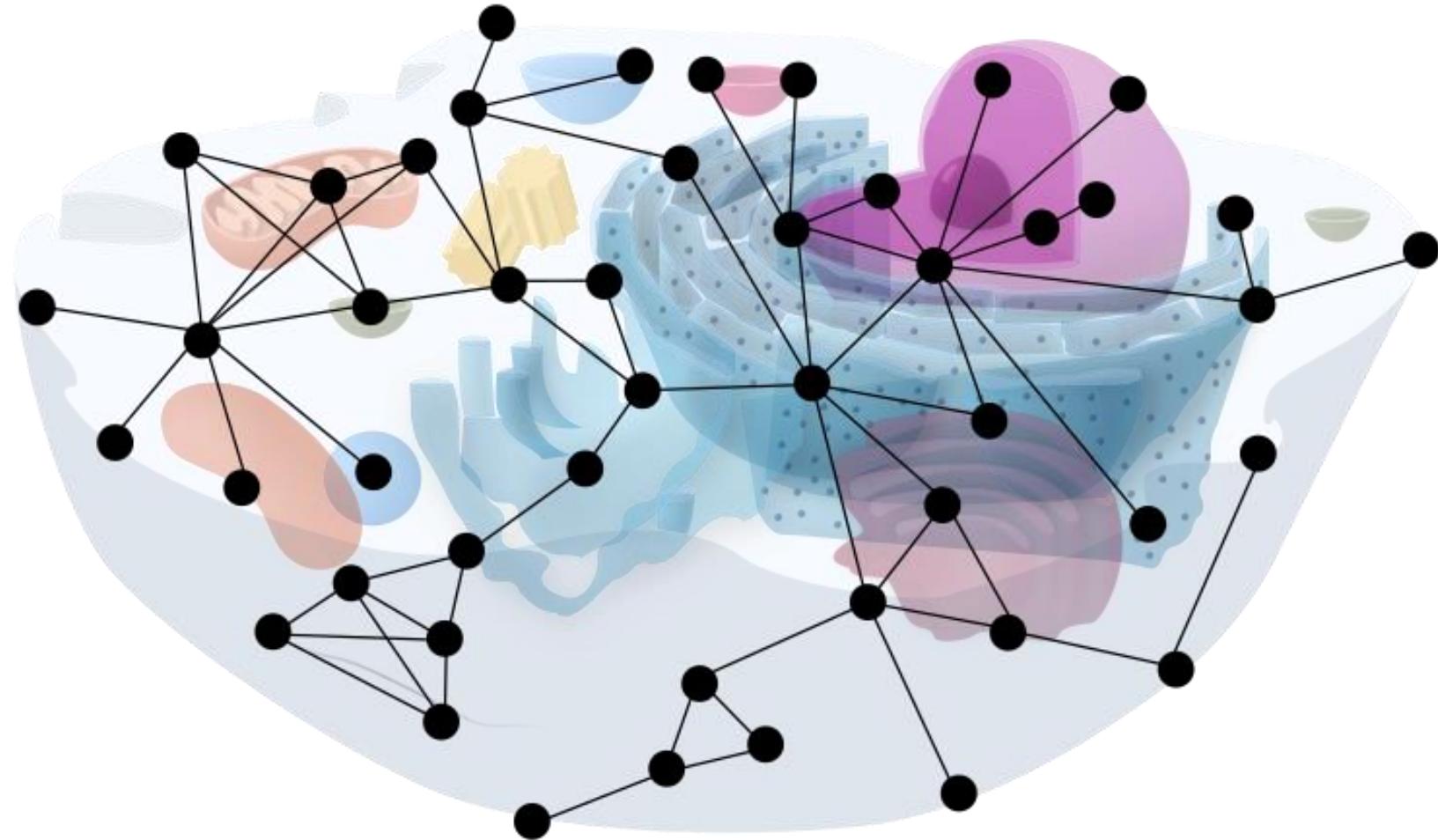


Connecting the Dots: Systems Biology at Every Scale



Ankit Agrawal
University of Würzburg
Germany

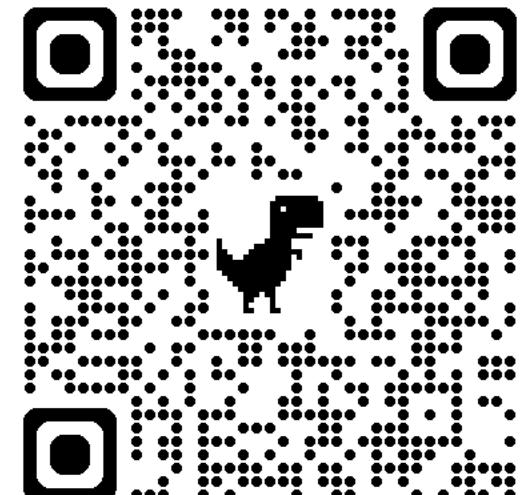
B403 Bioinformatics and Computational Biology

11. Systems Biology

(4 Lecture + 1 demo)

- Introduction, need for computers in system biology
- High-throughput and *omic* approaches, difference and application
- Graph theory (Network theory)
- Gene Ontology
- Demo: KEGG and gene ontology

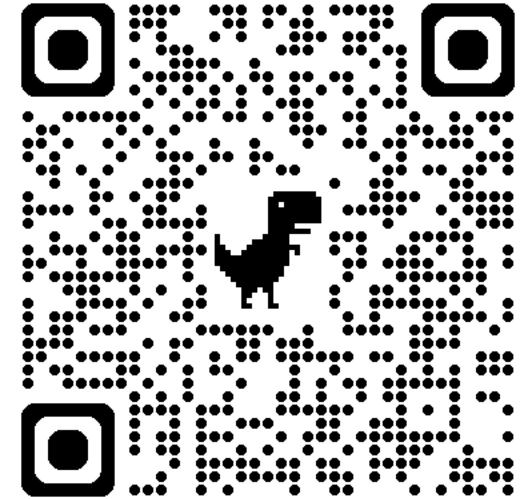
Scan me to download the presentation and the codes



Outline agenda

- Why do we need algorithms and computers
- What is systems biology
- Networks biology vs systems biology
- Biological networks are ubiquitous
- Basic component of a network
- Constructing and modeling networks

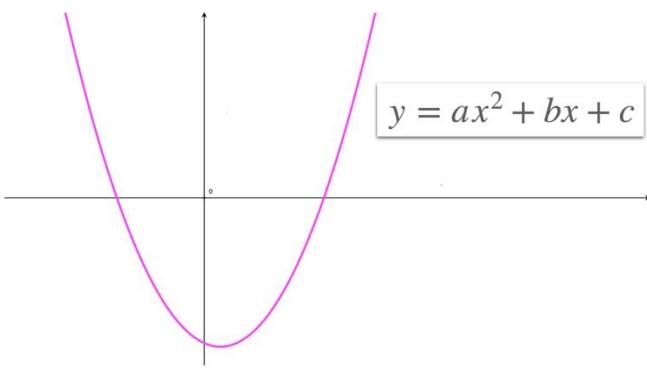
Scan me to download the presentation and the codes



What is a function?

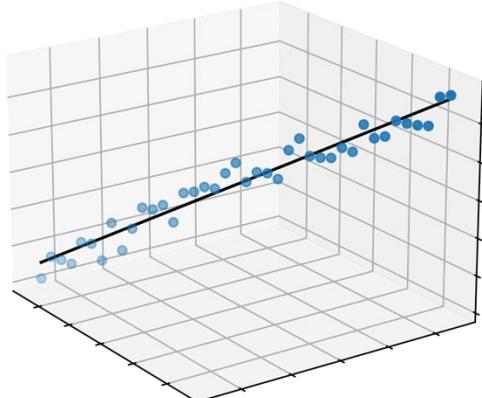


Parabola



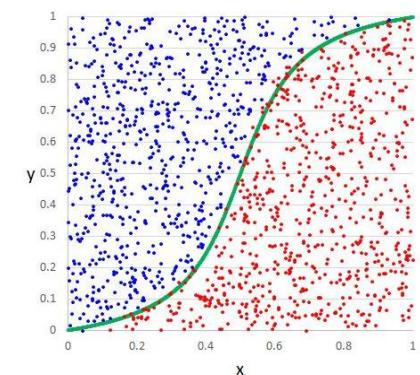
Linear Regression

$$y = a * x_1 + b * x_2 + \text{const}$$

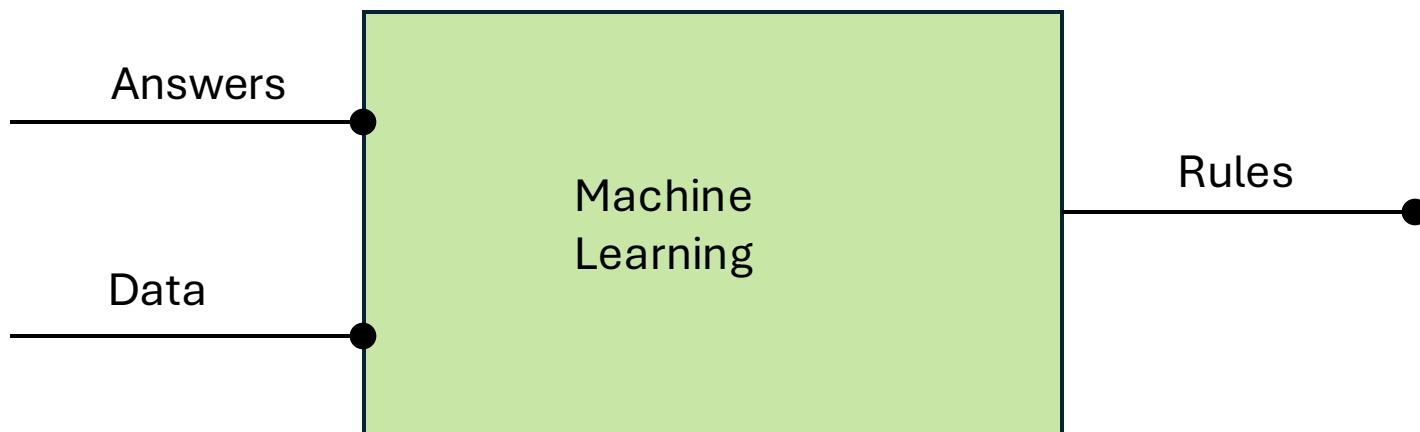


Logistic Regression

$$y = \frac{e^{(b_0 + b_1 x)}}{1 + e^{(b_0 + b_1 x)}}$$

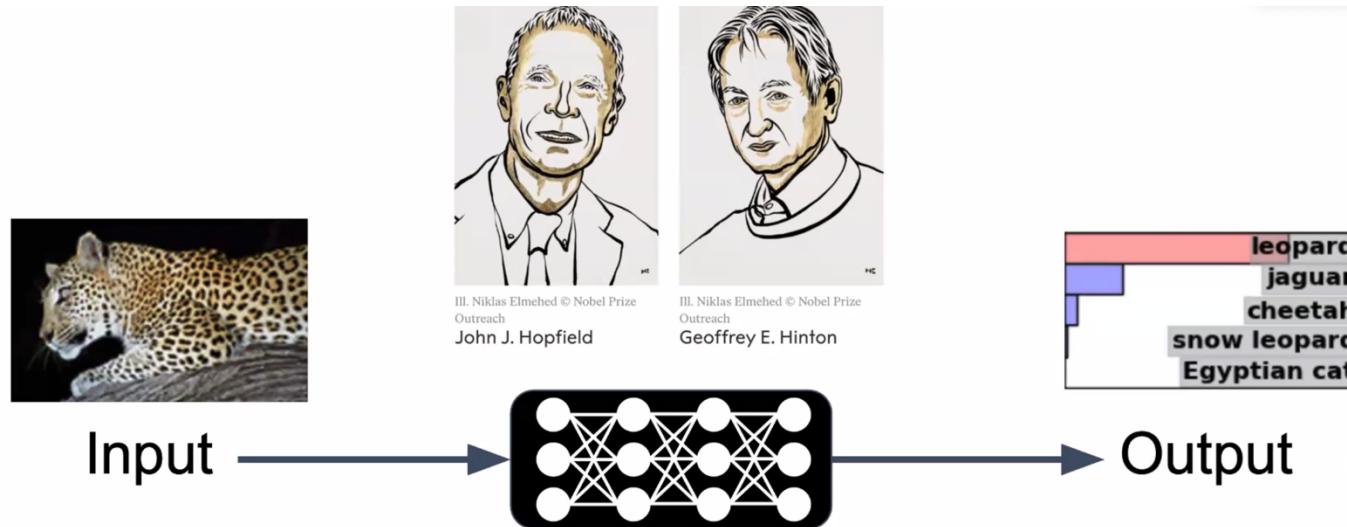


Programming vs Machine learning

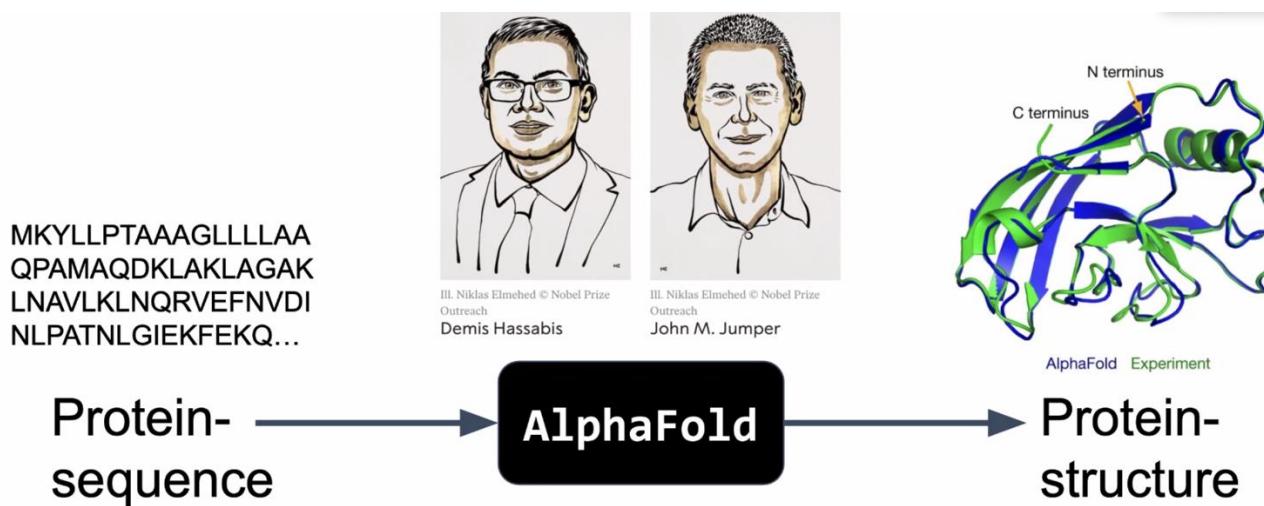


A machine learning task is to identify the underlying **function**, **rules** or **pattern** from the input data.

