/Network Science**

- Networks are real-world systems modeled using graphs.

G = (V, E), where V is the set of vertices and $E \subseteq (V*V)$ is the set of edges

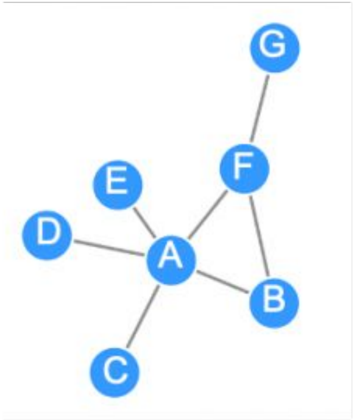


- In practice, the terms network and graph are often used interchangeably.

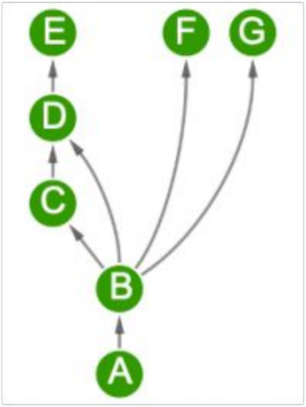
Key Concept 2: **Graph Theory/Network Science**

- Various types of networks…

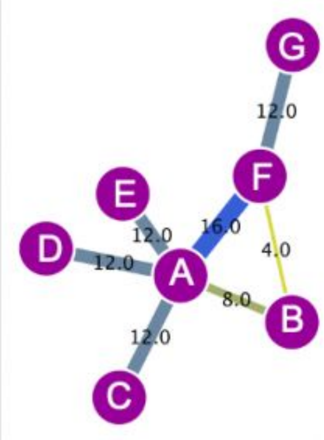
Undirected



Directed



Weighted



Source: EMBL-EBI. "Graph Theory: Graph Types and Edge Properties | Network Analysis of Protein Interaction Data." Accessed August 28, 2023. https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/introduction-to-graph-theory/graph-theory-graph-types-and-edge-properties/.

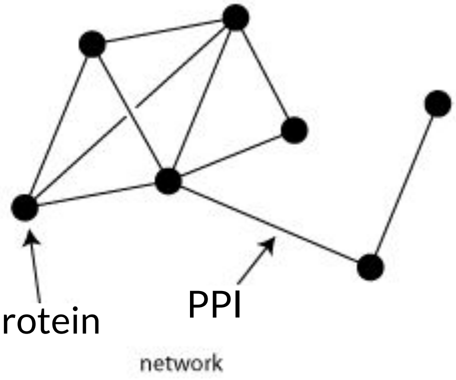
Network Science and **Systems Biology**

○ **Systems Biology**

Systems components: genes/proteins, …
Interactions: PPI, co-expression, …
Emerging properties: Phenotypes

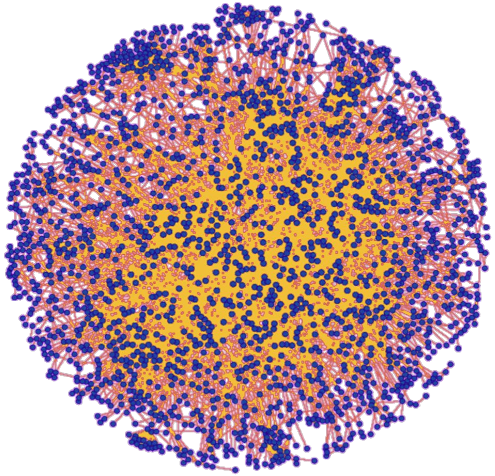
Phenotype does not emerge from isolated biological molecules but from their interactions

G = (V, E), where V is the set of vertices and E ⊆ (V*V) is the set of edges



Graph theory/network science: a powerful toolbox for representing and studying complex systems.

Outline



Source: https://www3.nd.edu/~tmilenko/research.html.

General Introduction, Key Concepts

Network Biology, Network Science, Complex Systems, …

Network Construction/Inference

Types of associations: from direct correlations, to directed causal interactions

Zoom on Bayesian Networks

Network Analysis

Measures

Degree, Centrality, Distances, …

Algorithms

Diffusion, Clustering, Embedding, …

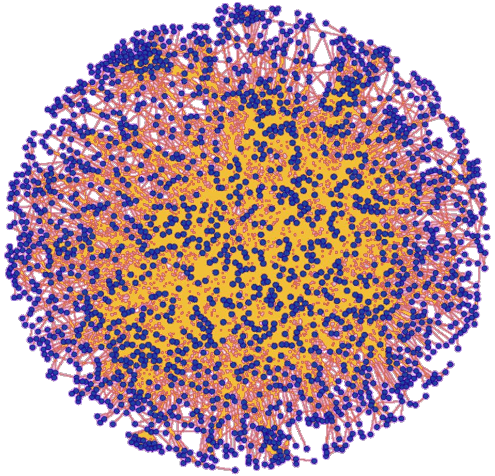
Networks and Integration

Multiplex Networks, Multilayer Networks, Knowledge Graphs, …

Network Modelling

Boolean networks, ODE models, …

Data sources

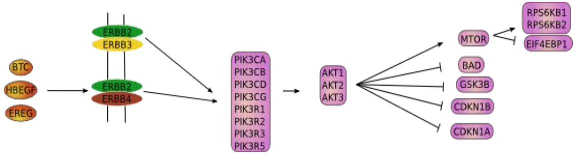


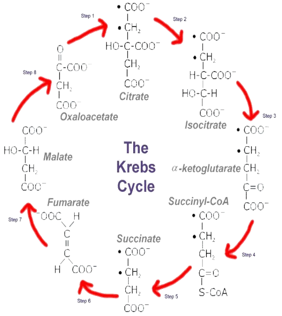
Source: https://www3.nd.edu/~tmilenko/research.html.

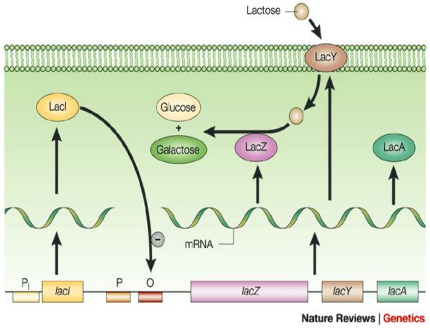
- ➢ Expert knowledge and literature
- ➢ High-throughput screening
- ➢ Inference from (omics) data

Source 1: **Expert knowledge and literature**

Biological Pathways





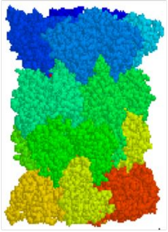


Metabolic Cycles

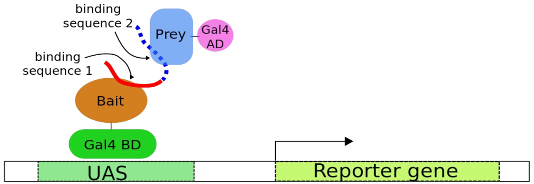
Operons / Regulatory Networks

Source: Courtesy of Anaïs Baudot

Source 2: **High-throughput screening**



Molecular complex interactions
-> GFP-trap



Protein-Protein interactions
-> Yeast 2-hybrids screens

Thousands of interactions in model organisms

STRING IntAct

MINT etc…

Source: Courtesy of Anaïs Baudot

Source 3: **Inference from (Omics) data**

- **Basically, you need two things:**
  - A dataset with the values of features for various samples
  - An algorithm to relate your features (e.g. gene network) or samples (e.g. patient network)

Source 3: **Inference from (Omics) data**

- **Basically, you need two things:**
    - A dataset with the values of features for various samples
    - An algorithm to relate your features (e.g. gene network) or samples (e.g. patient network)

- **The choice of the algorithm depends on your dataset and the type of interaction you want to model** (similarity, dependency, ...) **.** It can be :
    - Any type of similarity or distance

Source 3: **Inference from (Omics) data**

- **Basically, you need two things:**
  - A dataset with the values of features for various samples
  - An algorithm to relate your features (e.g. gene network) or samples (e.g. patient network)

- **The choice of the algorithm depends on your dataset and the type of interaction you want to model** (similarity, dependency, ...) **.** It can be :
  - Any type of similarity or distance
  - A regression model

relationship between j and i

$$X_j = \beta_0 + \boxed{\beta_1} X_i + \epsilon$$

Source 3: **Inference from (Omics) data**

- **Basically, you need two things:**
  - A dataset with the values of features for various samples
  - An algorithm to relate your features (e.g. gene network) or samples (e.g. patient network)

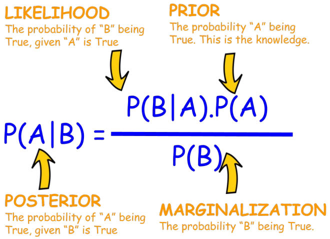
- **The choice of the algorithm depends on your dataset and the type of interaction you want to model** (similarity, dependency, …) . It can be :
  - Any type of similarity or distance
  - A regression model

relationship between j and i

$$X_j = \beta_0 + \boxed{\beta_1} X_i + \epsilon$$

  - A Bayesian model

$$P(X_j/X_i)$$



Source 3: **Inference from (Omics) data**

- **Basically, you need two things:**
    - A dataset with the values of features for various samples
    - An algorithm to relate your features (e.g. gene network) or samples (e.g. patient network)

- **The choice of the algorithm depends on your dataset and the type of interaction you want to model** (similarity, dependency, …) . It can be :
    - Any type of similarity or distance
    - A regression model

relationship between j and i

$$X_j = \beta_0 + \boxed{\beta_1} X_i + \epsilon$$

- A Bayesian model

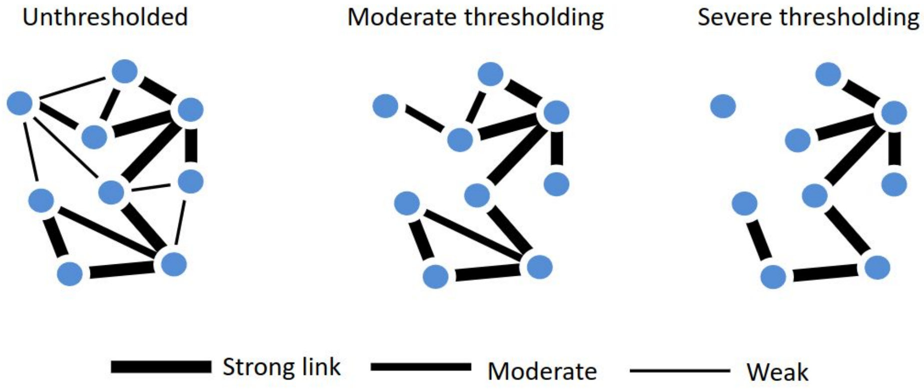
$$P(X_j/X_i)$$

- A correlation metric (more about this soon!)
- etc…

Source 3: **Inference from (Omics) data**

- **Basically, you need two things:**
  - A dataset with the values of features for various samples
  - An algorithm to relate your features (e.g. gene network) or samples (e.g. patient network)

- **Often, the similarity/distances/correlations/… are thresholded:**



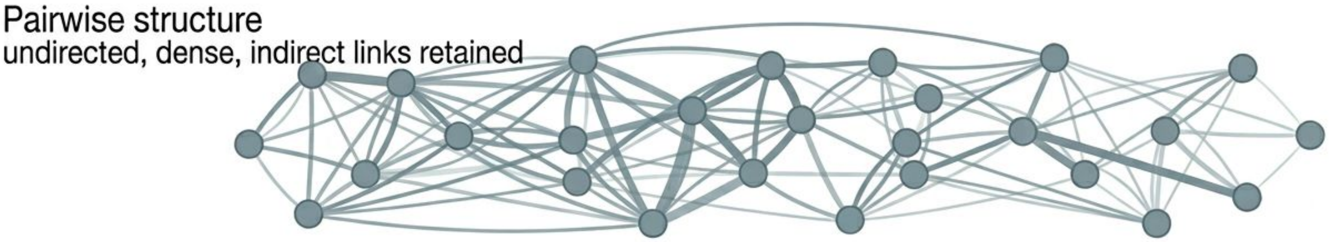
- Eliminating spurious (weak) associations
- Emphasizing topological properties
- Easing computational and storage burden of large graphs

Source: Zalesky, Andrew. “Network Statistics and Thresholding”

Topologies of network representations

Depending on the algorithm, three paradigms of representing interactions can be identified, corresponding to distinct topological structures.

1. Association networks

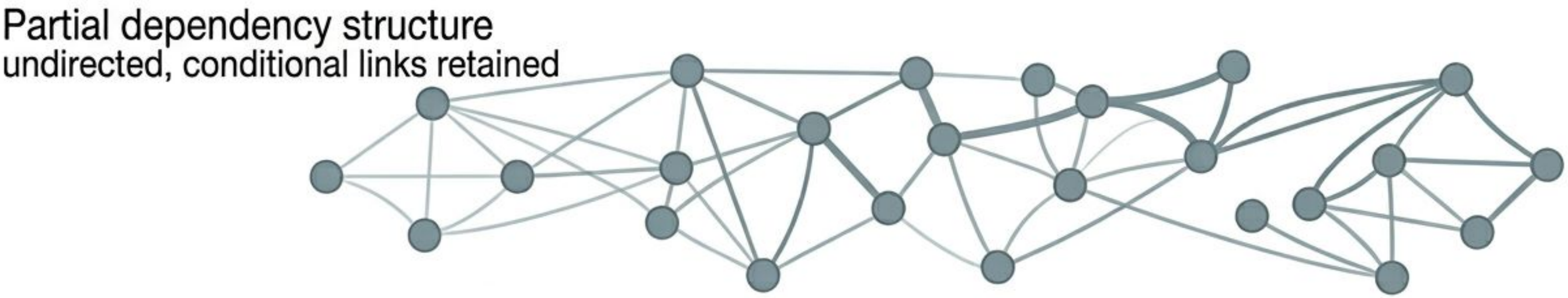


Pairwise structure
undirected, dense, indirect links retained

Topologies of network representations

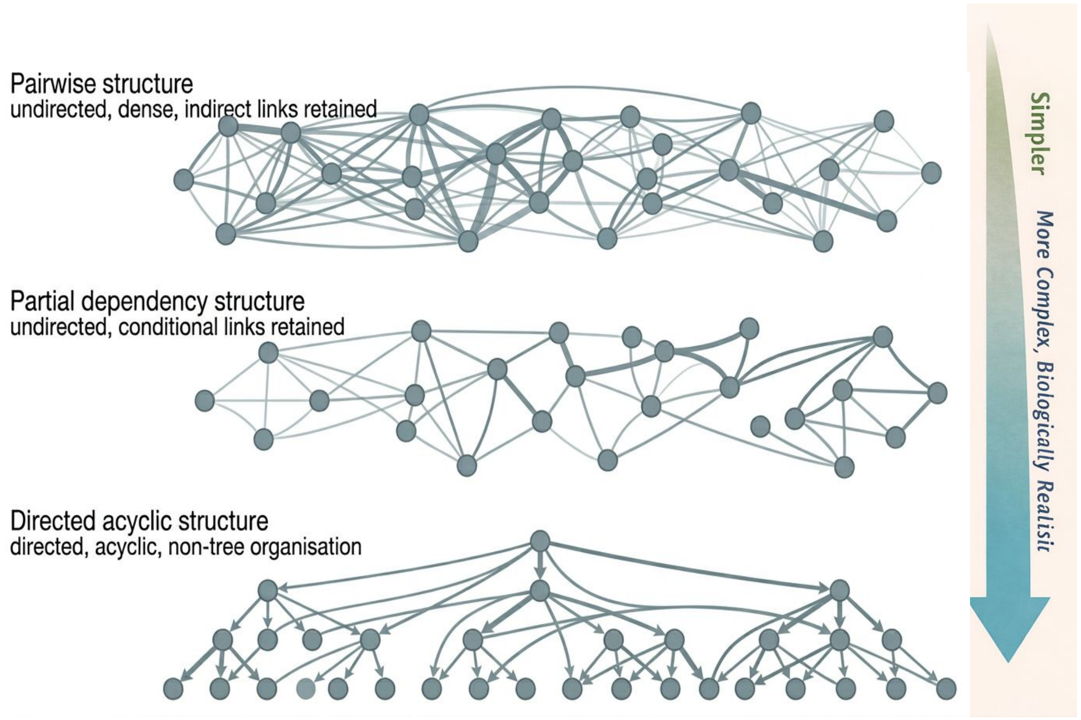
Depending on the algorithm, three paradigms of representing interactions can be identified, corresponding to distinct topological structures.

1. Association networks
2. **Dependency networks**

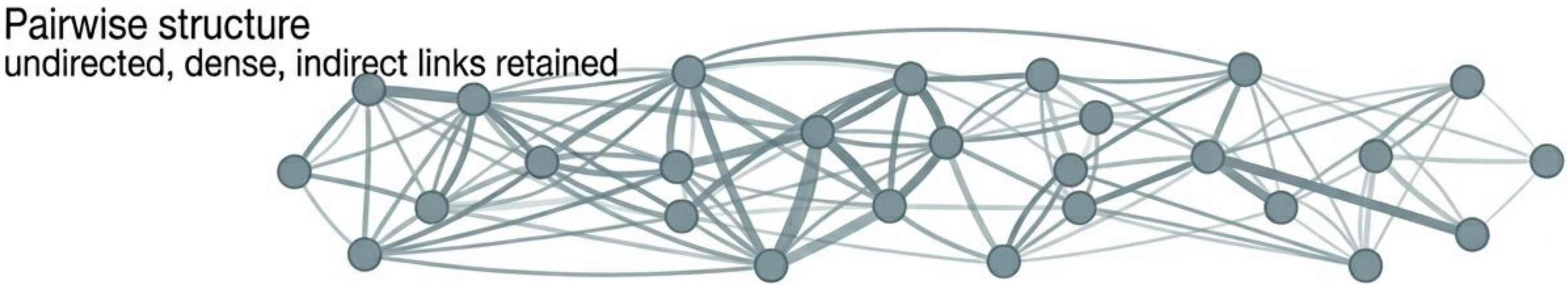


Topologies of network representations

1. Association networks
2. Dependency networks
3. **Regulatory networks**



1) Inference of association networks



Association networks

Association networks connect entities whose profiles **co-vary** across conditions or samples. They capture **similarity patterns**, not causal influence.

Depending on the level of representation:

- You can connect **features** (genes, proteins, metabolites, …) → *Feature–Feature network => see WGCNA R hands-on*
- Or connect **samples** (patients, cell lines, conditions) → *Sample–Sample network => see SNF R hands-on*

Association networks

Which metric to quantify the degree of similarity between two features?

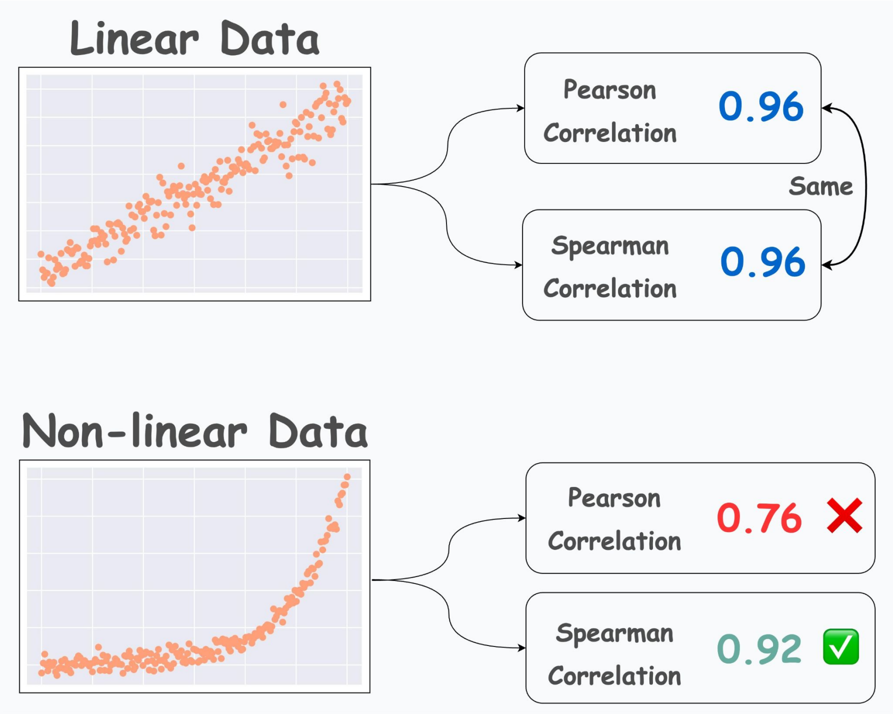
Association networks

Which metric to quantify the degree of similarity between two features?

- **Correlation-based** measures: Pearson, Kendall, …
- **Mutual information-based** measures
- Depends on the type of the inputs: continuous, discrete, ordinal, …

Correlation-based measures of association

Measure *how two variables co-vary* across samples. But edges are **undirected** (symmetry: corr(A,B) = corr(B,A))



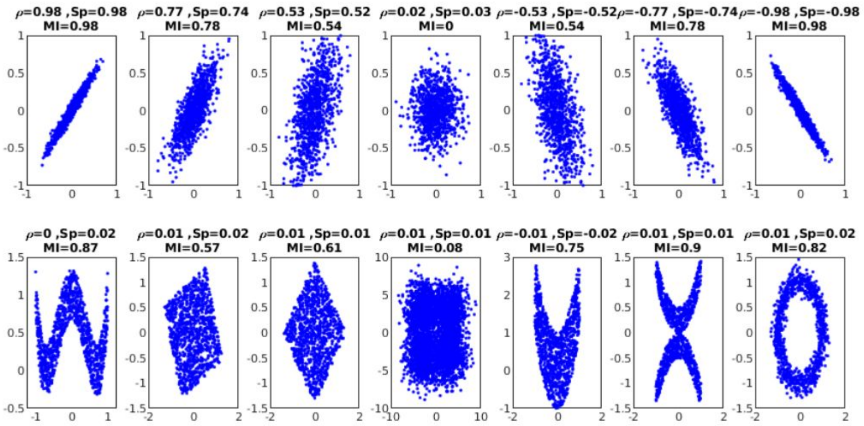
Pearson correlation
→ measures *linear* dependence between two vectors
→ sensitive to outliers

Spearman correlation
→ correlation of *ranks*
→ robust to outliers
→ detects non-linear but monotonic trends

Correlation versus Mutual Information

Correlation (Pearson / Spearman)

- Detects **linear** (Pearson) or **monotonic** (Spearman) co-variation.
- The edge has a **sign** (+ / −).
- But correlation also captures **indirect co-variation**



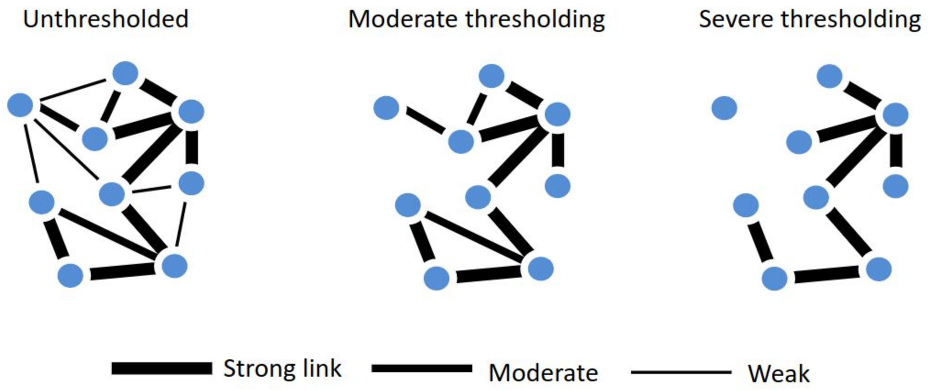
Mutual Information (MI)

- MI is always positive → the sign is missing.
- However, it's able to detect more complex dependence relationships, including non-linear and non-monotonic patterns.

