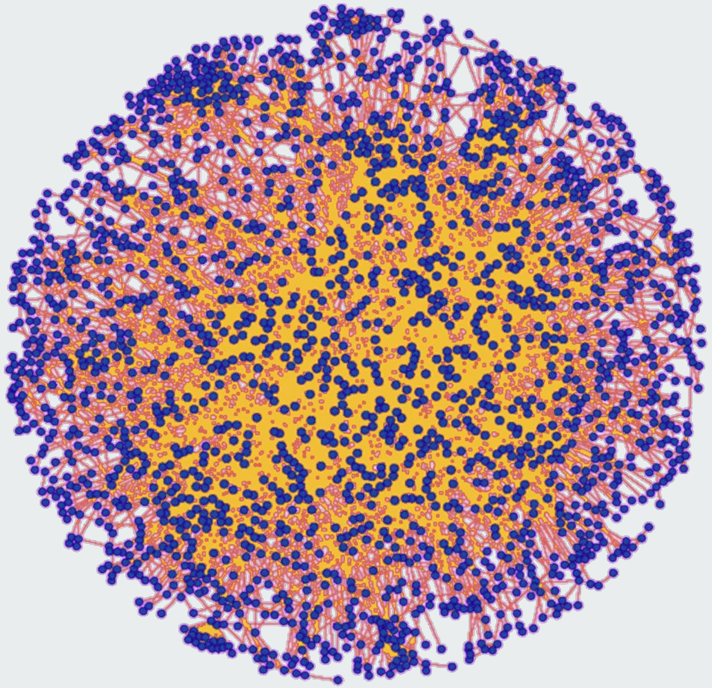
Network theory

Galadriel Brière, Morgane Térézol

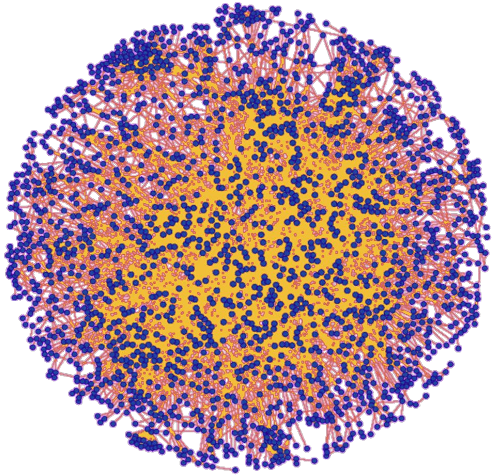
Inspired by the slides of Anaïs Baudot

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)



Swiss Institute of Bioinformatics
INSTITUT FRANÇAIS DE BIOINFORMATIQUE
elixir SWITZERLAND
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, …

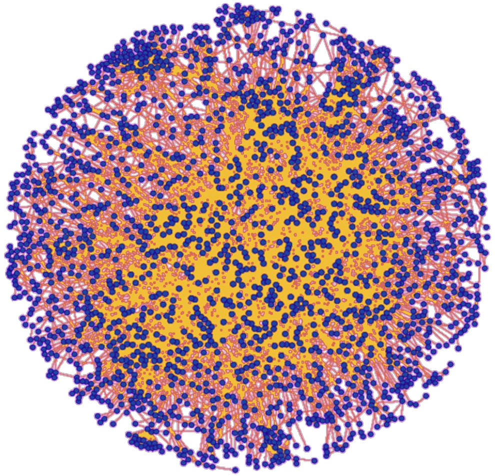
Algorithms

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


(a) (b)

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**

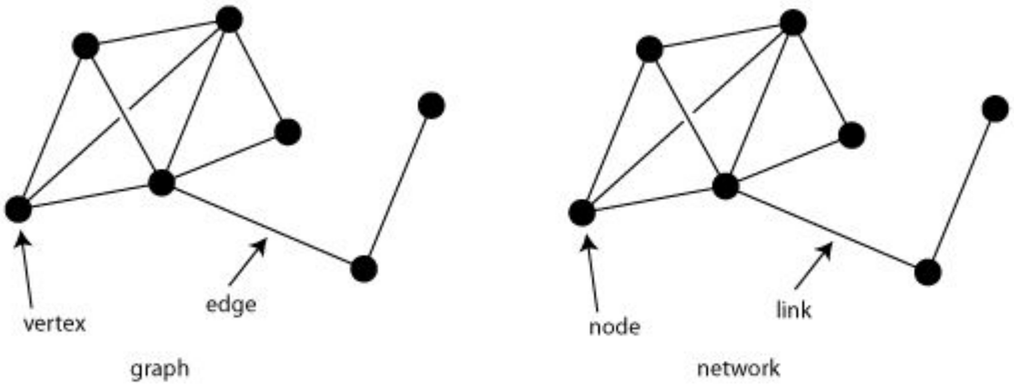
Systems components: neurons, axons, dendrites, …
Interactions: synaptic transmission, …
Emerging properties: memory, cognition, …

○ **etc….**

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

- Networks are real-world systems modeled using graphs.

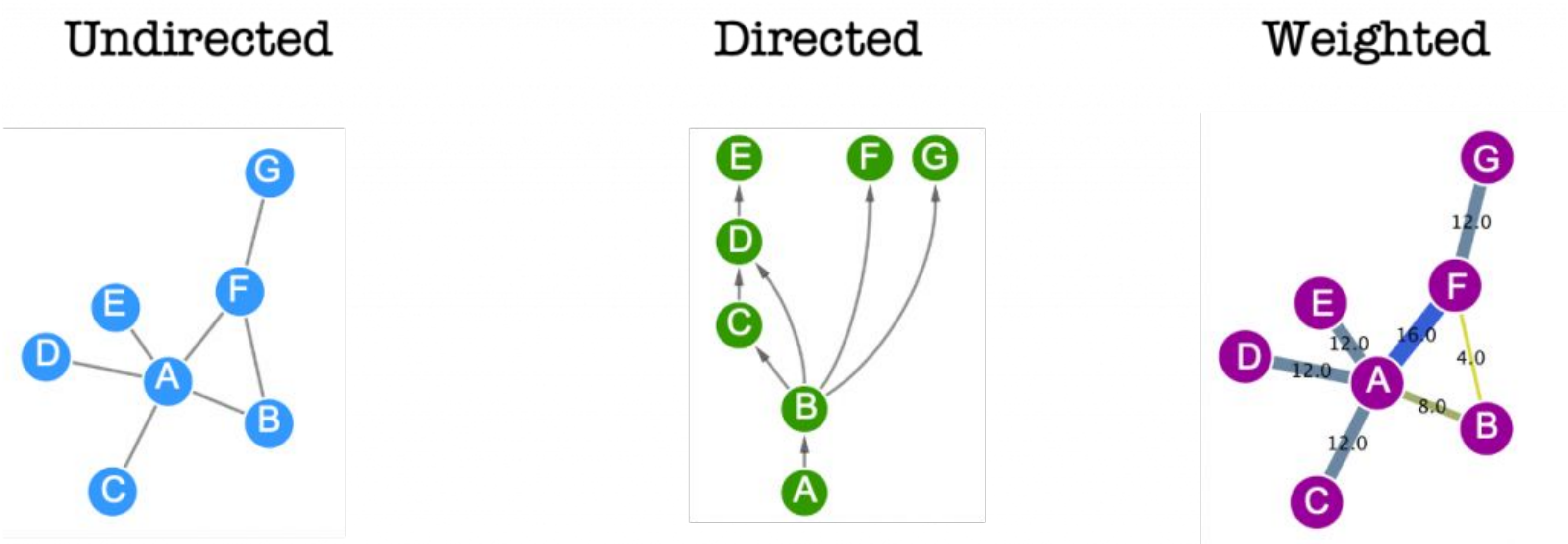
$G = (V, E)$, where V is the set of vertices and $E \subseteq (V*V)$ is the set of edges



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

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

- Various types of networks…



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/.

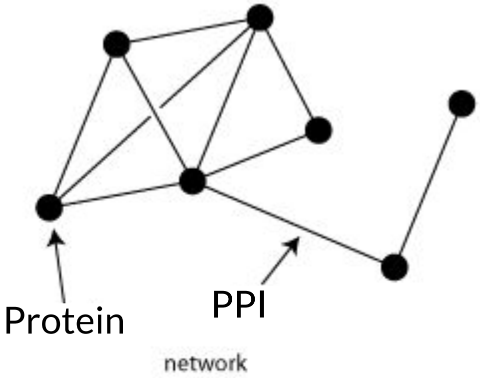
Network Science and **Systems Biology**

- ○ **Systems Biology**

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

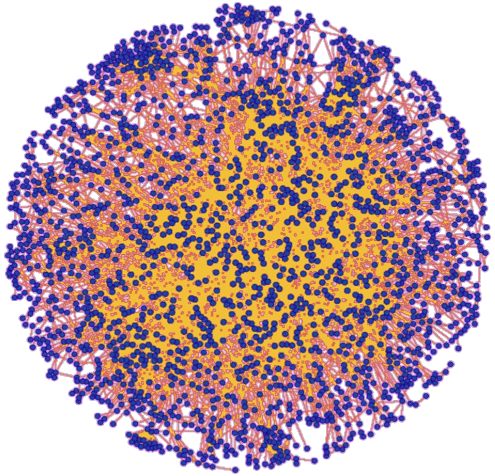
Phenotype does not emerge from isolated biological molecules but from their interactions

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

Data sources, Distances/Similarities, Correlation, …

Network Analysis

Measures

Degree, Centrality, Distances, …

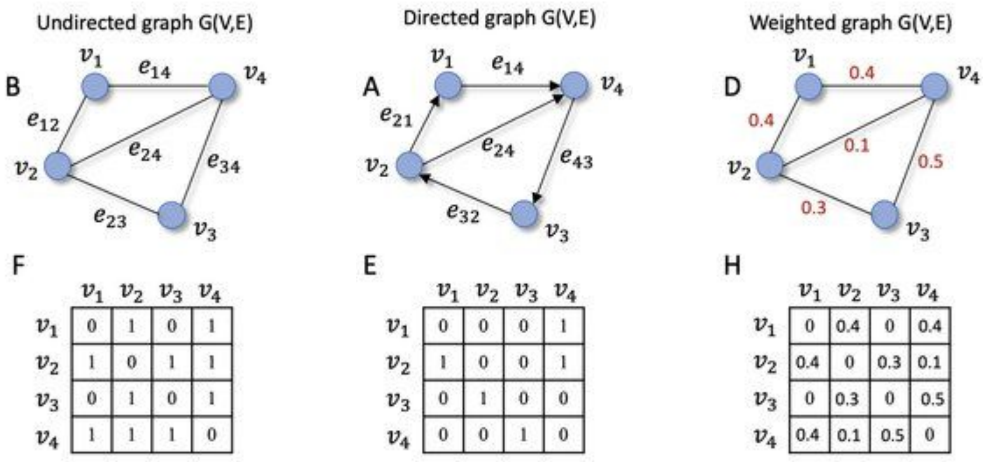
Algorithms

Diffusion, Clustering, Embedding, …

Networks and Integration

Multiplex Networks, Multilayer Networks, Knowledge Graphs, …

Adjacency matrix



F

	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

E

	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

H

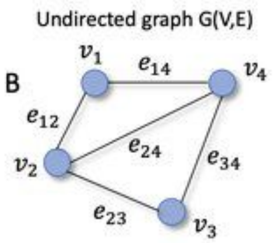
	v_1	v_2	v_3	v_4
v_1	0	0.4	0	0.4
v_2	0.4	0	0.3	0.1
v_3	0	0.3	0	0.5
v_4	0.4	0.1	0.5	0

Undirected -> **Symmetric adjacency**

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.

Incidence matrix



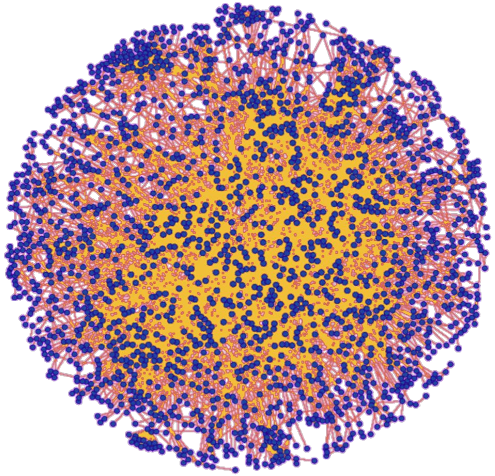
Edges

Nodes

	E1-2	E1-4	E2-3	E2-4	E3-4
V1	1	1	0	0	0
V2	1	0	1	1	0
V3	0	0	1	0	1
V4	0	1	0	1	1

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 sources

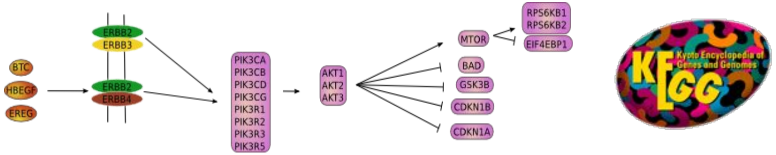


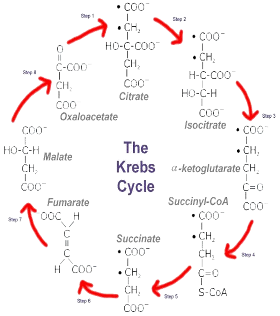
Source: https://www3.nd.edu/~tmilenko/research.html.

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

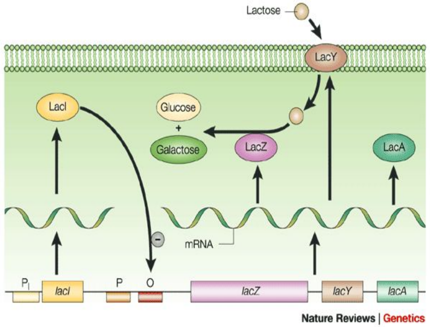
Source 1: **Expert knowledge and literature**

Biological Pathways





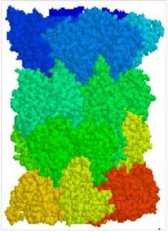
Metabolic Cycles



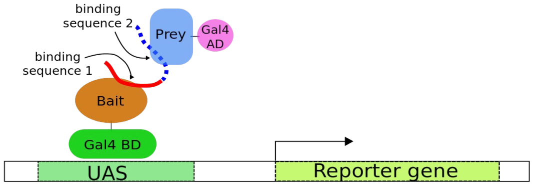
Operons / Regulatory Networks

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

$$X_j = \beta_0 + \enclose{circle}{\beta_1} X_i + \epsilon$$

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 Bayesian model

$$P(X_j/X_i)$$

  - A regression model

$$X_j = \beta_0 + \overset{\text{relationship between j and i}}{\boxed{\beta_1}} X_i + \epsilon$$

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

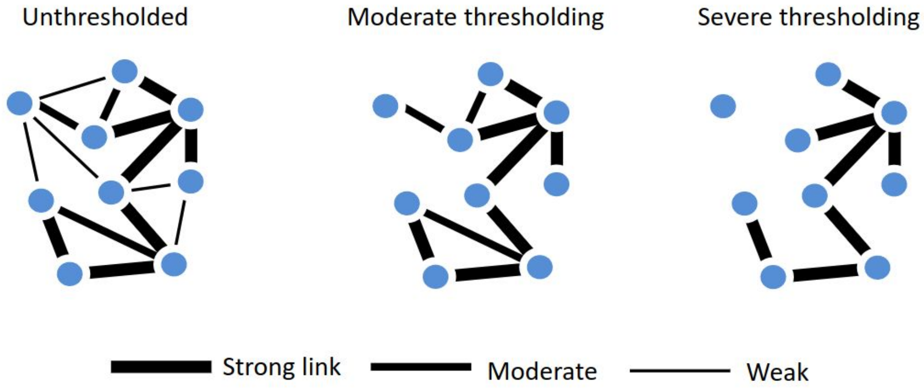
    $$P(X_j/X_i)$$

  - 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)

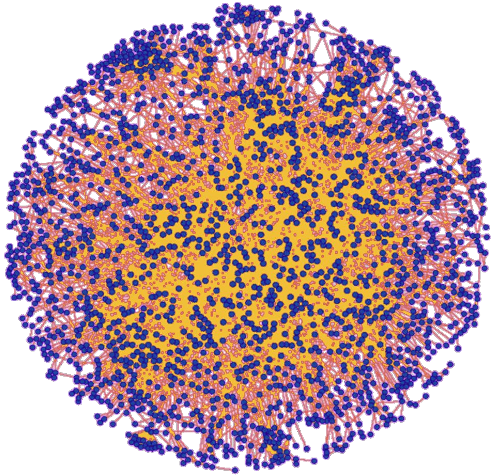
- **Often, the similarity/distances/correlations/… are thresholded:**



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

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

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, …

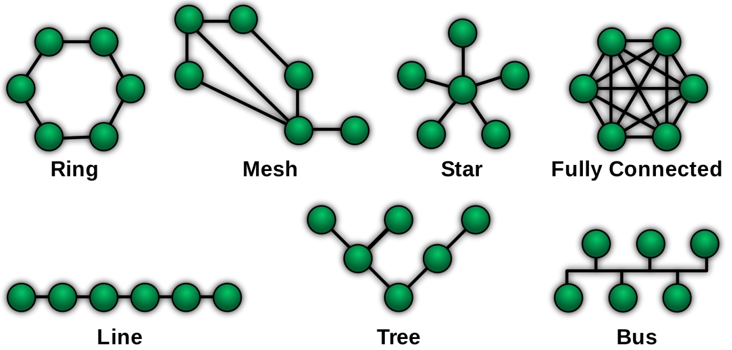
Networks and Integration

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.