What is Artificial intelligence (AI)?

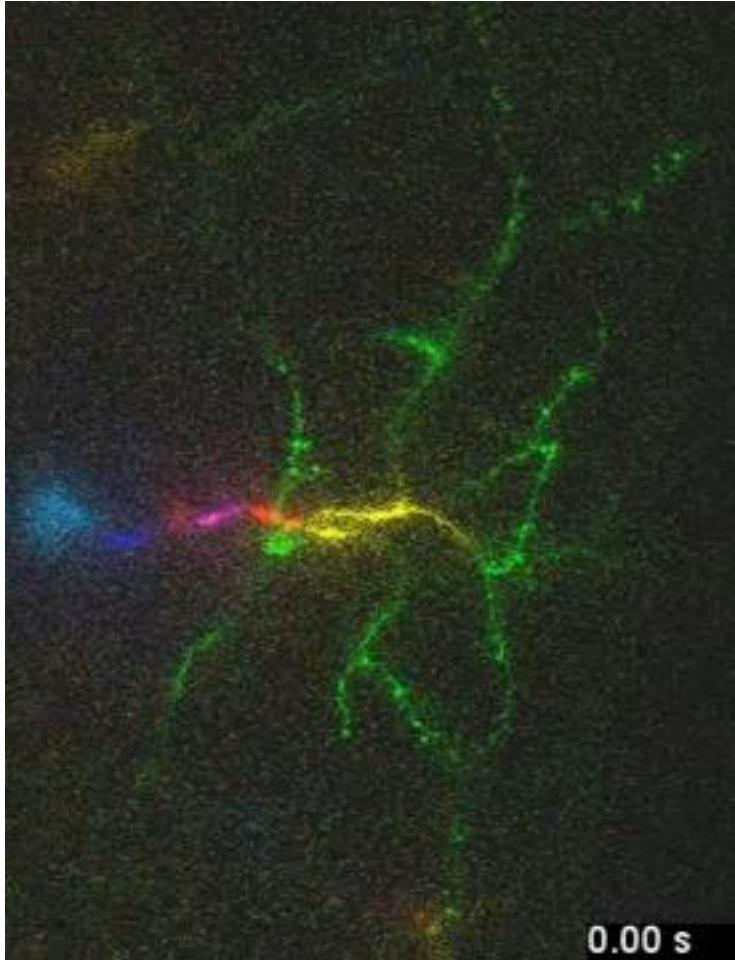


AI often uses machine learning as a method to understand data, find patterns, rules and **make decisions or predictions** that are **meaningful and relevant**.



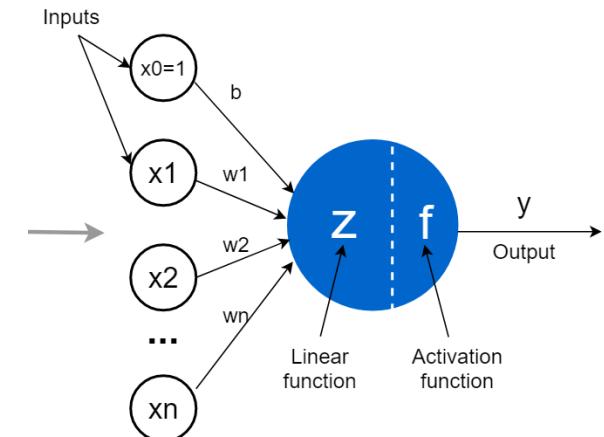
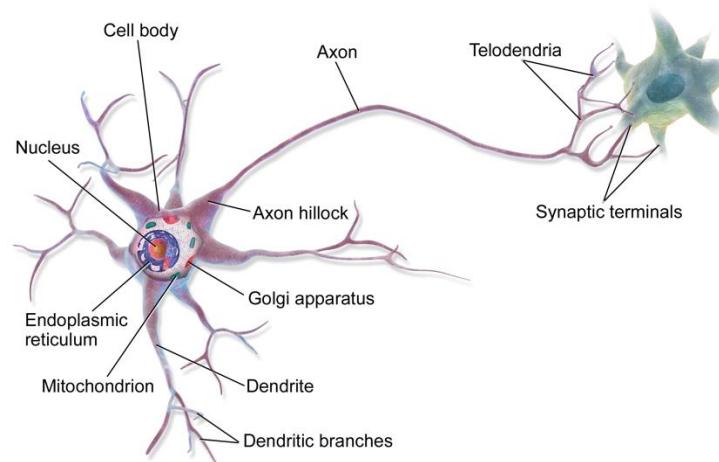
Some machine learning algorithms are biologically inspired

Coordination leads to collective intelligence



Youtube UC Berkeley Watch a neuron fire!

Swarm Intelligence



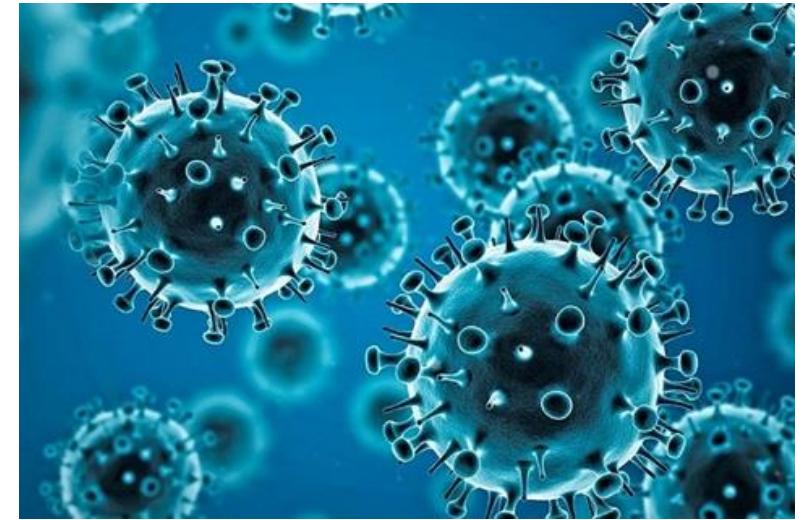
Emergent phenomenon



Smallest unit
Component or parts



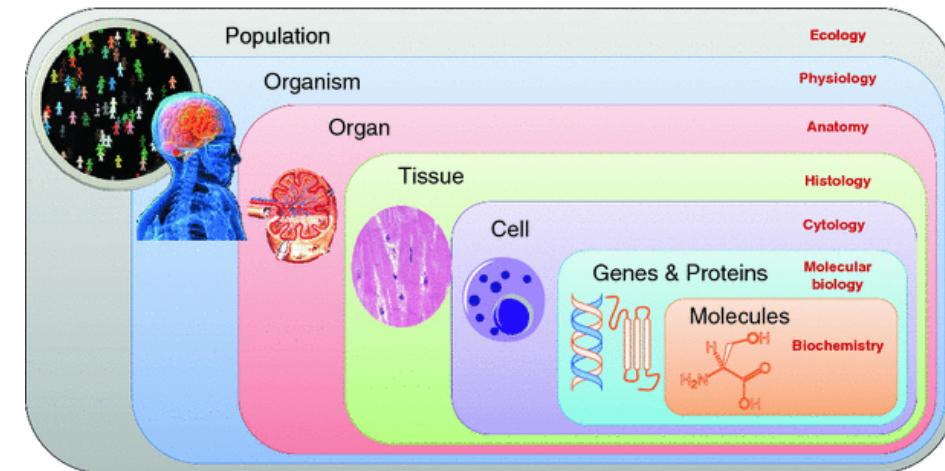
Smallest unit
Cells or molecules



Smallest unit
Single virus or protein

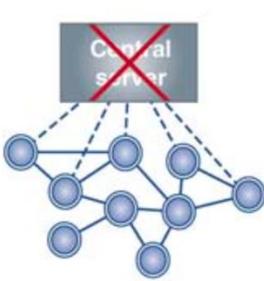
Systems biology (understand any biological system as a whole or system level)

- Systems Immunology
 - Seeks to understand the complex interplay between the immune system and the body across all scales.
- Systems-thinking skills
 - Describing a system's structure and organization
 - Reasoning about relationships within the system
 - Reasoning about the system as a whole
 - Analyzing emerges from studying the interactions of parts within the biological system.

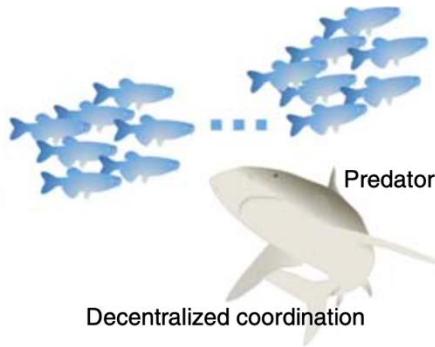


Shared principles between computational and biological systems

No central control: Molecules, cells, or organisms that interact, coordinate, and make decisions without central control.

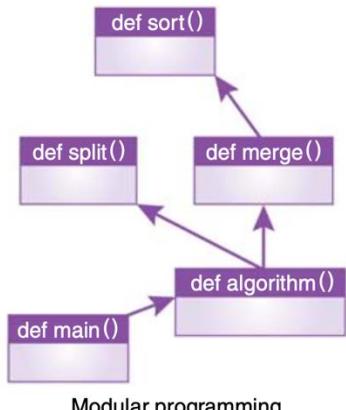


Decentralized systems

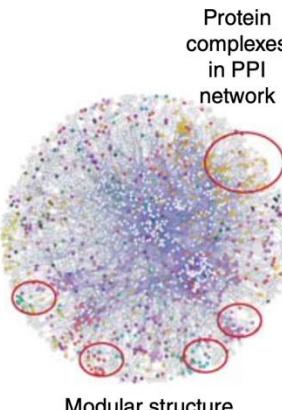


Decentralized coordination

Modular: they reuse certain components in multiple, and sometimes very different, applications.

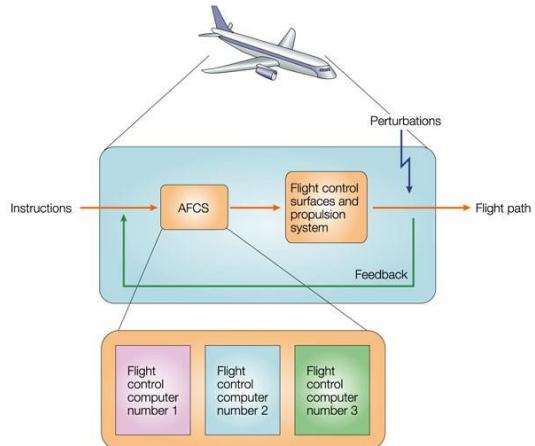


Modular programming

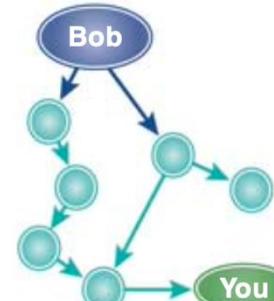


Modular structure

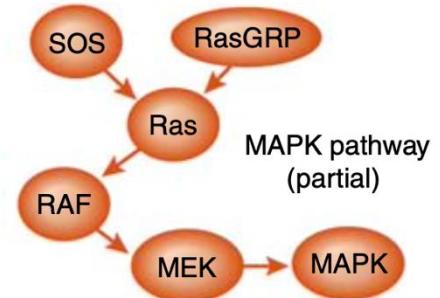
Robustness: biological processes need to be able to successfully handle failures and attacks to thrive.



Networks: serve as an important medium through which interactions occur, and information propagates.