Association networks represent co-variation between genes.
To identify **direct relationships**, we need to go beyond pairwise co-variation and **consider the gene set jointly**.
Correlation ≠ Causality

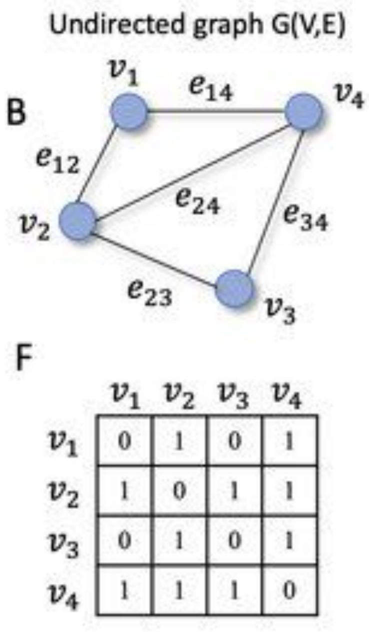
Source 3: **Inference from (Omics) data**

- **Often, the similarity/distances/correlations/… are thresholded.** Calibration **enables to:**
  - Reduce spurious (weak) associations
  - Emphasizing topological properties
  - Reducing storage burden of large graphs



Source: Zalesky, Andrew. “Network Statistics and Thresholding”

Data structure to store interactions: **Adjacency** matrix

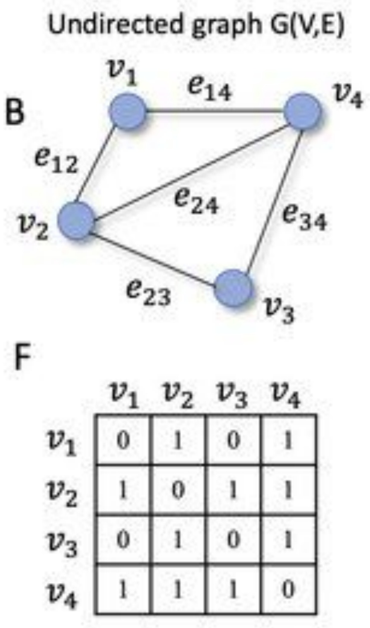


	v_1	v_2	v_3	v_4
v_1	0	1	0	1
v_2	1	0	1	1
v_3	0	1	0	1
v_4	1	1	1	0

Undirected -> Symmetric adjacency

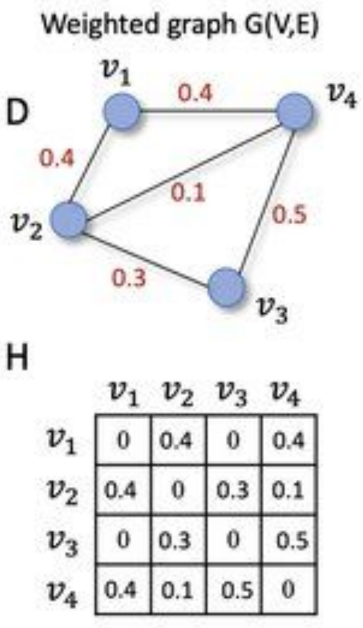
Source: Xu, Mengjia. "Understanding Graph Embedding Methods and Their Applications." *SIAM Review* 63 (November 4, 2021): 825–53. https://doi.org/10.1137/20M1386062.

Data structure to store interactions: **Adjacency** matrix



	v_1	v_2	v_3	v_4
v_1	0	1	0	1
v_2	1	0	1	1
v_3	0	1	0	1
v_4	1	1	1	0

Undirected -> **Symmetric adjacency**

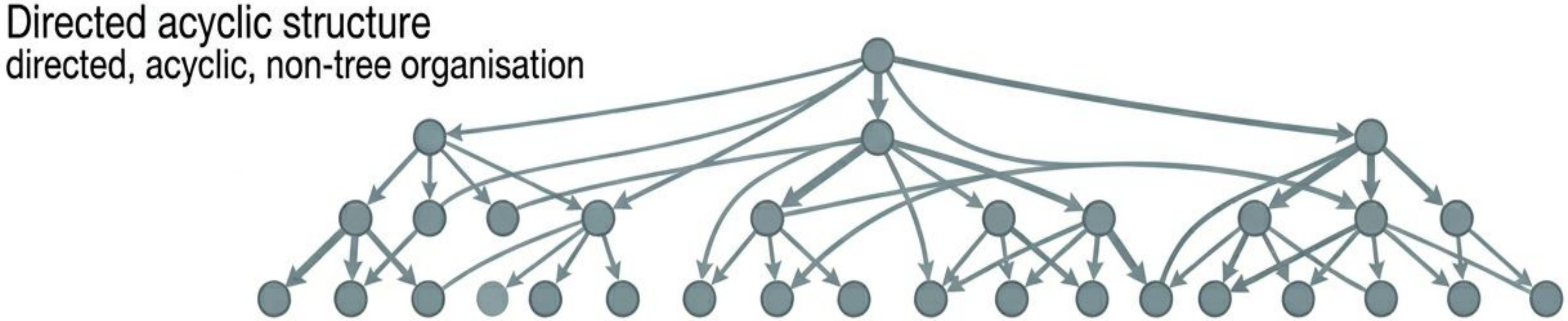


	v_1	v_2	v_3	v_4
v_1	0	0.4	0	0.4
v_2	0.4	0	0.3	0.1
v_3	0	0.3	0	0.5
v_4	0.4	0.1	0.5	0

Symmetric matrix, but with weights representing the strength of the interaction

Source: Xu, Mengjia. "Understanding Graph Embedding Methods and Their Applications." *SIAM Review* 63 (November 4, 2021): 825–53. https://doi.org/10.1137/20M1386062.

3) Inference of regulatory networks



Bayesian networks Table of Contents

1

Probability

Marginal, conditional,
and Bayes theorem

3

Bayesian networks

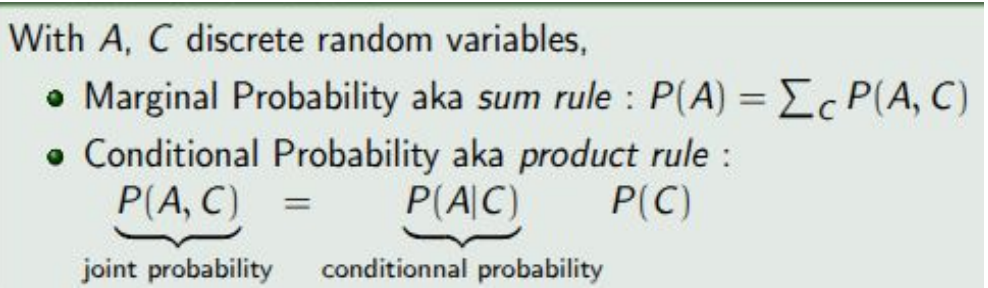
Factorisation,
d-separation

5

Inference of BNs

Score-based Versus
Constraint-based

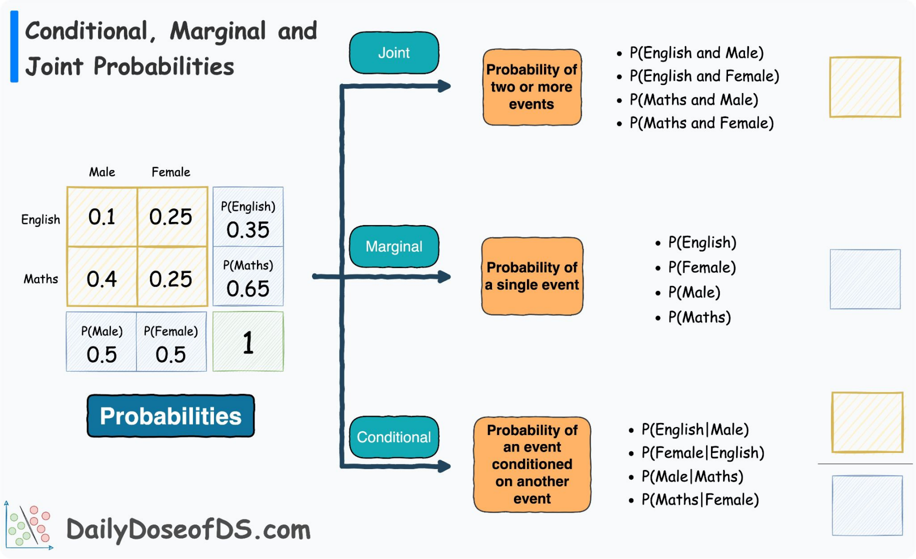
Probability reminders



With A, C discrete random variables,

- Marginal Probability aka *sum rule* : $P(A) = \sum_C P(A, C)$
- Conditional Probability aka *product rule* :

$$\underbrace{P(A, C)}_{\text{joint probability}} = \underbrace{P(A|C)}_{\text{conditionnal probability}} \quad P(C)$$



Conditional, Marginal and Joint Probabilities

	Male	Female	
English	0.1	0.25	P(English) 0.35
Maths	0.4	0.25	P(Maths) 0.65
	P(Male) 0.5	P(Female) 0.5	1

- P(English and Male)
- P(English and Female)
- P(Maths and Male)
- P(Maths and Female)

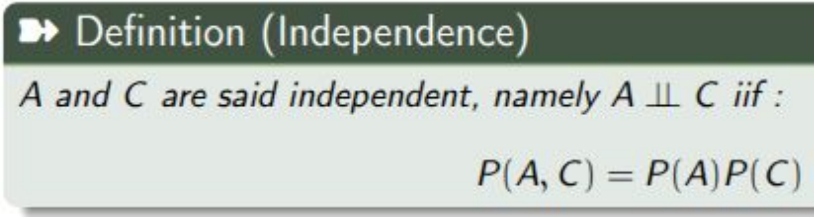
- P(English)
- P(Female)
- P(Male)
- P(Maths)

- P(English|Male)
- P(Female|English)
- P(Male|Maths)
- P(Maths|Female)

DailyDoseofDS.com

- **Marginalisation**, and **conditioning**, are the two core operations enabling to reconstruct any set of probability distributions.
- Extension to the continuous case using **integrals** instead of **sums**.

Probability reminders



Definition (Independence)

A and C are said independent, namely $A \perp\!\!\!\perp C$ iif :

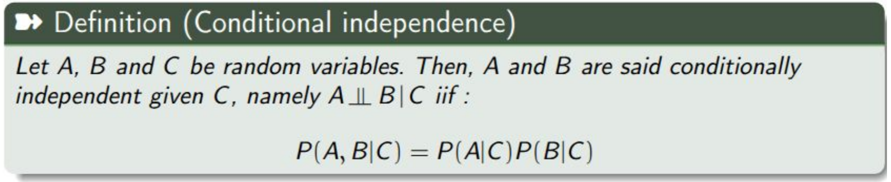
$$P(A, C) = P(A)P(C)$$

Independence advantages

- Decomposition of intractable joint multivariate distributions into simpler product of marginal terms
    - Faster inference of parameters
    - Straightforward estimation of the mean and variance estimates
- Most machine-learning algorithms assume independence between features

How to expand independence definition?

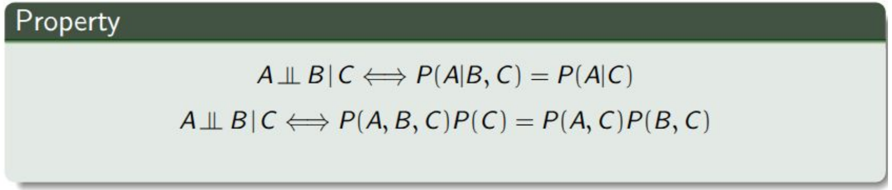
- In real-life, **full standard independence** is rarely observed -> underlying factors induce statistical dependencies (**flow of information** partially conveyed)
- **Conditional independence** solves this issue, by providing independence conditions after accounting for the effect of the remaining factors



Definition (Conditional independence)

Let A, B and C be random variables. Then, A and B are said conditionally independent given C, namely $A \perp\!\!\!\perp B \mid C$ iif :

$$P(A, B|C) = P(A|C)P(B|C)$$



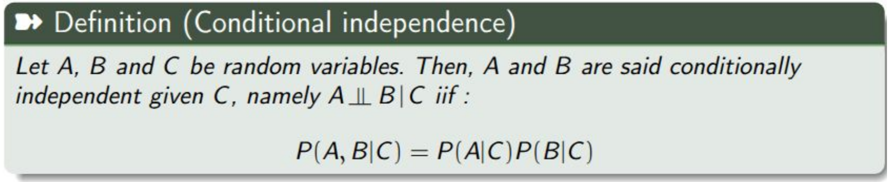
Property

$$A \perp\!\!\!\perp B \mid C \iff P(A|B, C) = P(A|C)$$

$$A \perp\!\!\!\perp B \mid C \iff P(A, B, C)P(C) = P(A, C)P(B, C)$$

Conditional Independence

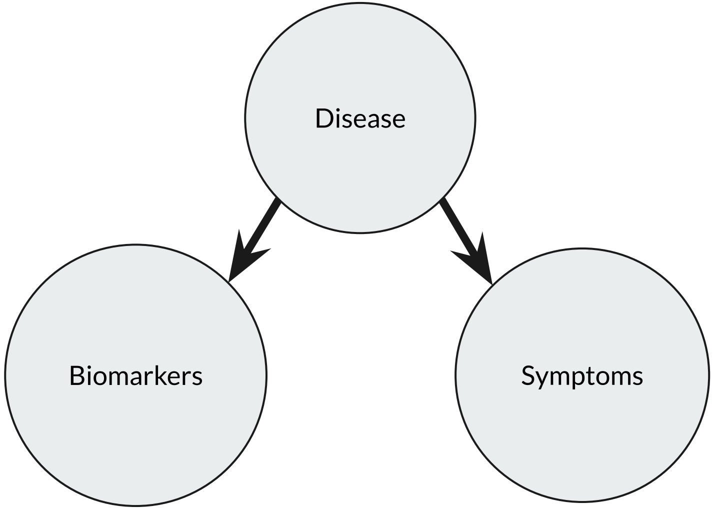
How to expand independence definition?



Definition (Conditional independence)

Let A, B and C be random variables. Then, A and B are said conditionally independent given C, namely $A \perp\!\!\!\perp B \mid C$ *iif :*

$$P(A, B|C) = P(A|C)P(B|C)$$



In exploratory research, we often try to pair biomarkers with the clinical symptoms

-> But the variations of biomarkers are rarely the core factors that trigger the clinical symptoms

Conditional Independence

Represent Conditional independencies

Definition of independence models

- An **independence model** I over a set **V:**
  - a set of triples ⟨X, Y | Z⟩, called i**ndependence relations**
  - X, Y and Z are disjoint subsets of V.
- The independence model *I* is said **probabilistic,** if it can be represented by a probabilistic distribution, compliant with basic probability rules

$$\langle \mathbf{X}, \mathbf{Y} \mid \mathbf{Z} \rangle \in I \iff \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \quad \text{w.r.t. } p.$$

Is there any insightful representation that would capture the set of independent clauses?

Represent Conditional independencies

Definition of independence models

- An **independence model** I over a set **V:**
    - a set of triples ⟨X, Y | Z⟩, called i**ndependence relations**
    - X, Y and Z are disjoint subsets of V.

- The independence model *I* is said **probabilistic,** if it can be represented by a probabilistic distribution, compliant with basic probability rules

$$\langle \mathbf{X}, \mathbf{Y} \mid \mathbf{Z} \rangle \in I \iff \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \quad \text{w.r.t. } p.$$

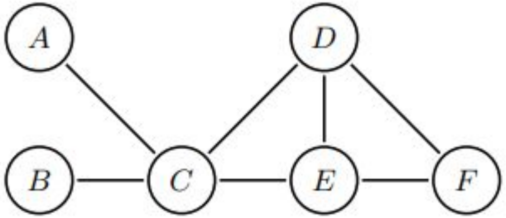
Is there any insightful representation that would capture the set of independent clauses?

Yes, **probabilistic graphical networks**

The two most popular PGMs frameworks

Markov Networks

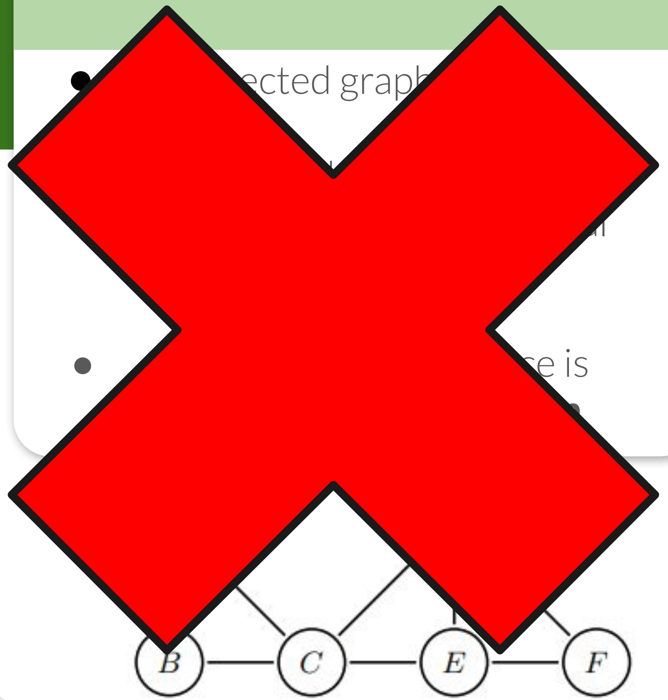
- Undirected graphs
- Also named Markov **Random fields -> GGMs** are a special case of it
- Criteria of independence is based on the ***u*-separation**



The two most popular PGMs frameworks

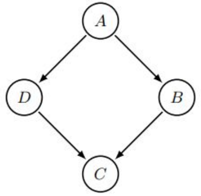
Markov Networks

- ected graph
- e is

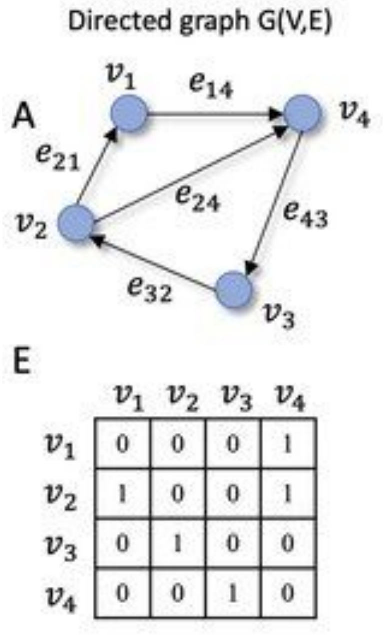


Bayesian Networks

- DAGs (no cycle, but directed)
- Factorises as a probability distribution
- Criteria of independence is based on the ***d*-separation**



Data structure to store interactions: **Adjacency** matrix



	v_1	v_2	v_3	v_4
v_1	0	0	0	1
v_2	1	0	0	1
v_3	0	1	0	0
v_4	0	0	1	0

Directed -> Asymmetric adjacency

Source: Xu, Mengjia. “Understanding Graph Embedding Methods and Their Applications.” *SIAM Review* 63 (November 4, 2021): 825–53. https://doi.org/10.1137/20M1386062.

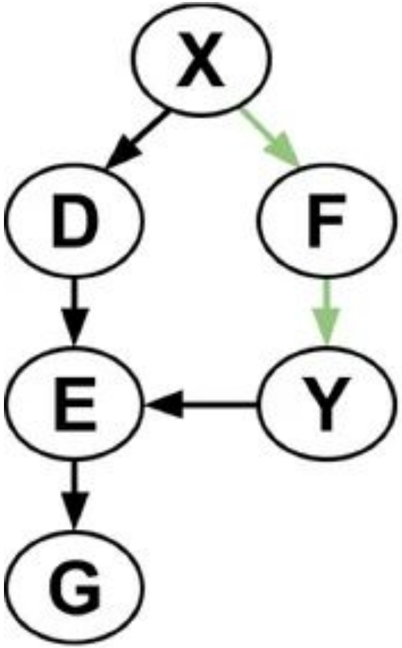
d-separation: Intuitive Interpretation as flow

D-separation for noobs

Consider a path between X and Y + disjoint subset Z:

1. Without conditioning node (Z=empty)
   a. any intermediate node between X and Y is open
   b. except if the node is a **collider,** or any of its **descendants -> in this case, it's closed**

To be independent, all paths between X and Y must be closed (no flux of information).



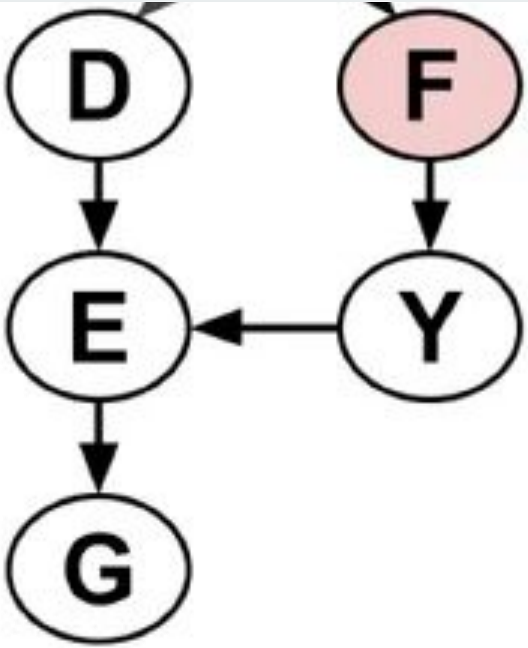
$X \not\perp Y$

D-separation for noobs

Consider a path between X and Y + disjoint subset Z:

1. Without conditioning node:
    a. any intermediate node between X and Y is open
    b. except if the node is a **collider,** or any of its **descendants -> in this case, it's closed**

Z d-separates X and Y if all the paths between X and Y are closed (no flux of information).



Are D and Y independent?

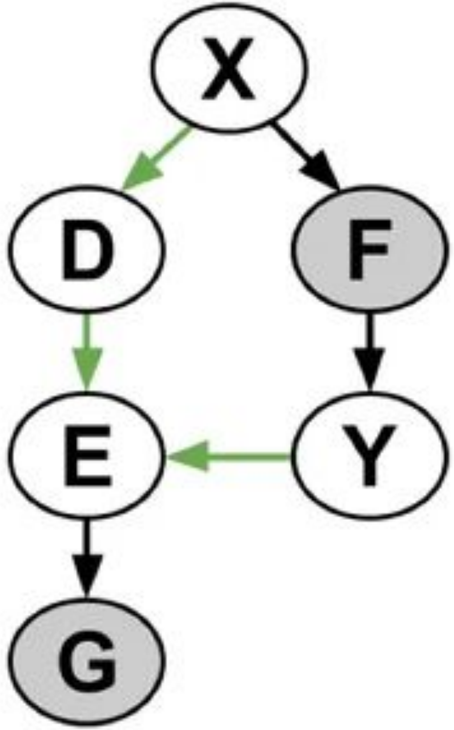
d-separation: Intuitive Interpretation as flow

D-separation for noobs

Consider a path between X and Y + disjoint subset Z:

1. **Conditioning** reverses the flow:
    a. Open -> close
    b. Close -> Open

Z d-separates X and Y if all the paths between X and Y are closed (no flux of information).



$X \not\perp Y \mid F, G$

d-separation: Intuitive Interpretation as flow

1. **First misconception:** only interested in direct pairwise interactions -> BNs are interested in modelling higher-order interactions
2. **Second misconception:** The direction of the arrows is irrelevant, except for V-structures
3. **Third misconception:** account for the topological context.
    a. V-structure are counter-intuitive
    b. Conditioning on a set of nodes may open paths that used to be closed, and reciprocally

3-node special BNs, and rehearsal

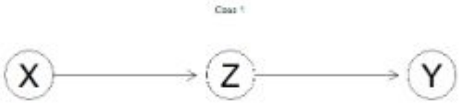
Measure the effect of variation of the **independent variable** X on the **response variable** Y (defined by expert knowledge) that is mediated by a **covariate** Z, with no direct effect of X upon Y.

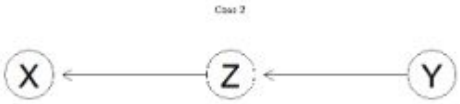
How many distinct configurations in total can you identify, being DAGs, and with node Z as the intermediate node?

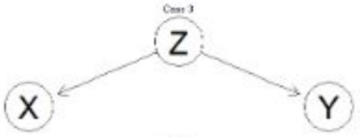
3-node special BNs, and rehearsal

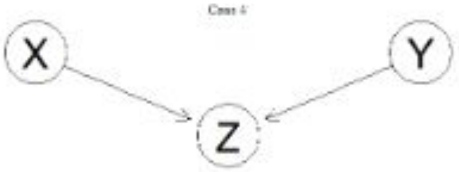
How many distinct configurations in total can you identify, being DAGs, and with node Z as the intermediate node?

4 distinct configurations can be identified, changing the orientation of edges.





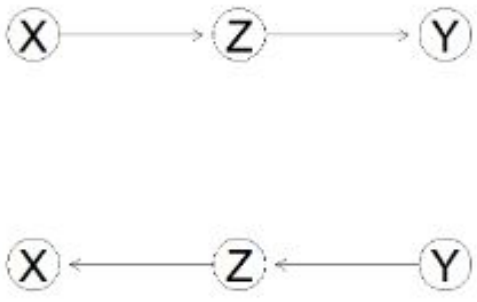




3-node special BNs, and rehearsal

Markov chains

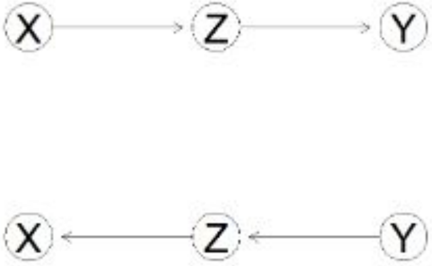
- Subsets of Bayesian networks.
- They model stochastic process, in which the next state depends only on the current state.
- Considering **dynamics**, you may interpret a given node as an observation at time *t*, and its parent an observation at time *t - 1*, after unrolling the BN.