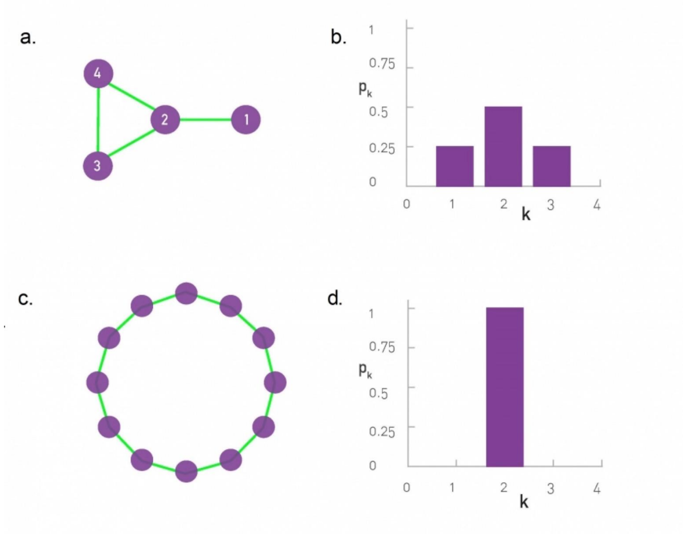
  - Robustness of a network
  - Hub nodes
  - Essential nodes
  - Paths/accessibility
  - Communities
  - ....



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**

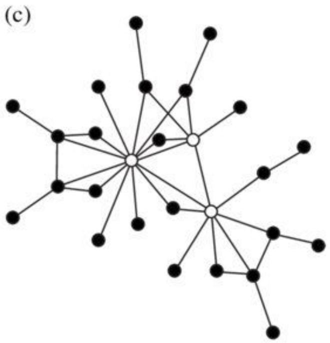
- The **degree of a node** in a network is the **number of edges that are incident to the node.**
- Observing the **degree distribution** of a network can reveal interesting properties:
  - Overall connectivity



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))



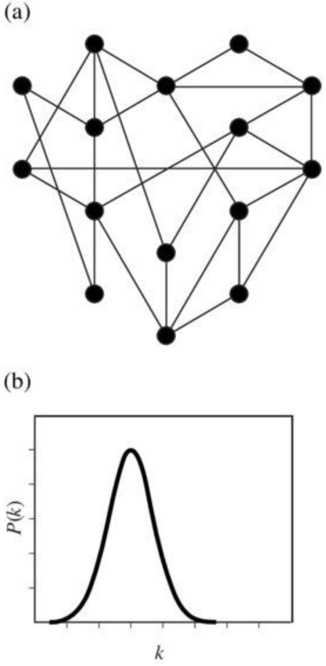
Source: Bentley, Barry. "Connectomics of Extrasynaptic Signalling: Applications to the Nervous System of Caenorhabditis Elegans," 2017. https://doi.org/10.17863/CAM.16873.

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

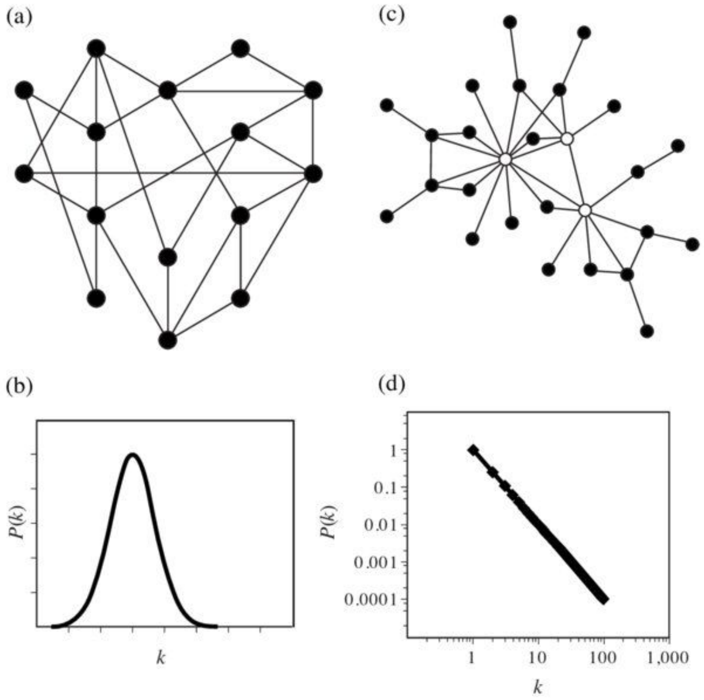


Random Network

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



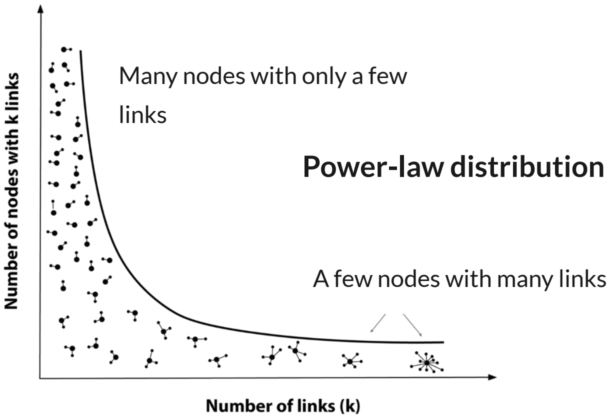
Random Network

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

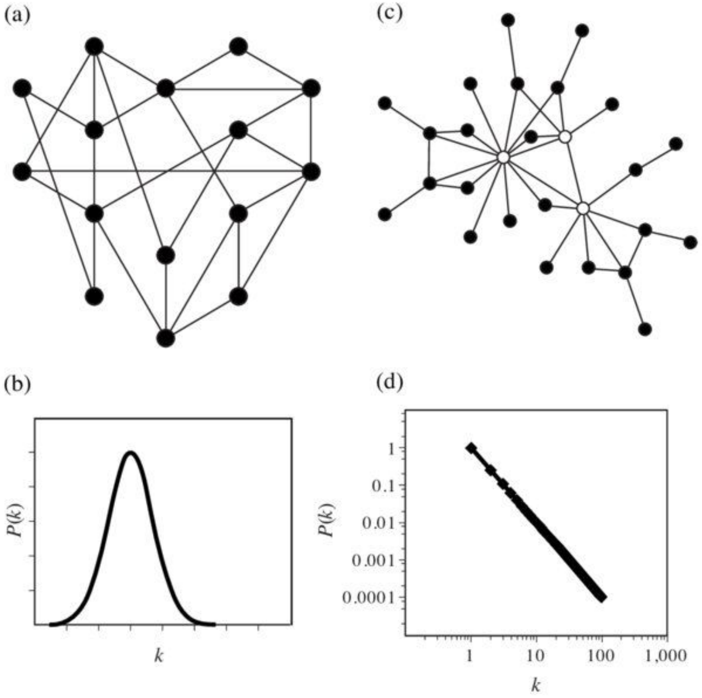
Source: Bentley, Barry. "Connectomics of Extrasynaptic Signalling: Applications to the Nervous System of Caenorhabditis Elegans," 2017. https://doi.org/10.17863/CAM.16873.

Metrics in network analysis: **Node Degree**

- The **degree of a node** in a network is the **number of edges 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



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



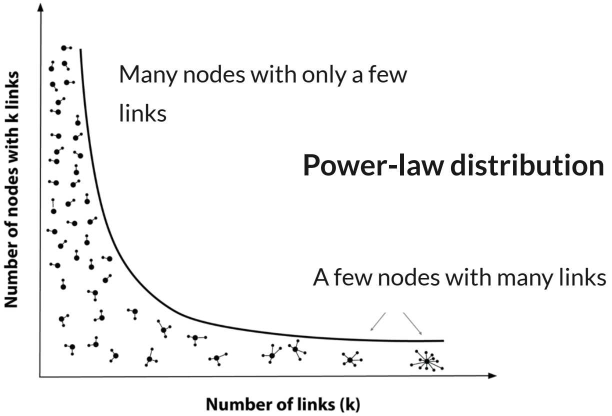
Random Network

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



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

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

Are biological networks scale-free ?

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

- PPI
- Metabolic networks
- Regulatory networks

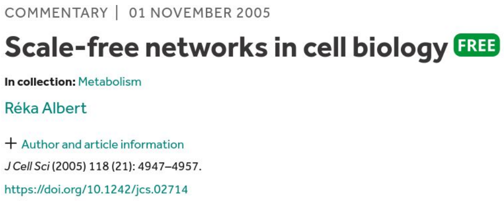
Are biological networks scale-free ? A debate!

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

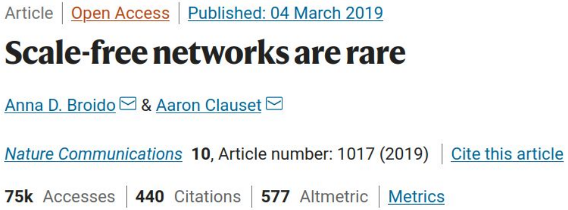
> J Comput Biol. 2006 Apr;13(3):810-8. doi: 10.1089/cmb.2006.13.810.

How scale-free are biological networks

Raya Khanin [1], Ernst Wit

Affiliations + expand

PMID: 16706727 DOI: 10.1089/cmb.2006.13.810



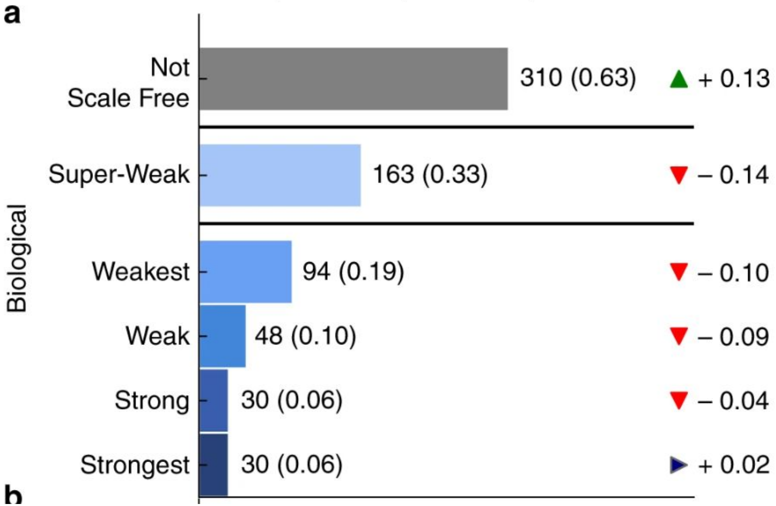
Article | Open Access | Published: 04 March 2019

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

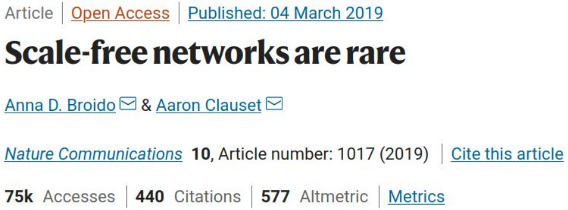
J Cell Sci (2005) 118 (21): 4947–4957.
https://doi.org/10.1242/jcs.02714

› 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



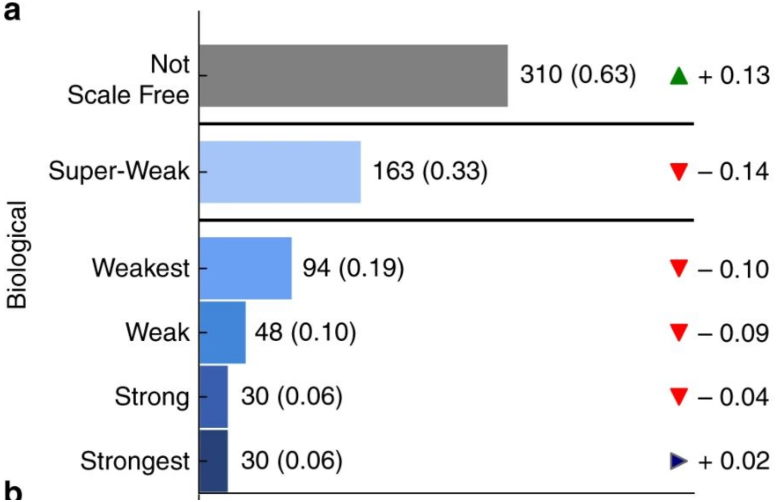
Article | Open Access | Published: 04 March 2019

Scale-free networks are rare

Anna D. Broido ✉ & Aaron Clauset ✉

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

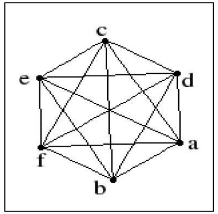
75k Accesses | **440** Citations | **577** Altmetric | Metrics



- ➢ 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

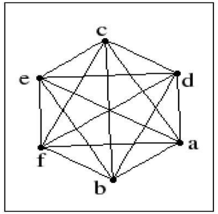
$$D(G) = \frac{|E|}{|E|_{Kn}}$$

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

Source: Courtesy of Anaïs Baudot

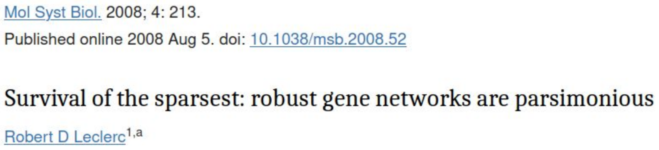
Metrics in network analysis: **Density**

- 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]

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

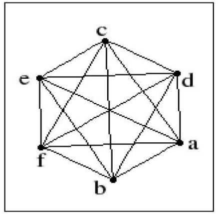
"**Robustness** implies a parsimonious network structure that is sparsely connected and not unnecessarily complex"

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

Source: Courtesy of Anaïs Baudot

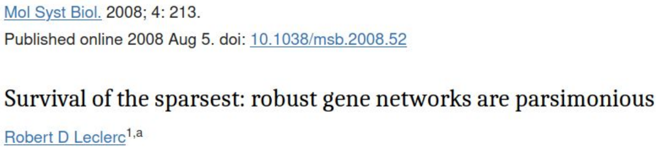
Metrics in network analysis: **Density**

- 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

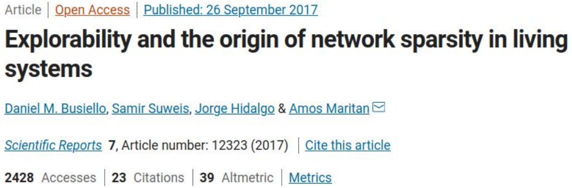
- Are biological networks dense or sparse?