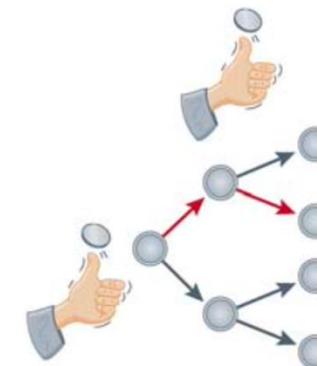
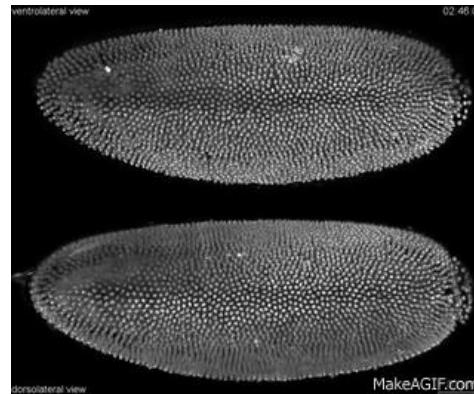


Rumor spreading

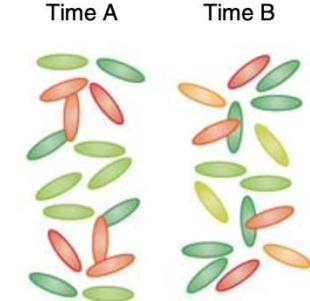


Signal propagation

Random: Biological processes are often stochastic



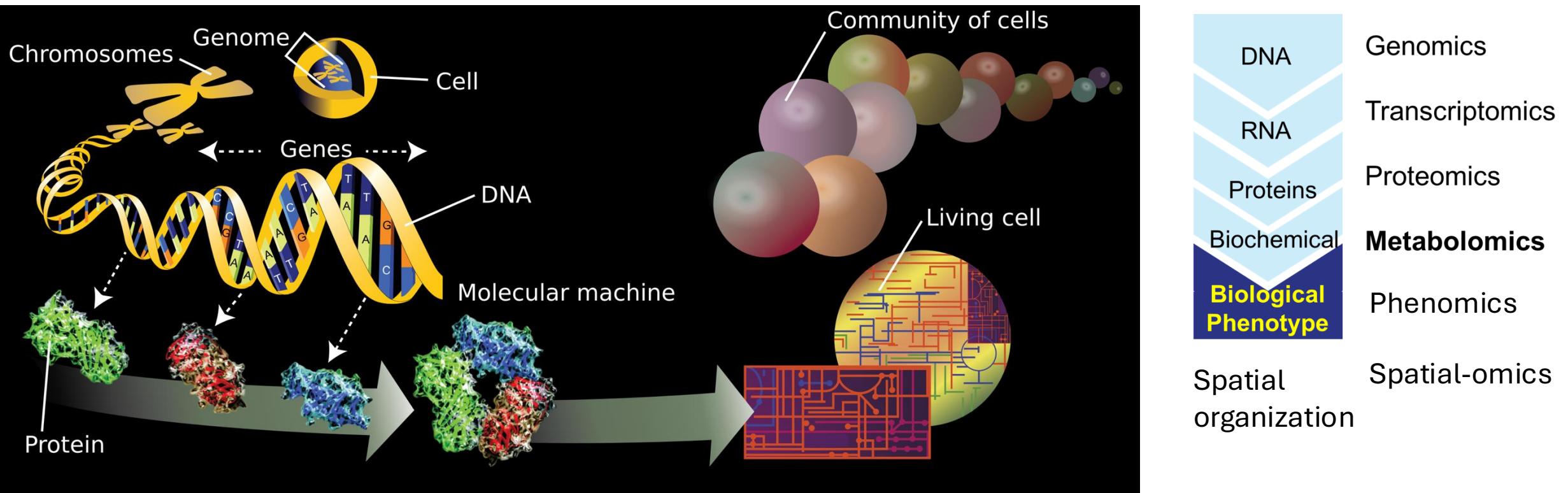
Randomized algorithms



Stochastic gene expression

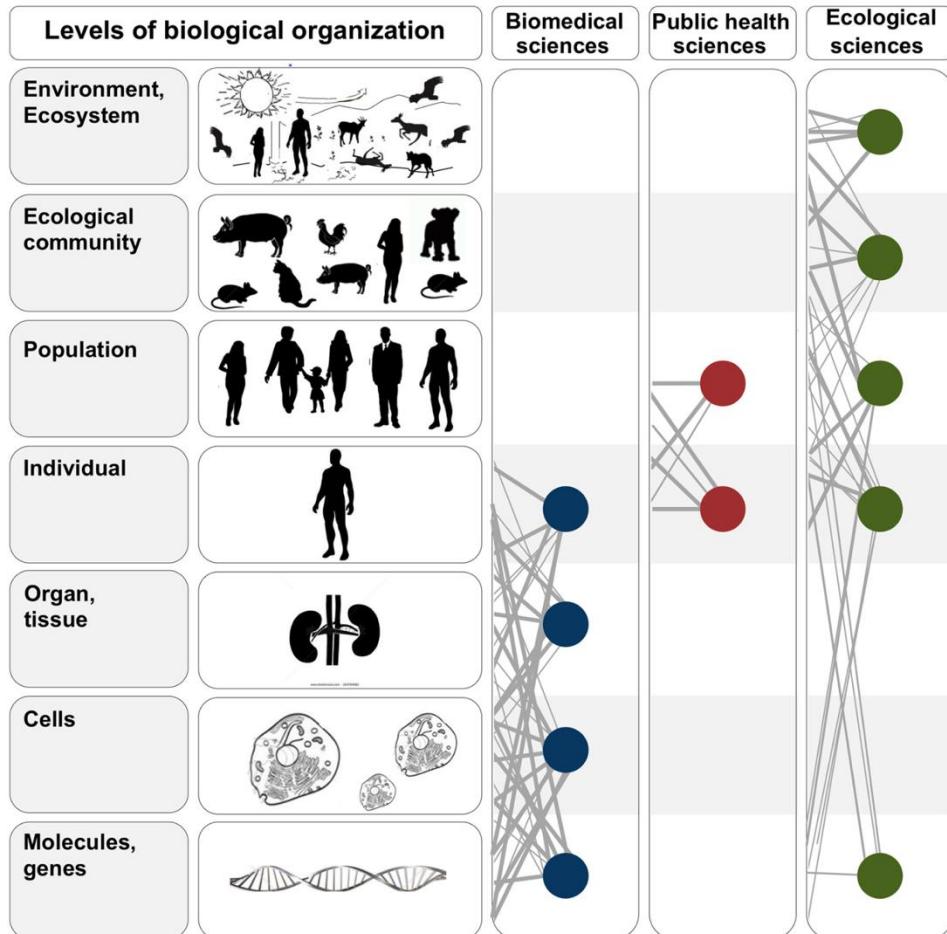
Systems biology vs omics

- Systems biology needs data to understand the systems as a whole
- Omics studies provide data

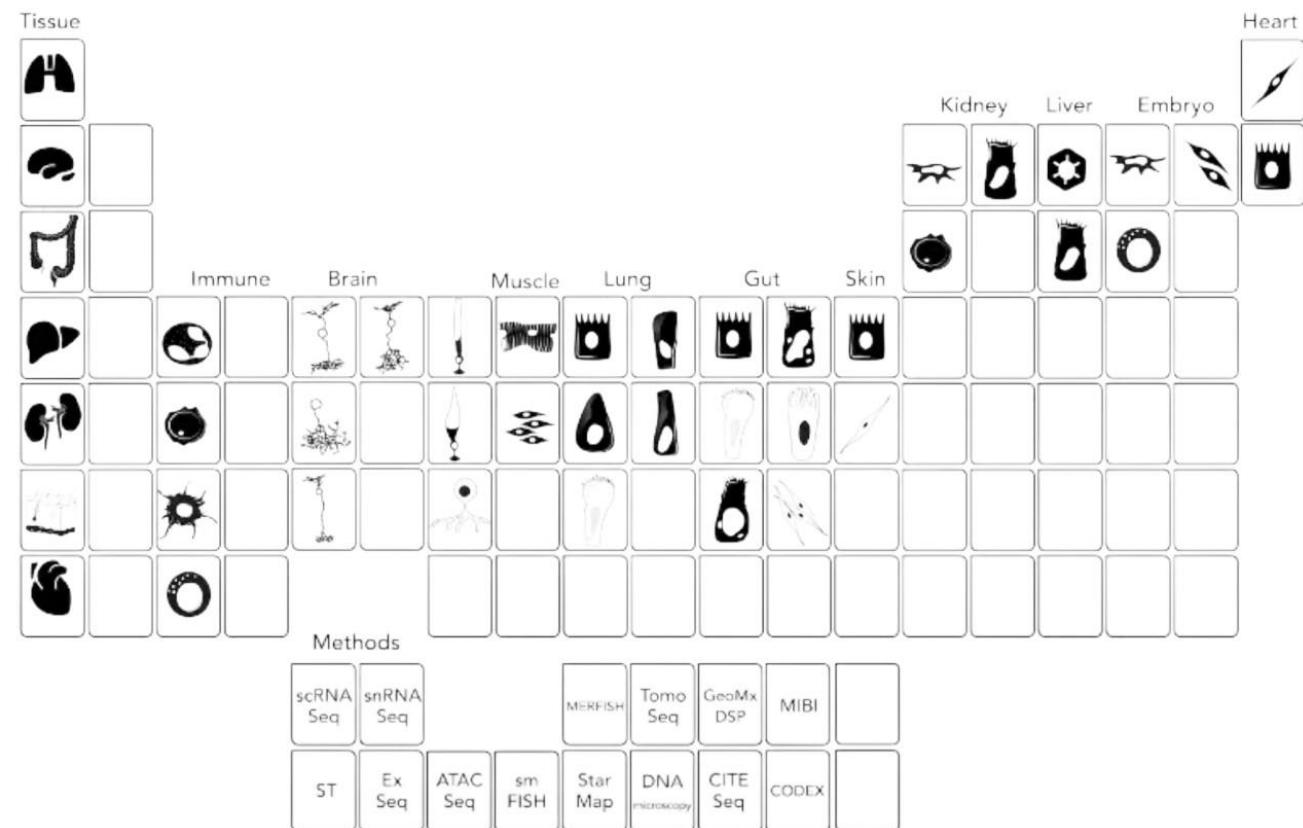


Network biology vs systems biology (Hierarchy of biological organization)

Concentrates on the structural and topological organization of biological systems



Focuses on the holistic understanding of biological systems, emphasizing how components interact to produce emergent behaviors



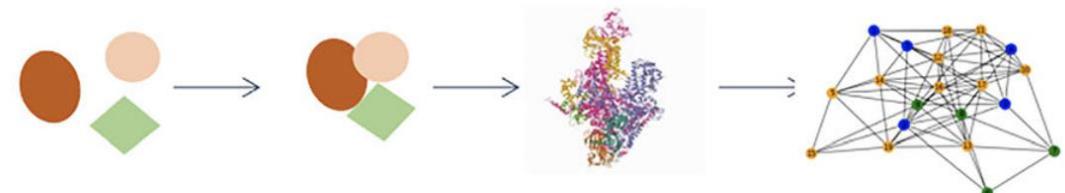
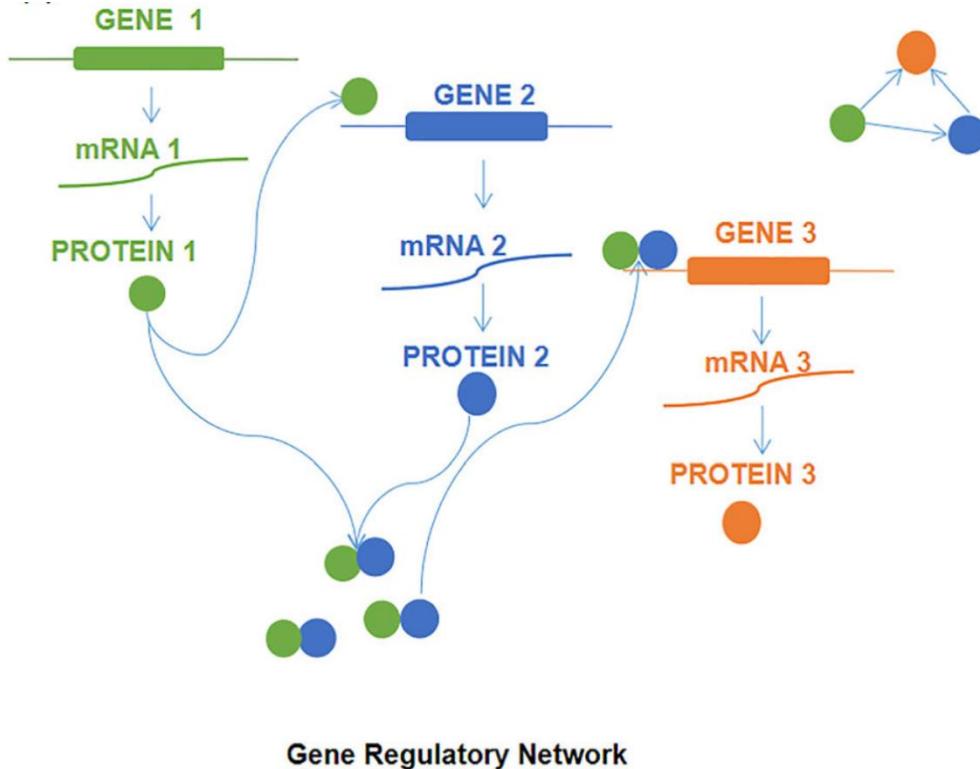
What is networks or graphs?