3-node special BNs: Markov Chains

Exercise: prove the two CIs

- Apply *d*-separation criterion.
- Apply Bayes theorem, and law of total probabilities



$X \not\perp\!\!\!\perp Y \mid \varnothing$

D-separation for noobs

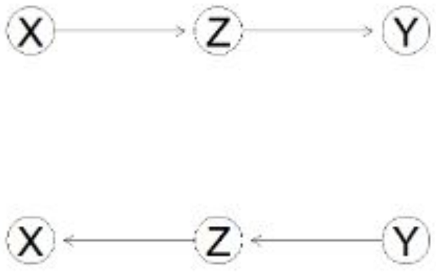
Consider a path between X and Y + disjoint subset Z:

1. Without conditioning node:
   a. any intermediate node between X and Y is open
   b. except if the node is a **collider,** or any of its **descendants -> in this case, it's closed**
2. Conditioning reverses the flow:
   a. Open -> close
   b. Close -> Open
3. **Z** d-separates X and Y if **all** the paths between X and Y are **closed** (no flux of information).

3-node special BNs: Markov Chains

Exercise: prove the two CIs

- Apply *d*-separation criterion.
- Apply Bayes theorem, and law of total probabilities



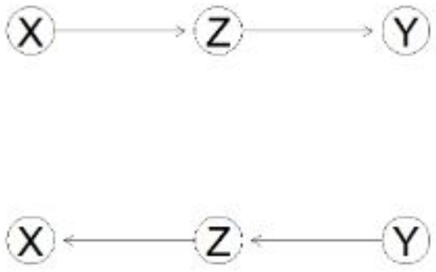
$X \not\perp\!\!\!\perp Y | \emptyset$

$$
\begin{aligned}
p(y \mid x) &= \frac{p(x,y)}{p(x)} \\
&= \frac{sum_z p(x,y,z)}{p(x)} \\
&= \frac{\cancel{p(x)}}{\cancel{p(x)}} \sum_z p(z \mid x) p(y \mid z) \\
&\neq p(y)
\end{aligned}
$$

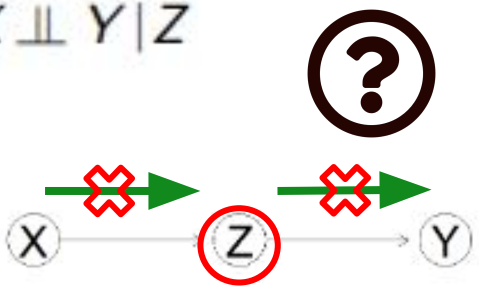
3-node special BNs: Markov Chains

Exercise: prove the two CIs

- Apply *d*-separation criterion.
- Apply Bayes theorem, and law of total probabilities

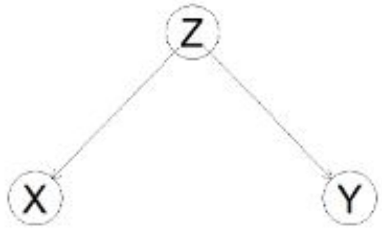


$$X \perp\!\!\!\perp Y \mid Z$$



$$\begin{aligned} p(x,y\mid z) &= \frac{p(x,y,z)}{p(z)} \\ &= \frac{p(z)p(y|z)p(x|z)}{p(z)} \\ &= p(y\mid z)p(x\mid z) \end{aligned}$$

3-node special BNs: Markov Chains



Confounder

- Z is a **latent cause**, that must be controlled
- If not observed, that may wrongly leads statisticians into setting up a spurious correlation between X and Y.

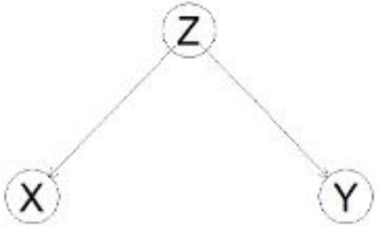
Exercise: prove the two CIs

- Apply *d*-separation criterion.
- Apply Bayes theorem, and law of total probabilities

$X \not\perp\!\!\!\perp Y | \emptyset$

$X \perp\!\!\!\perp Y | Z$

3-node special BNs: Markov Chains



Confounder

- Z is a **latent cause**, that must be controlled
- If not observed, that may wrongly leads statisticians into setting up a spurious correlation between X and Y.

Exercise: prove the two CIs

- Apply *d*-separation criterion.
- Apply Bayes theorem, and law of total probabilities

$$X \not\perp\!\!\!\perp Y \mid \emptyset$$

$$\begin{aligned} p(x,y) &= \textstyle\sum_z p(x,y,z) \\ &= \textstyle\sum_z p(x \mid z) p(y \mid z) \\ &\neq p(x)p(y) \end{aligned}$$

$$X \perp\!\!\!\perp Y \mid Z$$

$$\begin{aligned} p(x,y \mid z) &= \frac{p(x,y,z)}{p(z)} \\ &= \frac{p(z)p(y|z)p(x|z)}{p(z)} \\ &= p(y \mid z)p(x \mid z) \end{aligned}$$

V-structure (aka the least intuitive)

- Z is a **collider -> different strategy**
  - You should not control for it, as it may induce spurious correlation
  - Related to selection bias, and Simpson’s paradox
- Used for `XOR` encoding, such as genes activation, or toggle switch.

Exercise: prove the two CIs

- Apply *d*-separation criterion.
- Apply Bayes theorem, and law of total probabilities

$X \perp\!\!\!\perp Y | \emptyset$

$X \not\perp\!\!\!\perp Y | Z$

3-node special BNs: Markov Chains

Exercise: prove the two CIs

- Apply *d*-separation criterion.
- Apply Bayes theorem, and law of total probabilities

V-structure (aka the least intuitive)

$$X \perp\!\!\!\perp Y \mid \varnothing$$

$$\begin{aligned} p(x,y) &= \sum_z (p(x,y,z) \\ &= p(x)p(y)\sum_z p(z \mid x,y) \\ &= p(x)p(y) \end{aligned}$$

$$X \not\perp\!\!\!\perp Y \mid Z$$

$$\begin{aligned} p(x,y \mid z) &= \frac{p(x,y,z)}{p(z)} \\ &= \frac{p(x)p(y)p(z|x,y)}{p(z)} \\ &\neq p(x \mid z)p(y \mid z) \end{aligned}$$

Datasaurus toy model
Visual depiction of V-structure

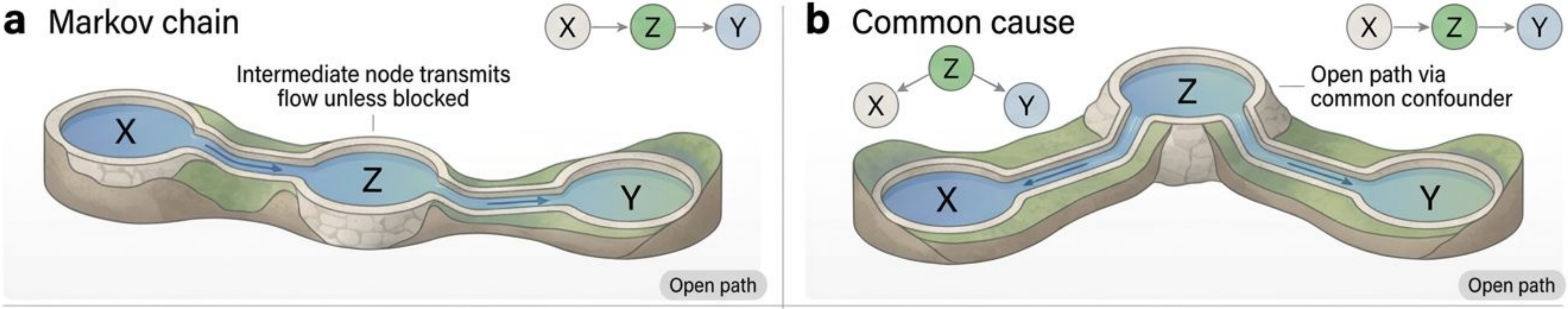
Add a **latent categorical** variable Z for each cloud of points will reveal an underlying connection between X and Y.

Always visualise the distribution of your datasets:

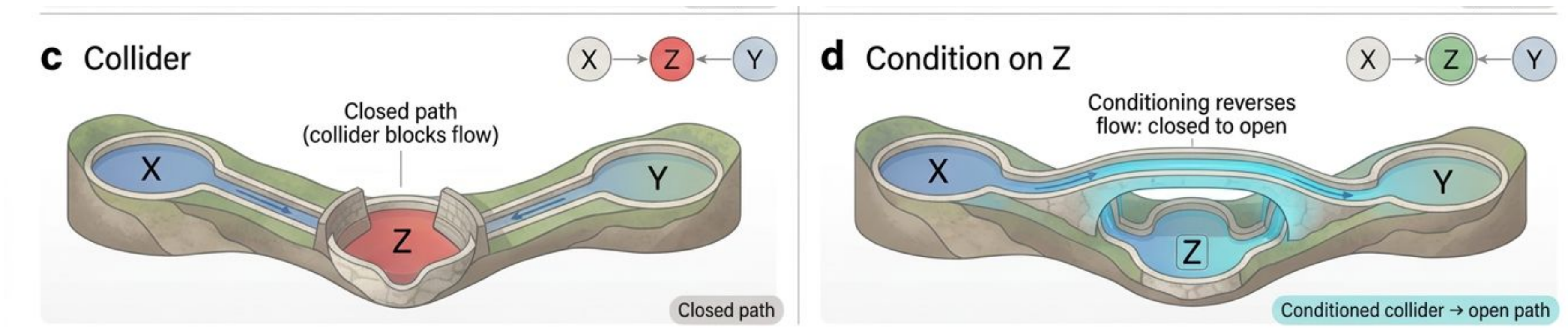
- Datasaurus: similar descriptive statistics
- But complete distinct distributions (visually easy to discriminate)

V-structures can unravel hidden dependencies.

A metaphor of conditional dependencies



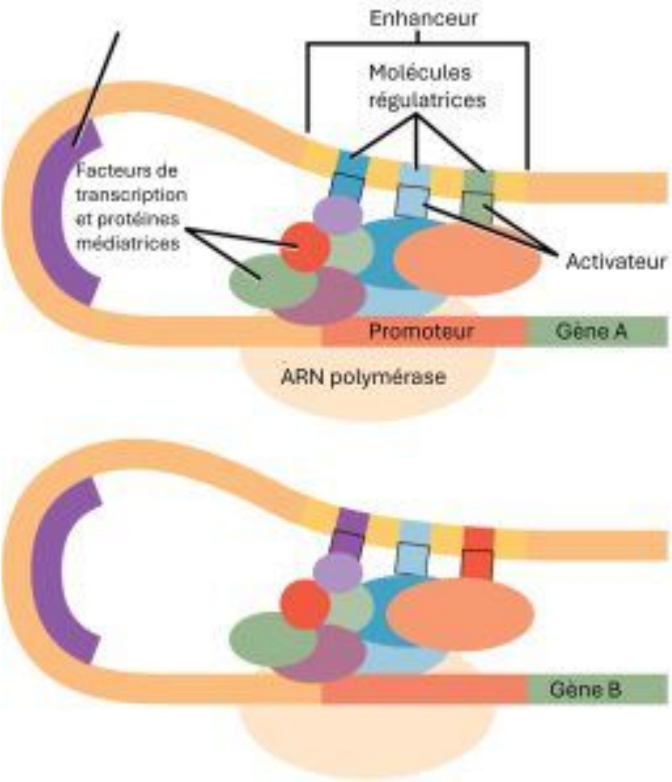
A metaphor of conditional dependencies



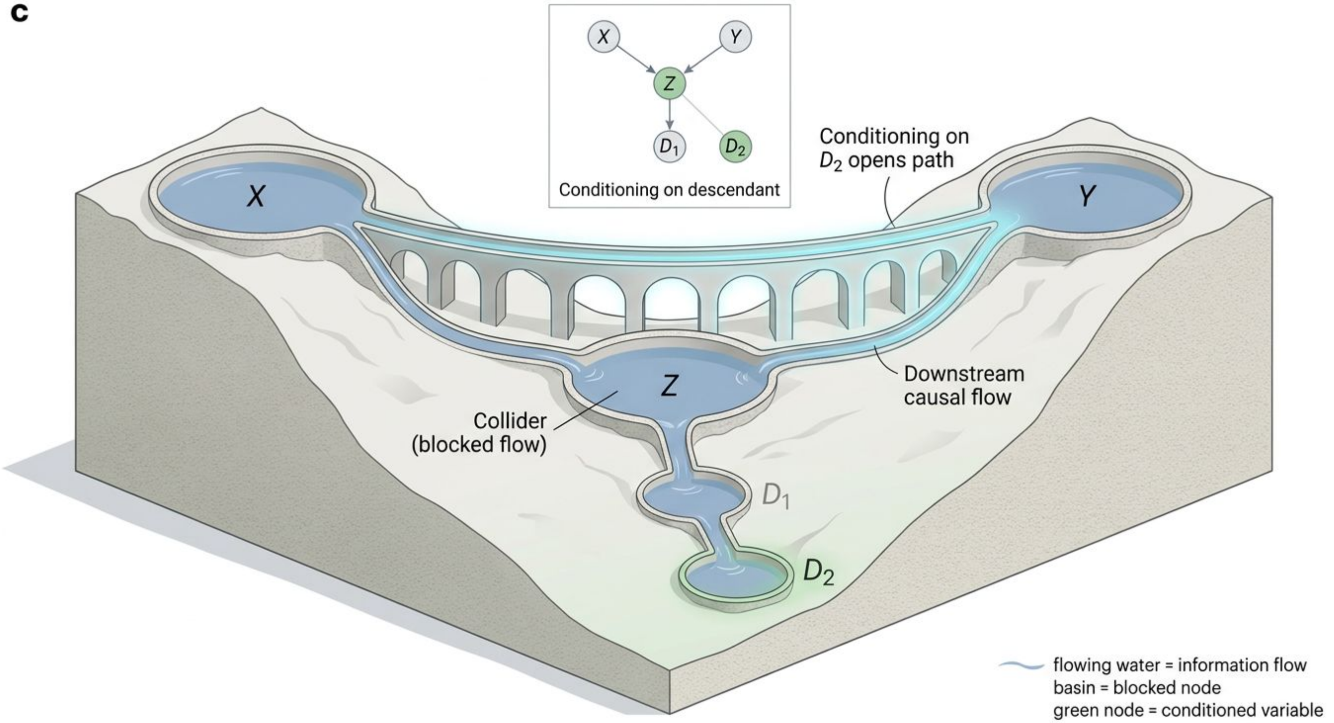
Z d-separates *X* and *Y* when all paths between them are closed (no information flow).
Conditioning can reverse path states: open paths may close, and colliders may open.

🟦 blue flowing water = information flow
🟫 blocked basin / dam = closed path
🟦 highlighted channel / gate / aqueduct = conditioning
🟩 DAG node *Z* (green: path open / red: path closed)

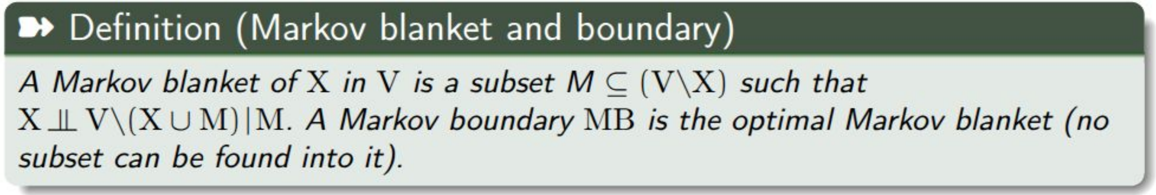
V-structure as complexes of transcription factors



Descendants can help in unravelling complex mechanistic interactions



Markov Blanket, and Boundary



Definition (Markov blanket and boundary)

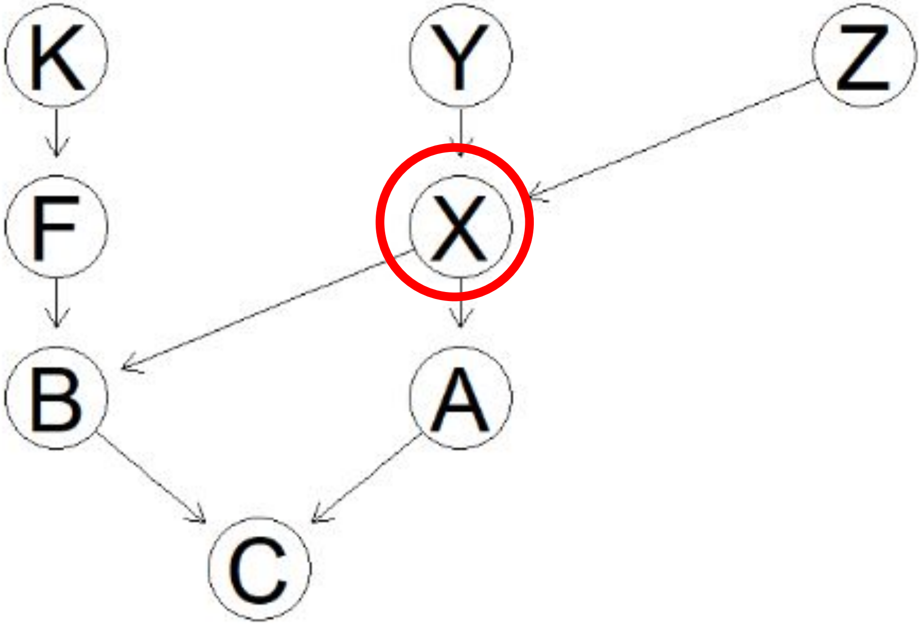
A Markov blanket of X *in* V *is a subset* $M \subseteq (\mathrm{V}\setminus\mathrm{X})$ *such that* $\mathrm{X} \perp\!\!\!\perp \mathrm{V}\setminus(\mathrm{X}\cup\mathrm{M})\,|\,\mathrm{M}$. *A Markov boundary* MB *is the optimal Markov blanket (no subset can be found into it).*

With:

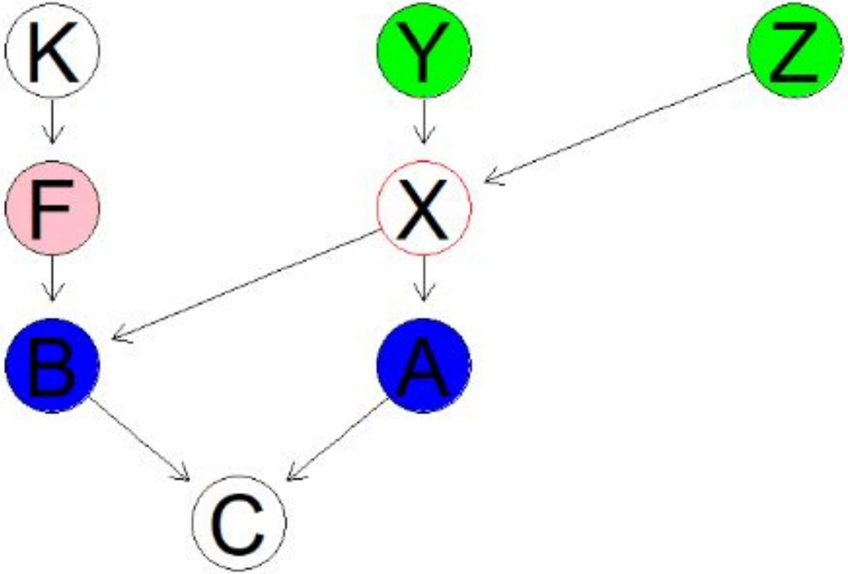
- V_x the **neighbours** of X (namely, its direct parents, and children)
- SP_x the **spouses** of X (nodes sharing the same children)

Markov Boundary: a biological application

Find the **Markov boundary** of node **X.**



Markov Boundary: a biological application

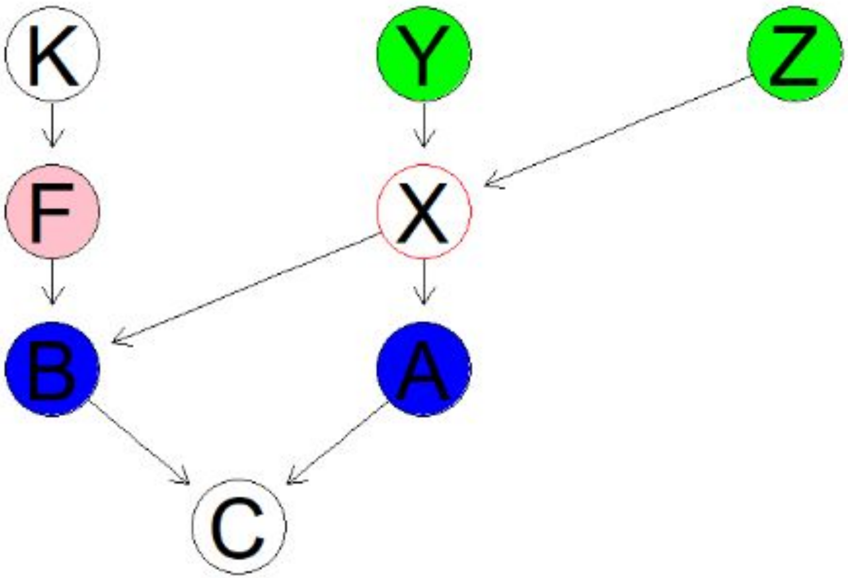


Solutions:

- a **Markov blanket** of node X

Find back the legend: what are the **spouses, children, and parents** of node *X*?

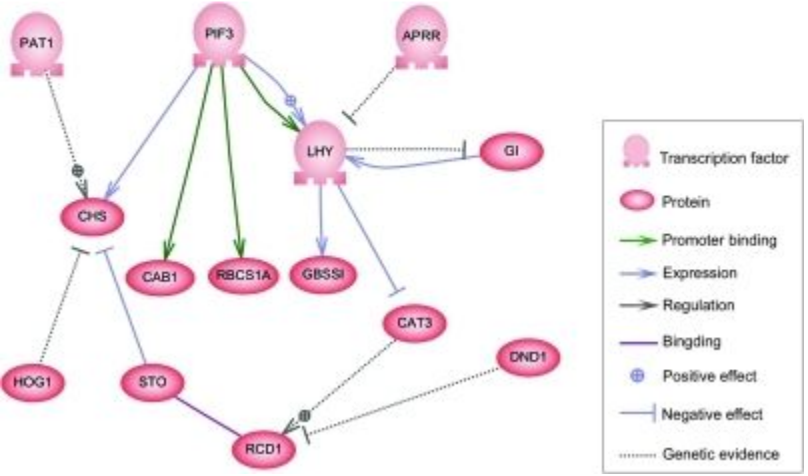
Markov Boundary: a biological application



Solutions:

- Which are the kinship relationships of X?
- **spouses** ●
- **children** 🔵
- **parents** 🟢

Markov Boundary: a biological application



In GRNs, the **Markov boundary** is the *smallest set of other genes* you need to look at to predict its behavior.

Example: to try to to predict whether a gene is “on” or “off”, instead of looking at *all* genes in the network, you only need to look at the genes covered by the **Markov boundary,** namely its:

1. **Direct regulators**, the parents, such as the **transcription factors**
2. The Direct downstream targets (children) **AND**
3. **Co-regulators** (spouses), aka the other genes that regulate the same targets as G (for example, **operons) ->** this information requires *V-structure* conditional structure

Bibliography

Books

- Patric Naim, Pierre-Henri Wuillemin (2007). *Réseaux bayésiens* (FR). Algorithmes
- Richard E. Neapolitan. *Learning Bayesian Networks* (2019).Pearson
- Koller, Daphne, and Nir Friedman. *Probabilistic Graphical Models*. MIT Press, 2009.

Packages

- bnlearn for BN inference with R
- PyAgrum for BN Inference with Python

Slides

- Alexandre Aussem on Probabilistic Graphical Models (Markov + Bayesian networks)
- Marco Scutari on Understanding Bayesian Networks with Examples in R

Biological applications

- Yupeng Li Naim, Yushi Liu (2019). A *Bayesian gene network reveals insight into the JAK-STAT pathway in systemic lupus erythematosus*. Plos One.
- Louis Verny and Hervé Isambert (2017). *Learning causal networks with latent variables from multivariate information in genomic data*. Plos Computational Biology

Acknowledgments

Alexandre Aussem

Hybrid-based BN inference

Pierre Henri Wuillemin

PyAgrum developper

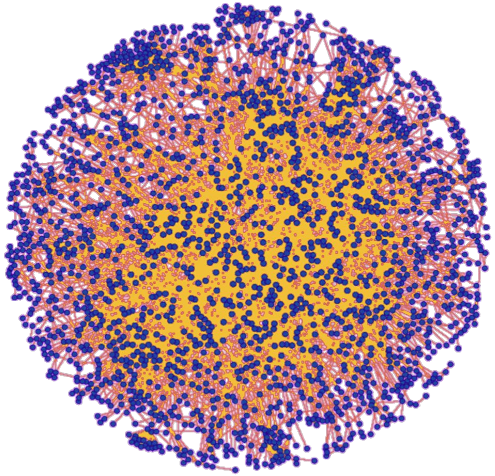
Herve Isambert

Curie's MIIC developper

Marco Scutari

Core bnlearn developper

Outline



Source: https://www3.nd.edu/~tmilenko/research.html.