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"

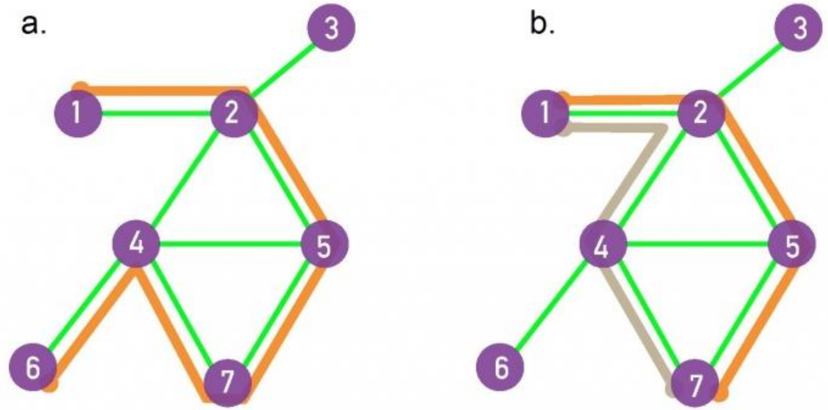
$$D_G = 0 \Leftrightarrow |E| = 0$$
$$D_G = 1 \Leftrightarrow G = Kn$$

"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**

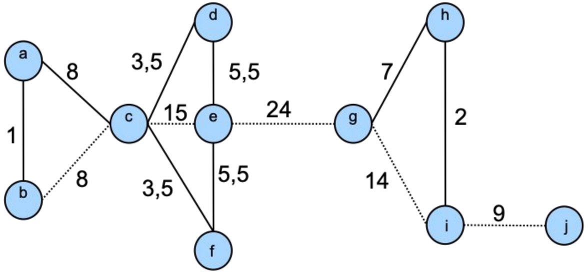
- A **path** is a **sequence of edges** which join a sequence of nodes
- The **distance** between two nodes in a network is the **length** (i.e. number of edges) of the **shortest path** connecting them.
- The **diameter** of a network is the **greatest distance between any pair of nodes** in the network.

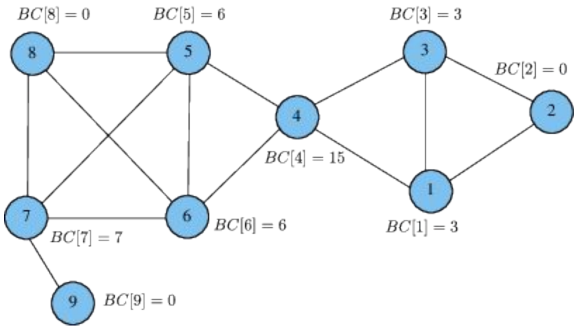


Source: Courtesy of Anaïs Baudot

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**”





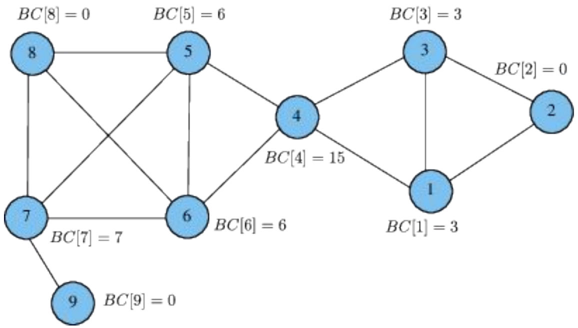
Source: Courtesy of Anaïs Baudot

Metrics in network analysis: **Betweenness**

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

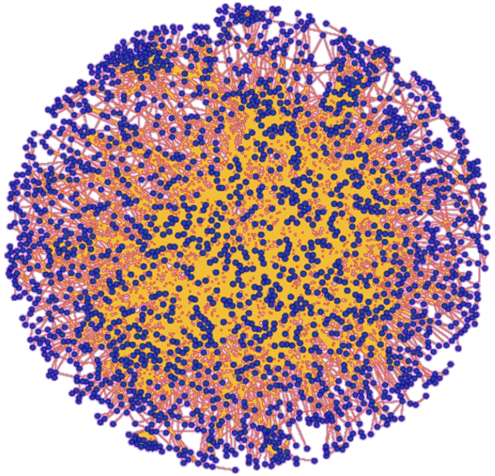
Haiyuan Yu, Philip M Kim, Emmett Sprecher, Valery Trifonov, Mark Gerstein

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

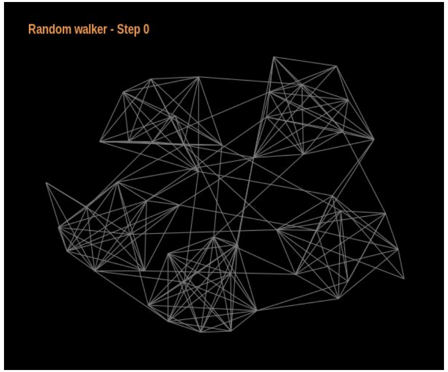
Diffusion, Clustering, Embedding, …

Networks and Integration

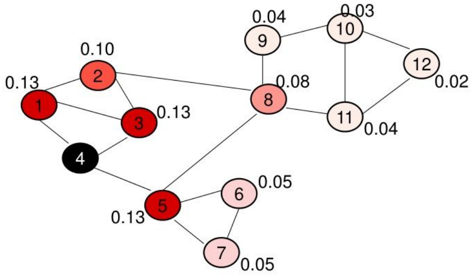
Multiplex Networks, Multilayer Networks, Knowledge Graphs, …

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

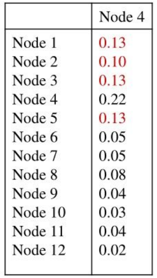
- ➢ State of the art “**guilt-by-association**” approach
- ➢ Similarity between a **seed node** and all other nodes in the network
- ➢ Used for node prioritisation, network clustering, network embedding, etc...



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

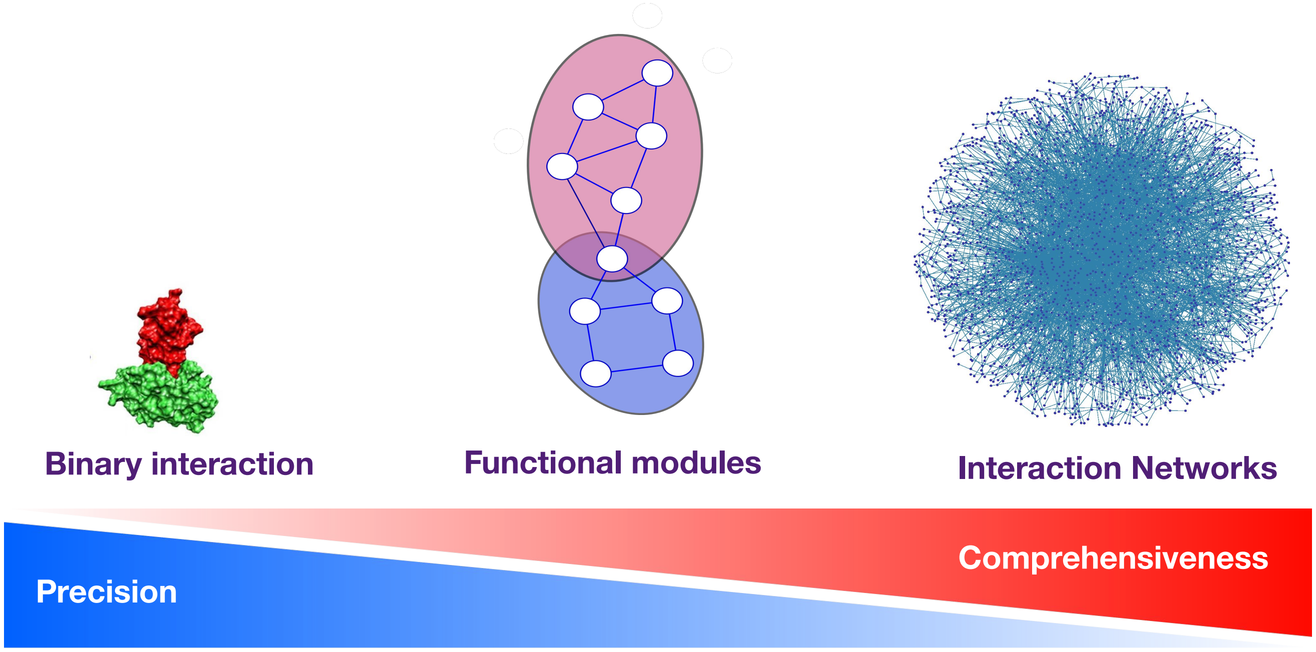


	Node 4
Node 1	0.13
Node 2	0.10
Node 3	0.13
Node 4	0.22
Node 5	0.13
Node 6	0.05
Node 7	0.05
Node 8	0.08
Node 9	0.04
Node 10	0.03
Node 11	0.04
Node 12	0.02

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

Precision

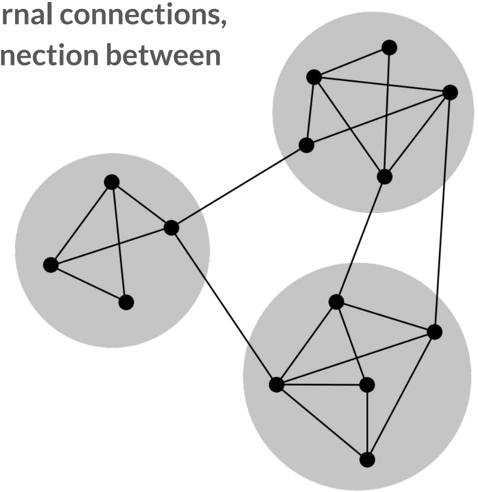
Comprehensiveness

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

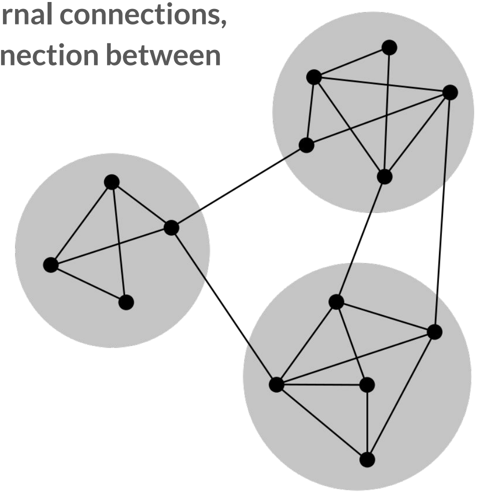


Source: “Community Structure.” In Wikipedia, August 26, 2023.

Algorithms for network analysis: **Community detection**

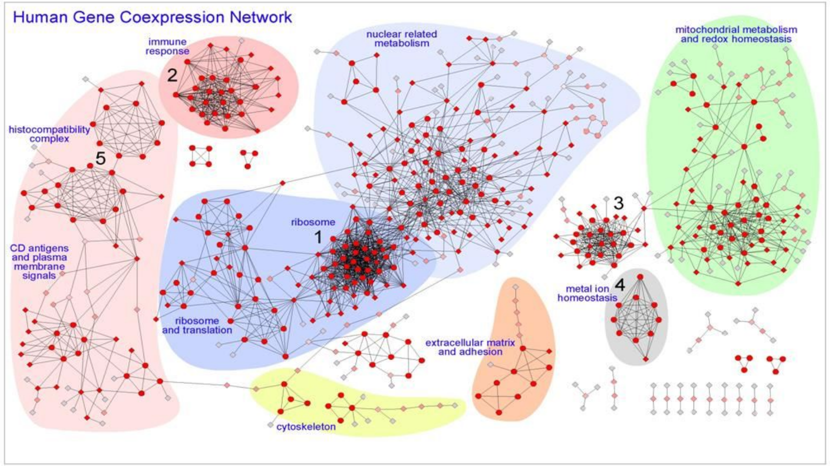
- Community detection = **clustering** for networks!
- Find groups of nodes (**communities/modules/clusters**) that are **more similar to each other than to the other nodes**

Dense internal connections, sparse connection between groups



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

Why performing Community detection in biological networks ?



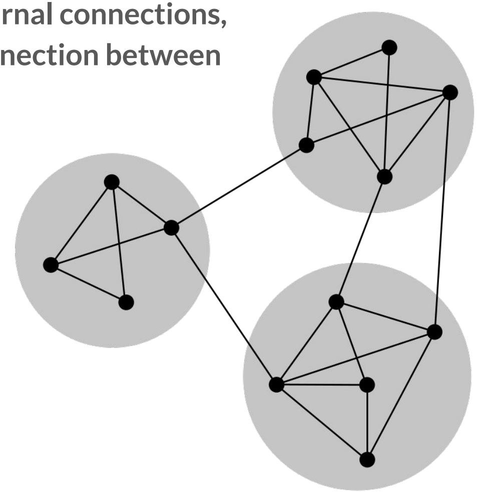
Source: “Human Gene Coexpression.” Accessed August 29, 2023. http://bioinfow.dep.usal.es/pages/coexpression/index.html.

Algorithms for network analysis: **Community detection**

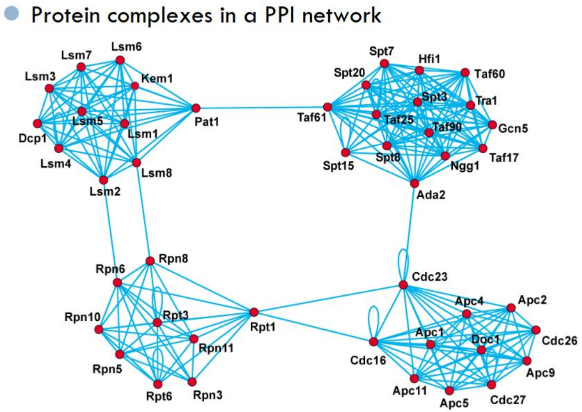
- ➢ Community detection = **clustering** for networks!
- ➢ Find groups of nodes (**communities/modules/clusters**) that are **more similar to each other than to the other nodes**

Why performing Community detection in biological networks ?

Dense internal connections, sparse connection between groups



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



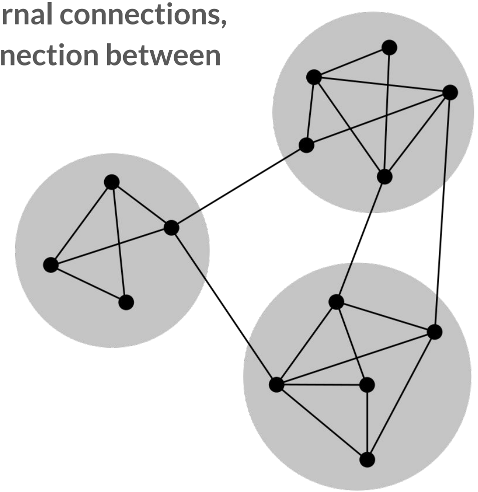
Source: Ngom, Dr Alioune. “NETWORK CLUSTERING METHODS,” n.d.