Prompt

Create a cartoon-style illustration showing the evolution of a person's social network from infancy to adulthood. Start with a baby in the center who first connects with their parents and siblings. As the child grows, add connections to extended family (grandparents, aunts, uncles), followed by friends from school, teachers, and neighbors. Then, show young adult connections branching out to university friends, mentors, and colleagues. Make the network web grow larger and more interconnected with each stage of life, symbolizing social expansion



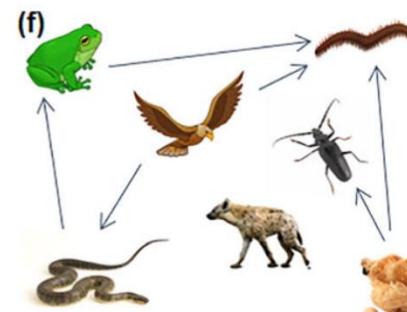
Types of Biological Network



Protein-Protein Interaction Network



Signal Transduction Network

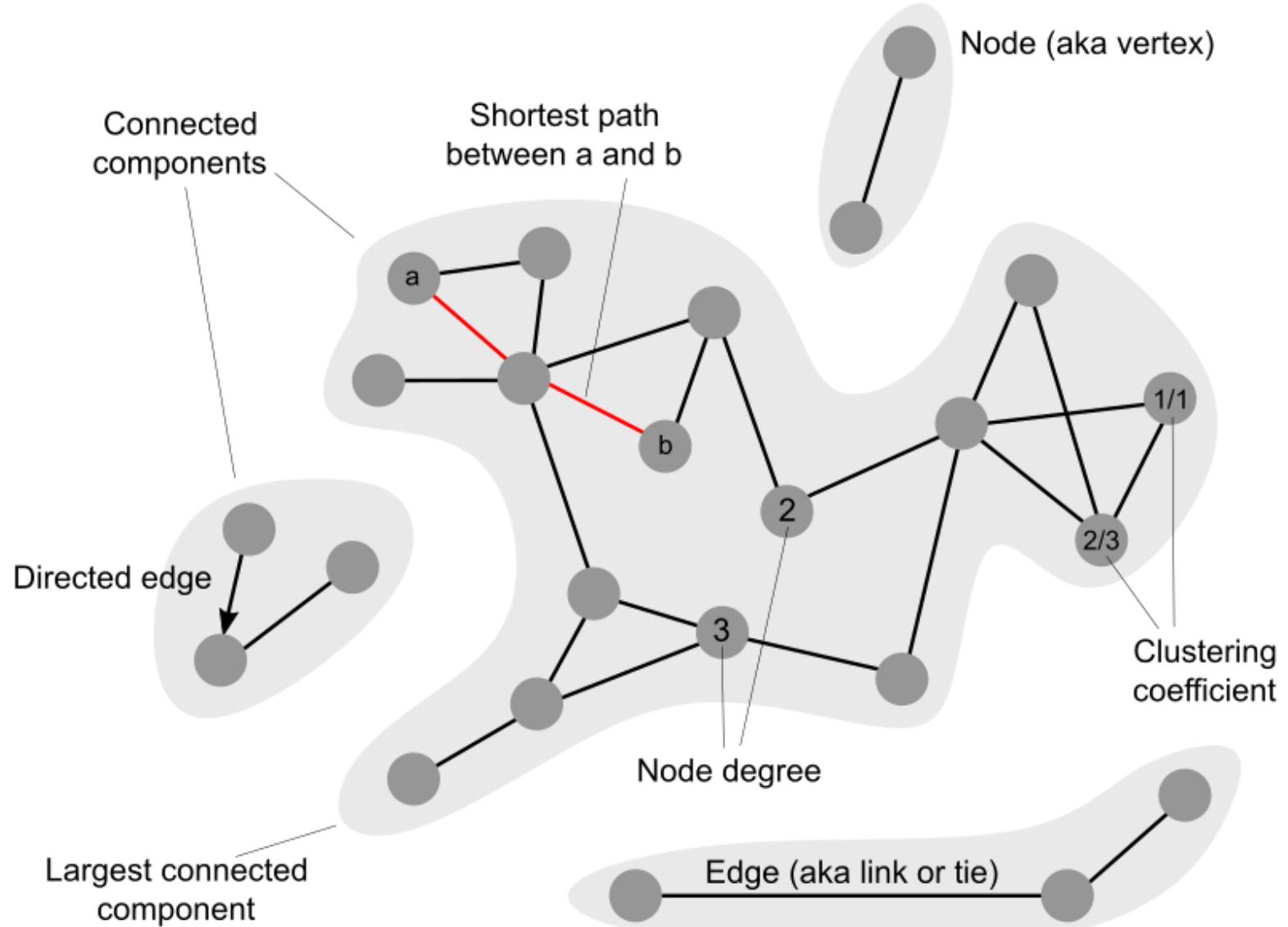


Ecological Network

Properties of network

Protein interactions
Nodes: amino acids, TF, proteins

Gene regulatory network
Gene → Gene
Gene → Protein → Gene
Protein → Gene



Graph representation

$G = (V, E)$,

V is a set of vertices

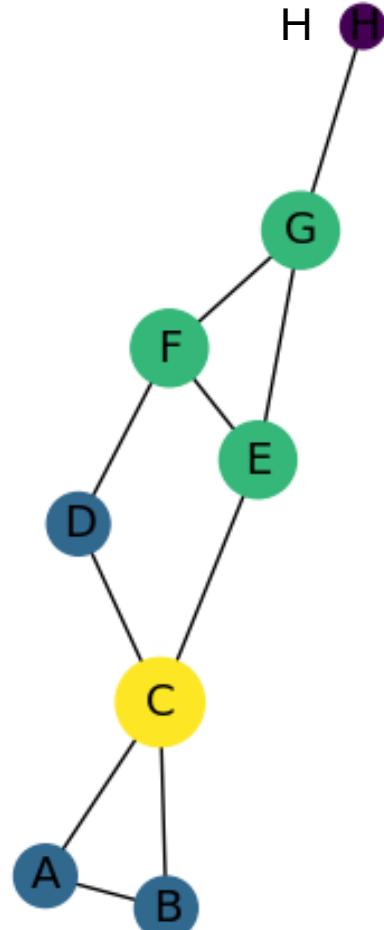
E is a set of unordered pairs of vertices, whose elements are called edges or links.

$G = (V, E)$

$V = \{A, B, C, D, E, F, G, H\}$

And

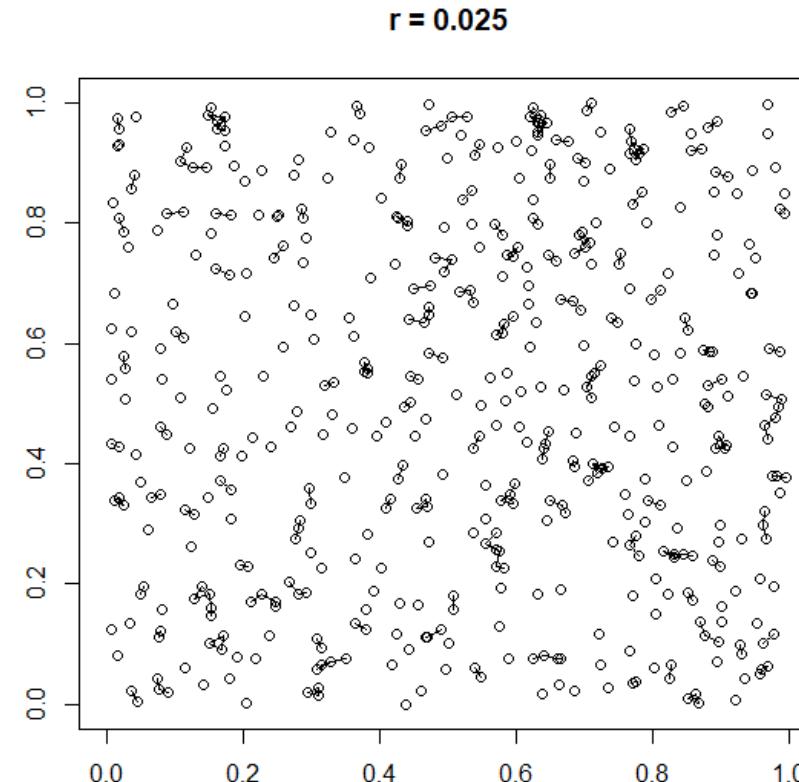
$E = \{(A, B), (A, C), (B, C), (C, D), (D, F), (C, E), (E, F), (F, G), (E, G), (G, H)\}$



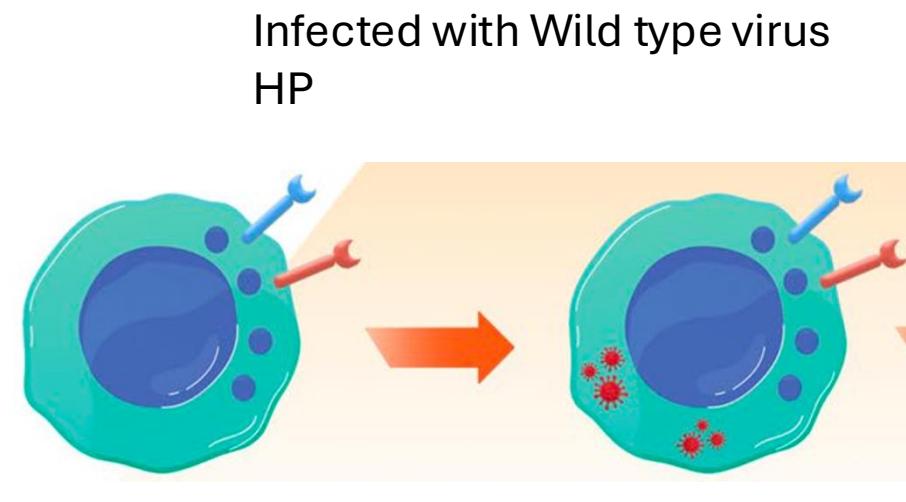
	A	B	C	D	E	F	G	H
A		1	1					
B	1				1			
C	1	1			1	1		
D				1			1	
E					1		1	1
F					1	1	1	
G						1	1	1
H								1

Random graph

- A graph is constructed by connecting nodes randomly.
- Each edge is included in the graph with **probability r** , independently from every other edge.