General Introduction, Key Concepts

Network Biology, Network Science, Complex Systems, …

Network Construction/Inference

Data sources, Distances/Similarities, Correlation, …

Network Analysis

Measures

Degree, Centrality, Distances, …

Algorithms

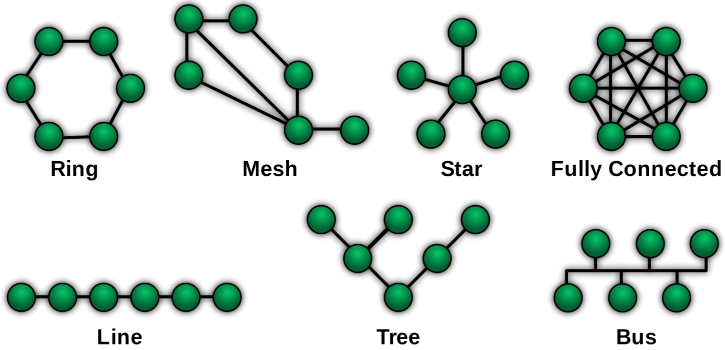
Diffusion, Clustering, Embedding, …

Networks and Integration

Multiplex Networks, Multilayer Networks, Knowledge Graphs, …

Studying Network **Topology**

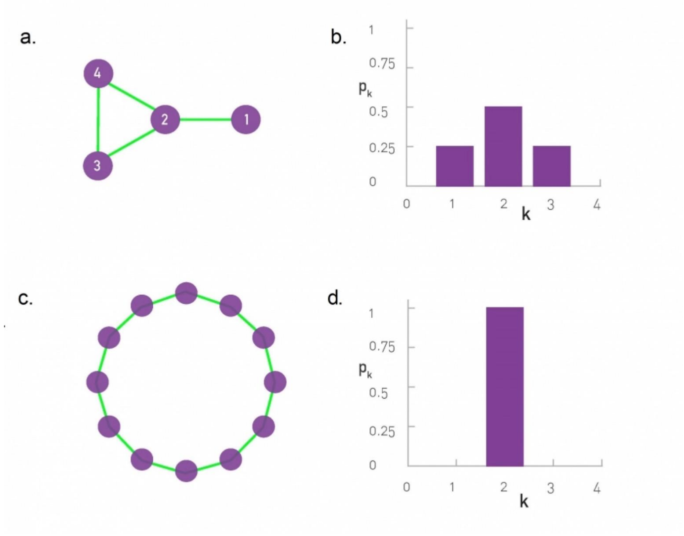
- ➢ **Topology** is the way in which the **nodes and the edges are arranged** within a network.
  - ○ Robustness of a network
  - ○ Hub nodes
  - ○ Essential nodes
  - ○ Paths/accessibility
  - ○ Communities
  - ○ ….



Source: “Network Topology.” In Wikipedia, August 30, 2023. https://en.wikipedia.org/w/index.php?title=Network_topology&oldid=1172927659.

Metrics in network analysis: **Node Degree**

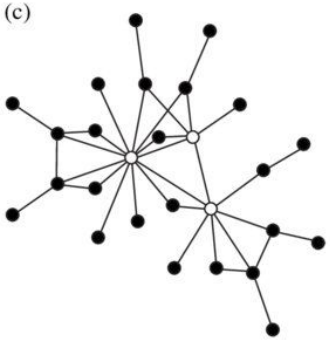
- The **degree of a node** in a network is the **number of edges that are incident to the node.**
- Observing the **degree distribution** of a network can reveal interesting properties:
  - Overall connectivity



Source: Courtesy of Anaïs Baudot

Metrics in network analysis: **Node Degree**

- The **degree of a node** in a network is the **number of edges that are incident to the node.**
- Observing the **degree distribution** of a network can reveal interesting properties:
  - Overall connectivity
  - Presence of hubs (i.e. nodes with #links >> avg(#links))



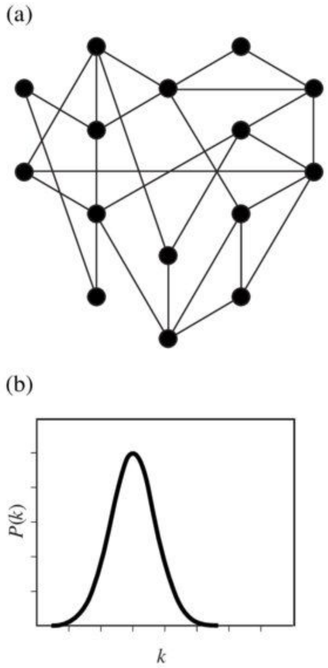
Source: Bentley, Barry. "Connectomics of Extrasynaptic Signalling: Applications to the Nervous System of Caenorhabditis Elegans," 2017. https://doi.org/10.17863/CAM.16873.

Metrics in network analysis: **Node Degree**

- The **degree of a node** in a network is the **number of edges that are incident to the node.**
- Observing the **degree distribution** of a network can reveal interesting properties:
  - Overall connectivity
  - Presence of hubs (i.e. nodes with #links >> avg(#links))
  - Scale-free property

Metrics in network analysis: **Node Degree**

- The **degree of a node** in a network is the **number of edges that**
- Observing the **degree distribution** of a network can reveal int
  - Overall connectivity
  - Presence of hubs (i.e. nodes with #links >> avg(#links))
  - Scale-free property

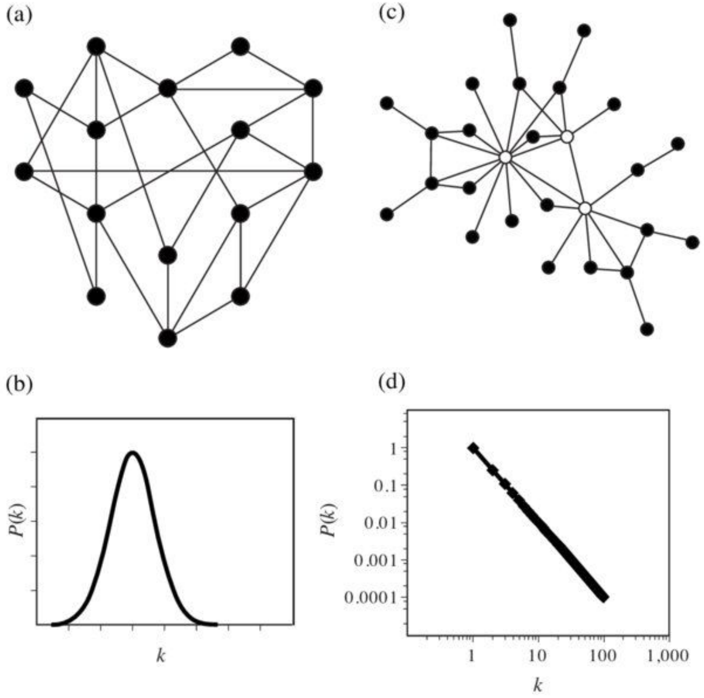


Random Network

Source: Bentley, Barry. “Connectomics of Extrasynaptic Signalling: Applications to the Nervous System of Caenorhabditis Elegans,” 2017. https://doi.org/10.17863/CAM.16873.

Metrics in network analysis: **Node Degree**

- The **degree of a node** in a network is the **number of edges tha**
- Observing the **degree distribution** of a network can reveal int
  - Overall connectivity
  - Presence of hubs (i.e. nodes with #links >> avg(#links))
  - Scale-free property



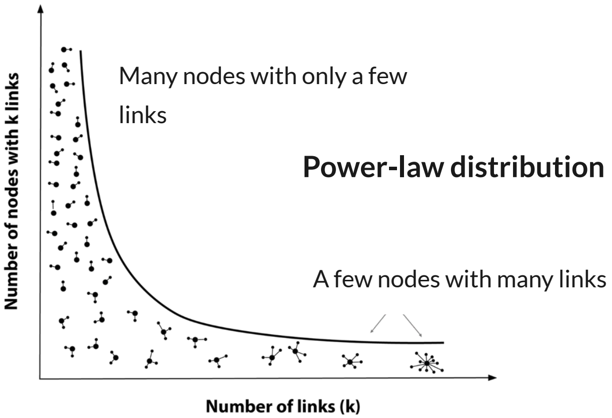
Random Network

Scale-free Network
(plot in log-log scale)

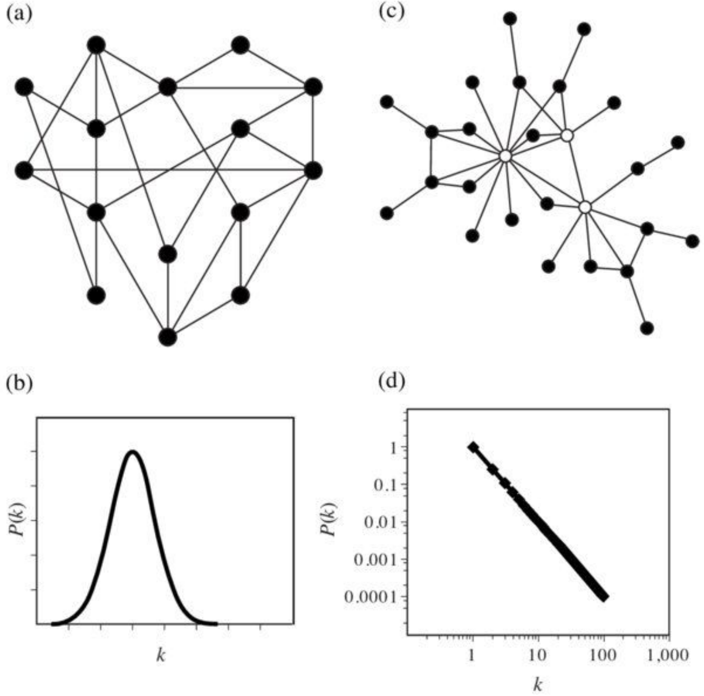
Source: Bentley, Barry. "Connectomics of Extrasynaptic Signalling: Applications to the Nervous System of Caenorhabditis Elegans," 2017. https://doi.org/10.17863/CAM.16873.

Metrics in network analysis: **Node Degree**

- The **degree of a node** in a network is the **number of edges that**
- Observing the **degree distribution** of a network can reveal int
  - Overall connectivity
  - Presence of hubs (i.e. nodes with #links >> avg(#links))
  - Scale-free property



Source: nesta. "The Rise of the Platform Economy." Accessed August 28, 2023. https://www.nesta.org.uk/blog/rise-platform-economy/.



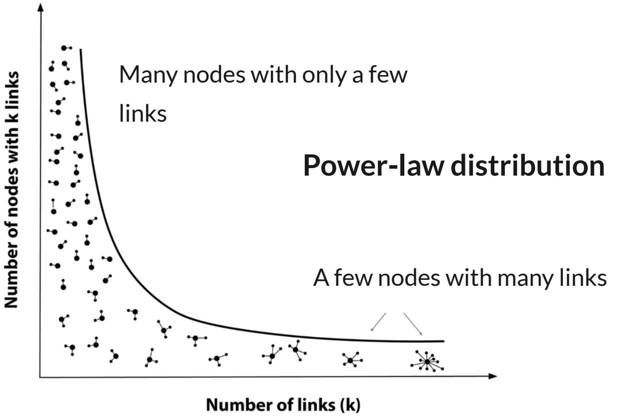
Random Network

Scale-free Network
(plot in log-log scale)

Source: Bentley, Barry. "Connectomics of Extrasynaptic Signalling: Applications to the Nervous System of Caenorhabditis Elegans," 2017. https://doi.org/10.17863/CAM.16873.

Metrics in network analysis: **Node Degree**

- The **degree of a node** in a network is the **number of edges that are incident to the node.**
- Observing the **degree distribution** of a network can reveal interesting properties:
  - Overall connectivity
  - Presence of hubs (i.e. nodes with #links >> avg(#links))
  - Scale-free property



Source: nesta. "The Rise of the Platform Economy." Accessed August 28, 2023. https://www.nesta.org.uk/blog/rise-platform-economy/.

- Biological interpretation ?
  - Growth with preferential attachment ("rich get richer") -> hub nodes
  - Robust to random attack, sensitive to targeted attacks

Are biological networks scale-free ?

Scale-Freeness and Biological Networks

Masanori Arita

The Journal of Biochemistry, Volume 138, Issue 1, Jul. 2005, Pages 1–4,
https://doi.org/10.1093/jb/mvi094
Published: 01 July 2005 **Article history** ▾

COMMENTARY | 01 NOVEMBER 2005

Scale-free networks in cell biology FREE

In collection: Metabolism

Réka Albert

+ Author and article information

J Cell Sci (2005) 118 (21): 4947–4957.
https://doi.org/10.1242/jcs.02714

› J Comput Biol. 2006 Apr;13(3):810-8. doi: 10.1089/cmb.2006.13.810.

How scale-free are biological networks

Raya Khanin [1], Ernst Wit

Affiliations + expand

PMID: 16706727 DOI: 10.1089/cmb.2006.13.810

- ➢ PPI
- ➢ Metabolic networks
- ➢ Regulatory networks

Are biological networks scale-free ? A debate!

Scale-Freeness and Biological Networks

Get access

Masanori Arita

The Journal of Biochemistry, Volume 138, Issue 1, Jul. 2005, Pages 1–4, https://doi.org/10.1093/jb/mvi094

Published: 01 July 2005 **Article history** ▾

COMMENTARY | 01 NOVEMBER 2005

Scale-free networks in cell biology FREE

In collection: Metabolism

Réka Albert

+ Author and article information

J Cell Sci (2005) 118 (21): 4947–4957.

https://doi.org/10.1242/jcs.02714

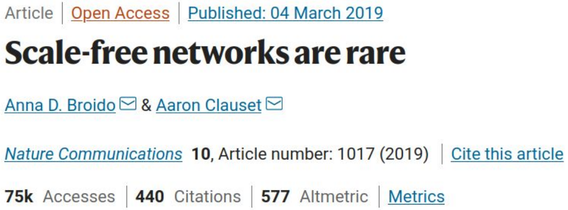
> J Comput Biol. 2006 Apr;13(3):810-8. doi: 10.1089/cmb.2006.13.810.

How scale-free are biological networks

Raya Khanin [1], Ernst Wit

Affiliations + expand

PMID: 16706727 DOI: 10.1089/cmb.2006.13.810



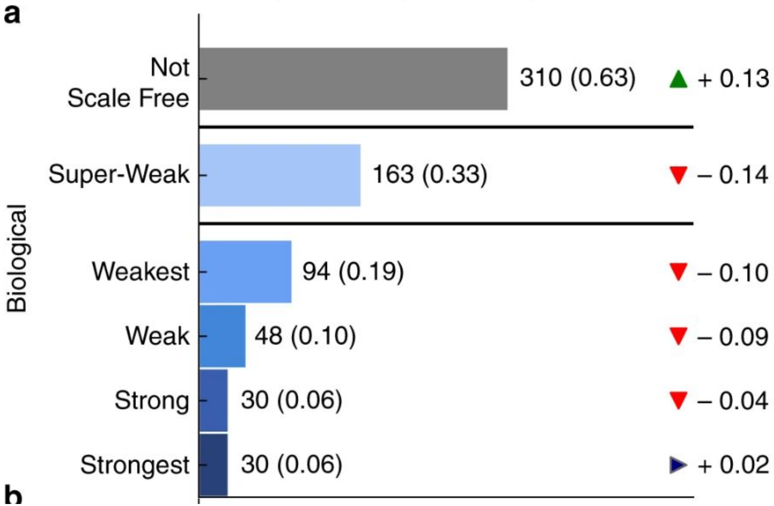
Article | Open Access | Published: 04 March 2019

Scale-free networks are rare

Anna D. Broido & Aaron Clauset

Nature Communications **10**, Article number: 1017 (2019) | Cite this article

75k Accesses | **440** Citations | **577** Altmetric | Metrics



Are biological networks scale-free ? A debate!

Scale-Freeness and Biological Networks

Get access

Masanori Arita

The Journal of Biochemistry, Volume 138, Issue 1, Jul. 2005, Pages 1–4, https://doi.org/10.1093/jb/mvi094

Published: 01 July 2005 **Article history** ▾

COMMENTARY | 01 NOVEMBER 2005

Scale-free networks in cell biology FREE

In collection: Metabolism

Réka Albert

+ Author and article information

J Cell Sci (2005) 118 (21): 4947–4957.

https://doi.org/10.1242/jcs.02714

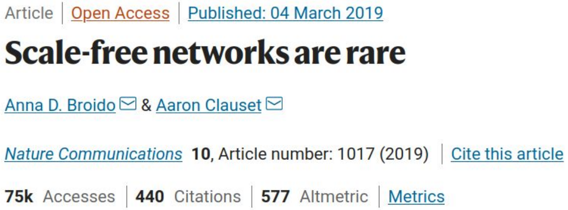
› J Comput Biol. 2006 Apr;13(3):810-8. doi: 10.1089/cmb.2006.13.810.

How scale-free are biological networks

Raya Khanin [1], Ernst Wit

Affiliations + expand

PMID: 16706727 DOI: 10.1089/cmb.2006.13.810



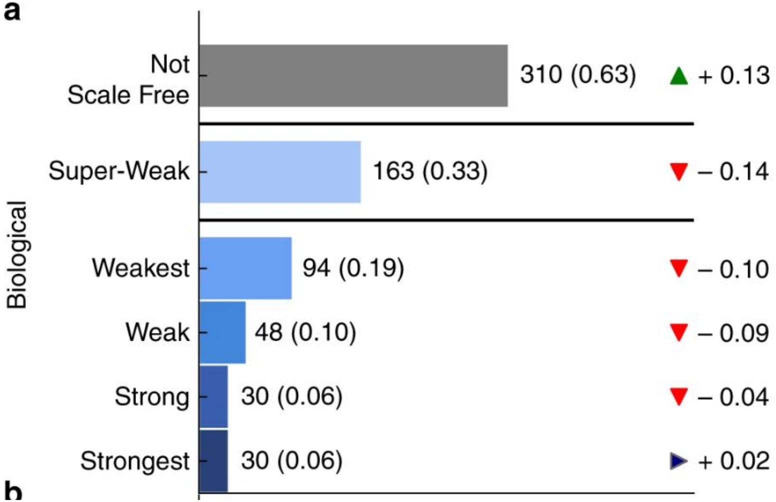
Article | Open Access | Published: 04 March 2019

Scale-free networks are rare

Anna D. Broido & Aaron Clauset

Nature Communications **10**, Article number: 1017 (2019) | Cite this article

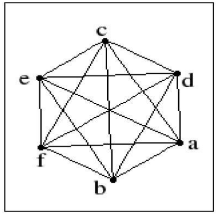
75k Accesses | **440** Citations | **577** Altmetric | Metrics



- ➢ In practice, when inferring molecular networks (especially co-expression networks), researchers often aim to obtain scale-free networks! You'll do it yourself in the WGCNA hands-on!

Metrics in network analysis: **Density**

- The **density** of a network is the **ratio of the number of edges** with respect to the **maximum possible number of edges** (i.e. the number of edges if the network was complete).



Graphe complet Kn

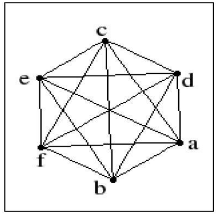
$$D(G) = \frac{|E|}{|E|_{Kn}}$$

$$D_G = 0 \Leftrightarrow |E| = 0$$
$$D_G = 1 \Leftrightarrow \mathrm{G} = \mathrm{Kn}$$

Source: Courtesy of Anaïs Baudot

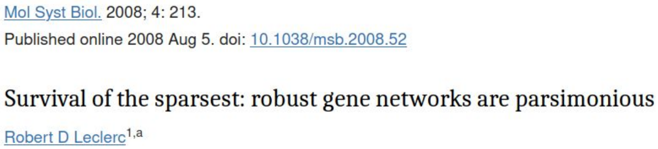
Metrics in network analysis: **Density**

- The **density** of a network is the **ratio of the number of edges** with respect to the **maximum possible number of edges** (i.e. the number of edges if the network was complete).



Graphe complet Kn

- Are biological networks dense or sparse?



Mol Syst Biol. 2008; 4: 213.
Published online 2008 Aug 5. doi: 10.1038/msb.2008.52

Survival of the sparsest: robust gene networks are parsimonious

Robert D Leclerc[1,a]

$$D(G) = \frac{|E|}{|E|_{Kn}}$$

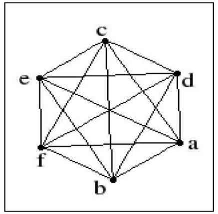
"**Robustness** implies a parsimonious network structure that is sparsely connected and not unnecessarily complex"

$$D_G = 0 \Leftrightarrow |E| = 0$$
$$D_G = 1 \Leftrightarrow \mathrm{G} = \mathrm{Kn}$$

Source: Courtesy of Anaïs Baudot

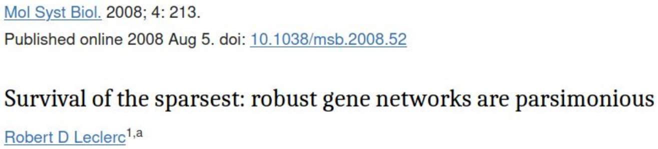
Metrics in network analysis: **Density**

- The **density** of a network is the **ratio of the number of edges** with respect to the **maximum possible number of edges** (i.e. the number of edges if the network was complete).



Graphe complet Kn

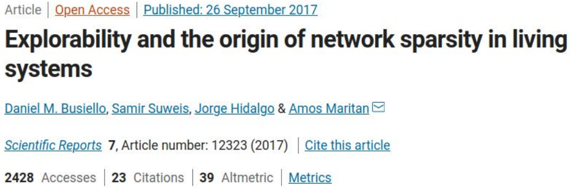
- Are biological networks dense or sparse?



Mol Syst Biol. 2008; 4: 213.
Published online 2008 Aug 5. doi: 10.1038/msb.2008.52

Survival of the sparsest: robust gene networks are parsimonious

Robert D Leclerc[1,a]



Article | Open Access | Published: 26 September 2017

Explorability and the origin of network sparsity in living systems

Daniel M. Busiello, Samir Suweis, Jorge Hidalgo & Amos Maritan

Scientific Reports 7, Article number: 12323 (2017) | Cite this article

2428 Accesses | **23** Citations | **39** Altmetric | Metrics

$$D(G) = \frac{|E|}{|E|_{Kn}}$$

“**Robustness** implies a parsimonious network structure that is sparsely connected and not unnecessarily complex”

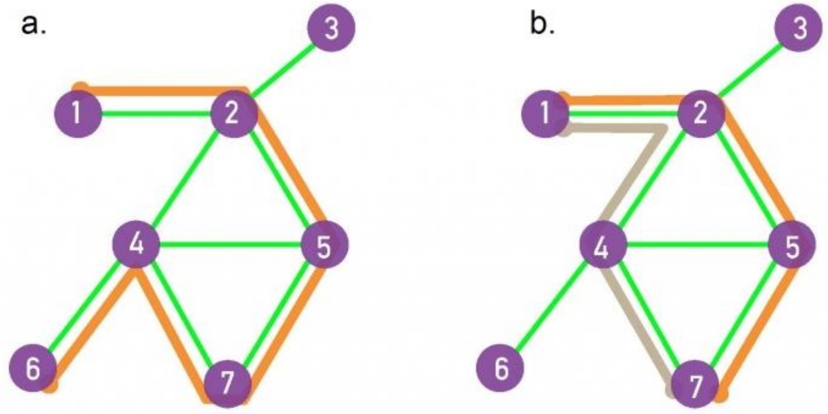
$$D_G = 0 \Leftrightarrow |E| = 0$$
$$D_G = 1 \Leftrightarrow G = Kn$$

“We show that sparsity is an emergent property resulting from optimising both **explorability and dynamical robustness**, i.e. the capacity of the system to remain stable after perturbations of the underlying dynamics”

Source: Courtesy of Anaïs Baudot

Metrics in network analysis: **Distances and paths**

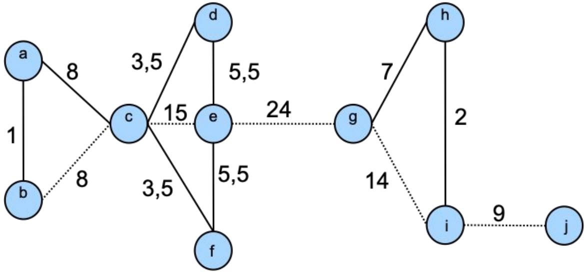
- A **path** is a **sequence of edges** which join a sequence of nodes
- The **distance** between two nodes in a network is the **length** (i.e. number of edges) of the **shortest path** connecting them.
- The **diameter** of a network is the **greatest distance between any pair of nodes** in the network.