Algorithms for network analysis: **Community detection**

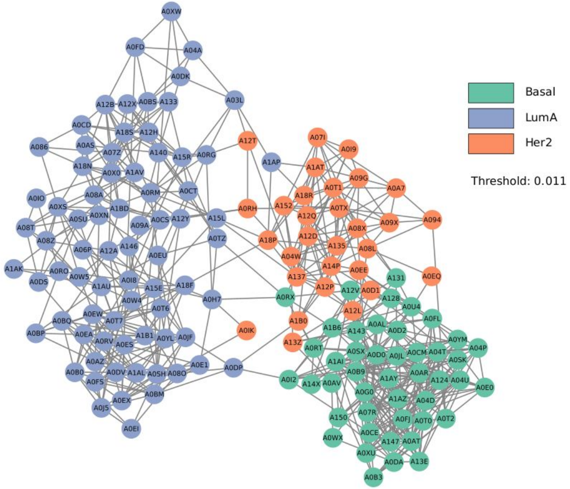
- ➢ Community detection = **clustering** for networks!
- ➢ Find groups of nodes (**communities/modules/clusters**) that are **more similar to each other than to the other nodes**

Why performing Community detection in biological networks ?

Dense internal connections, sparse connection between groups



Source: "Community Structure." In Wikipedia, August 26, 2023.

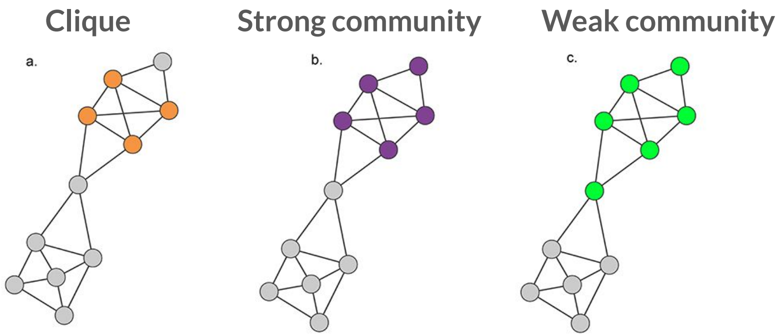


Source: Morgane Térézol - SNF tutorial

Algorithms for network analysis: **Community detection**

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

Cliques are **complete subgraphs.**

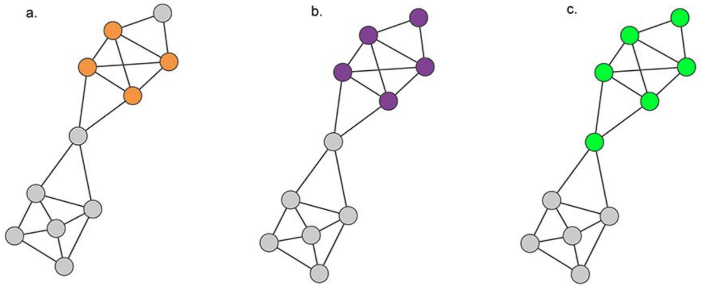


Source: Network Science by Albert-László Barabási. Accessed August 29, 2023.
http://networksciencebook.com/.

Algorithms for network analysis: **Community detection**

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

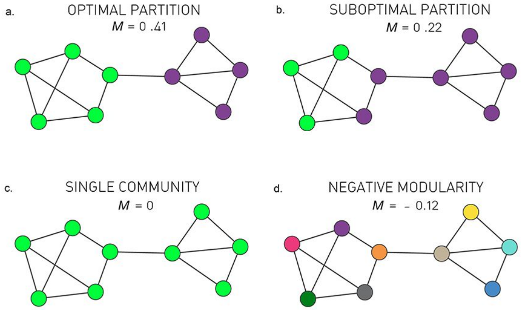
  Cliques are **complete subgraphs.**


Clique
Strong community
Weak community

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.

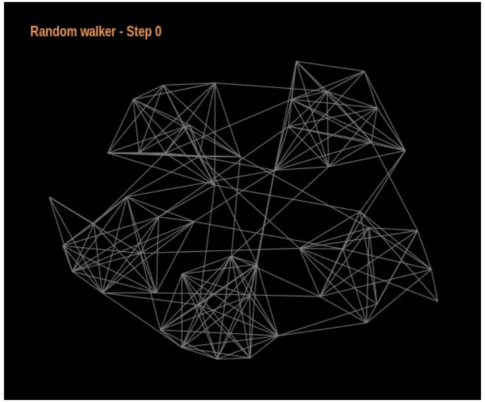


Source: Network Science by Albert-László Barabási. Accessed August 29, 2023. http://networksciencebook.com/.

Algorithms for network analysis: **Community detection**

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

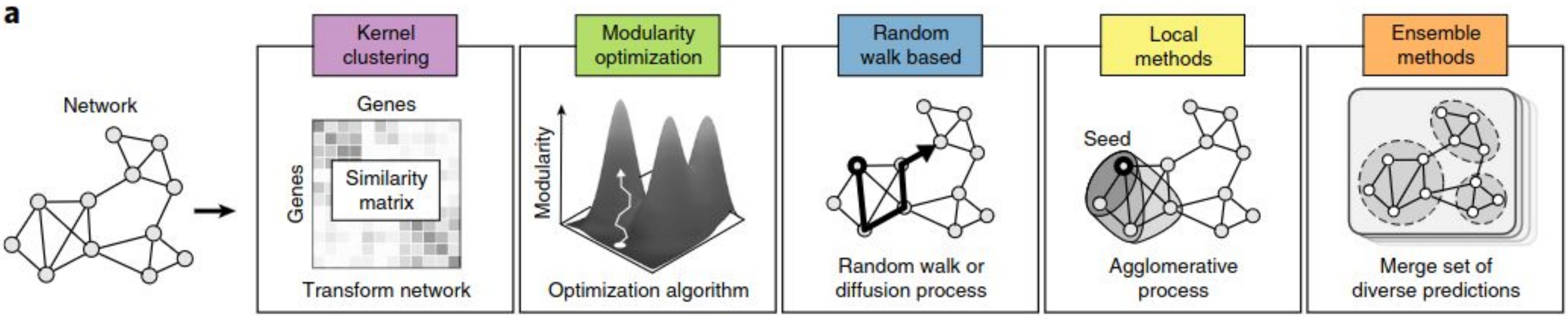
Random walks on a graph tend to get **“trapped” into densely connected parts** corresponding to communities.

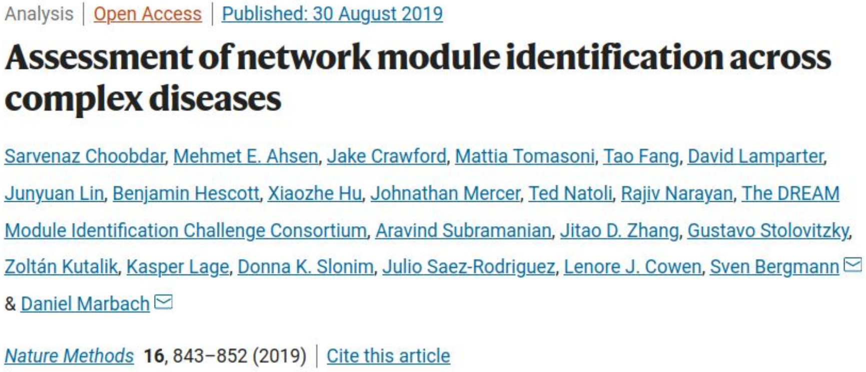


➢ **and many other…**

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.

Algorithms for network analysis: **Community detection**





Analysis | Open Access | Published: 30 August 2019

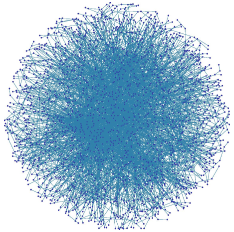
Assessment of network module identification across complex diseases

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

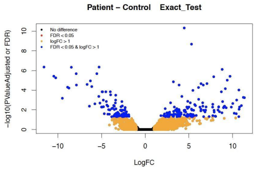
Algorithms for network analysis: **Active Modules**

Biological Network



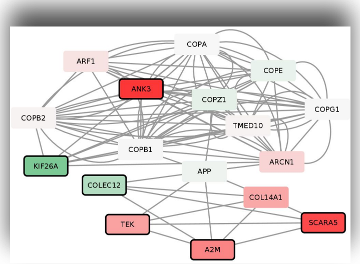
+

RNA-seq transcriptomics data



↓

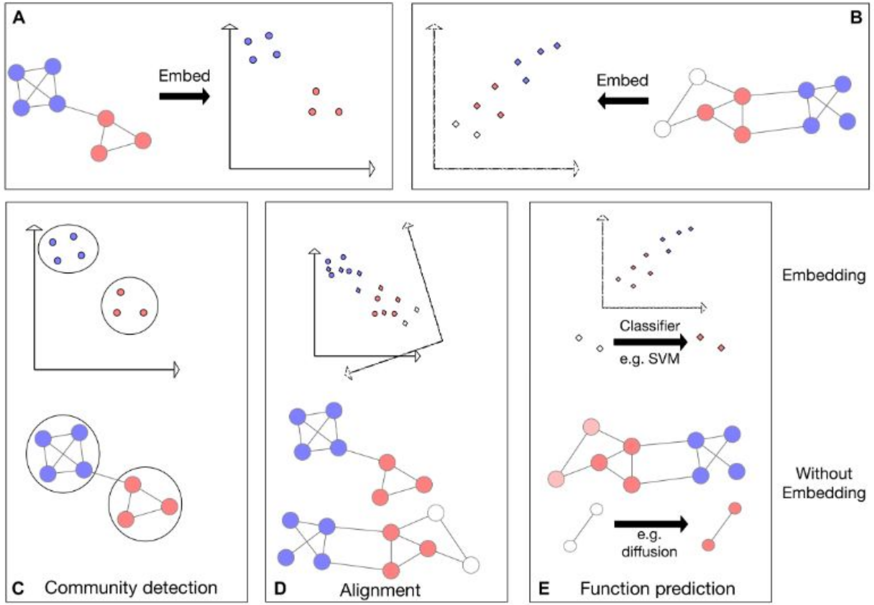
Find subnetworks of interest
aka “active modules”



Source: Courtesy of Anaïs Baudot

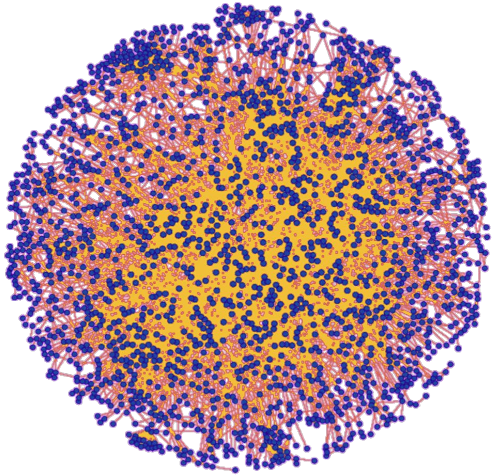
Algorithms for network analysis: **Network Embedding**

- 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

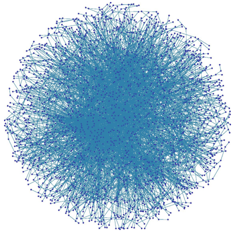
Diffusion, Clustering, Embedding, …

Networks and Integration

Multiplex Networks, Multilayer Networks, Knowledge Graphs, …

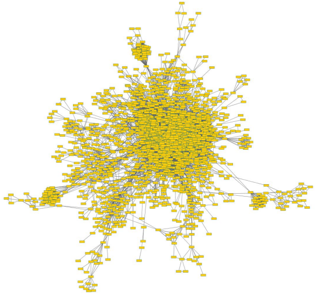
Many Biological Networks

PPI



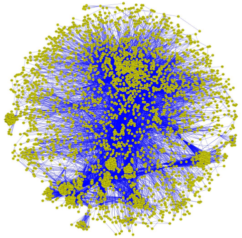
~60 000 edges

Complexes



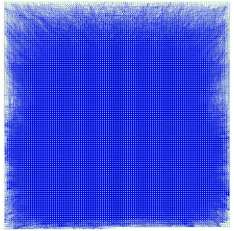
~40 000 edges

Pathways



~250 000 edges

Correlation of expression



~1 400 000 edges

Experimental networks

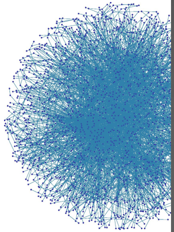
Curated networks

Inferred networks

Source: Courtesy of Anaïs Baudot

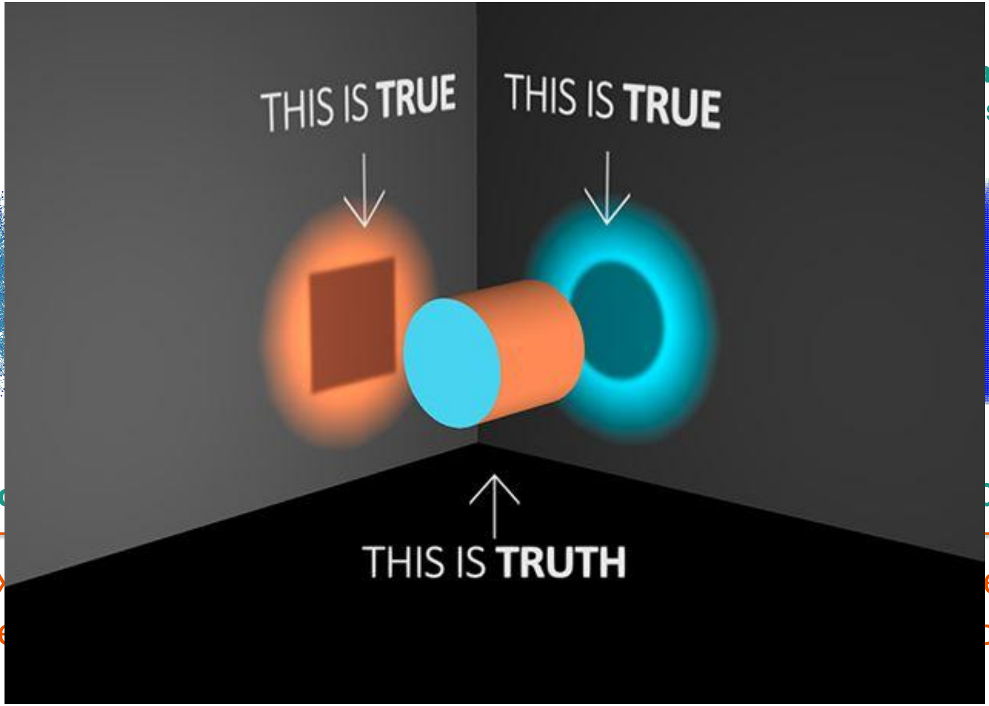
Many Biological Networks

PPI



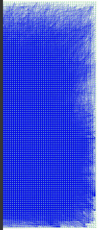
~60 000 e

Ex
ne



Source: Courtesy of Anaïs Baudot

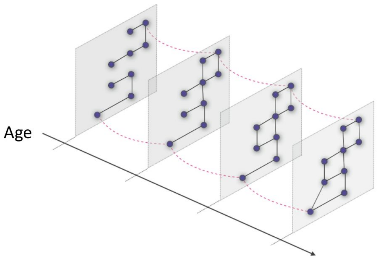
**ation of
ssion**



00 edges

ed
rks

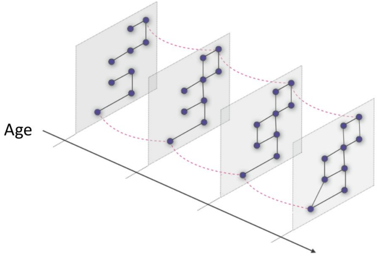
Networks can easily be **combined**

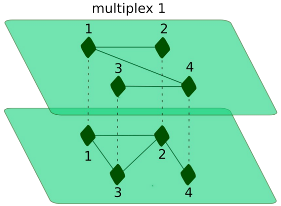


- **Temporal networks**
  - Same nodes
  - Same type of interactions
  - Various time-points