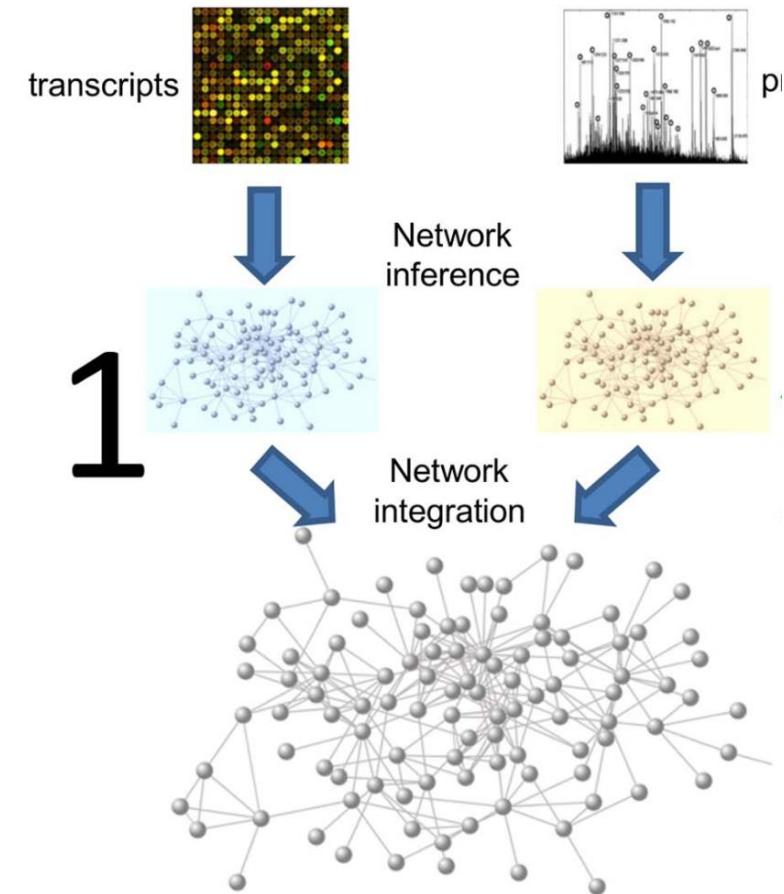


Experiment

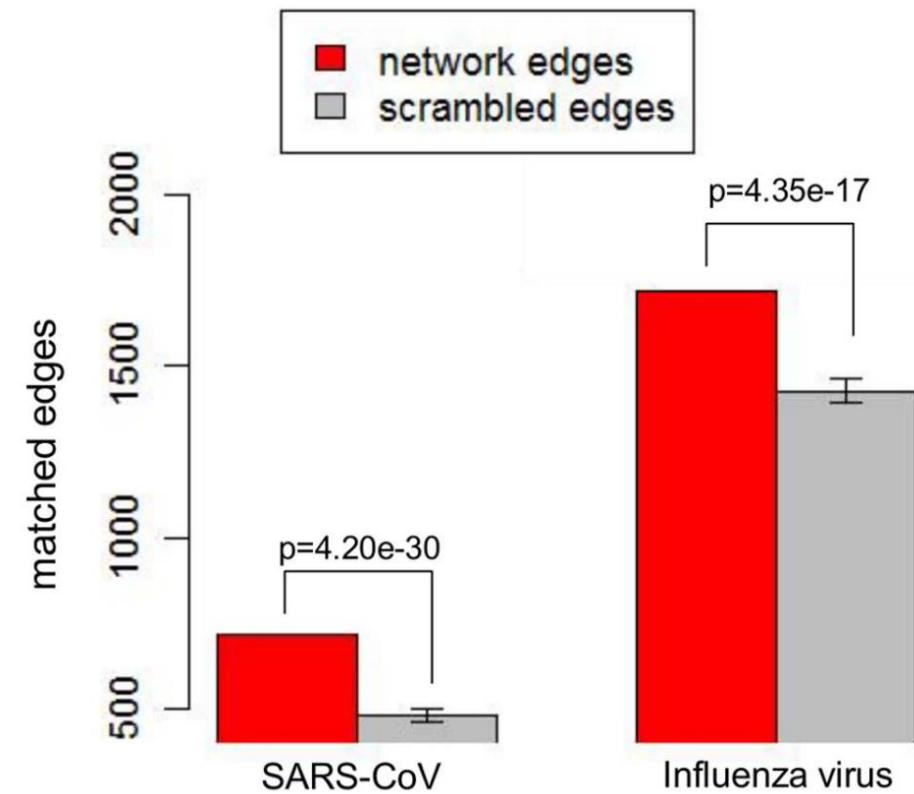
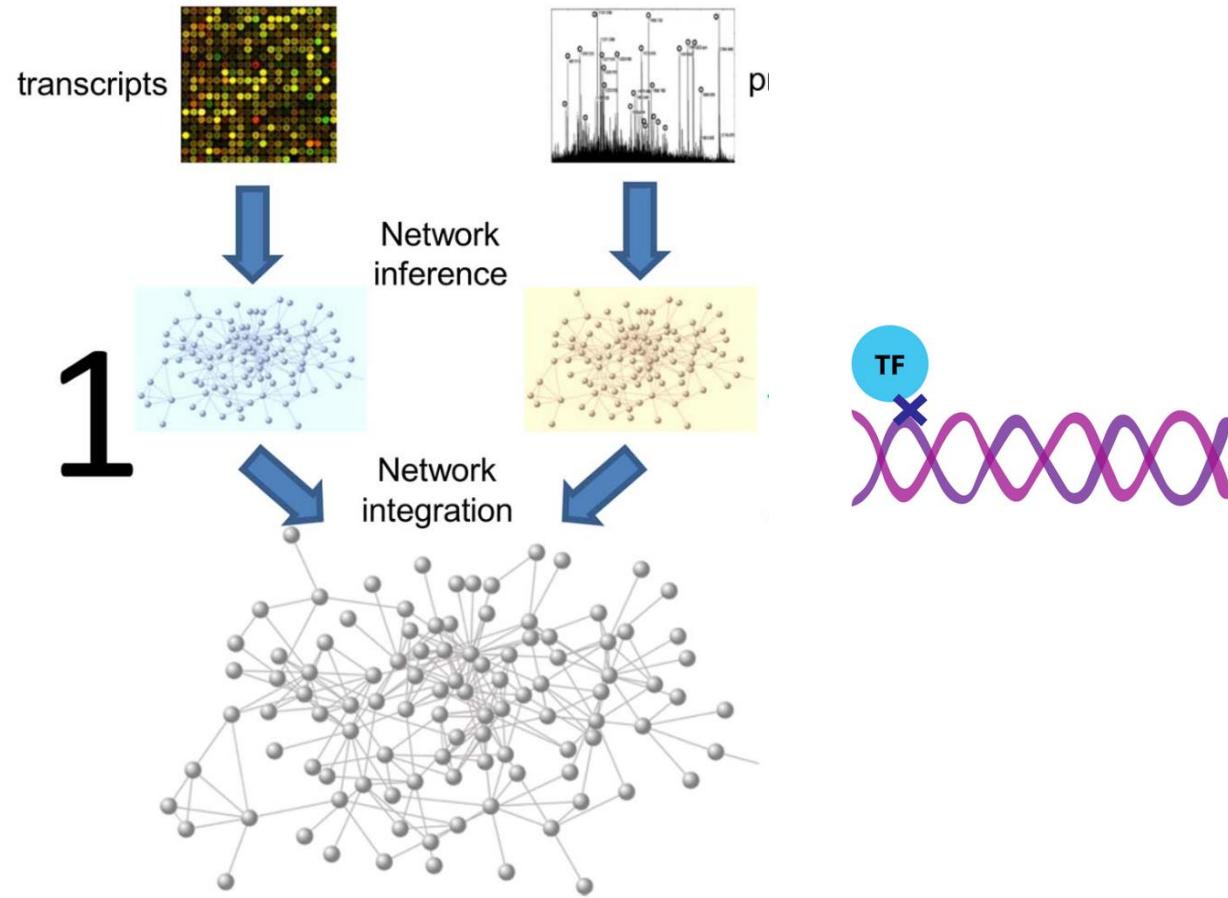


72 hours → Mircoarray RNA
And
Proteomics data

Can we predict genes that regulate
the host response to viral infection
based on their topological position?



Inferred network edge validation

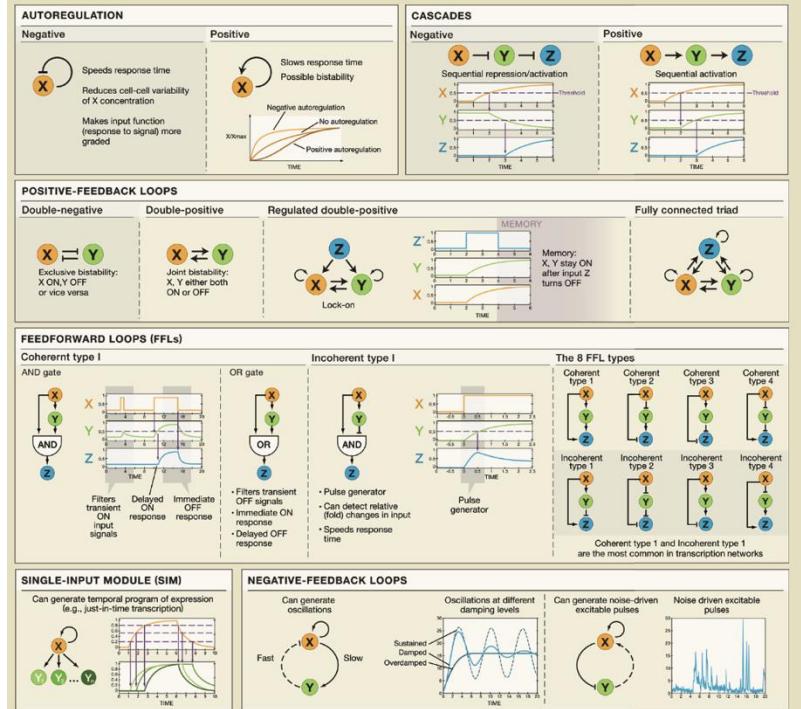


Let's have some tutorial with python
programming to make our first graph

Resources

SnapShot: Network Motifs

Oren Shoval and Uri Alon
Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel

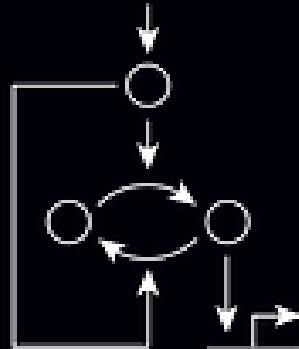


Oren Shoval and Uri Alon, Cell 2010

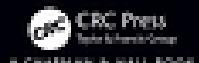
SECOND EDITION

An Introduction to Systems Biology

Design Principles of Biological Circuits



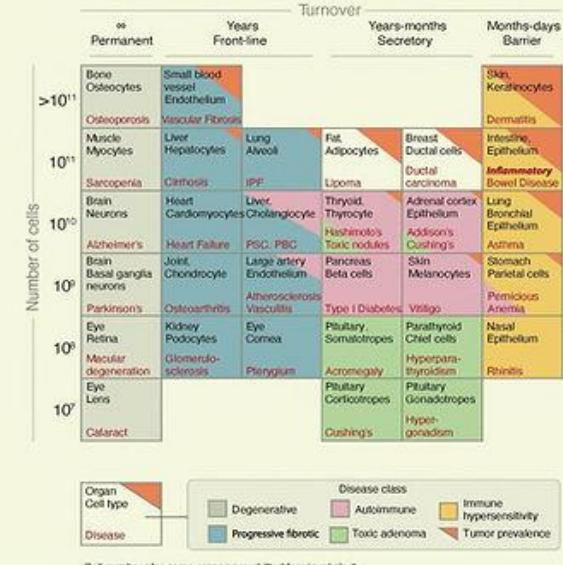
Uri Alon



A CHAPMAN & HALL BOOK

7 New Chapters

Systems Medicine



Cell numbers for some organs are shifted for visual clarity

Physiological Circuits
and the Dynamics of Disease

Uri Alon

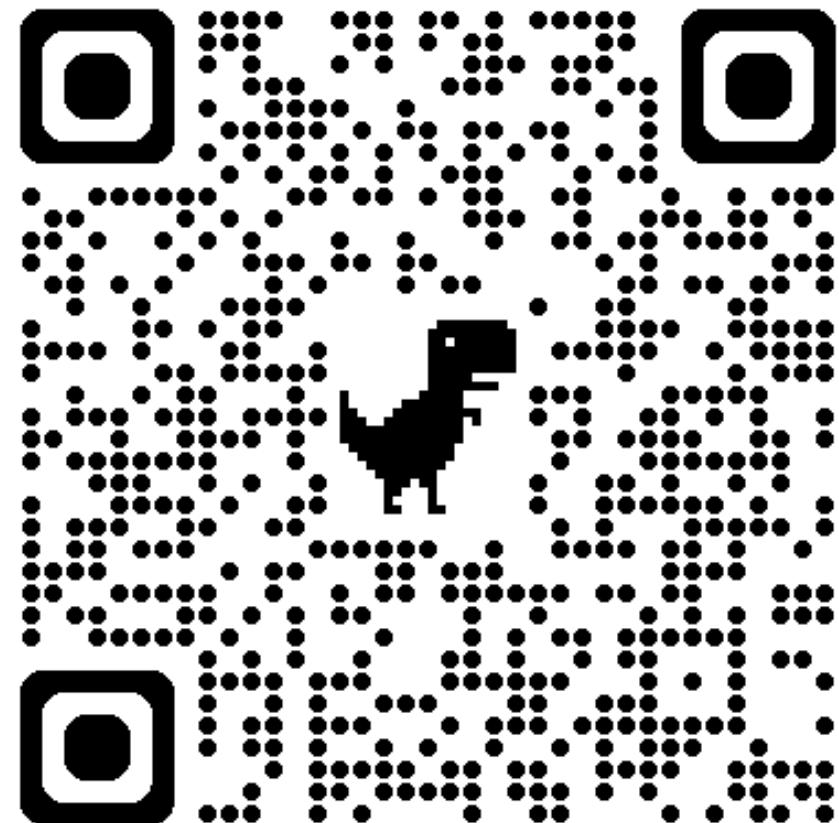


A CHAPMAN & HALL BOOK

Copyrighted material

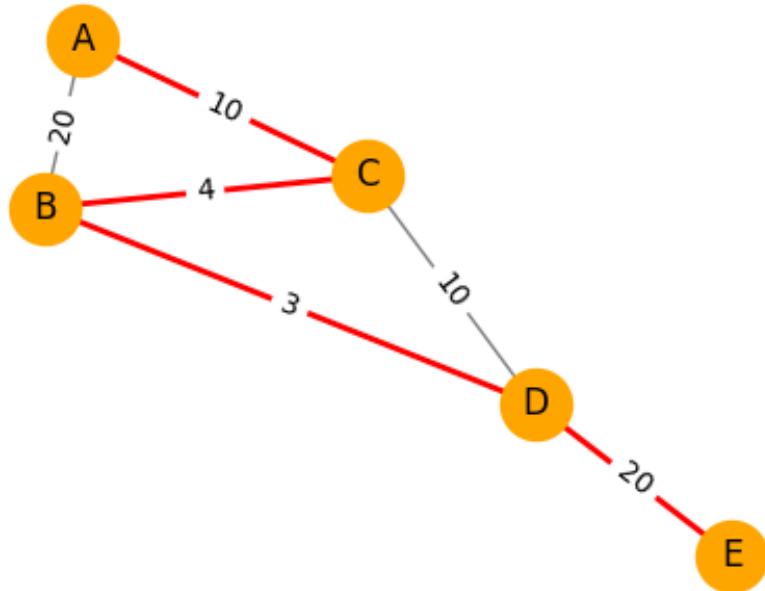
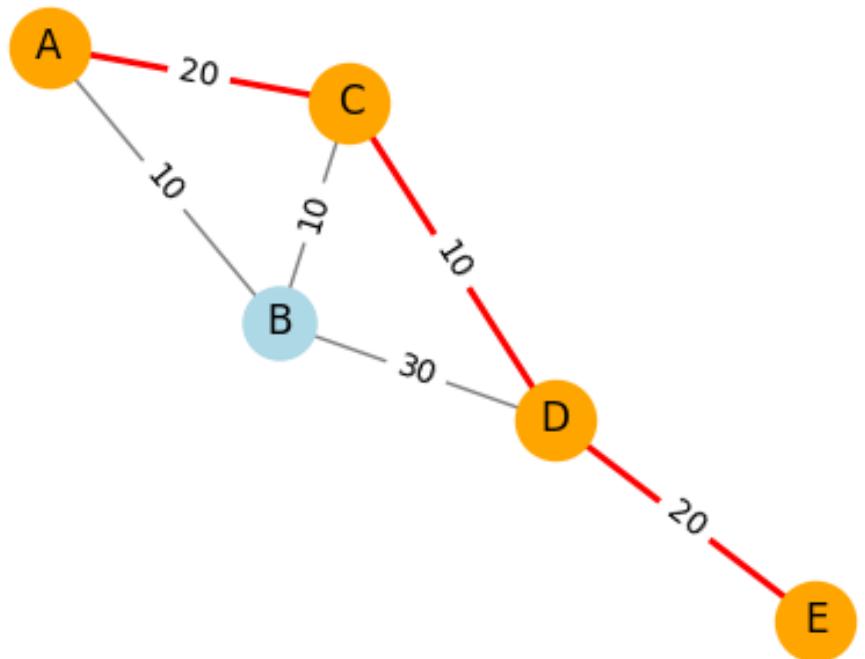
Thank you