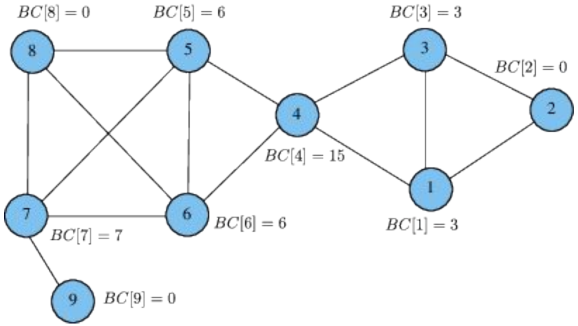


Source: Courtesy of Anaïs Baudot

Metrics in network analysis: **Betweenness**

- The **edge betweenness** is the **number of shortest path** running through an edge
- The **node betweenness** is the **number of shortest path** running through a node
- High edge/node betweenness -> "**bottleneck**"





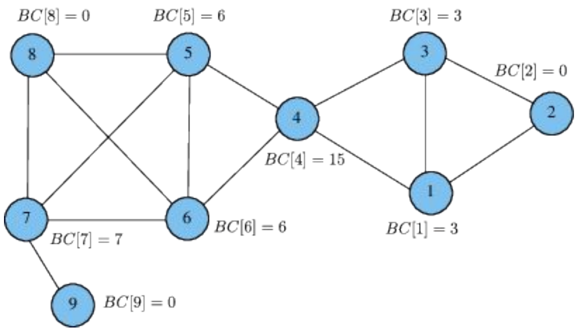
Source: Courtesy of Anaïs Baudot

Metrics in network analysis: **Betweenness**

The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics

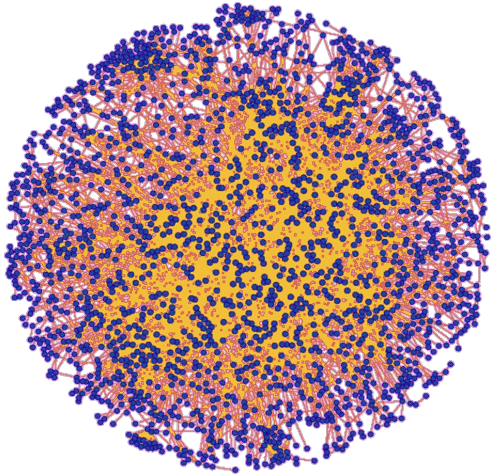
Haiyuan Yu, Philip M Kim, Emmett Sprecher, Valery Trifonov, Mark Gerstein

Published: April 20, 2007 • https://doi.org/10.1371/journal.pcbi.0030059



- ➢ Bottleneck proteins are **more likely to be essential proteins.**
- ➢ Bottleneck-ness (betweenness) is a much more **significant indicator of essentiality** than hub-ness (degree).
- ➢ Bottleneck proteins are significantly less well coexpressed with their neighbors -> **the network's topology is intricately linked with the dynamics of gene expression.**

Outline



Source: https://www3.nd.edu/~tmilenko/research.html.

General Introduction, Key Concepts

Network Biology, Network Science, Complex Systems, …

Network Construction/Inference

Data sources, Distances/Similarities, Correlation, …

Network Analysis

Measures

Degree, Centrality, Distances, …

Algorithms

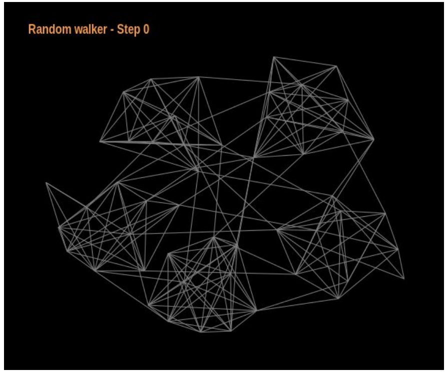
Diffusion, Clustering, Embedding, …

Networks and Integration

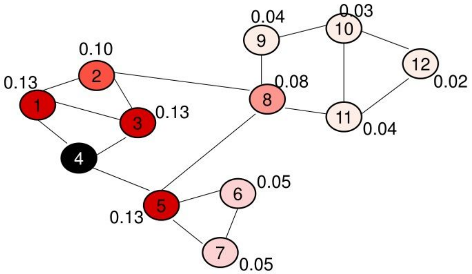
Multiplex Networks, Multilayer Networks, Knowledge Graphs, …

Algorithms for network analysis: **Random Walk with Restart**

- State of the art “**guilt-by-association**” approach
- Similarity between a **seed node** and all other nodes in the network
- Used for node prioritisation, network clustering, network embedding, etc...



Source: “Script to Animate the Path of a Random Walk across a Graph - A Walker Starts at a Node and Takes Random Steps through the Graph, Tending to Get ‘Stuck’ in Dense Subgraphs.” Gist. Accessed August 29, 2023 https://gist.github.com/clairemcwhite/7fb348acca2c84c464d751ba38ce72e1.



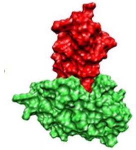
Nearby nodes, higher scores
More red, more relevant

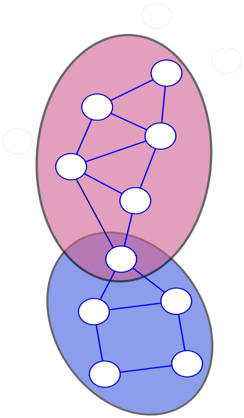
	Node 4
Node 1	0.13
Node 2	0.10
Node 3	0.13
Node 4	0.22
Node 5	0.13
Node 6	0.05
Node 7	0.05
Node 8	0.08
Node 9	0.04
Node 10	0.03
Node 11	0.04
Node 12	0.02

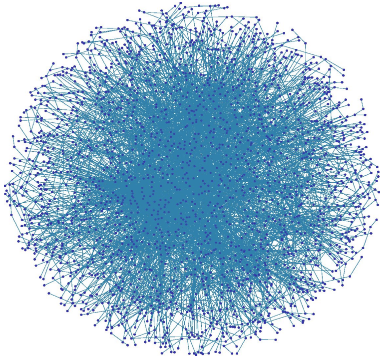
Ranking vector
$\vec{r}_4$

Source: Fast Random Walk with Restart and Its Applications. Hanghang Tong, Christos Faloutsos and Jia-Yu (Tim) Pan. ICDM 2006 Dec. 18-22, HongKong

Algorithms for network analysis: **Community detection**







Binary interaction

Functional modules

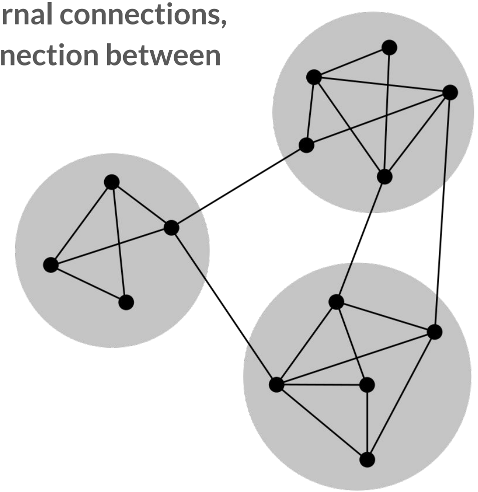
Interaction Networks

Inspired from P. Aloy, ECCB 2014
Courtesy of Anaïs Baudot

Algorithms for network analysis: **Community detection**

- Community detection = **clustering** for networks!
- Find groups of nodes (**communities/modules/clusters**) that are **more similar to each other than to the other nodes**

Dense internal connections, sparse connection between groups



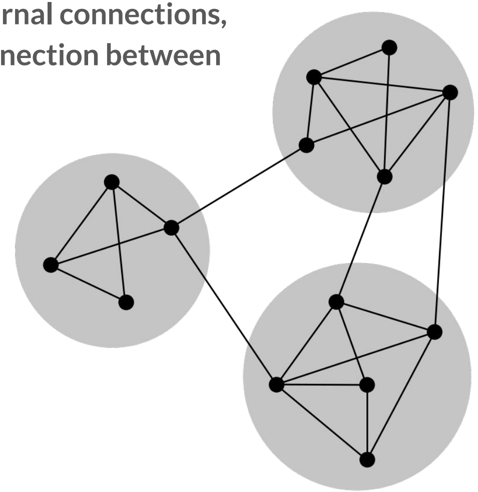
Source: “Community Structure.” In Wikipedia, August 26, 2023.

Algorithms for network analysis: **Community detection**

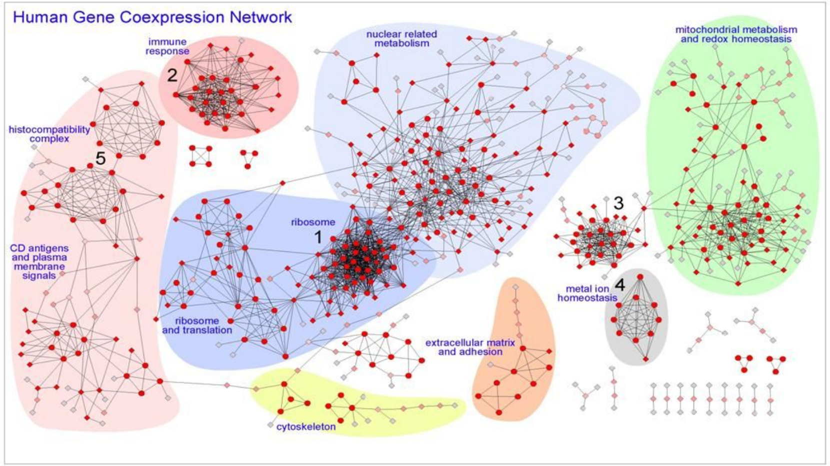
- Community detection = **clustering** for networks!
- Find groups of nodes (**communities/modules/clusters**) that are **more similar to each other than to the other nodes**

Why performing Community detection in biological networks ?

Dense internal connections, sparse connection between groups



Source: “Community Structure.” In Wikipedia, August 26, 2023.

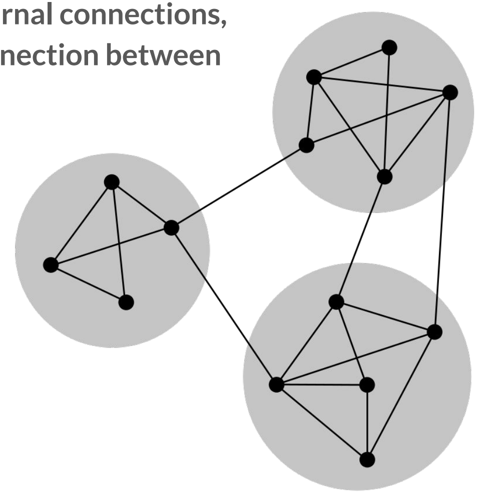


Source: “Human Gene Coexpression.” Accessed August 29, 2023.
http://bioinfow.dep.usal.es/pages/coexpression/index.html.

Algorithms for network analysis: **Community detection**

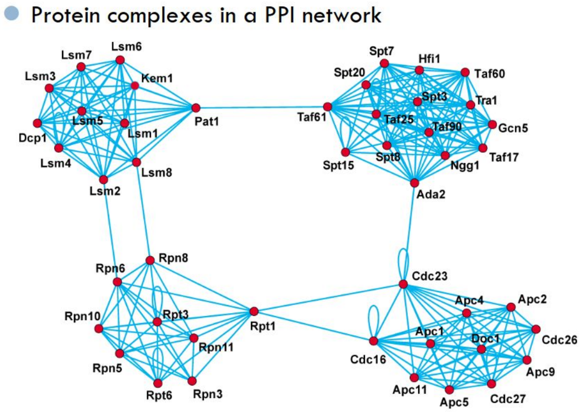
- Community detection = **clustering** for networks!
- Find groups of nodes (**communities/modules/clusters**) that are **more similar to each other than to the other nodes**

Dense internal connections, sparse connection between groups



Source: “Community Structure.” In Wikipedia, August 26, 2023.

Why performing Community detection in biological networks ?



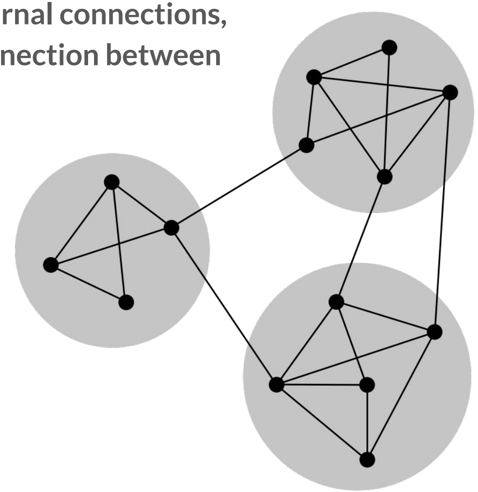
Source: Ngom, Dr Alioune. “NETWORK CLUSTERING METHODS,” n.d.

Algorithms for network analysis: **Community detection**

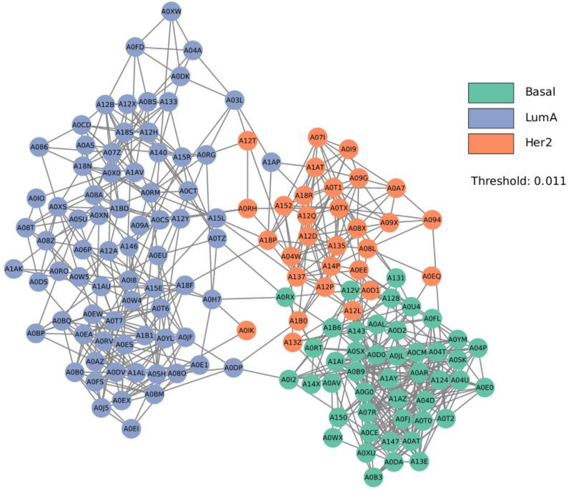
- ➢ Community detection = **clustering** for networks!
- ➢ Find groups of nodes (**communities/modules/clusters**) that are **more similar to each other than to the other nodes**

Why performing Community detection in biological networks ?

Dense internal connections, sparse connection between groups



Source: “Community Structure.” In Wikipedia, August 26, 2023.



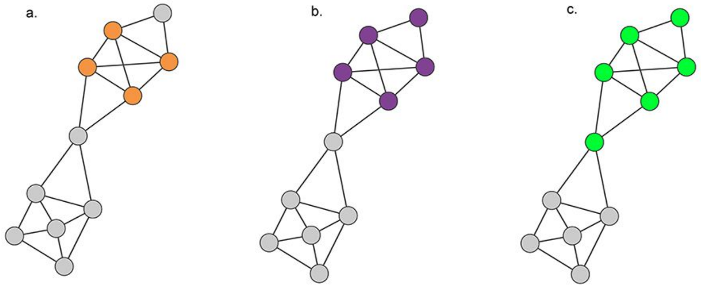
Source: Morgane Térézol - SNF tutorial

Algorithms for network analysis: **Community detection**

- **Clique-based methods** (e.g. CliquePercolation)

Cliques are **complete subgraphs.**

Clique **Strong community** **Weak community**

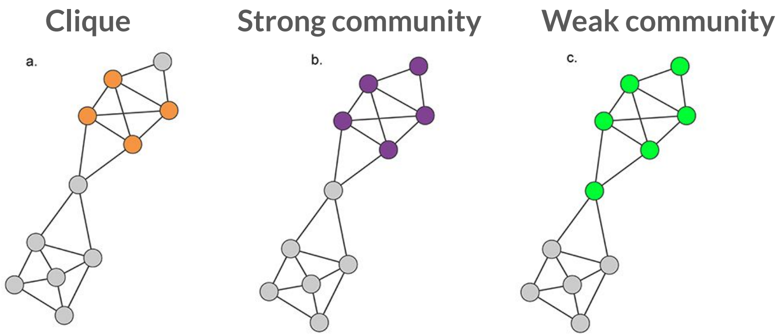


Source: Network Science by Albert-László Barabási. Accessed August 29, 2023. http://networksciencebook.com/.

Algorithms for network analysis: **Community detection**

- **Clique-based methods** (e.g. CliquePercolation)

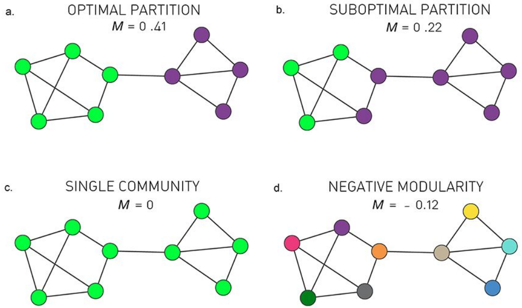
  Cliques are **complete subgraphs.**



Source: Network Science by Albert-László Barabási. Accessed August 29, 2023. http://networksciencebook.com/.

- **Modularity optimisation** (e.g. Louvain Algorithm)

  How much more **densely connected** the nodes within a community are, compared to how connected they would be in a random network.

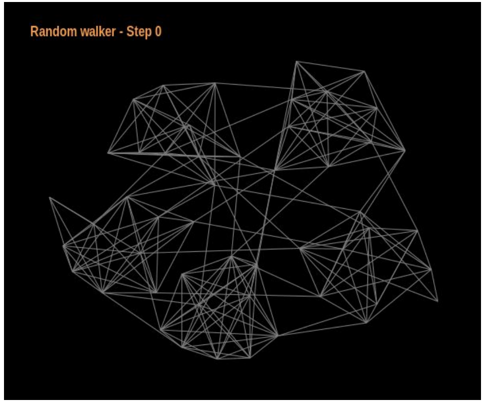


Source: Network Science by Albert-László Barabási. Accessed August 29, 2023. http://networksciencebook.com/.

Algorithms for network analysis: **Community detection**

- **Random-Walk based methods** (e.g. Walktrap)

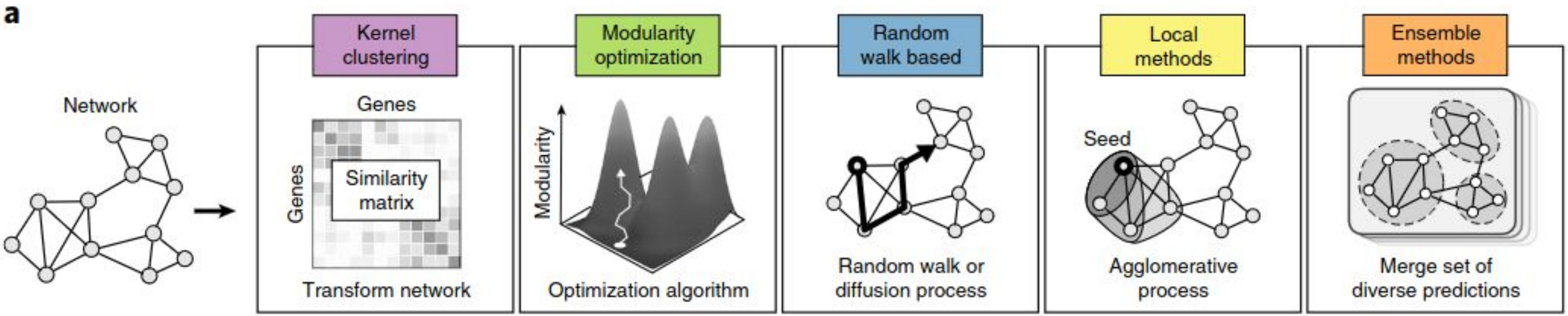
  Random walks on a graph tend to get **“trapped” into densely connected parts** corresponding to communities.

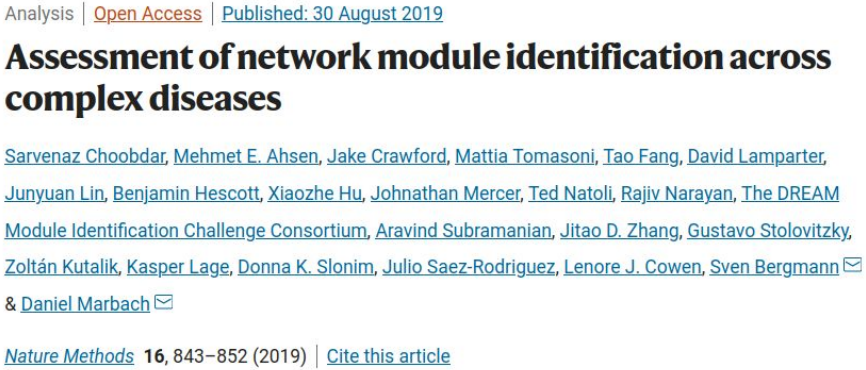


Source: “Script to Animate the Path of a Random Walk across a Graph - A Walker Starts at a Node and Takes Random Steps through the Graph, Tending to Get ‘Stuck’ in Dense Subgraphs.” Gist. Accessed August 29, 2023
https://gist.github.com/clairemcwhite/7fb348acca2c84c464d751ba38ce72e1.

- **and many other…**

Algorithms for network analysis: **Community detection**




Analysis | Open Access | Published: 30 August 2019

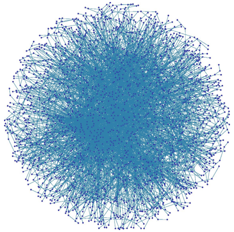
Assessment of network module identification across complex diseases

Sarvenaz Choobdar, Mehmet E. Ahsen, Jake Crawford, Mattia Tomasoni, Tao Fang, David Lamparter, Junyuan Lin, Benjamin Hescott, Xiaozhe Hu, Johnathan Mercer, Ted Natoli, Rajiv Narayan, The DREAM Module Identification Challenge Consortium, Aravind Subramanian, Jitao D. Zhang, Gustavo Stolovitzky, Zoltán Kutalik, Kasper Lage, Donna K. Slonim, Julio Saez-Rodriguez, Lenore J. Cowen, Sven Bergmann ✉ & Daniel Marbach ✉

Nature Methods **16**, 843–852 (2019) | Cite this article

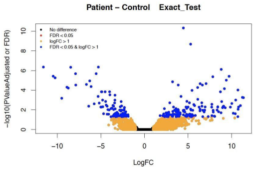
Algorithms for network analysis: **Active Modules**

Biological Network



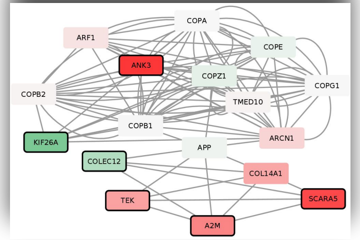
+

RNA-seq transcriptomics data



↓

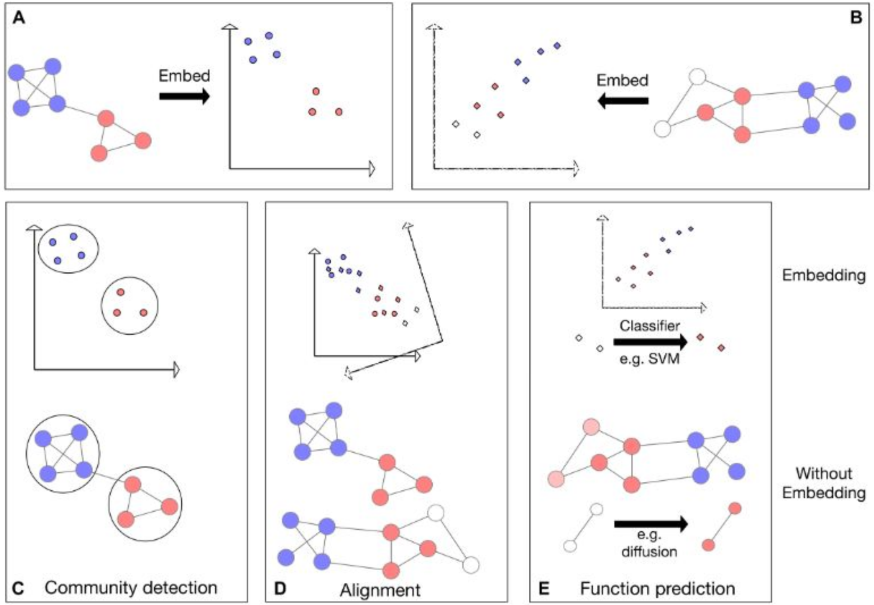
Find subnetworks of interest
aka “active modules”



Source: Courtesy of Anaïs Baudot

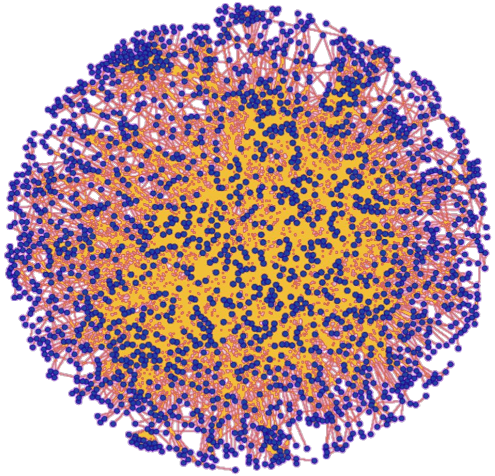
Algorithms for network analysis: **Network Embedding**

- Find a **low-dimensional representation of the nodes of a network**, while preserving the structural and relational information present in the original network.
    - Matrix factorisation, random-walk based methods, auto-encoders, …
    - A variety of downstream analysis
    - Easier analysis (low-dimensional, vector space)



Source: Nelson, Walter, et al. "To embed or not: network embedding as a paradigm in computational biology." Frontiers in genetics 10 (2019): 381.

Outline



Source: https://www3.nd.edu/~tmilenko/research.html.