Source: Courtesy of Anaïs Baudot

Networks can easily be **combined**



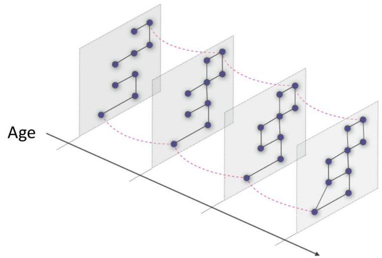


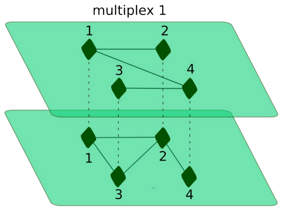
- ➢ **Multiplex networks**
  - ○ Same nodes
  - ○ Different types of interactions

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

Source: Courtesy of Anaïs Baudot

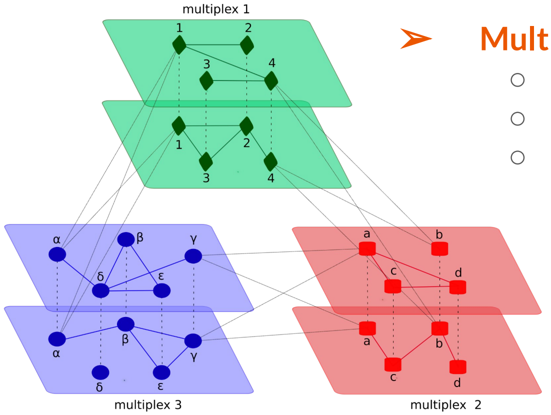
Networks can easily be **combined**





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

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

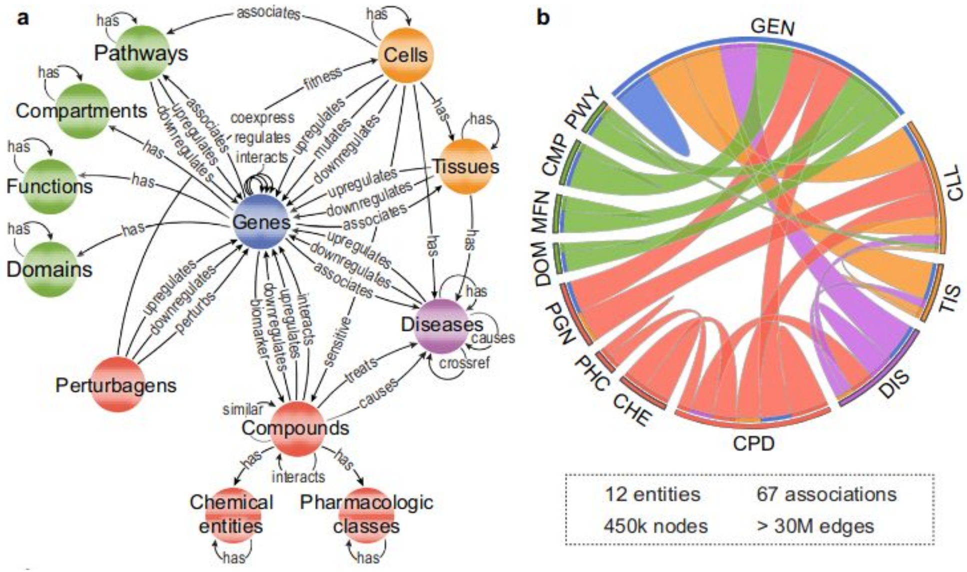


- ➢ **Multilayer networks**
  - ○ Various node types
  - ○ Different types of interactions
  - ○ Bipartite edges (across node types)

Source: Courtesy of Anaïs Baudot

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

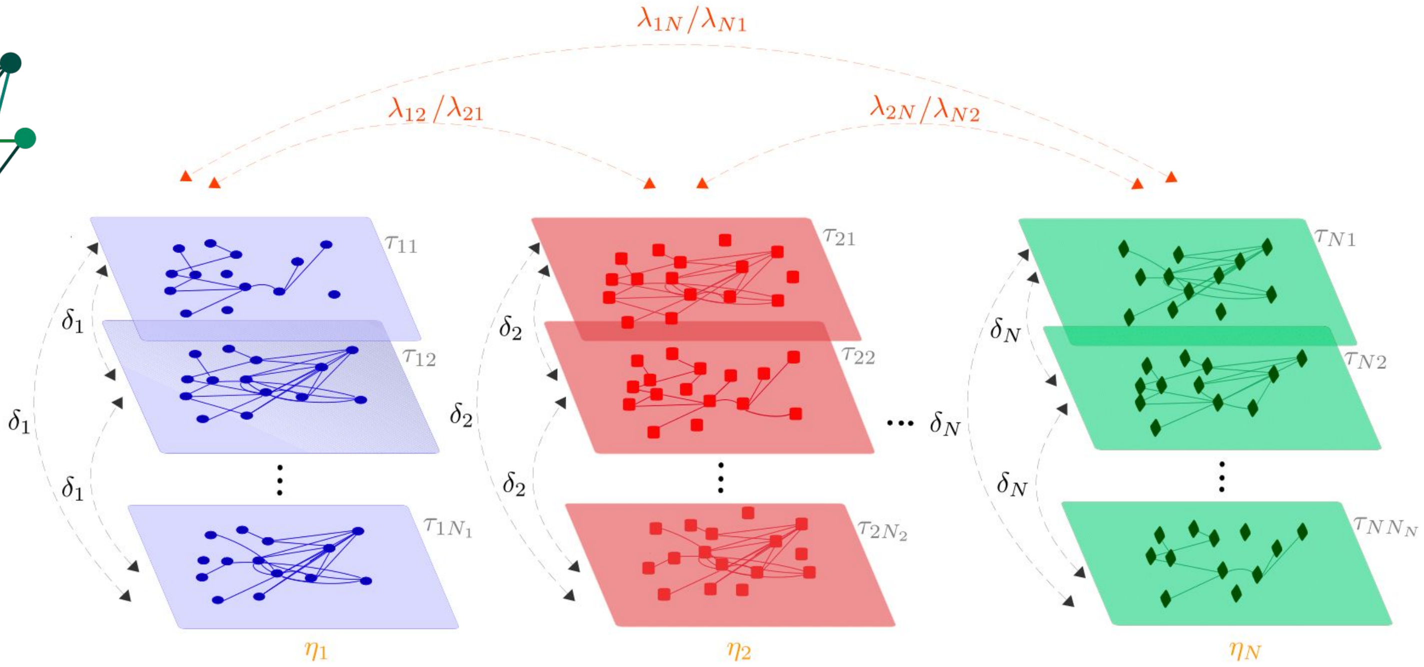
- **Knowledge graphs** … are basically networks, but with different semantics
  - Represent real-world **information** as a set of entities and relationships



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

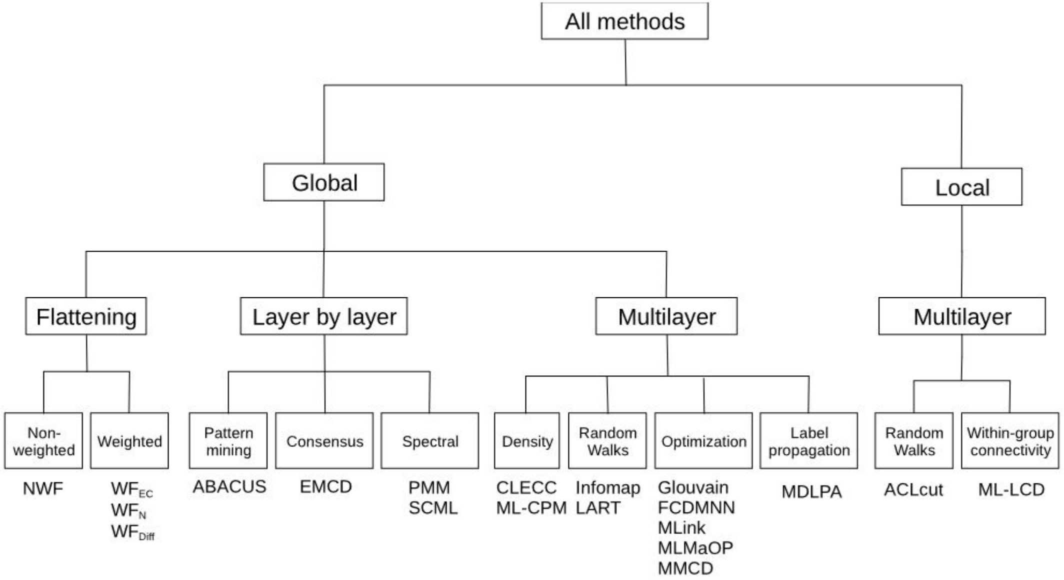
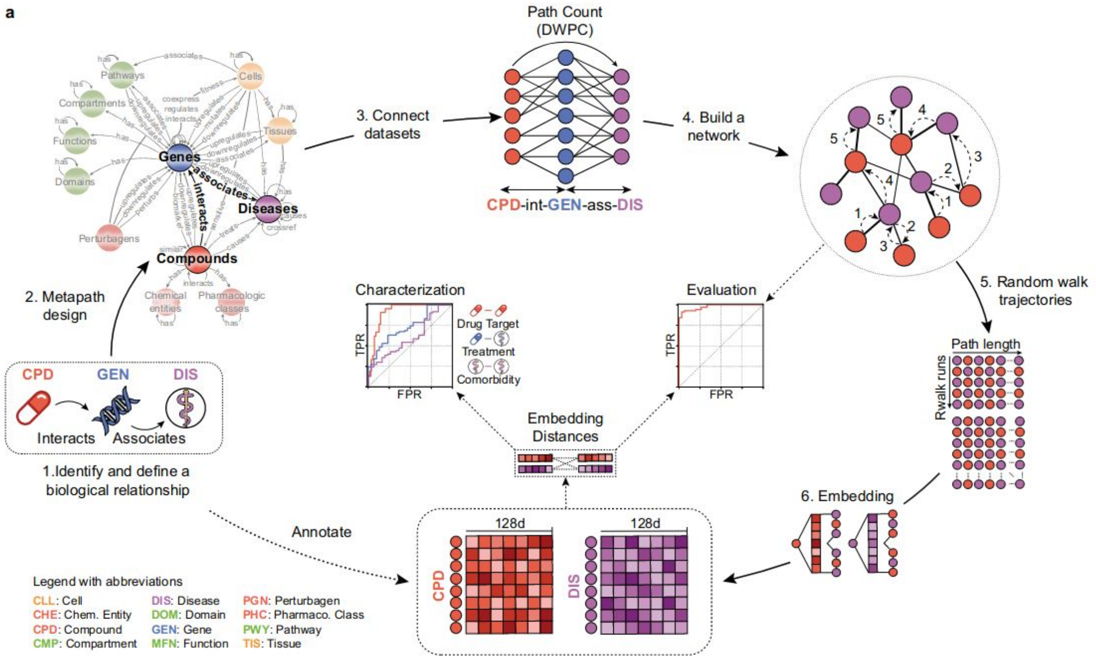


Figure 4: A taxonomy of multiplex community detection algorithms

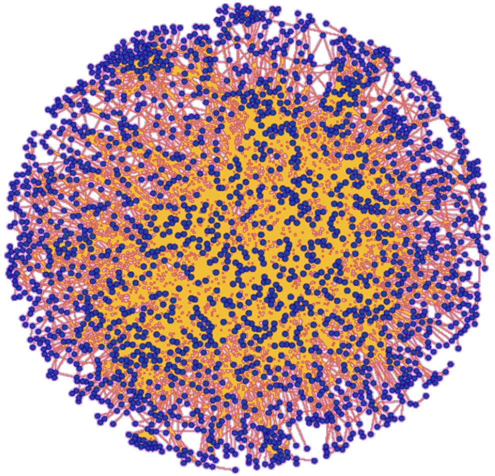
Source: Magnani, Matteo, Obaida Hanteer, Roberto Interdonato, Luca Rossi, and Andrea Tagarelli. “Community Detection in Multiplex Networks.” ACM Computing Surveys 54, no. 3 (April 30, 2022): 1–35. https://doi.org/10.1145/3444688.

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

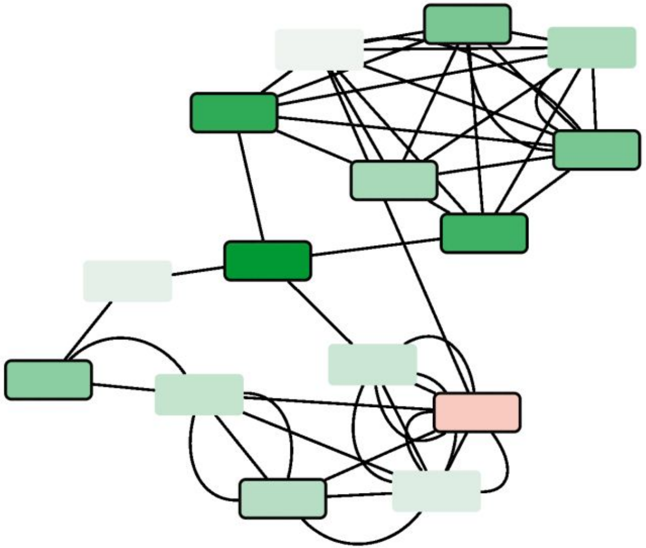
Diffusion, Clustering, Embedding, …

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

Acknowledgments

Summer School

Florence Mehl
Olivier Sand
Grégoire Rossier
Jimmy Vandel
Guillemette Marot
Anaïs Baudot
Maxime Delmas
Jean-Clément Gallardo
Marco Pagni
Arnaud Gloaguen
Lucie Khamvongsa Charbonnier

Slides

Anaïs Baudot

Hands-on

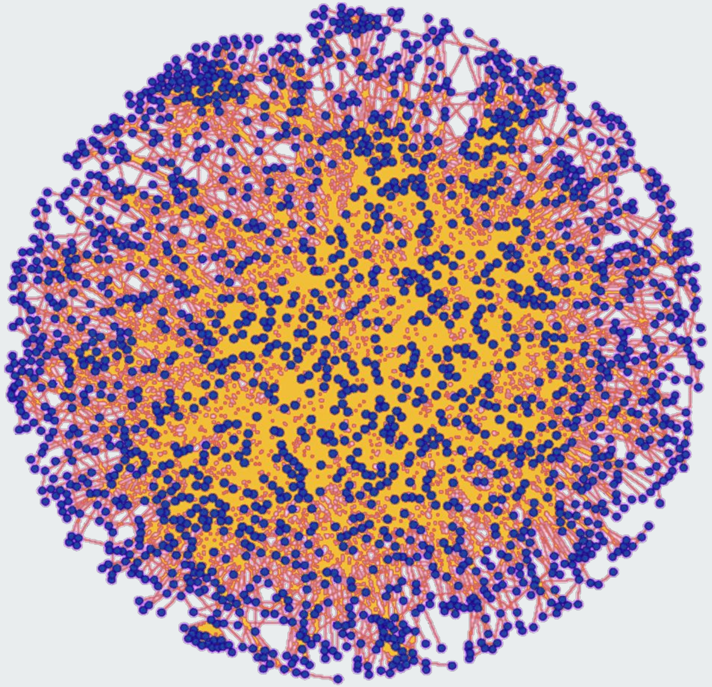
Samuel Chaffron

MMG/I2M

Anaïs Baudot
Élisabeth Rémy
Nadine Ben Boina
Céline Chevalier
David Hirst
Judith Lambert
Sevda Rafatov
Laurent Tichit
Jose Sanchez Villanueva
Daniel Lepe Soltero
Paul Villoutreix
Ozan Ozisik
Matéo Melki
Solène Song
Malek Senoussi

Hands-on : WGCNA

It’s your turn now! Have fun!