Scan me to download the presentation and the codes



Shortest path in weighted graph

- Shortest Path (AE):
['A', 'C', 'D', 'E']
- Shortest Path Length: 50



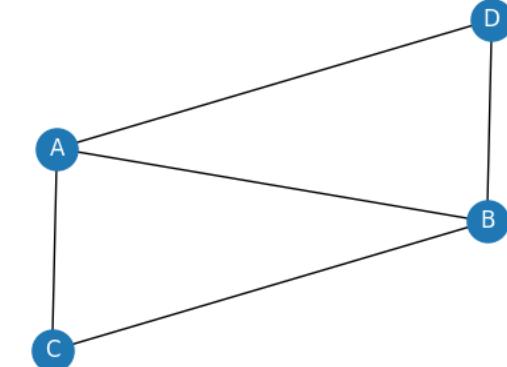
- Shortest Path (AE):
['A', 'C', 'B', 'D', 'E']
- Shortest Path Length: 37

Centrality

- Degree: number of nearest neighbors / (n-1)

n is maximal possible degree

{'A': 1.0, 'B': 1.0, 'C': 0.66, 'D': 0.66}



- Closeness: centrality of a node u is the reciprocal of the average shortest path distance to u over all (n-1) reachable nodes.

$$C(u) = \frac{n - 1}{\sum_{v=1}^{n-1} d(v, u)}$$

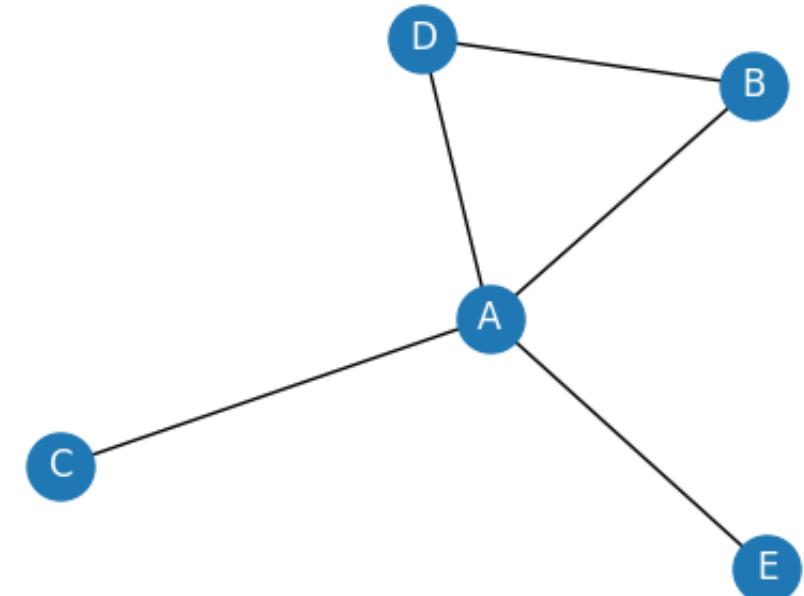
$$A = 1.0 = 4 / (1+1+1+1) = 4/4$$

$$B = 0.66 = 4 / (1+1+2+2) = 4/6$$

$$C = 0.57 = 4 / (1+2+2+2) = 4/7$$

D = 0.66 same as B

E = 0.57 same as C



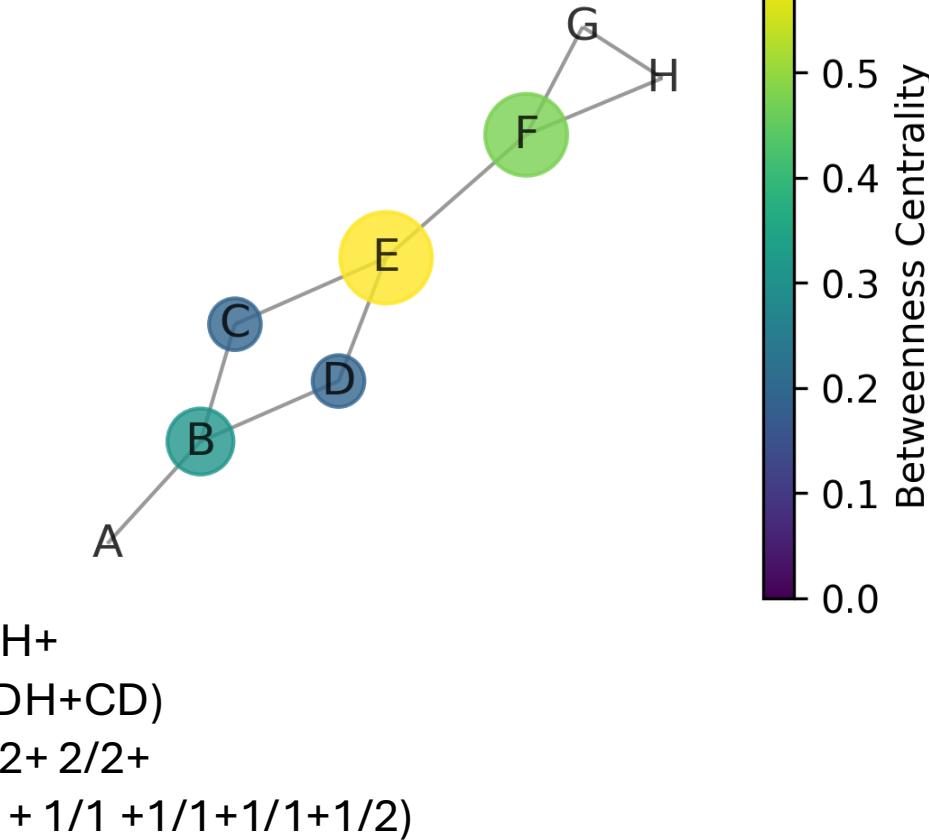
Betweenness Centrality

- Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that passes through v .

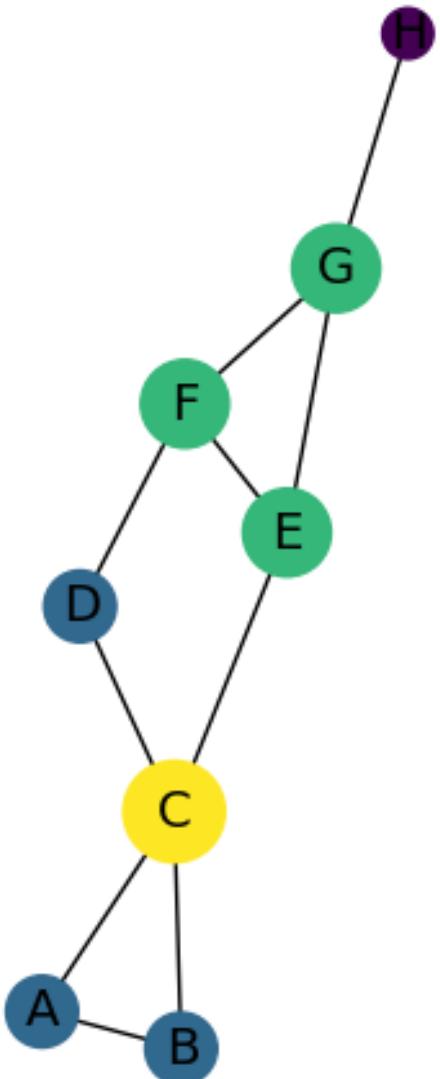
$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s, t|v)}{\sigma(s, t)}$$

A = 0.0
B = 6.5
C = 4
D = 4
E = 12.5
F = 10
G = 0.0
H = 0.0

Node E = (AF+AG+AH+BF+BG+BH+
CF+CG+CH+DF+DG+DH+CD)
=(2/2+ 2/2+ 2/2+ 2/2+ 2/2+ 2/2+
1/1+ 1/1 + 1/1+ 1/1 +1/1+1/1+1/2)



Clustering coefficient



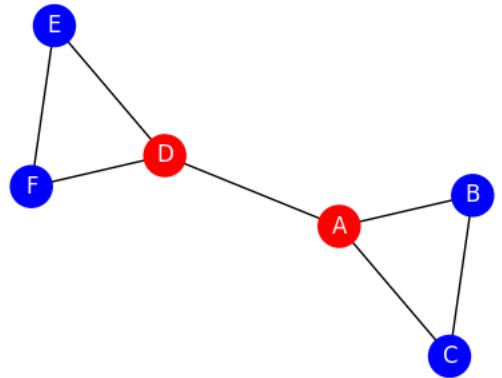
Clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together.

Neighbors of a given node are connected to each other.

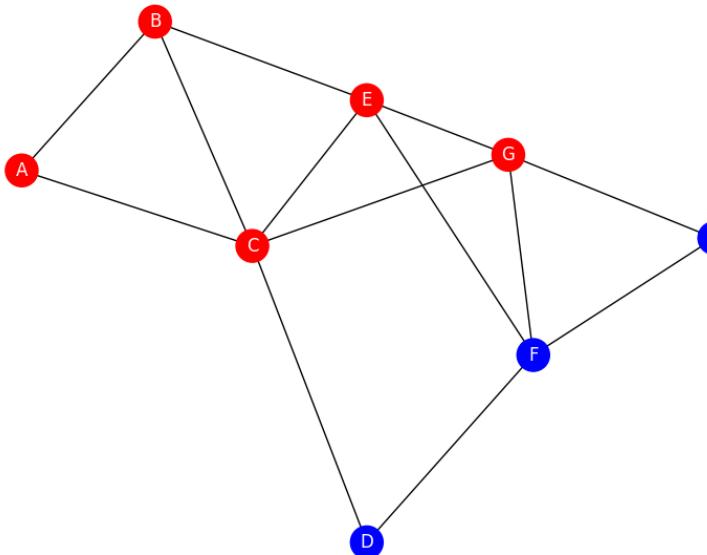
Node	# of neigh	# possible conne	clustering coef
A	2	1	(1/1)=1
B	2	1	(1/1)=1
C	4	6	(1/6)=0.167
D	2	1	(0/1)=0
E	3	3	(1/3)=0.33
F	3	3	(1/3)=0.33
G	3	3	(1/3)=0.33
Avg coef.		= 0.40	

Examples

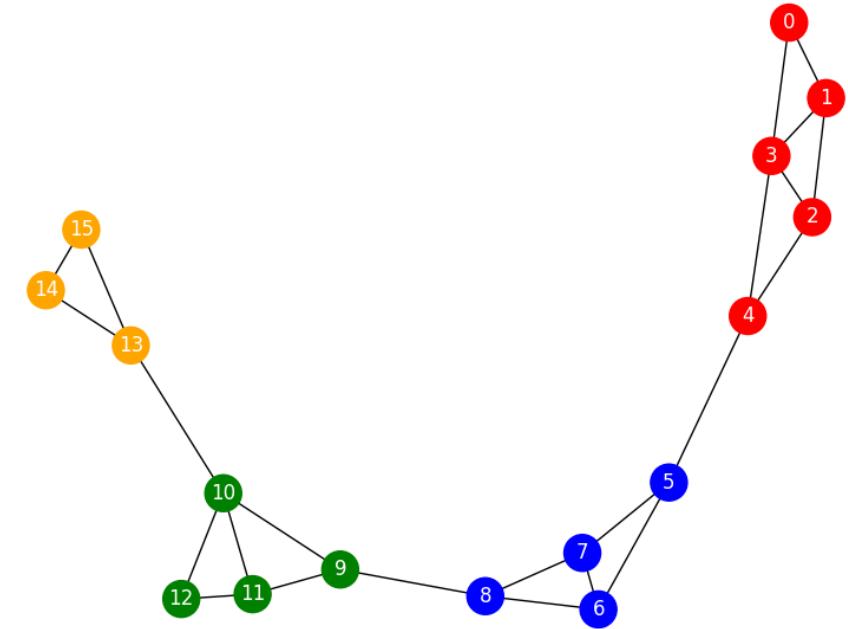
```
mod = nx.community.modularity(G,  
nx.community.label_propagation_communities(G))
```



Mod=-0.08



Mod=0.12



Mod=0.60

Network modularity

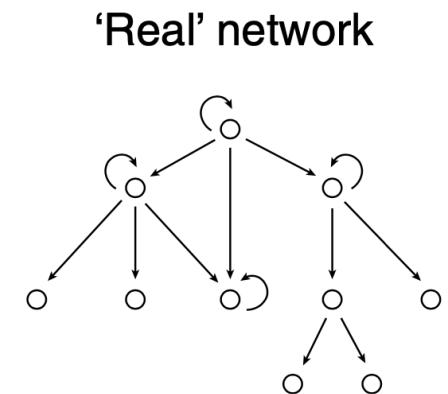
- Modularity measures the strength of division of a network into modules (also called clusters or communities).
- Values
 - closer to 1 indicate strong community structures.
 - near 0 indicate no significant community structure.
 - Negative values means worse-than-random divisions.