General Introduction, Key Concepts

Network Biology, Network Science, Complex Systems, …

Network Construction/Inference

Data sources, Distances/Similarities, Correlation, …

Network Analysis

Measures

Degree, Centrality, Distances, …

Algorithms

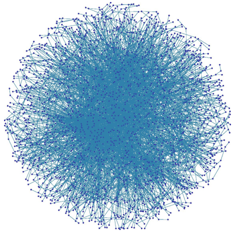
Diffusion, Clustering, Embedding, …

Networks and Integration

Multiplex Networks, Multilayer Networks, Knowledge Graphs, …

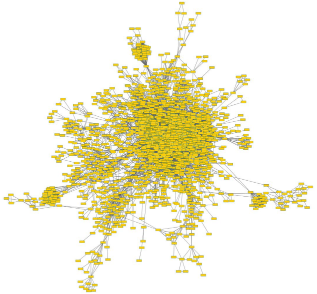
Many Biological Networks

PPI



~60 000 edges

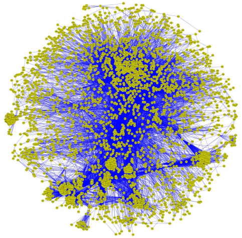
Complexes



~40 000 edges

Experimental networks

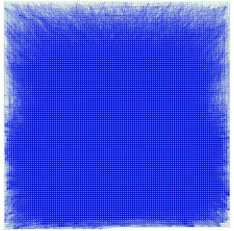
Pathways



~250 000 edges

Curated networks

Correlation of expression



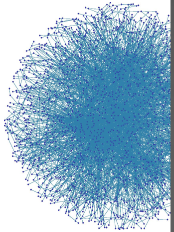
~1 400 000 edges

Inferred networks

Source: Courtesy of Anaïs Baudot

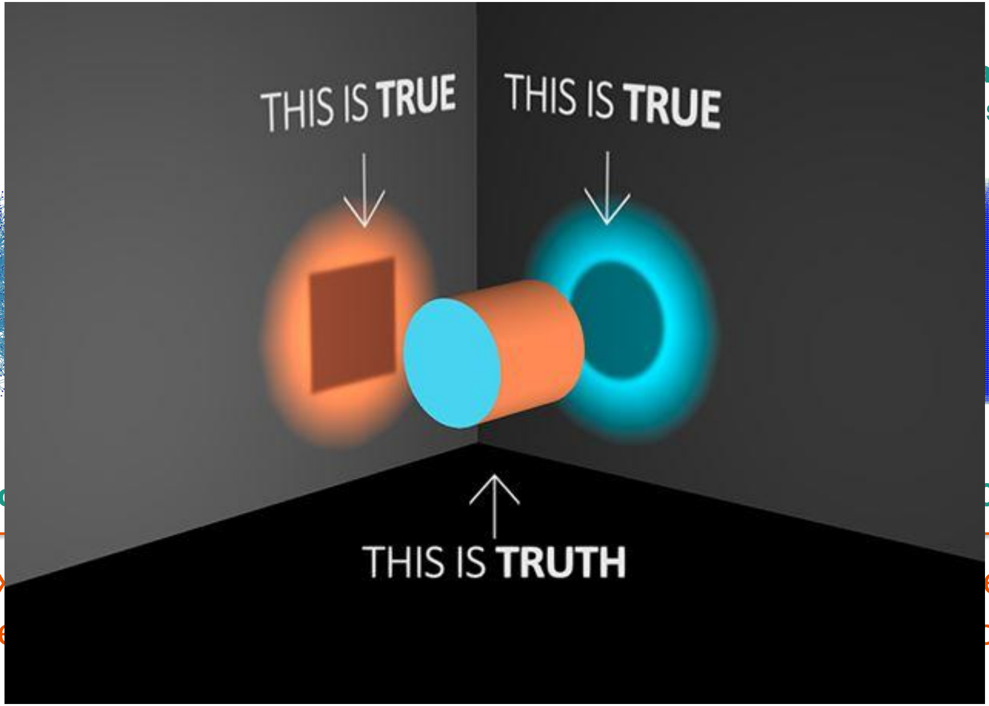
Many Biological Networks

PPI



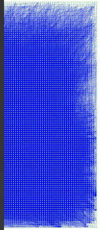
~60 000 ed

Ex
ne



Source: Courtesy of Anaïs Baudot

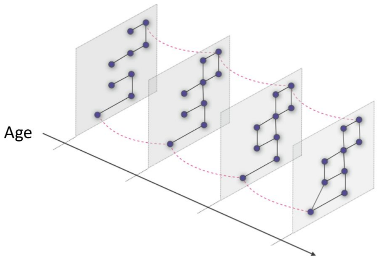
ation of
ssion



00 edges

ed
orks

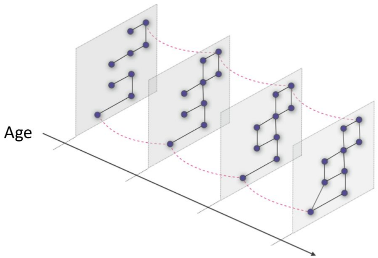
Networks can easily be **combined**

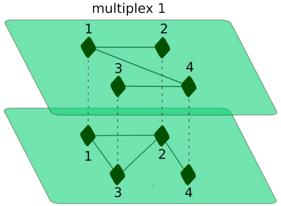


- **Temporal networks**
  - Same nodes
  - Same type of interactions
  - Various time-points

Source: Courtesy of Anaïs Baudot

Networks can easily be **combined**



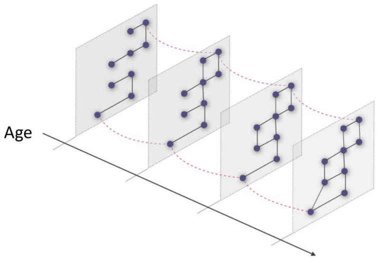


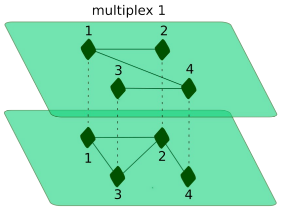
- ➢ **Multiplex networks**
  - ○ Same nodes
  - ○ Different types of interactions

- ➢ **Temporal networks**
  - ○ Same nodes
  - ○ Same type of interactions
  - ○ Various time-points

Source: Courtesy of Anaïs Baudot

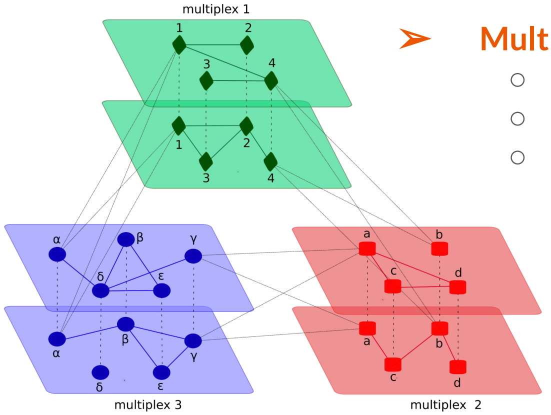
Networks can easily be **combined**





- ➢ **Multiplex networks**
  - ○ Same nodes
  - ○ Different types of interactions

- ➢ **Temporal networks**
  - ○ Same nodes
  - ○ Same type of interactions
  - ○ Various time-points

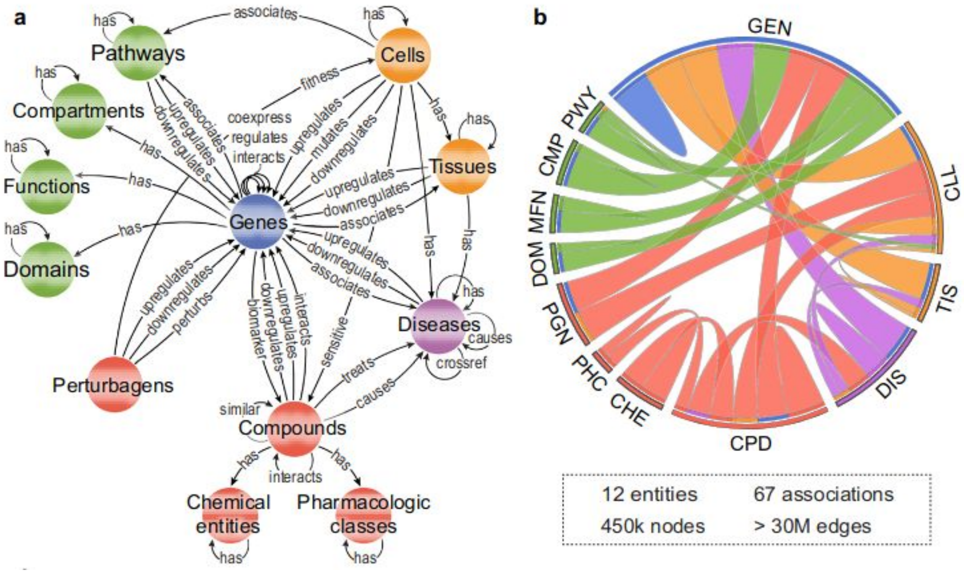


- ➢ **Multilayer networks**
  - ○ Various node types
  - ○ Different types of interactions
  - ○ Bipartite edges (across node types)

Source: Courtesy of Anaïs Baudot

Networks can easily be combined: **Knowledge Graphs**

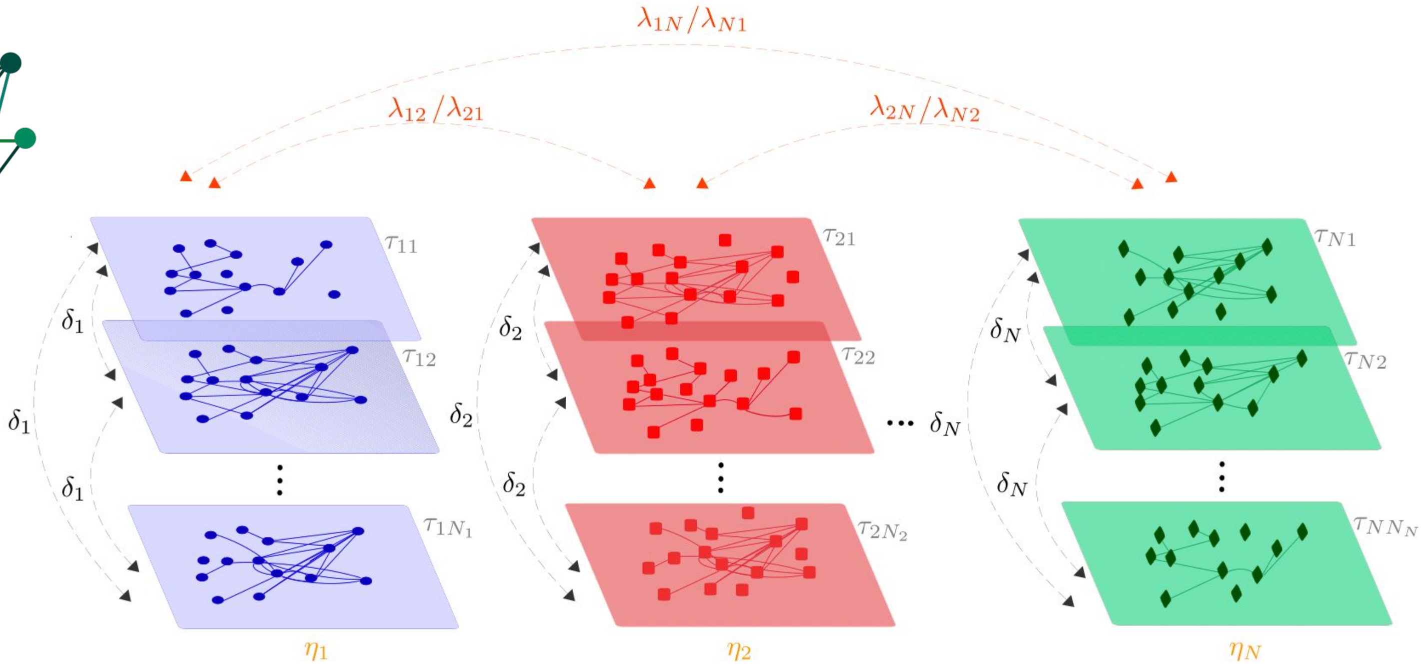
- **Knowledge graphs** … are basically networks, but with semantic information
  - Represent real-world **facts** as a set of entities and relationships



Source: Fernández-Torras, Adrià, Miquel Duran-Frigola, Martino Bertoni, Martina Locatelli, and Patrick Aloy. "Integrating and Formatting Biomedical Data as Pre-Calculated Knowledge Graph Embeddings in the Bioteque." Nature Communications 13, no. 1 (September 9, 2022): 5304. https://doi.org/10.1038/s41467-022-33026-0.

Random Walk with Restart for multilayer network

multiXrank



Source: Baptista, Anthony, Aitor Gonzalez, and Anaïs Baudot. "Universal Multilayer Network Exploration by Random Walk with Restart." Communications Physics 5, no. 1 (July 1, 2022): 1–9. https://doi.org/10.1038/s42005-022-00937-9.

Multiplex Community Detection

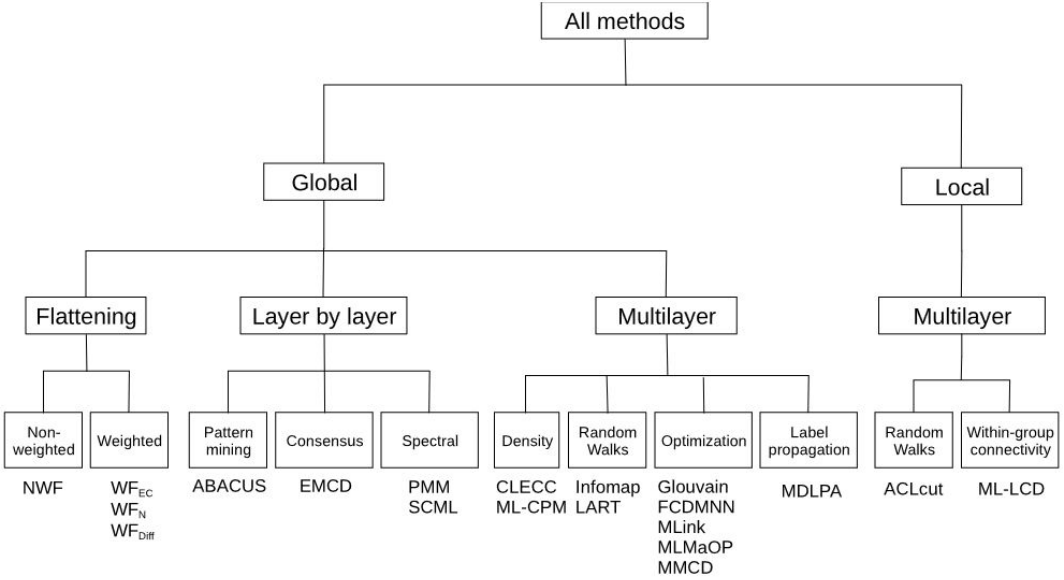
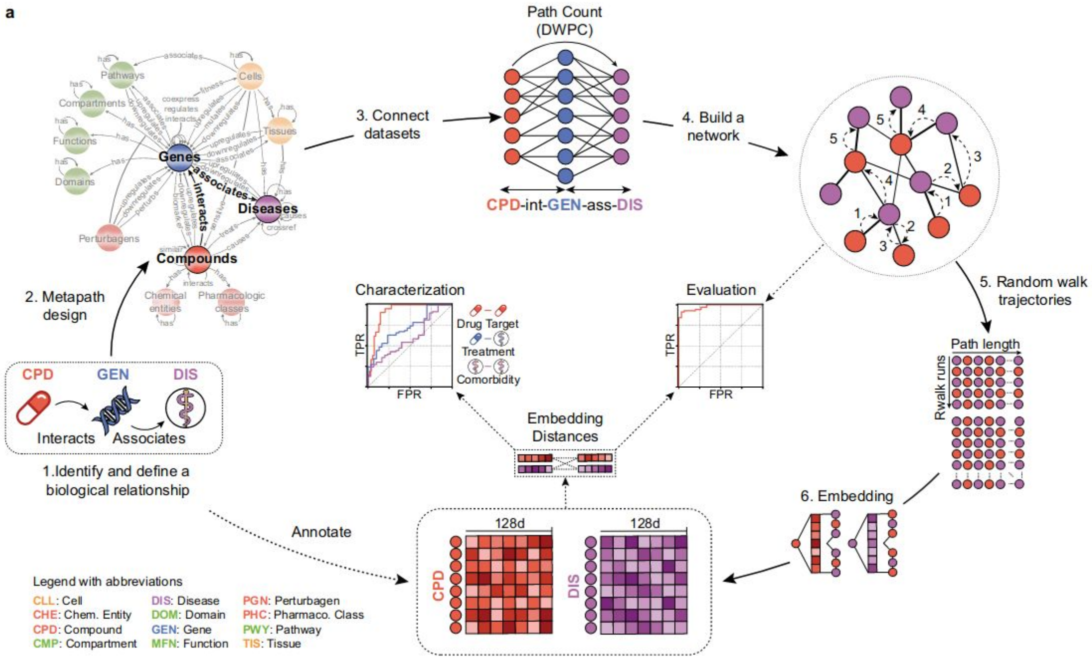


Figure 4: A taxonomy of multiplex community detection algorithms

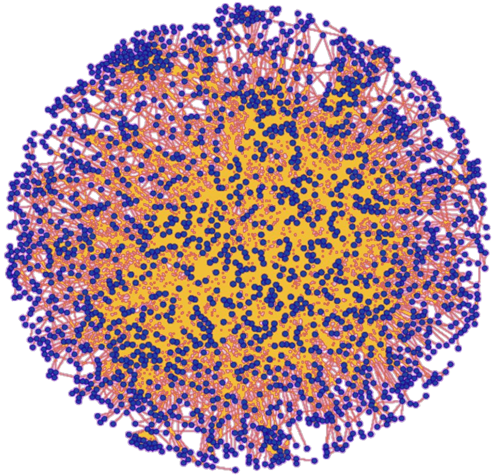
Source: Magnani, Matteo, Obaida Hanteer, Roberto Interdonato, Luca Rossi, and Andrea Tagarelli. "Community Detection in Multiplex Networks." ACM Computing Surveys 54, no. 3 (April 30, 2022): 1–35. https://doi.org/10.1145/3444688.

Embedding for Knowledge Graphs



Source: Fernández-Torras, Adrià, Miquel Duran-Frigola, Martino Bertoni, Martina Locatelli, and Patrick Aloy. "Integrating and Formatting Biomedical Data as Pre-Calculated Knowledge Graph Embeddings in the Bioteque." Nature Communications 13, no. 1 (September 9, 2022): 5304. https://doi.org/10.1038/s41467-022-33026-0.

Outline



Source: https://www3.nd.edu/~tmilenko/research.html.

General Introduction, Key Concepts

Network Biology, Network Science, Complex Systems, …

Network Construction/Inference

Data sources, Distances/Similarities, Correlation, …

Network Analysis

Measures

Degree, Centrality, Distances, …

Algorithms

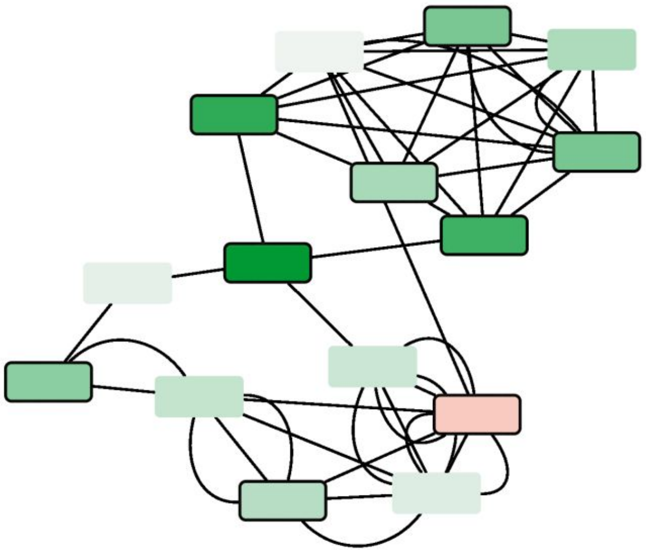
Diffusion, Clustering, Embedding, …

Networks and Integration

Multiplex Networks, Multilayer Networks, Knowledge Graphs, …

Conclusion

Networks are great

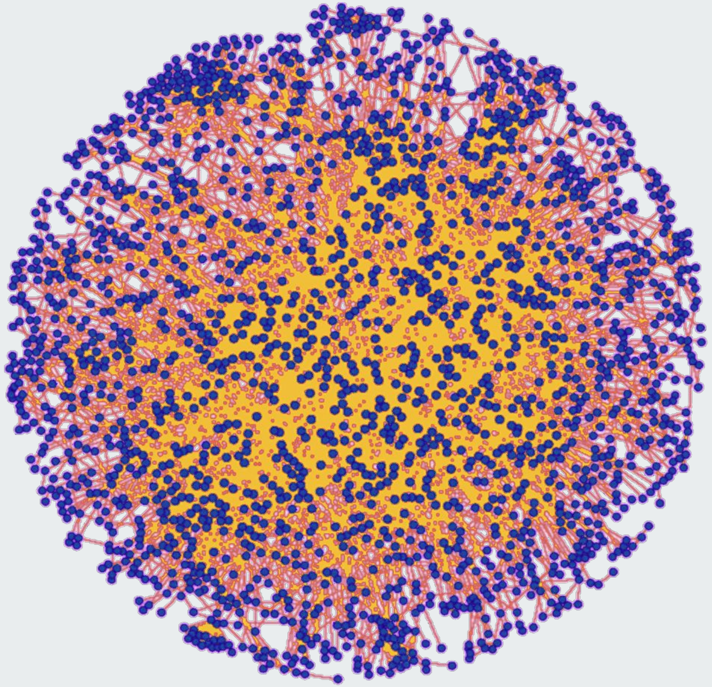


- Relationships
- Interpretable/Intuitive
- Scaffold for integration
  - heterogeneous data
  - previous knowledge
- Toolbox of graph theory

Source: Courtesy of Anaïs Baudot

Hands-on : WGCNA

It’s your turn now! Have fun!