Weighted Gene Correlation Network Analysis (WGCNA)

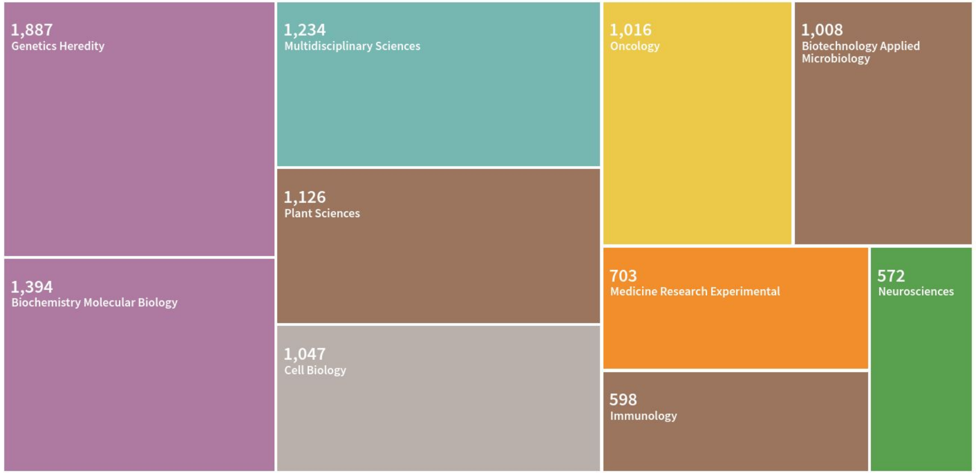
Software | Open Access | Published: 29 December 2008

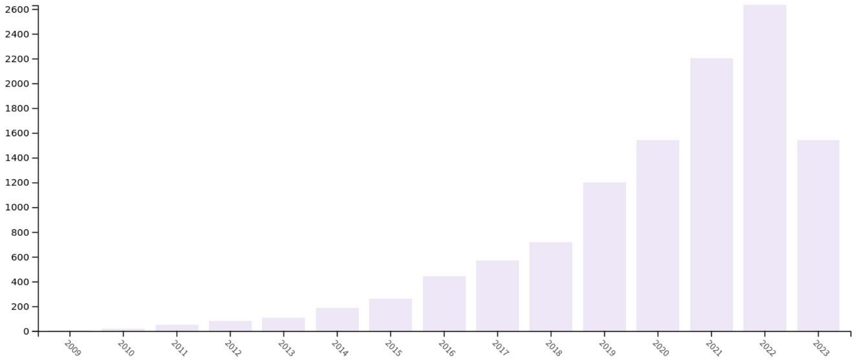
WGCNA: an R package for weighted correlation network analysis

Peter Langfelder & Steve Horvath

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

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

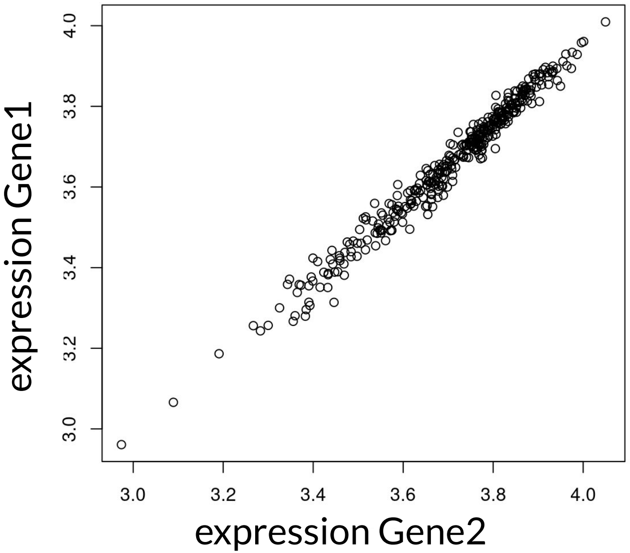




Source: www.webofscience.com

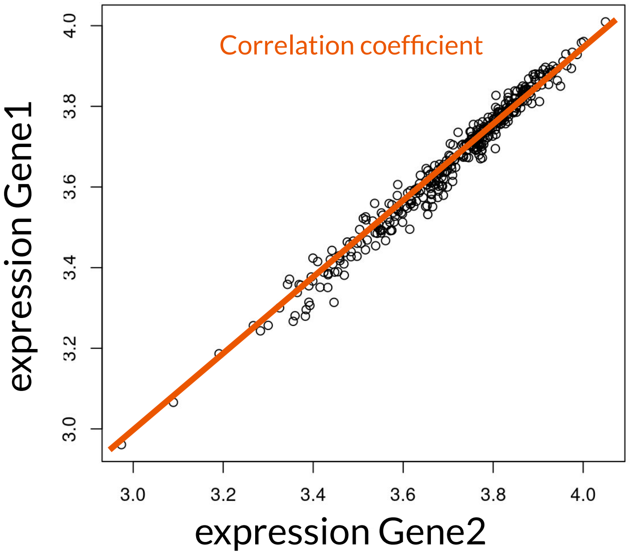
Weighted Gene Correlation Network Analysis (WGCNA)

- Gene co-expression



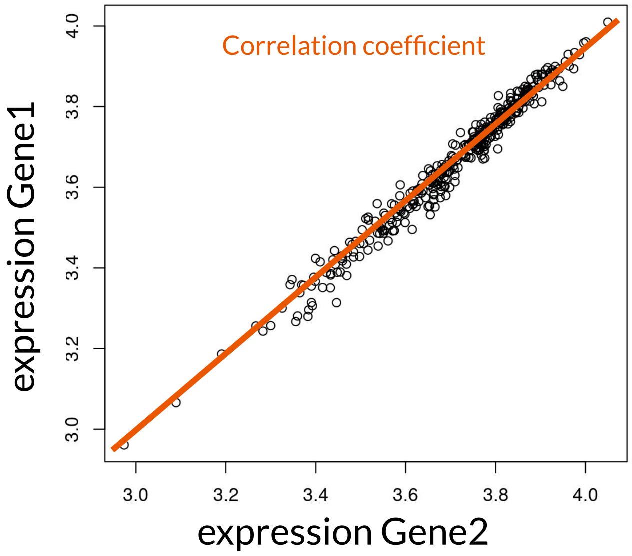
Weighted Gene Correlation Network Analysis (WGCNA)

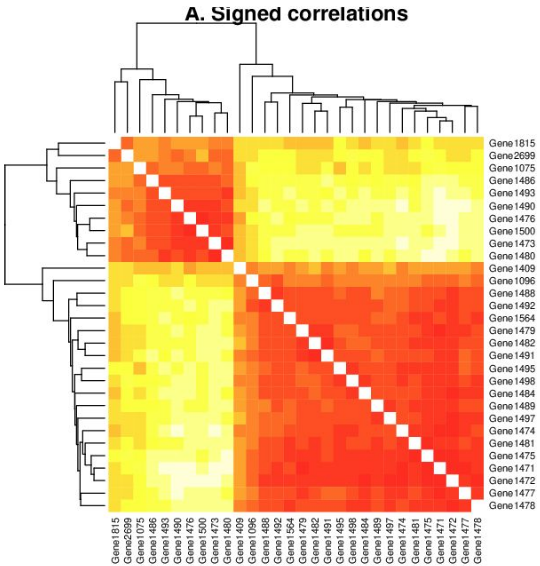
- **Gene co-expression**



Weighted Gene Correlation Network Analysis (WGCNA)

- ➢ Gene co-expression

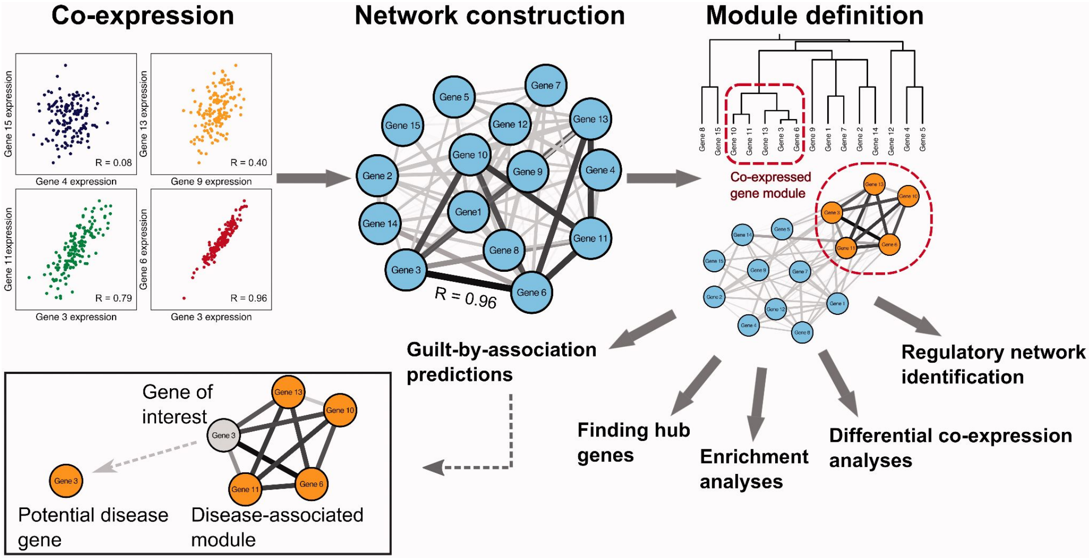




Source: Horvath, S., and P. Langfelder. “Tutorial for the WGCNA Package for R ” 2011.

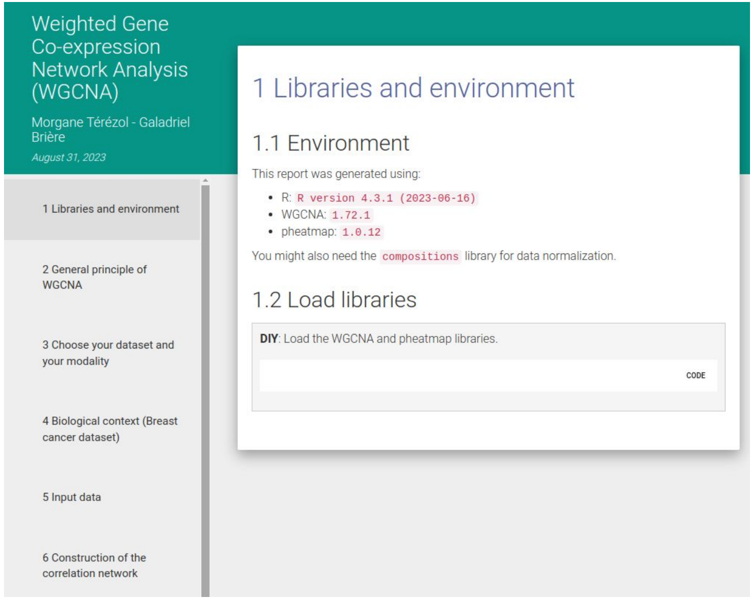
Weighted Gene Correlation Network Analysis (WGCNA)

➢ 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
- 2 General principle of WGCNA
- 3 Choose your dataset and your modality
- 4 Biological context (Breast cancer dataset)
- 5 Input data
- 6 Construction of the correlation network

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

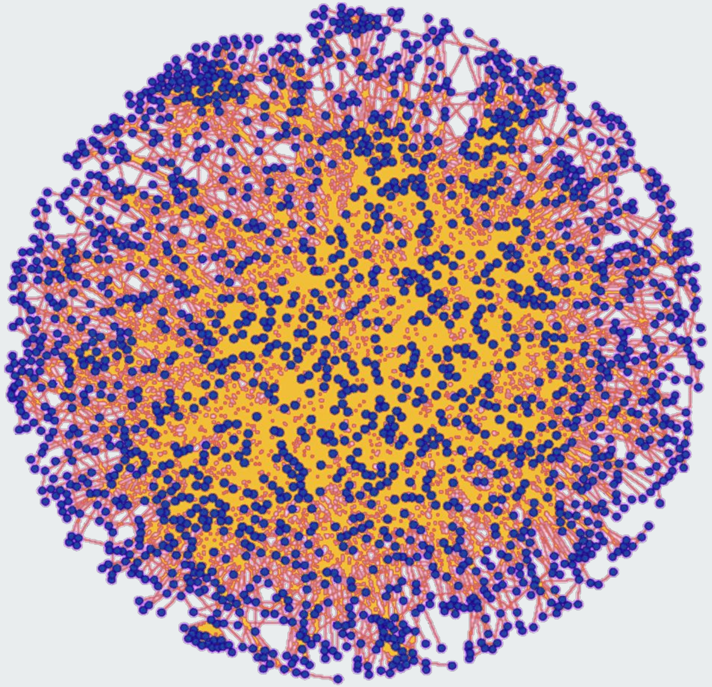
DIY: Load the WGCNA and pheatmap libraries.

https://github.com/sib-swiss/summer-school-multiomics-data-analysis-and-integration

- ➢ Tutorial and data are available in the Github repository
- ➢ R and Rstudio (or your favorite IDE)

Hands-on : SNF

Galadriel Brière, Morgane Térézol, Anaïs Baudot

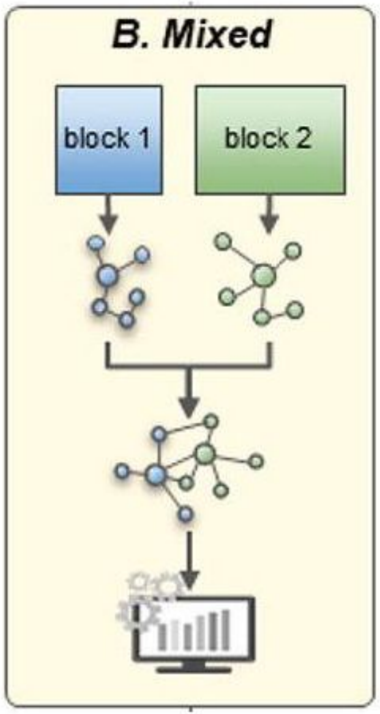


Similarity Network Fusion (SNF)

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

Similarity Network Fusion (SNF)

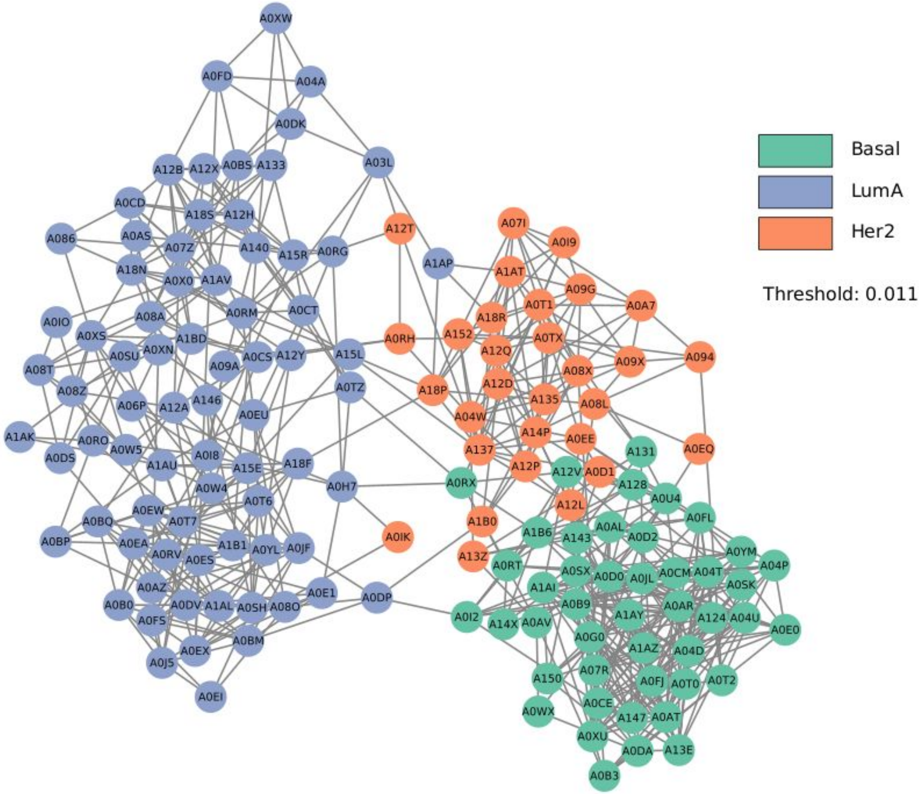
- Bo Wang *et al.*, Nature Methods. 2014
- **Mixed integration** method:
  - Sample **network** creation of each **data type**
  - **Fusion** of each data type networks