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \gamma \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

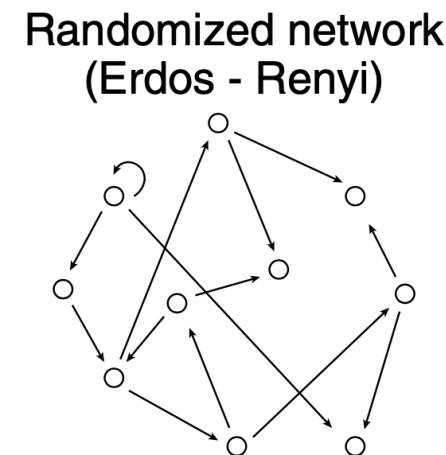
- m is the number of edges (or sum of all edge weights)
- A is adjacency matrix of graph G
- k_i is the (weighted) degree of node i
- γ is the resolution parameter
- $\delta(c_i, c_j)$ is 1 if i and j are in same cluster else 0.

Network motifs

- A recurring **patterns** that occur in the real network significantly more often than in randomized networks with the same characteristics (number of nodes, number of edges) are called **network motifs**.
- Edges in the network motifs must be constantly selected in order to survive randomization forces in unexpected high amounts.



N=10 nodes
A=14 arrows
N_{self}=4 self-arrows

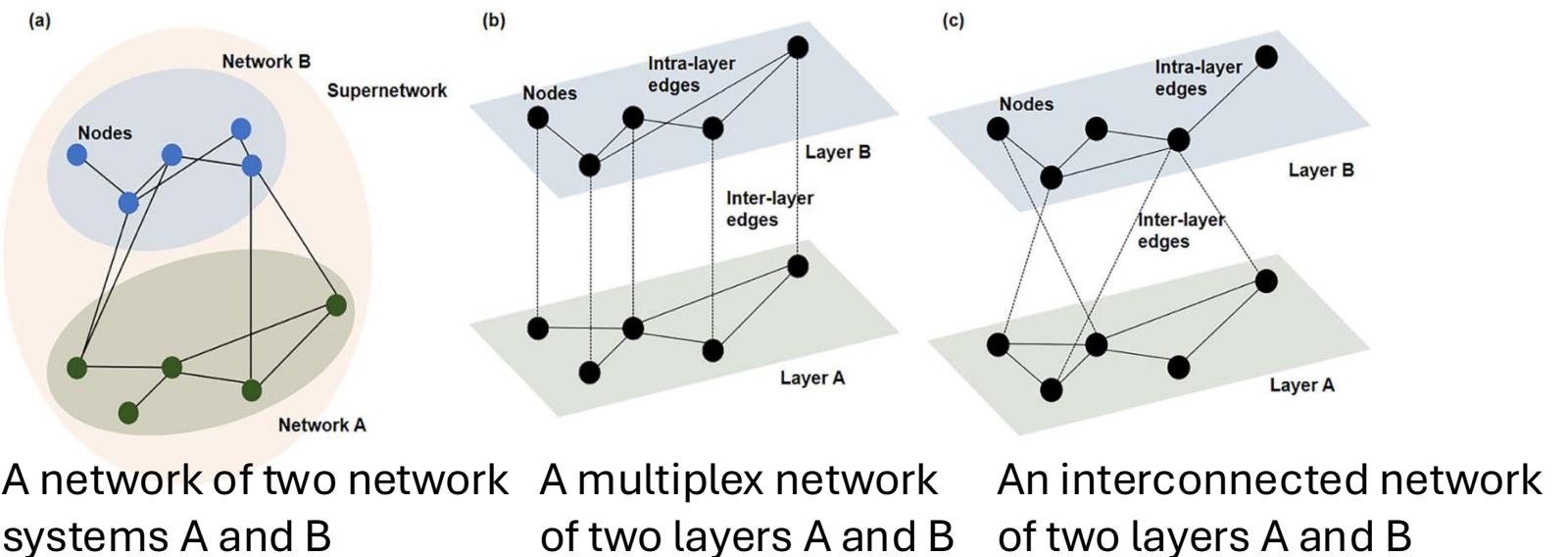


N=10 nodes
A=14 arrows
N_{self}=1 self-arrow

Let's have some tutorial with python
programming to find motifs

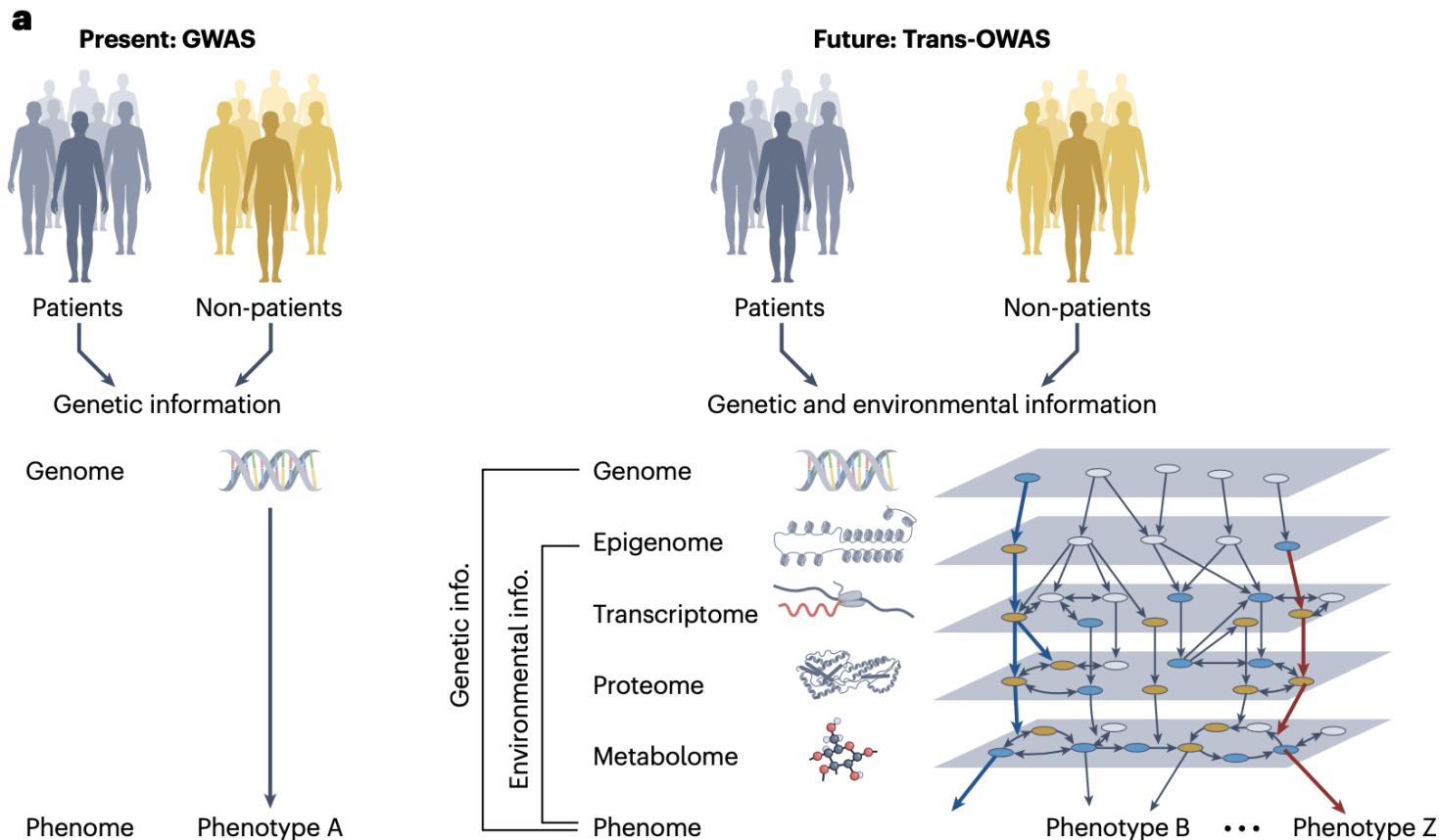
Multilayer networks

- Multilayer networks provide the framework to capture the complexity typical of systems of systems.
- It enable the analysis of biophysical, social and human-made networks from an integrated perspective.
- Layered structure of multilayer networks is responsible for phenomena that cannot be observed from the analysis of subsystems in isolation or from their aggregation, such as emergent mesoscale organization.



Manlio De Domenico "More is different in real-world multilayer networks" Nat Phys 2023

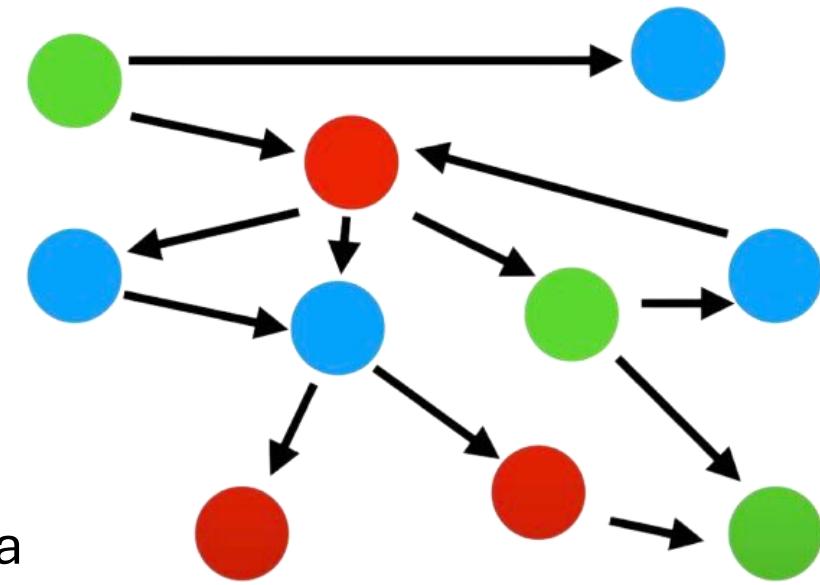
Multilayer networks



- Genetic information collected from healthy and unhealthy groups is generally used to capture information about macroscopic effects, such as a disease.
- Multi-omics makes available richer datasets that can be integrated within a multilayer network, where layers also encode interactions at the scale of the transcriptome, the proteome and the metabolome

Knowledge graph

- A knowledge graph is knowledge base that is a graph and represents facts about the world.
- What is a knowledge base?
 - A knowledge base is a technology used to store complex **structured and unstructured information**.
 - A **collection of knowledge** expressed using some formal knowledge representation language.
 - A **store of information or data** that is available to draw on.
 - The underlying **set of facts, assumptions, and rules** which a computer system has available to solve a problem.
- What is a graph?



Evolution of Google search algorithm

- **PageRank** 1998: The order of search results returned by Google is based on this.
- **Knowledge Graph** 2012: Added semantic search by organizing information into entities and relationships, providing direct answers and improving contextual understanding.
- **RankBrain** 2015: Integrated machine learning to interpret user intent, handling queries never seen before.
- **BERT** 2019: Leveraged deep learning to understand the nuances of natural language, particularly context in search queries.