Weighted Gene Correlation Network Analysis (WGCNA)

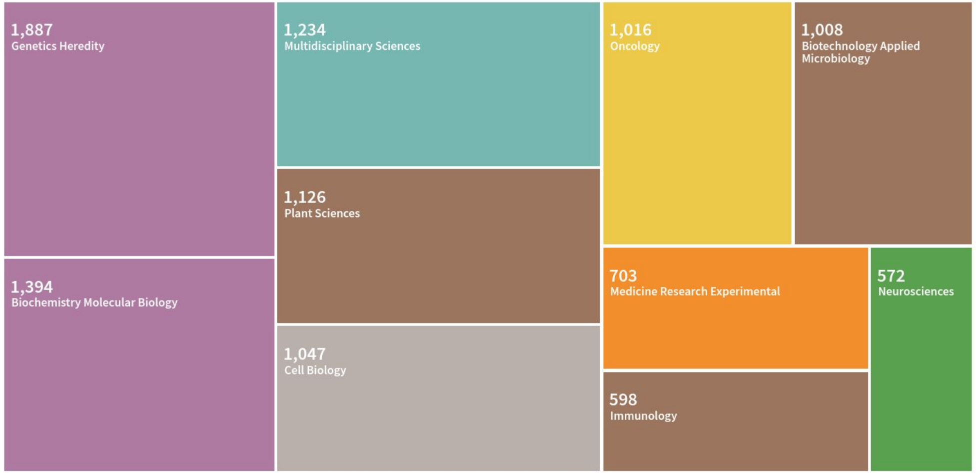
Software | Open Access | Published: 29 December 2008

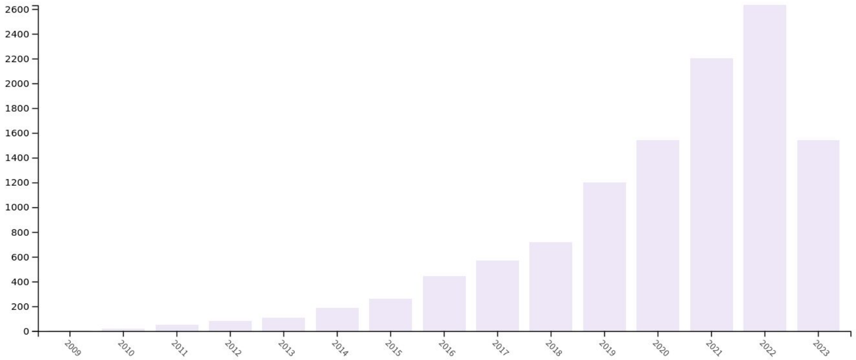
WGCNA: an R package for weighted correlation network analysis

Peter Langfelder & Steve Horvath ✉

BMC Bioinformatics **9**, Article number: 559 (2008) | Cite this article

385k Accesses | **11547** Citations | **88** Altmetric | Metrics

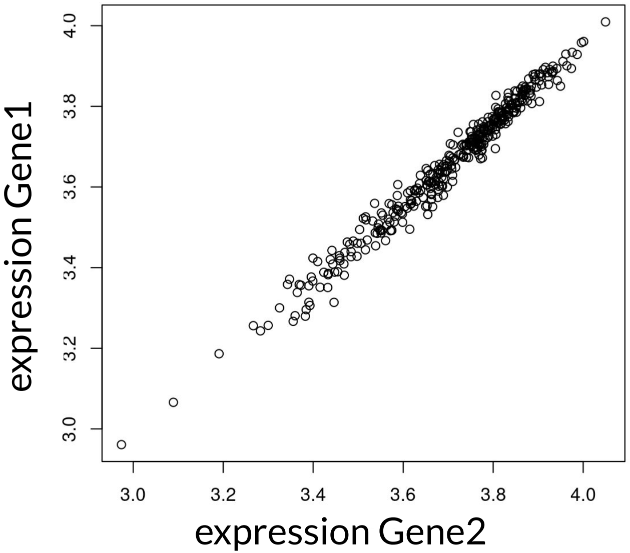




Source: www.webofscience.com

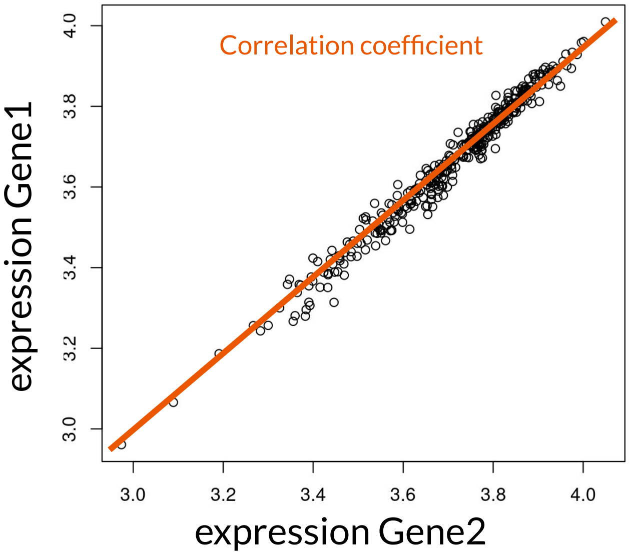
Weighted Gene Correlation Network Analysis (WGCNA)

➢ Gene co-expression



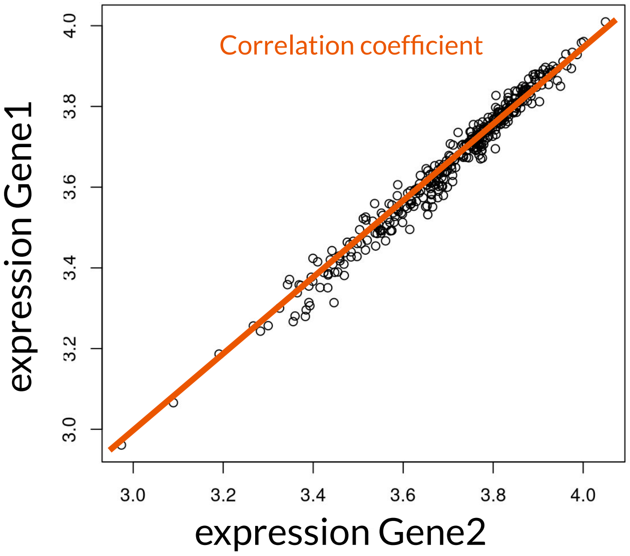
Weighted Gene Correlation Network Analysis (WGCNA)

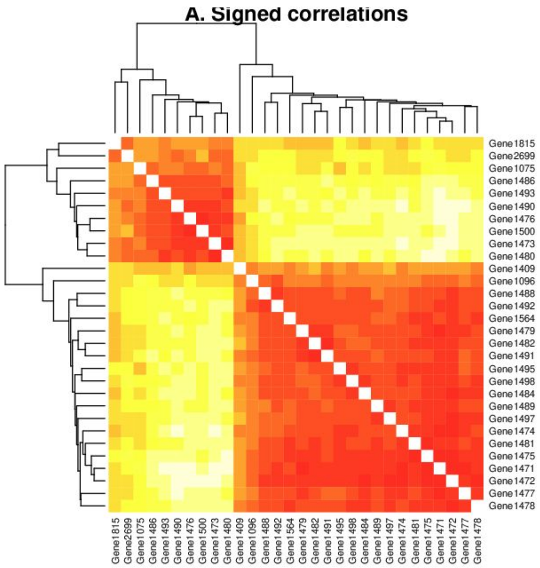
➢ Gene co-expression



Weighted Gene Correlation Network Analysis (WGCNA)

➢ Gene co-expression

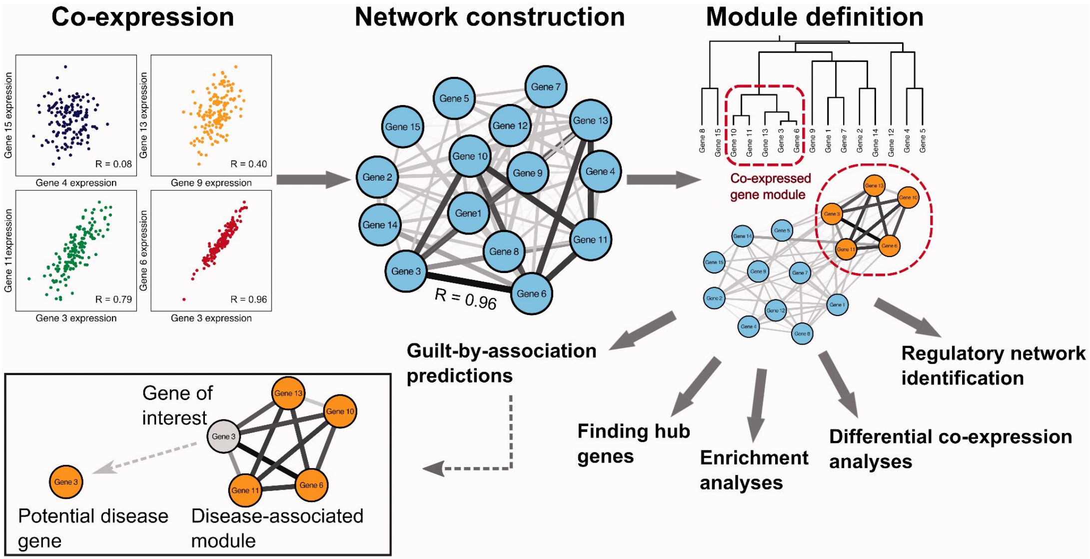




Source: Horvath, S., and P. Langfelder. "Tutorial for the WGCNA Package for R" 2011.

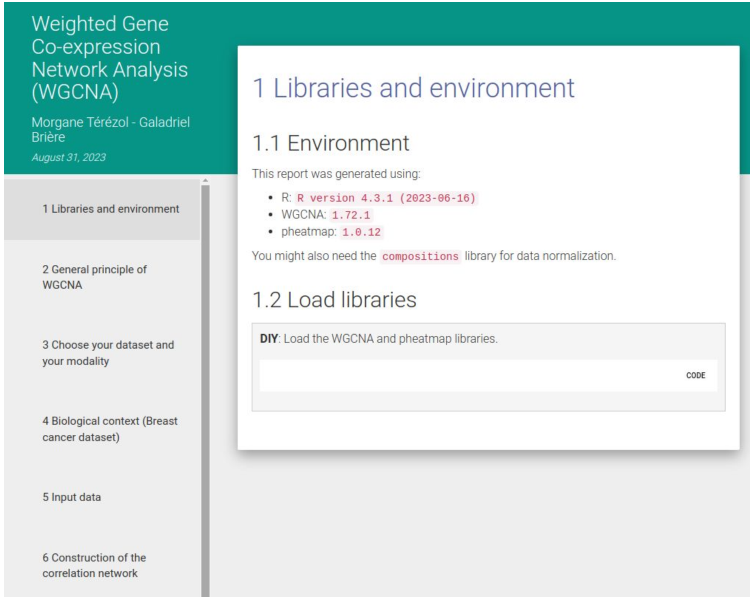
Weighted Gene Correlation Network Analysis (WGCNA)

- Co-expressed gene modules



Source: Dam, Sipko van, Urmo Võsa, Adriaan van der Graaf, Lude Franke, and João Pedro de Magalhães. "Gene Co-Expression Analysis for Functional Classification and Gene–Disease Predictions." Briefings in Bioinformatics 19, no. 4 (July 20, 2018): 575–92. https://doi.org/10.1093/bib/bbw139.

Hands-on



Weighted Gene Co-expression Network Analysis (WGCNA)

Morgane Térézol - Galadriel Brière

August 31, 2023

1 Libraries and environment

1.1 Environment

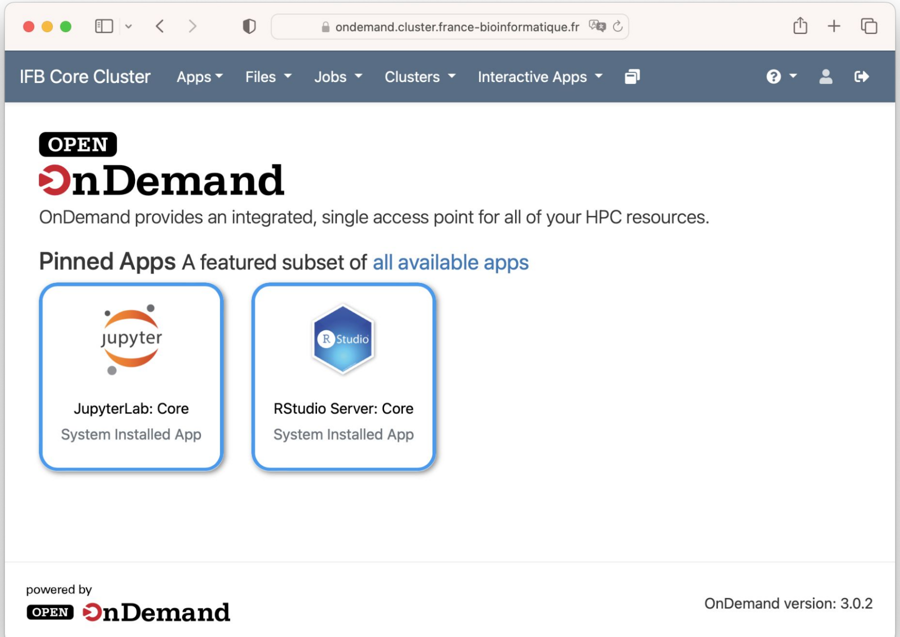
This report was generated using:

- R: `R version 4.3.1 (2023-06-16)`
- WGCNA: `1.72.1`
- pheatmap: `1.0.12`

You might also need the `compositions` library for data normalization.

1.2 Load libraries

DIY: Load the WGCNA and pheatmap libraries.



OPEN OnDemand

OnDemand provides an integrated, single access point for all of your HPC resources.

Pinned Apps A featured subset of all available apps

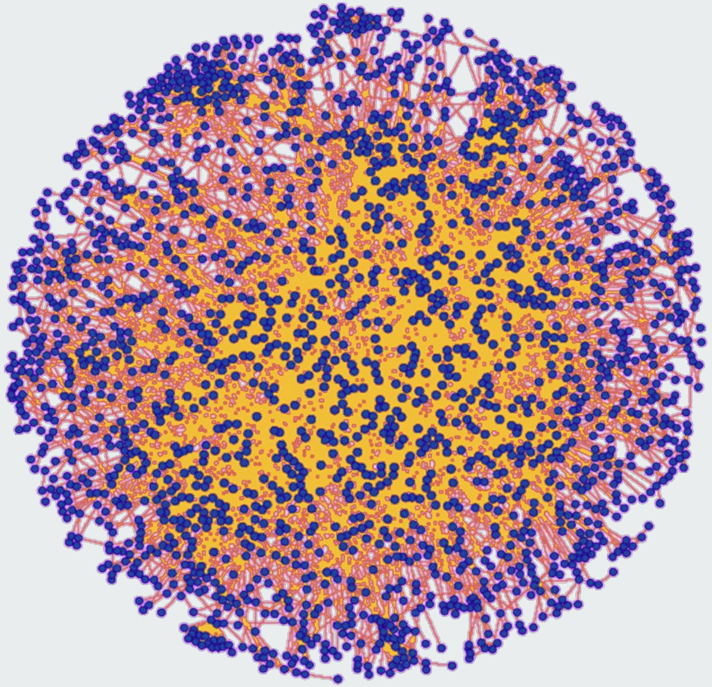
JupyterLab: Core
System Installed App

RStudio Server: Core
System Installed App

powered by OPEN OnDemand

OnDemand version: 3.0.2

Hands-on : SNF



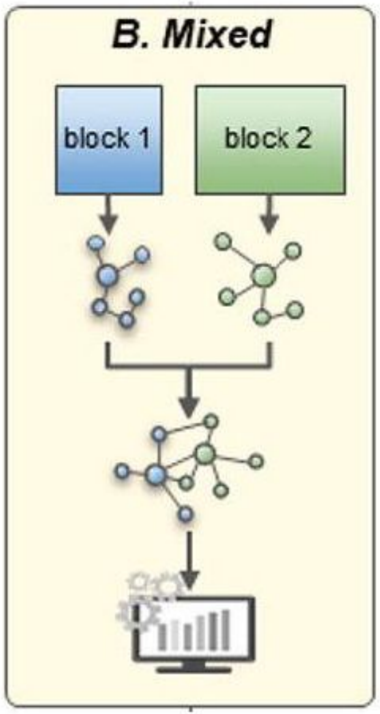
Galadriel Brière, Morgane Térézol, Benjamin Loire,
Anaïs Baudot, Samuel Chaffron

Similarity Network Fusion (SNF)

- Bo Wang *et al.*, Nature Methods. 2014

Similarity Network Fusion (SNF)

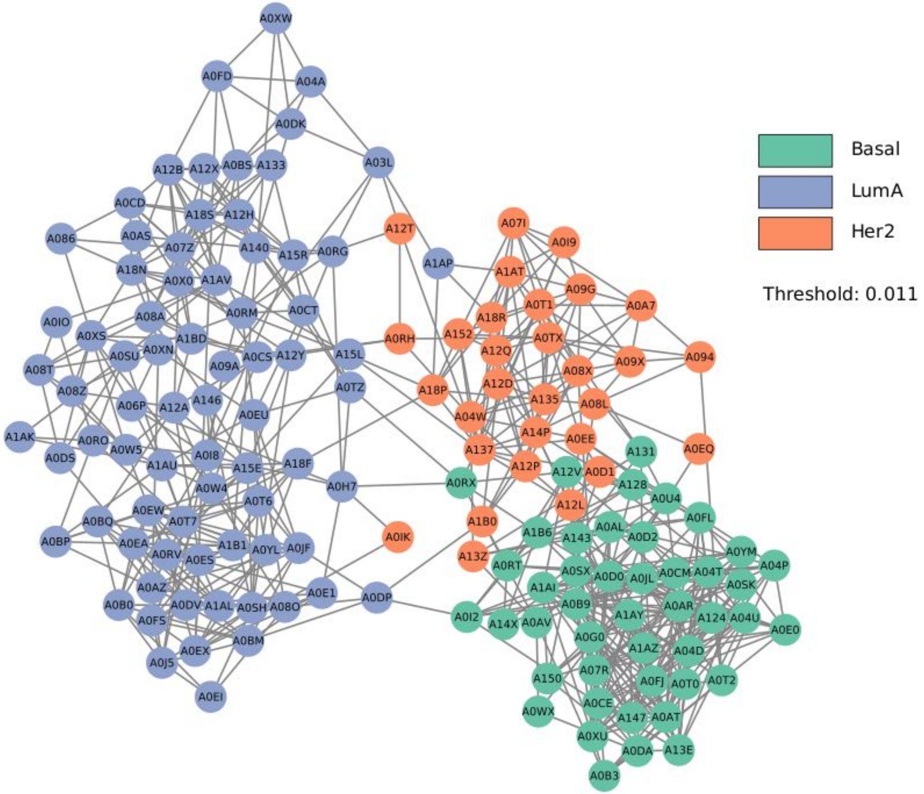
- Bo Wang *et al.*, Nature Methods. 2014
- **Mixed integration** method:
  - Sample **network** creation of each **data type**
  - **Fusion** of each data type networks



Picard M. et al., Comput Struct Biotechnol J. 2021

Similarity Network Fusion (SNF)

- Bo Wang *et al.*, Nature Methods. 2014
- **Mixed integration** method:
  - Sample **network** creation of each **data type**
  - **Fusion** of each data type networks
- **Sample** network



Similarity Network Fusion (SNF)

- Bo Wang *et al.*, Nature Methods. 2014
- **Mixed integration** method:
  - Sample **network** creation of each **data type**
  - **Fusion** of each data type networks
- **Sample** network
- Can manage:
  - **small number** of samples
  - noise
  - data **heterogeneity**
  - large number of features

Similarity Network Fusion (SNF)

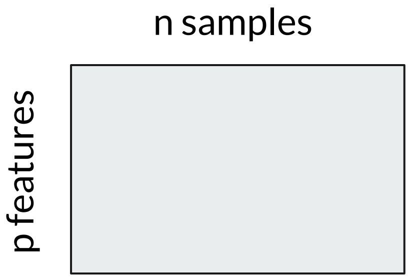
- Bo Wang *et al.*, Nature Methods. 2014
- **Mixed integration** method:
    - Sample **network** creation of each **data type**
    - **Fusion** of each data type networks
- **Sample** network
- Can manage:
    - **small number** of samples
    - noise
    - data **heterogeneity**
    - large number of features
- Apply **measures** and **algorithms** for network on the fused network

Similarity Network Fusion (SNF): Data preparation

- The most **important** step

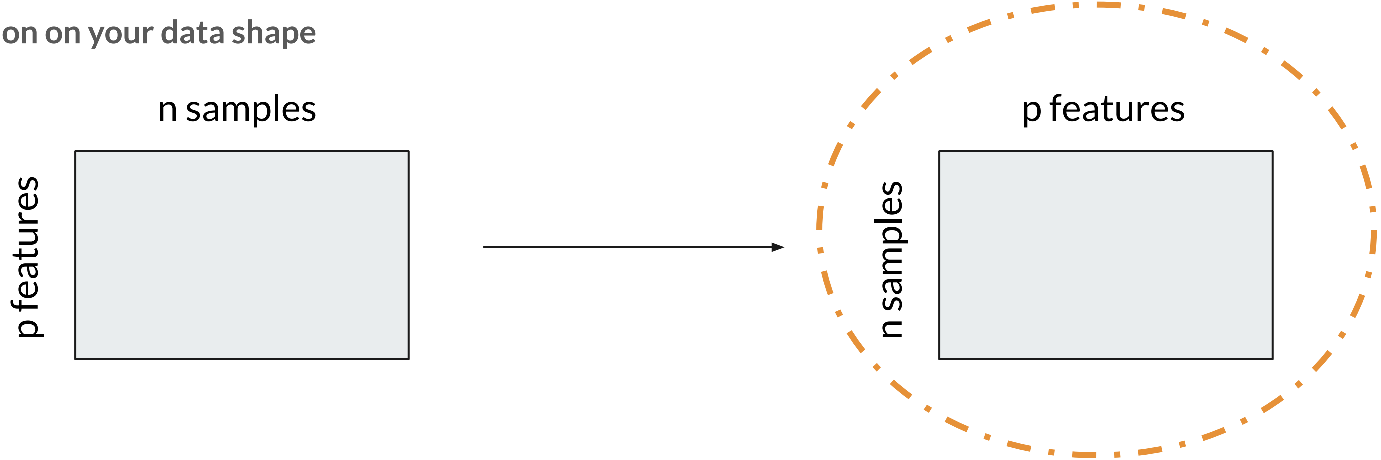
Similarity Network Fusion (SNF): Data preparation

- ➢ The most **important** step
- ➢ **Pay attention on your data shape**



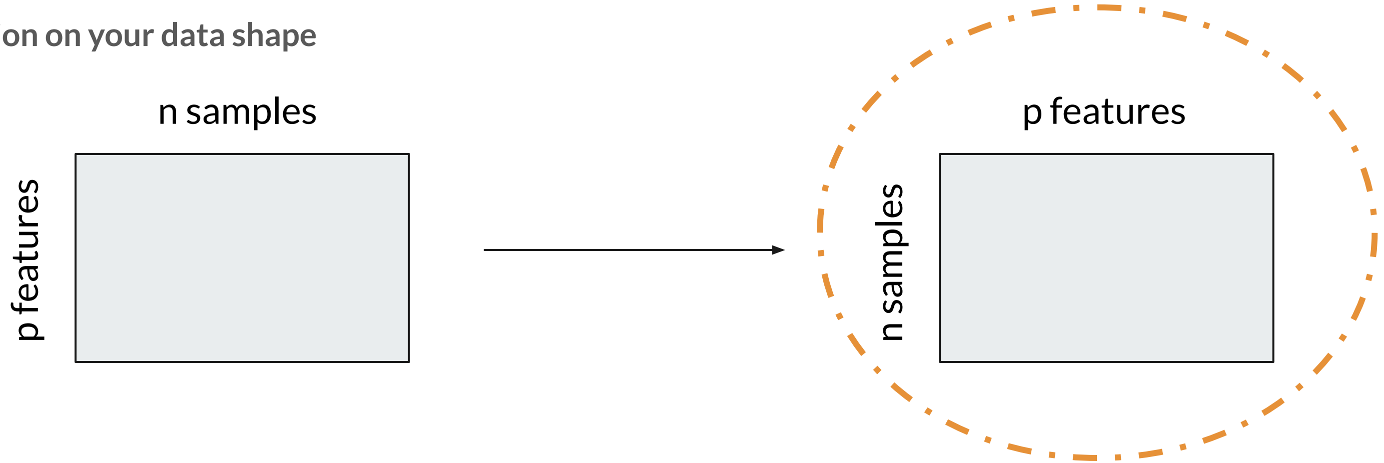
Similarity Network Fusion (SNF): Data preparation

- ➢ The most **important** step
- ➢ **Pay attention on your data shape**



Similarity Network Fusion (SNF): Data preparation

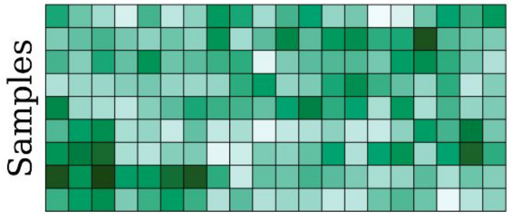
- ➢ The most **important** step
- ➢ **Pay attention on your data shape**



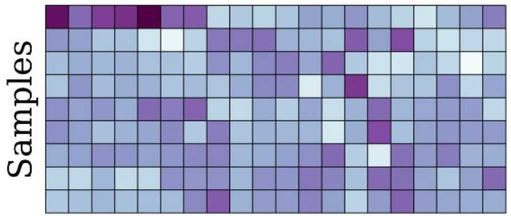
- ➢ **Normalization**: data should be normalized according to their type