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)

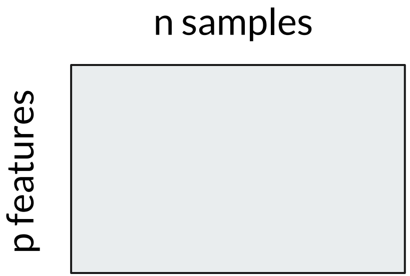
- Bo Wang *et al.*, Nature Methods. 2014
- **Mixed integration** method:
  - Sample **network** creation of each **data type**
  - **Fusion** of each data type networks
- **Sample** network
- Can manage:
  - **small number** of samples
  - noise
  - data **heterogeneity**
  - large number of features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

Similarity Network Fusion (SNF): Data preparation

- The most **important** step

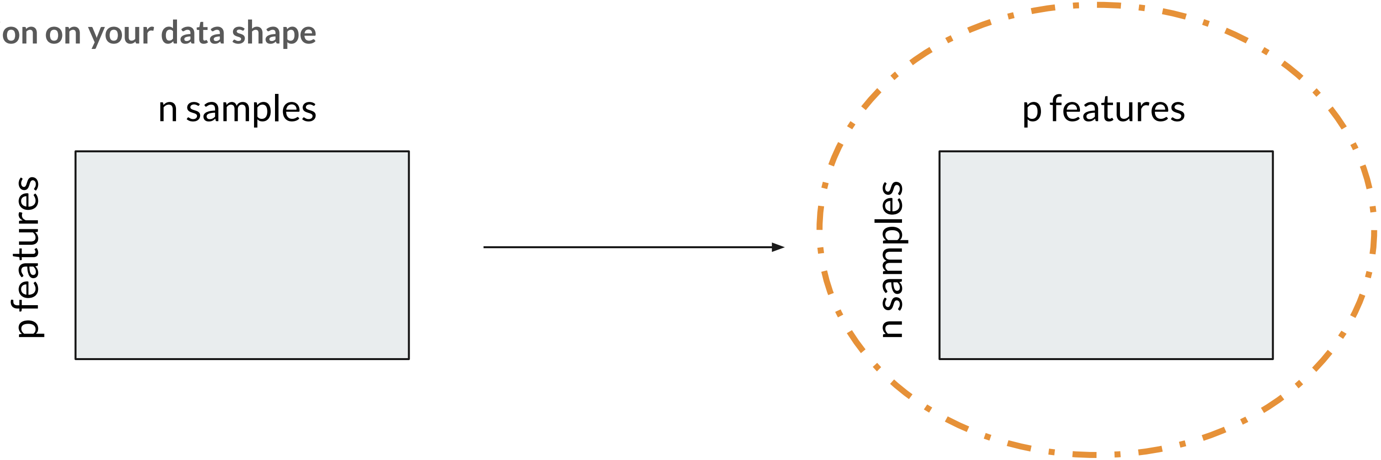
Similarity Network Fusion (SNF): Data preparation

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



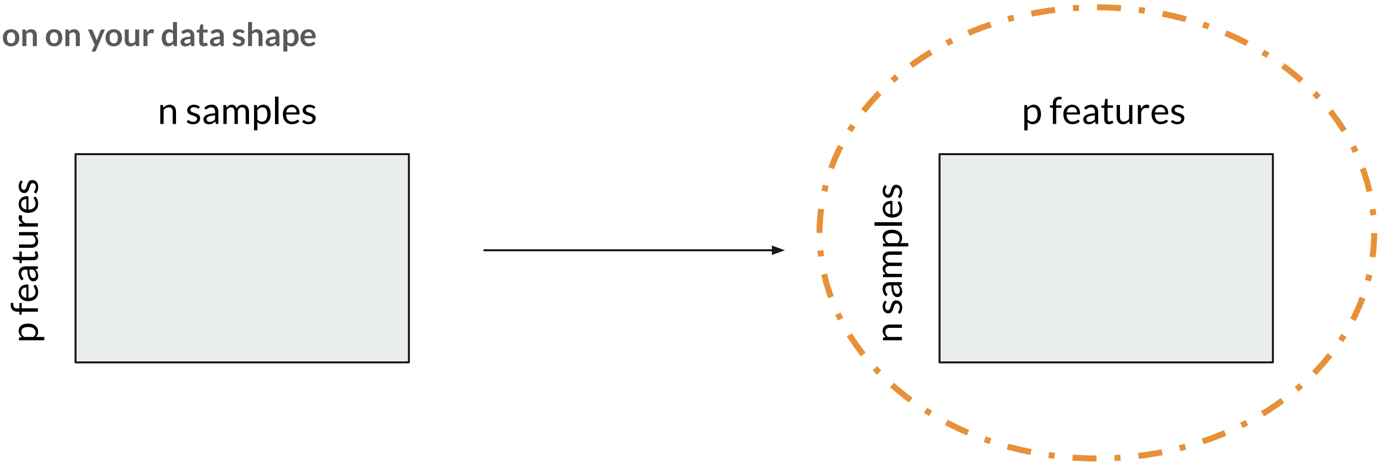
Similarity Network Fusion (SNF): Data preparation

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



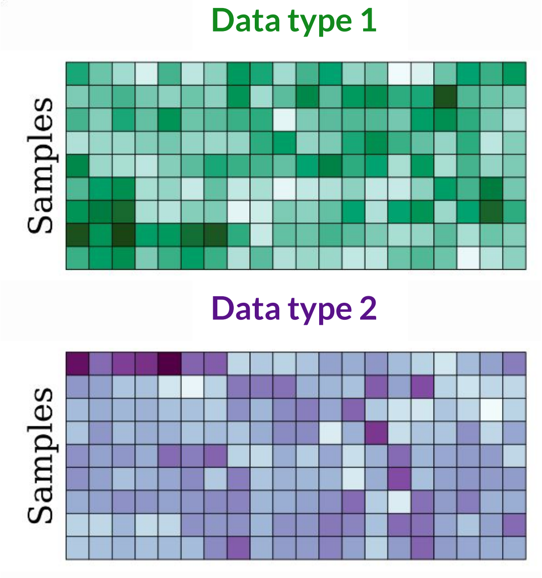
Similarity Network Fusion (SNF): Data preparation

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

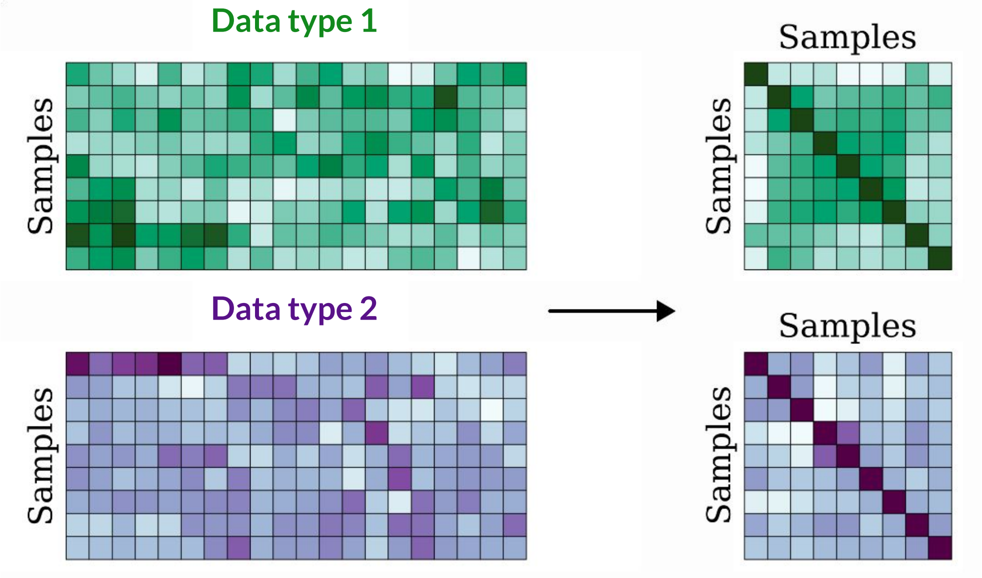


- ➢ **Normalization**: data should be normalized according their type

Similarity Network Fusion (SNF): Network creation

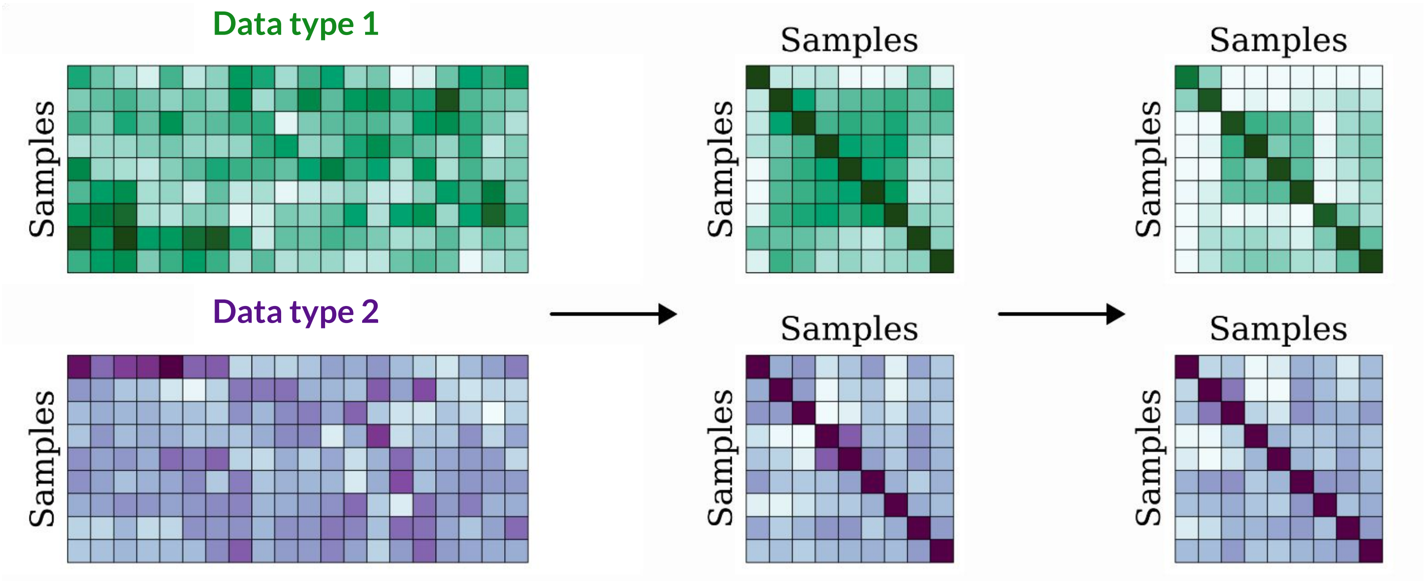


Similarity Network Fusion (SNF): Network creation



D
Distance matrix

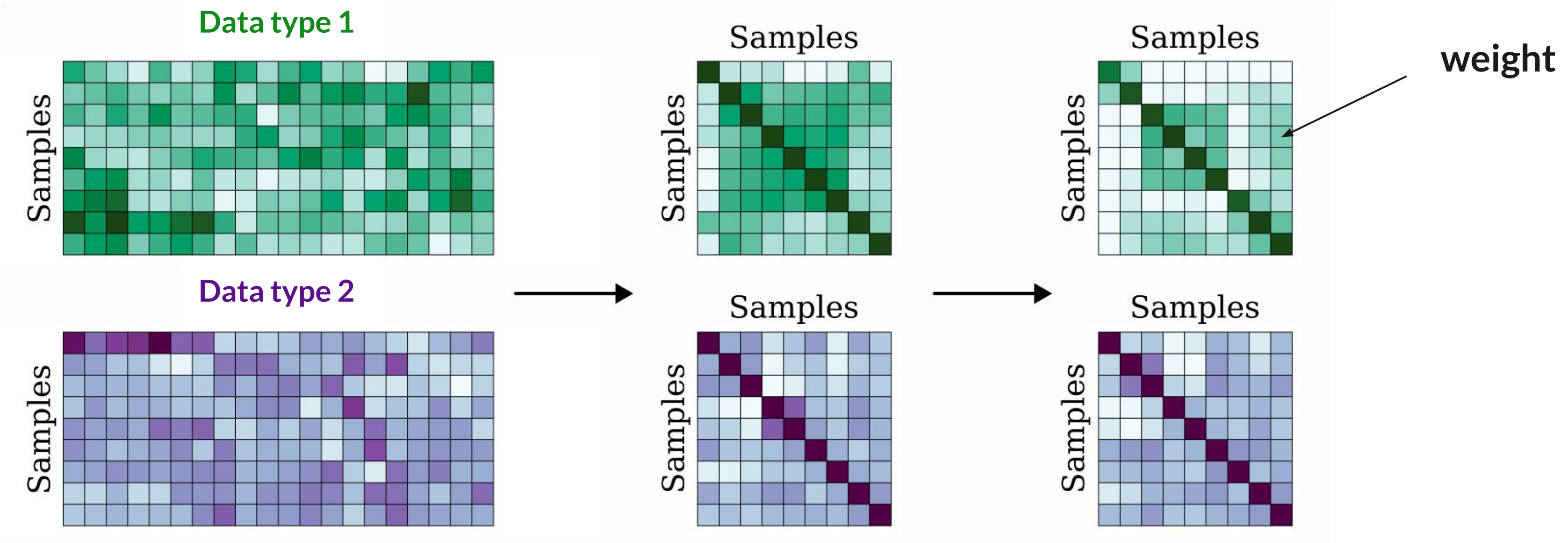
Similarity Network Fusion (SNF): Network creation



D
Distance matrix

W
Similarity matrix

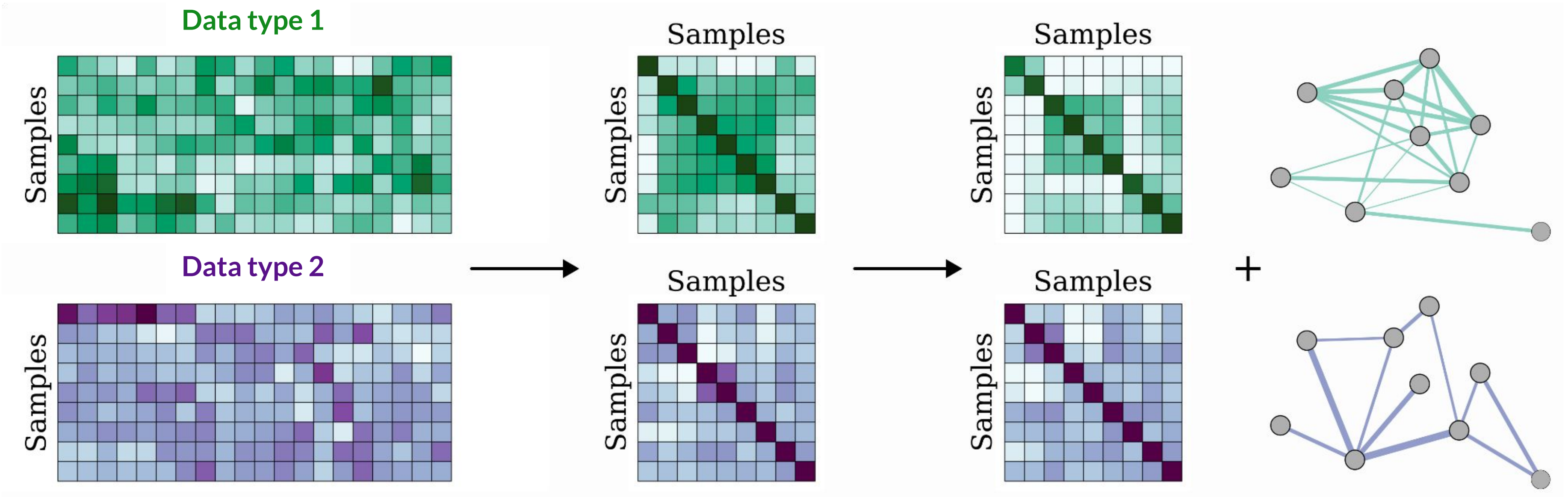
Similarity Network Fusion (SNF): Network creation