Similarity Network Fusion (SNF): Network creation

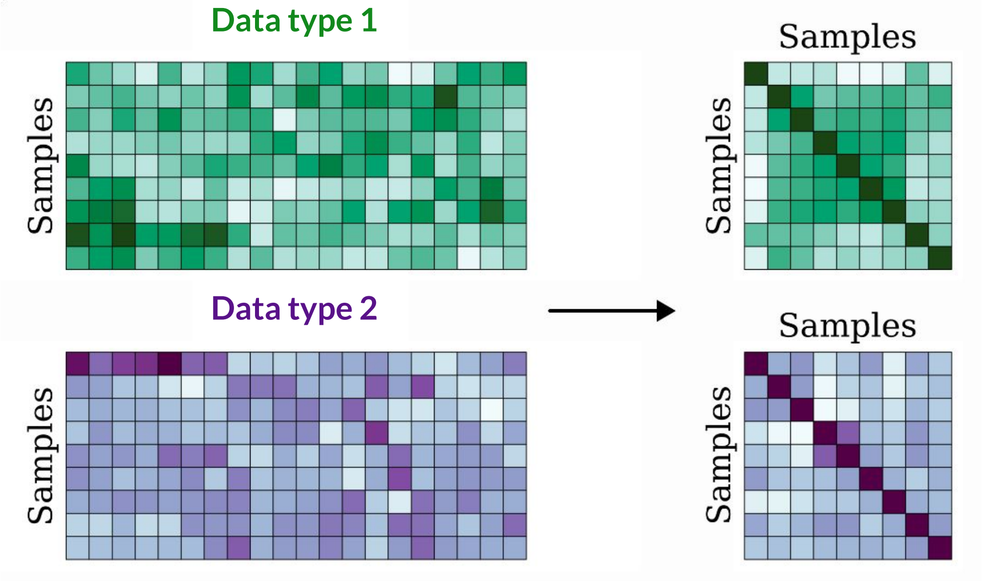
Data type 1



Data type 2

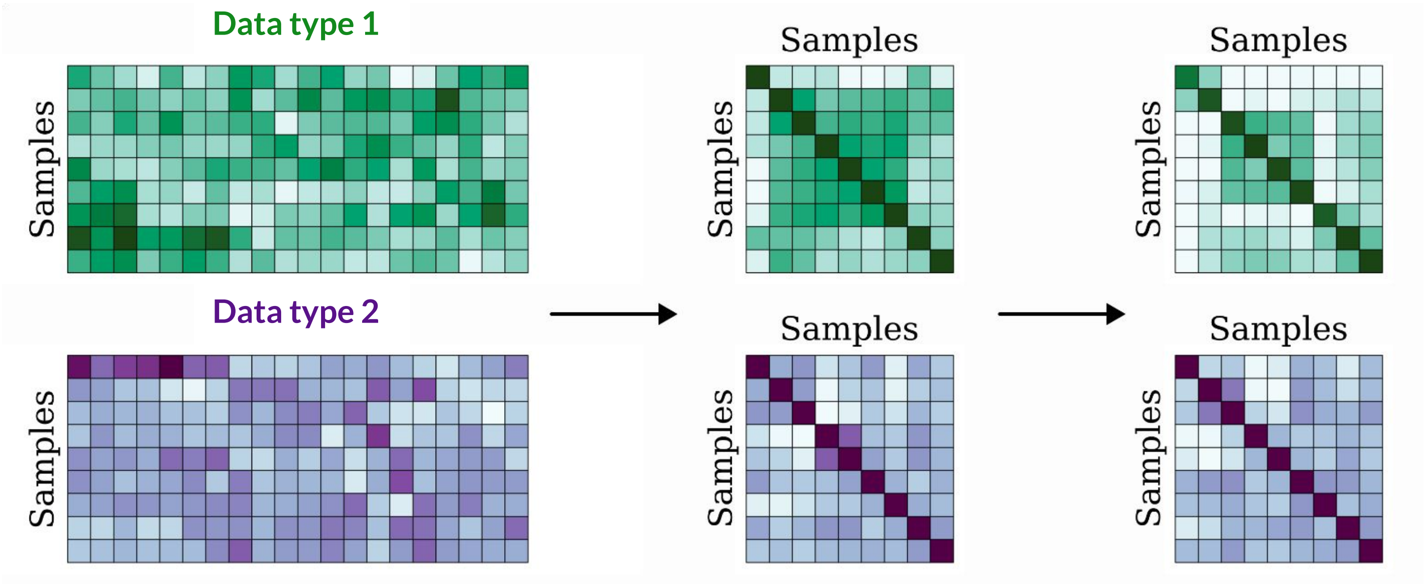


Similarity Network Fusion (SNF): Network creation



D
Distance matrix

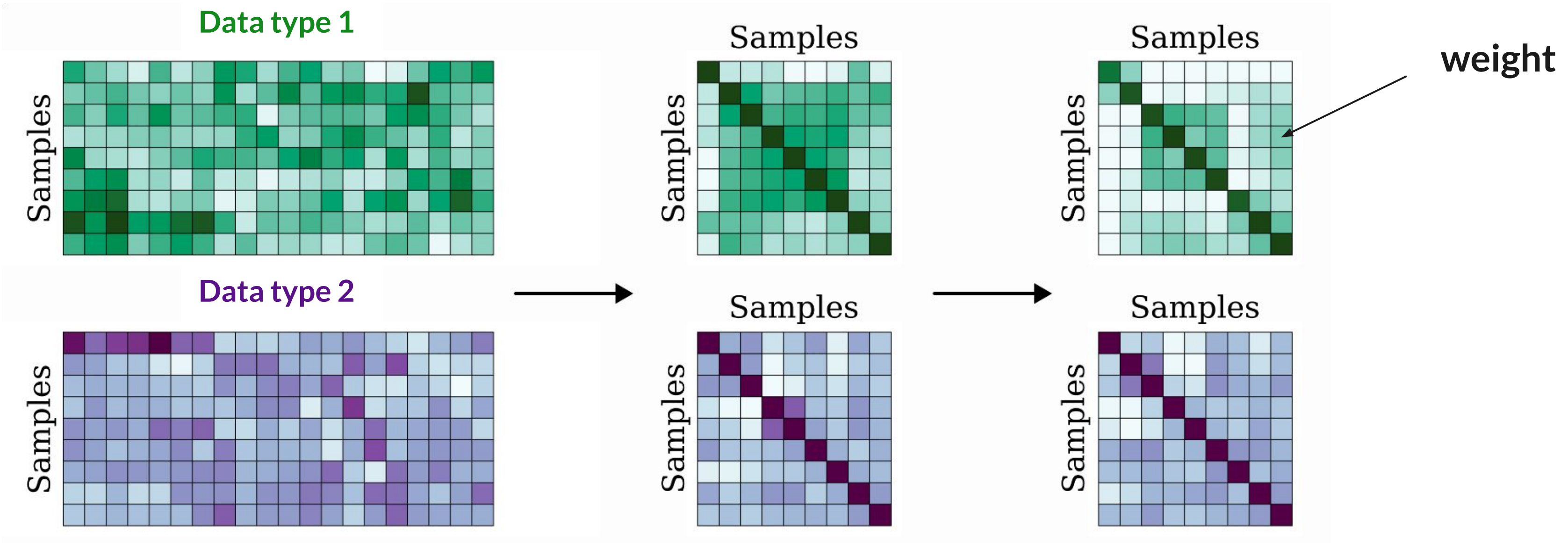
Similarity Network Fusion (SNF): Network creation



D
Distance matrix

W
Similarity matrix

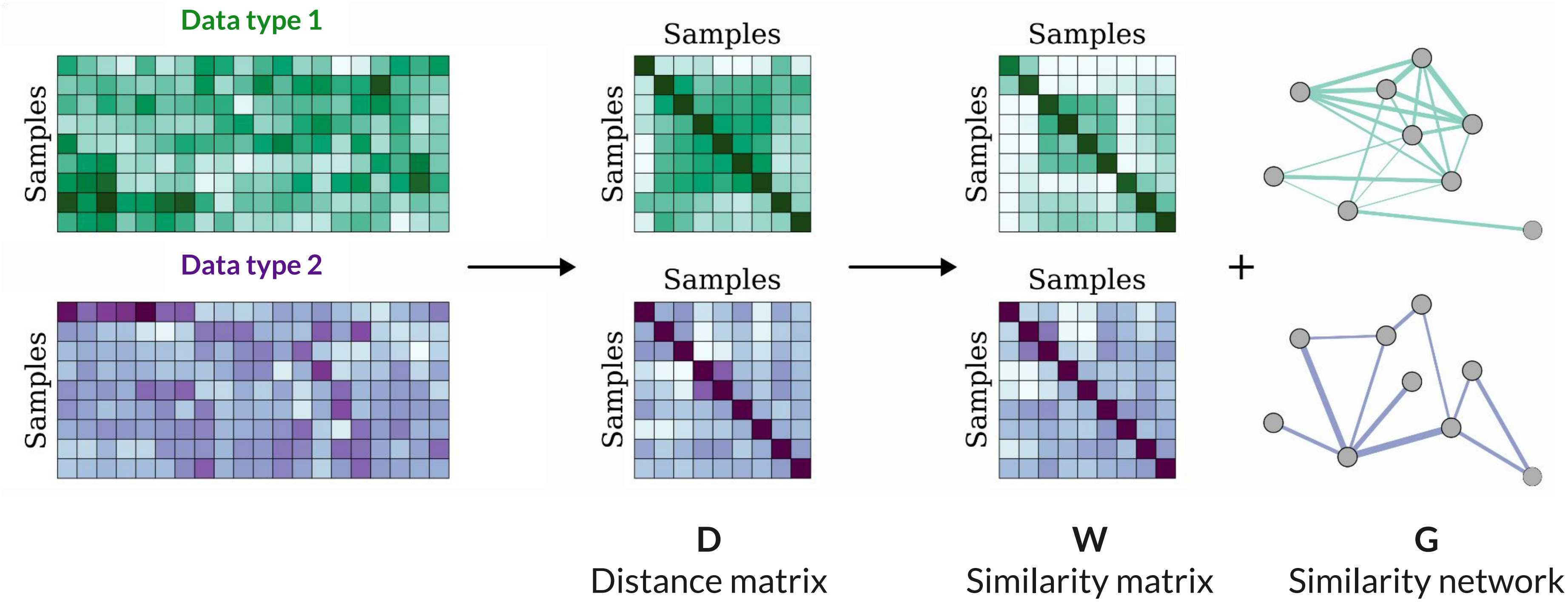
Similarity Network Fusion (SNF): Network creation



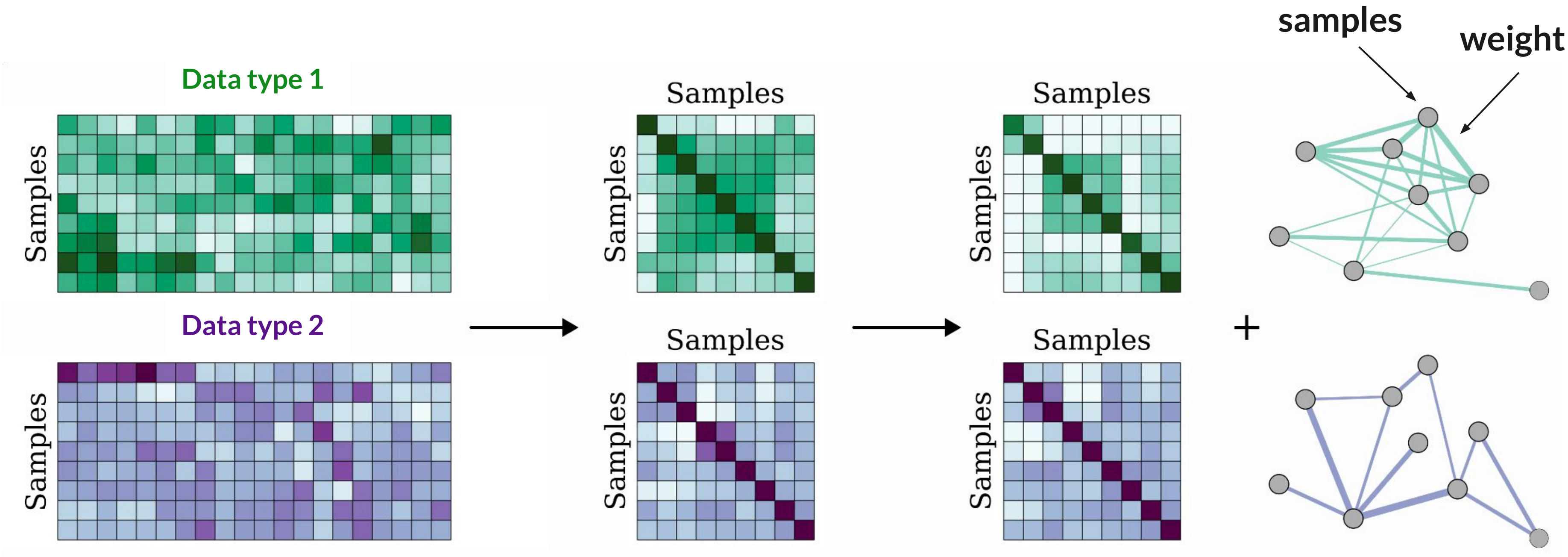
D
Distance matrix

W
Similarity matrix

Similarity Network Fusion (SNF): Network creation



Similarity Network Fusion (SNF): Network creation

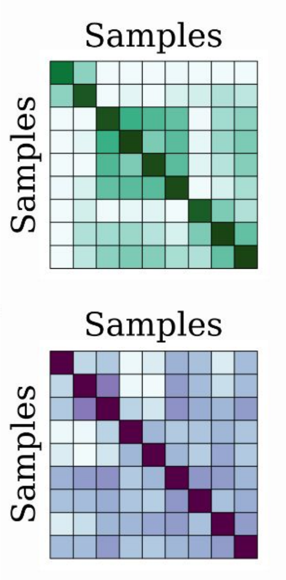


D
Distance matrix

W
Similarity matrix

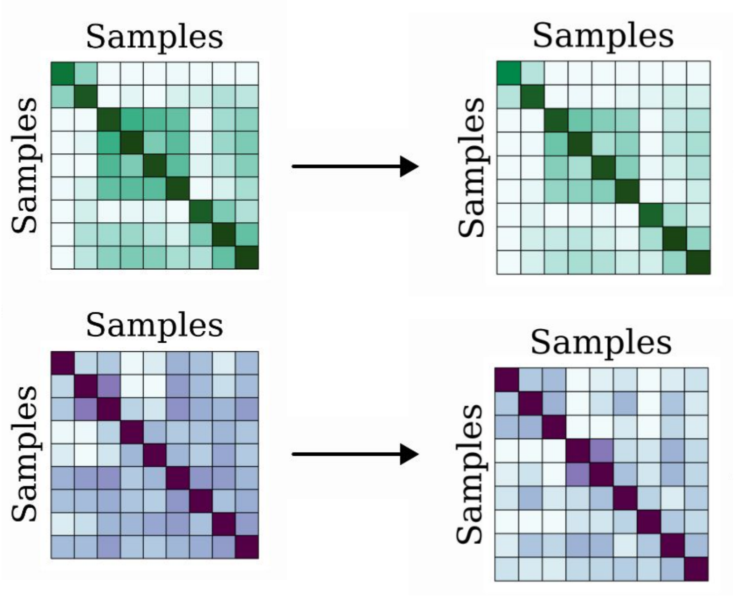
G
Similarity network

Similarity Network Fusion (SNF): Fusion



W
Similarity matrix

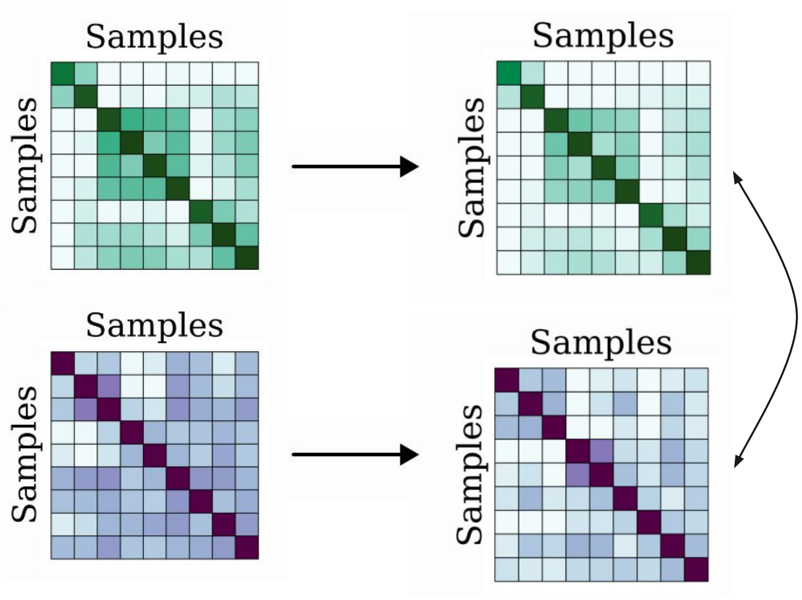
Similarity Network Fusion (SNF): Fusion



W
Similarity matrix

P(t)
Status matrix

Similarity Network Fusion (SNF): Fusion

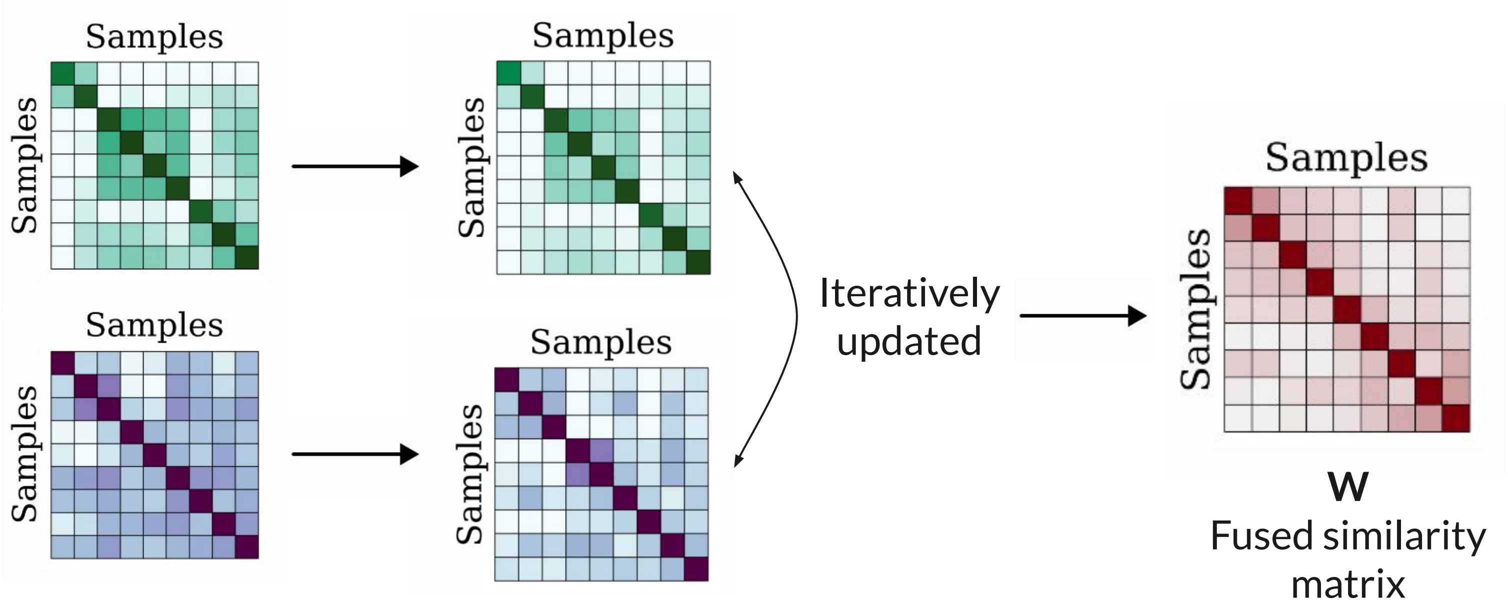


Iteratively updated

W
Similarity matrix

P(t)
Status matrix

Similarity Network Fusion (SNF): Fusion



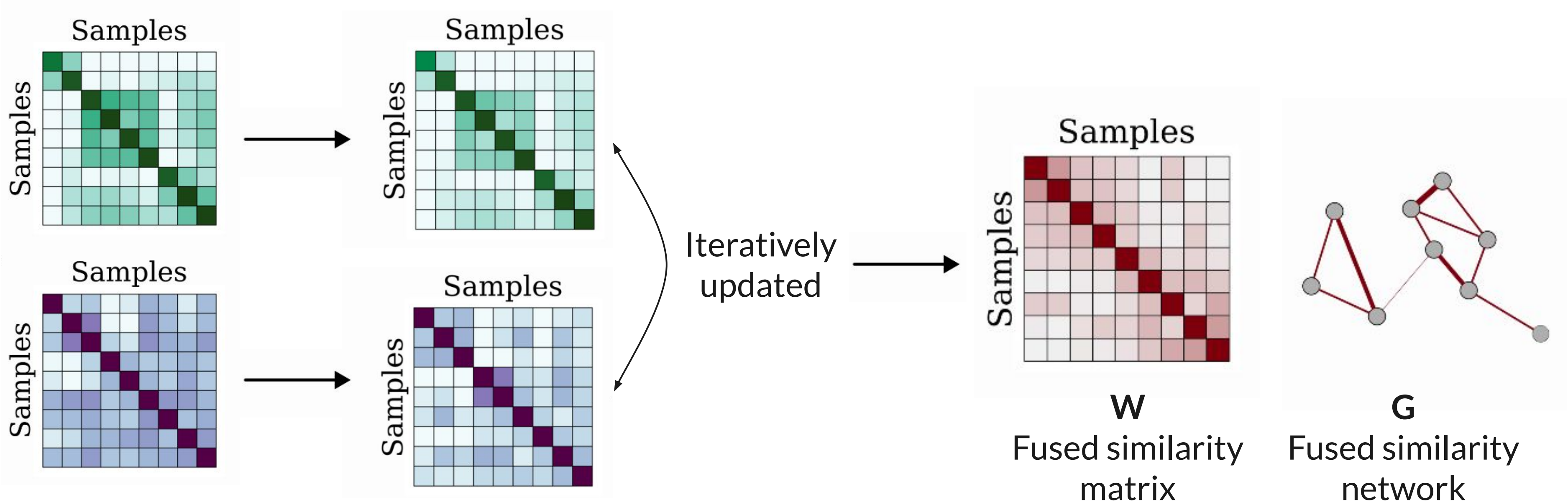
Iteratively updated

W
Similarity matrix

P(t)
Status matrix

W
Fused similarity matrix

Similarity Network Fusion (SNF): Fusion



Iteratively updated

W
Similarity matrix

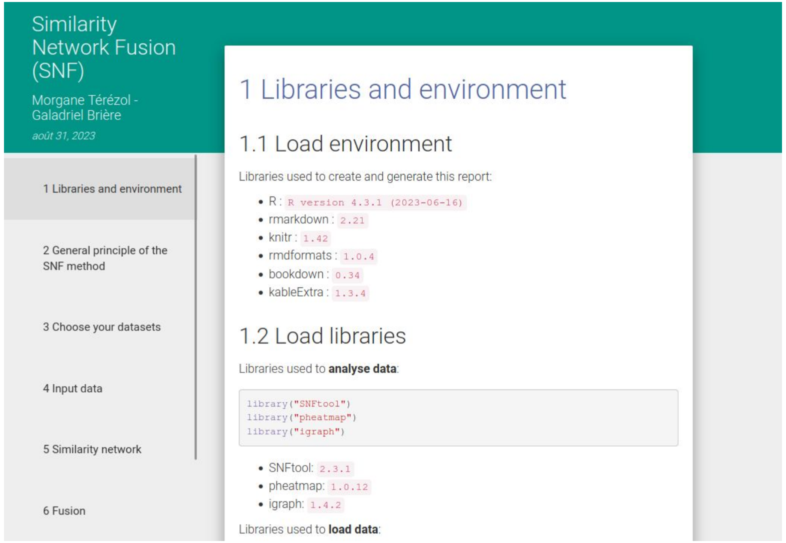
P(t)
Status matrix

W
Fused similarity matrix

G
Fused similarity network

Hands-on

Hands-on in **ETBII_2026_SNF.html**



Similarity Network Fusion (SNF)

Morgane Térézol - Galadriel Brière

août 31, 2023

1 Libraries and environment

1.1 Load environment

Libraries used to create and generate this report:

- R : `R version 4.3.1 (2023-06-16)`
- rmarkdown : `2.21`
- knitr : `1.42`
- rmdformats : `1.0.4`
- bookdown : `0.34`
- kableExtra : `1.3.4`

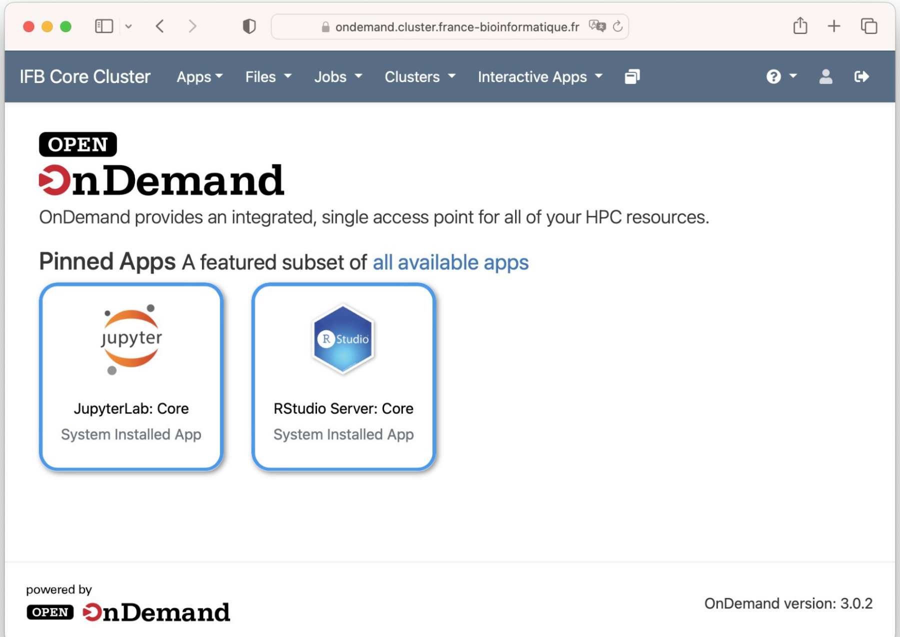
1.2 Load libraries

Libraries used to **analyse data**:

```
library("SNFtool")
library("pheatmap")
library("igraph")
```

- SNFtool: `2.3.1`
- pheatmap: `1.0.12`
- igraph: `1.4.2`

Libraries used to **load data**:



OPEN OnDemand

OnDemand provides an integrated, single access point for all of your HPC resources.

Pinned Apps A featured subset of all available apps

powered by OPEN OnDemand

OnDemand version: 3.0.2