D
Distance matrix

W
Similarity matrix

Similarity Network Fusion (SNF): Network creation

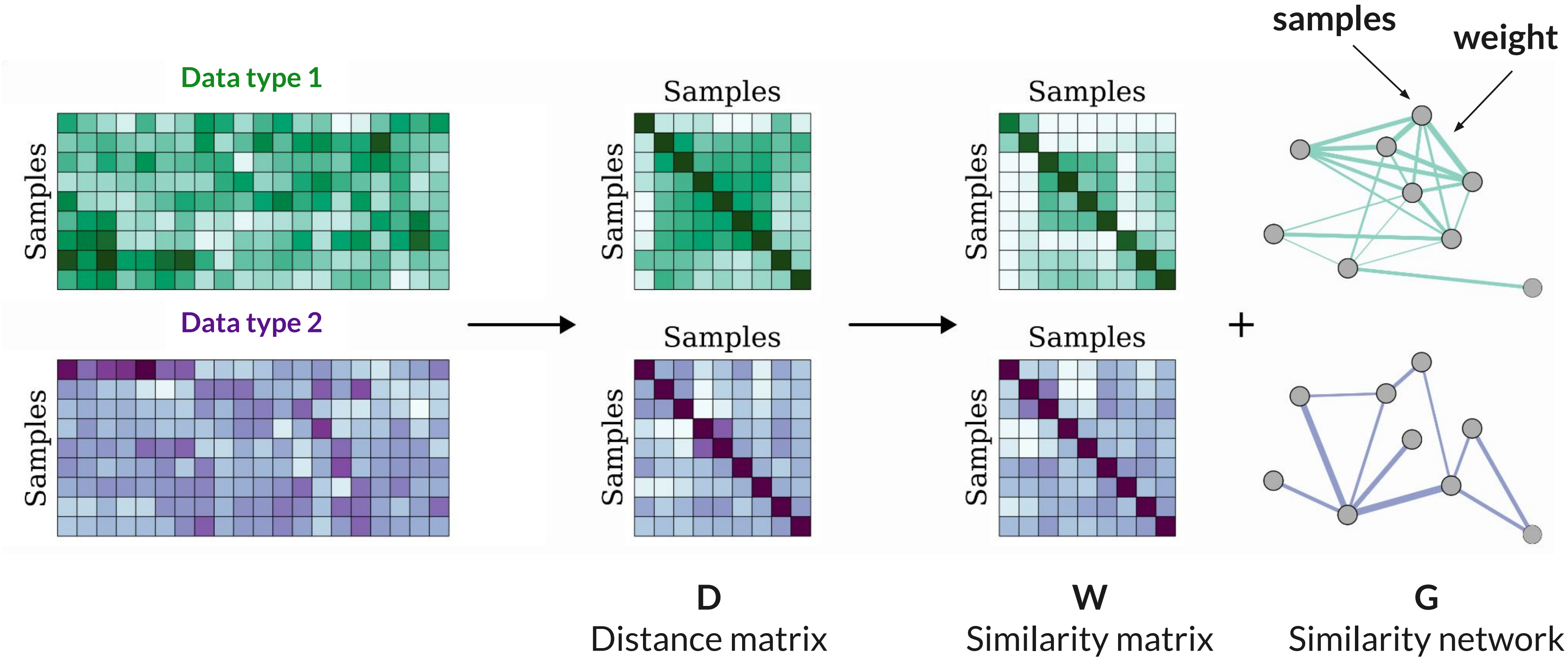


D
Distance matrix

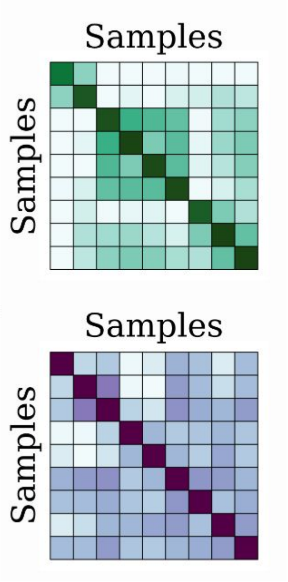
W
Similarity matrix

G
Similarity network

Similarity Network Fusion (SNF): Network creation



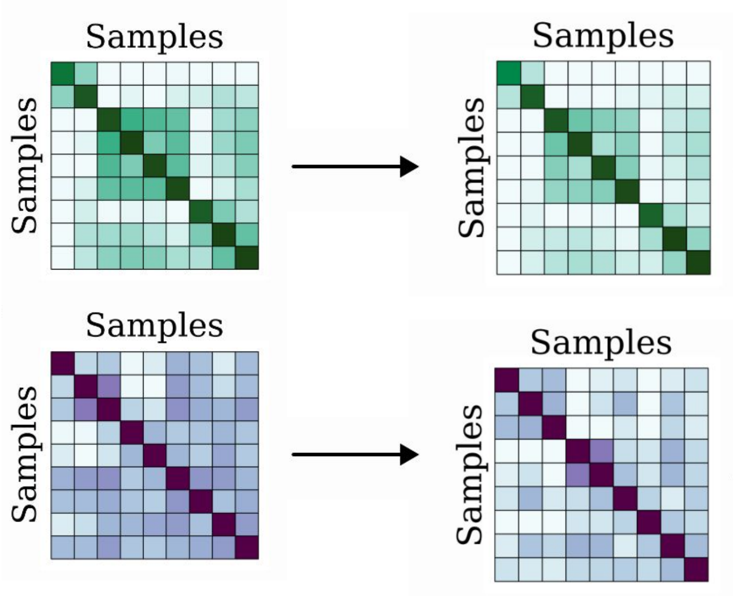
Similarity Network Fusion (SNF): Fusion



W

Similarity matrix

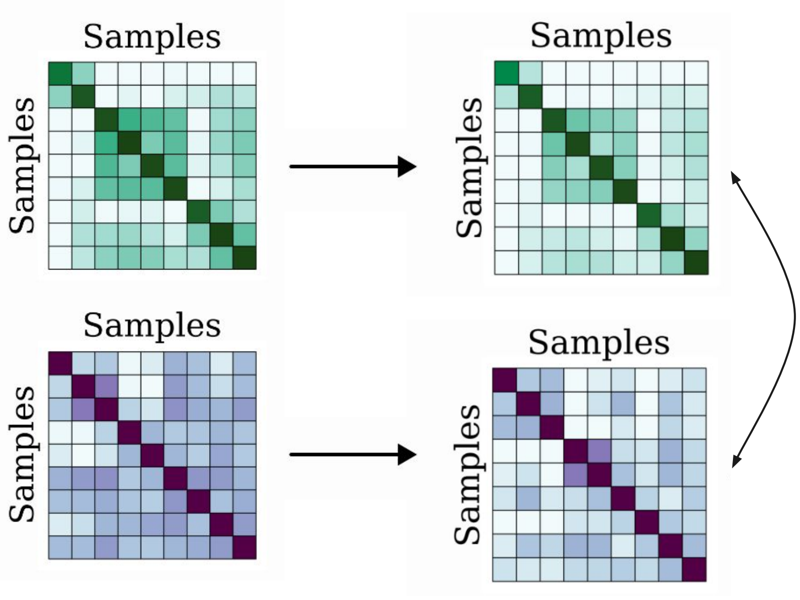
Similarity Network Fusion (SNF): Fusion



W
Similarity matrix

P(t)
Status matrix

Similarity Network Fusion (SNF): Fusion

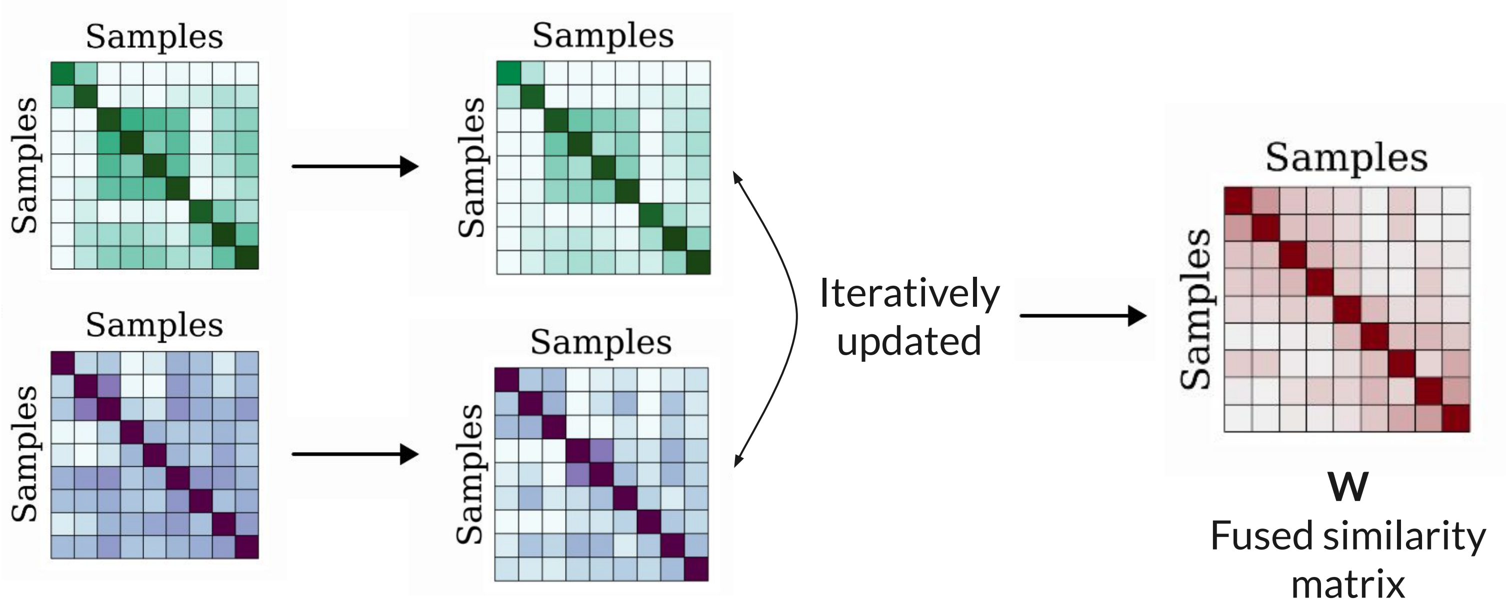


Iteratively
updated

W
Similarity matrix

P(t)
Status matrix

Similarity Network Fusion (SNF): Fusion

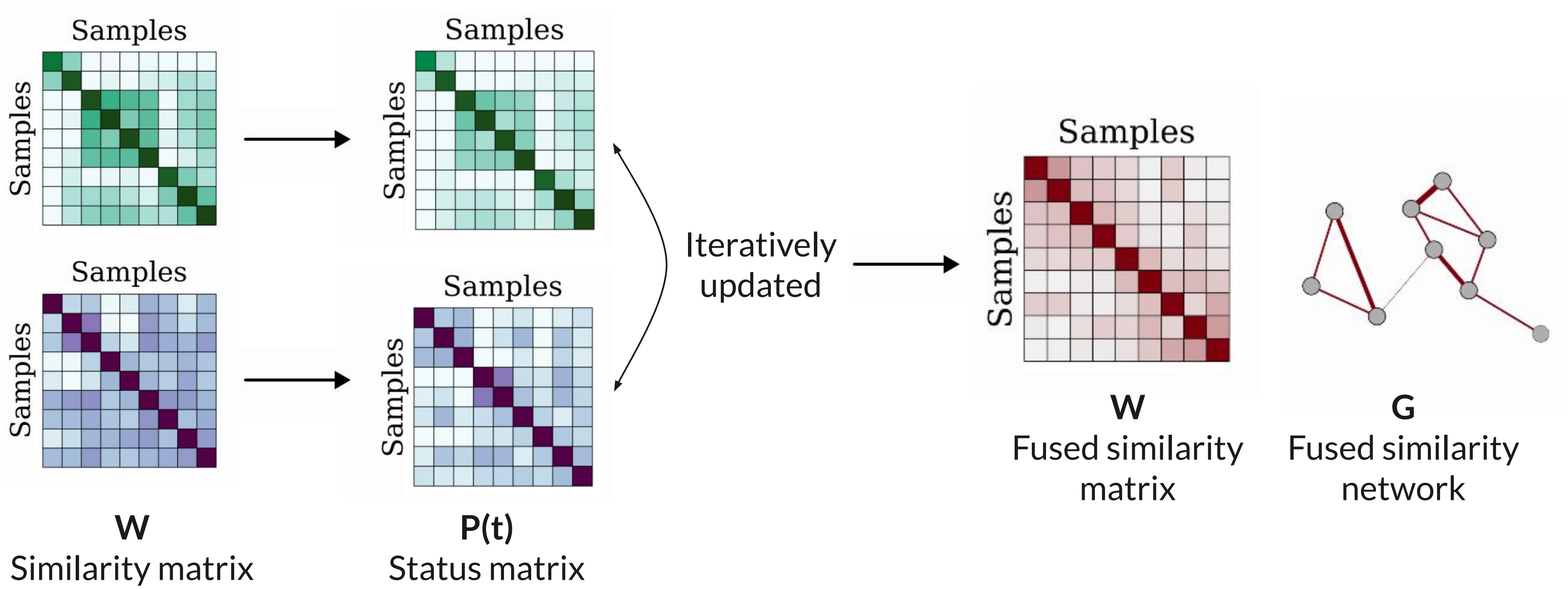


W
Similarity matrix

P(t)
Status matrix

W
Fused similarity matrix

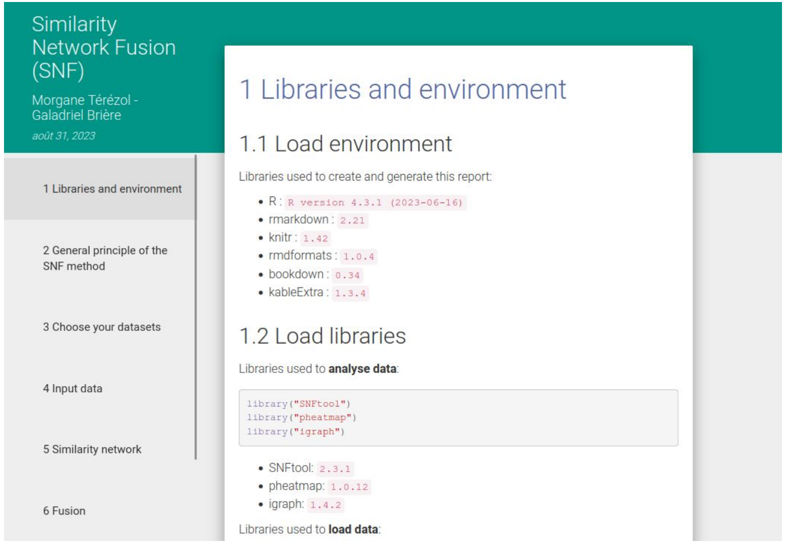
Similarity Network Fusion (SNF): Fusion



Hands-on

https://github.com/sib-swiss/summer-school-multiomics-data-analysis-and-integration

Hands-on in **SIB_SummerSchool_SNF.html**



Similarity Network Fusion (SNF)

Morgane Térézol - Galadriel Brière

août 31, 2023

- 1 Libraries and environment
- 2 General principle of the SNF method
- 3 Choose your datasets
- 4 Input data
- 5 Similarity network
- 6 Fusion

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**:

- ➢ Data are available in the Github repository
- ➢ R and Rstudio (or your favorite IDE)
- ➢ Visualization using Cytoscape