Biological molecule is also context dependent

“apple” is a **polysemic** word...



grow an apple

buy an apple|



buy an apple|

grow an apple

grow an apple tree
grow an apple tree from seed
grow an apple tree in a pot
grow an apple tree indoors

buy an apple watch
buy an apple gift card
buy an apple tv



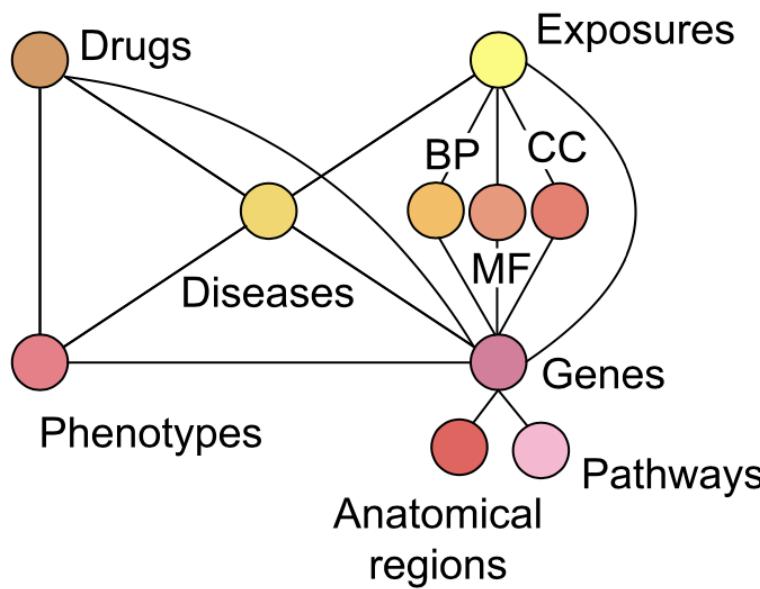
Stem	Cell	Plant	Conserved regions in the virus
Protein	Activate or repress signalling pathway	Stabilize its active confirmation	Coordinates the iron in the heme group of haemoglobin

Knowledge graph in precision medicine

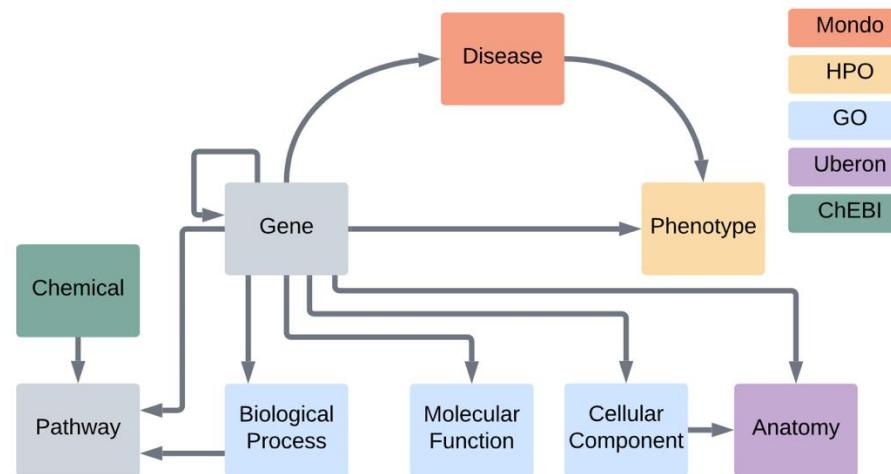
- Developing personalized diagnostic strategies and targeted treatments requires a deep understanding of disease biology and the ability to dissect the relationship between molecular and genetic factors and their phenotypic consequences.
- Such knowledge is fragmented across publications, non- standardized repositories, databases and evolving ontologies describing various scales of biological organization between genotypes and clinical phenotypes.
- Knowledge graph
 - Drug repurpose
 - Clinical trial optimization
 - Patient Stratification

Existing knowledge graph in biomedical

It connects molecular, genetic, phenotypic, and drug-related information, enabling analyses like **drug-disease interactions**, **disease mechanisms**, and **therapeutic predictions**.



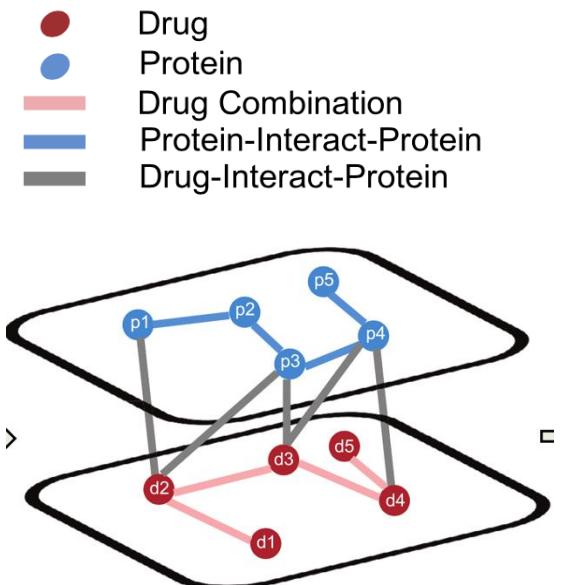
It is a translational research platform that integrates and analyzes genetic, phenotypic, and disease data across species using open ontologies, semantic data models, and knowledge graphs, enabling applications like **variant prioritization**, **deep phenotyping**, and **patient profile matching**.



PrimeKG: Payal Chandak et al
Scientific Data 2023

Monarch: Tim E. Putman et al
NAR 2024

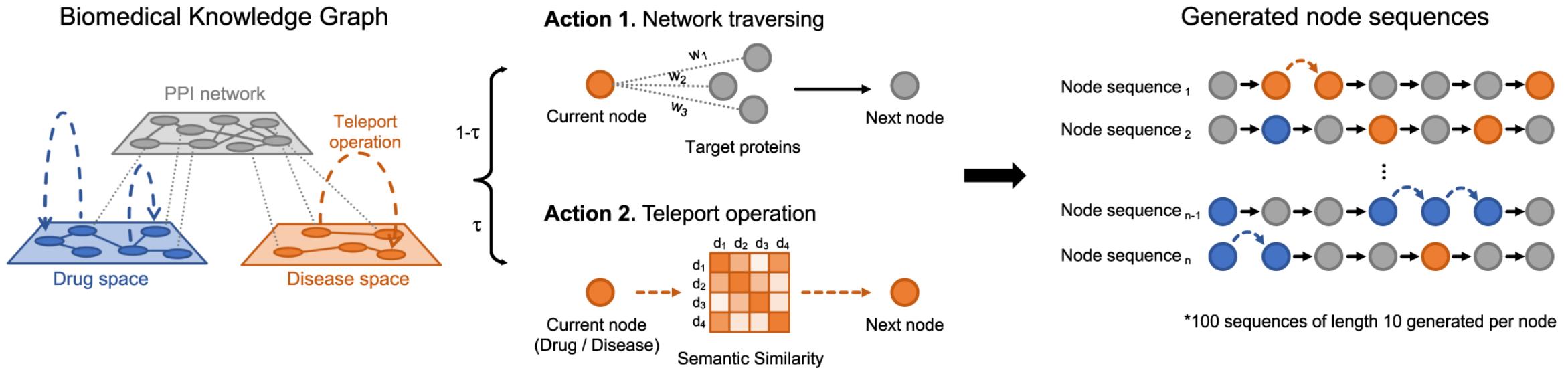
Predict Drug combination therapies (DCBs) for infectious diseases



Qing Ye et al Cell Reports 2023

Biomedical knowledge graph for drug repurposing

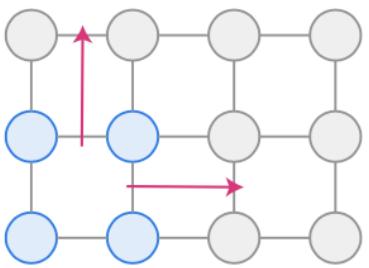
DREAMwalk map drugs and diseases into a unified embedding space, improving prediction of drug-disease associations.



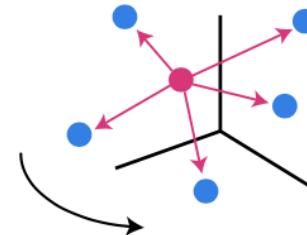
Last Tutorial

Networks in ML

Convolutional



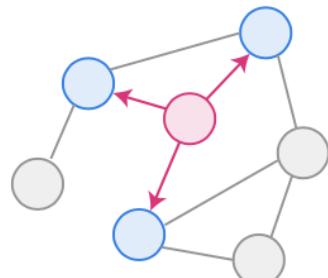
Group equivariant



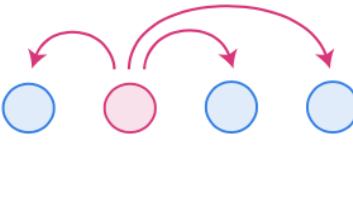
Recurrent



Relational



Attentional



Translational invariance

Rotational invariance

Repeating dynamics

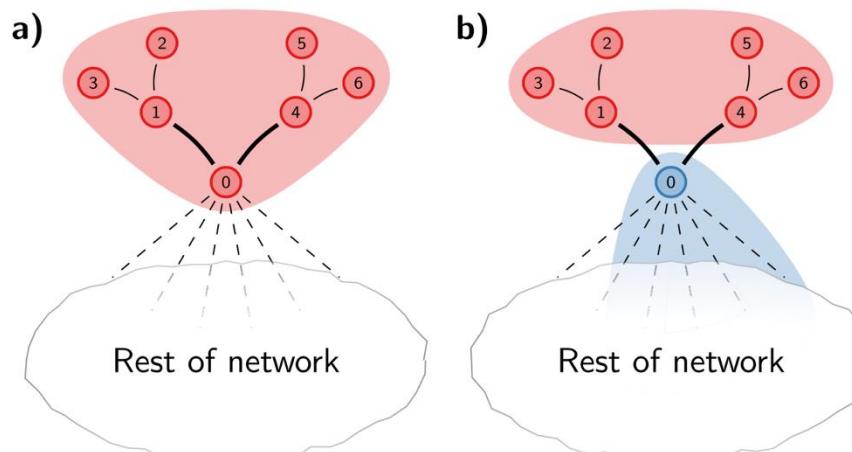
Non-locality

Locality

Unordered

Application of modularity

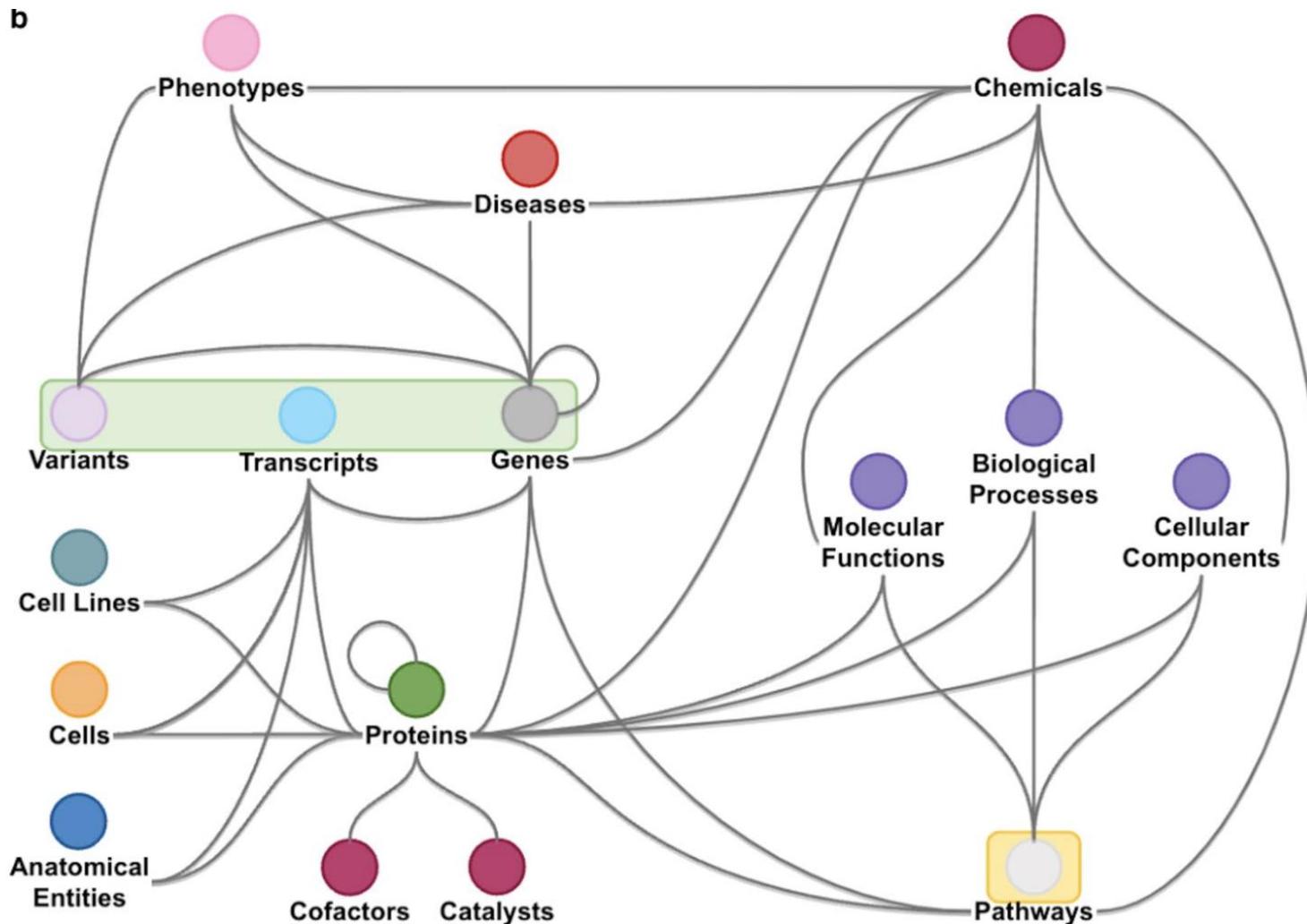
- Louvain clustering 2008
 - Start with a single partition where all nodes are in their own community
 - Iterate repeatedly
 - Local node movement: Move nodes to communities to maximize modularity
 - Network aggregation: Aggregate communities into new nodes within network
- Leiden clustering 2019
 - Faster and address an issue where communities may become internally disconnected but remain as one community



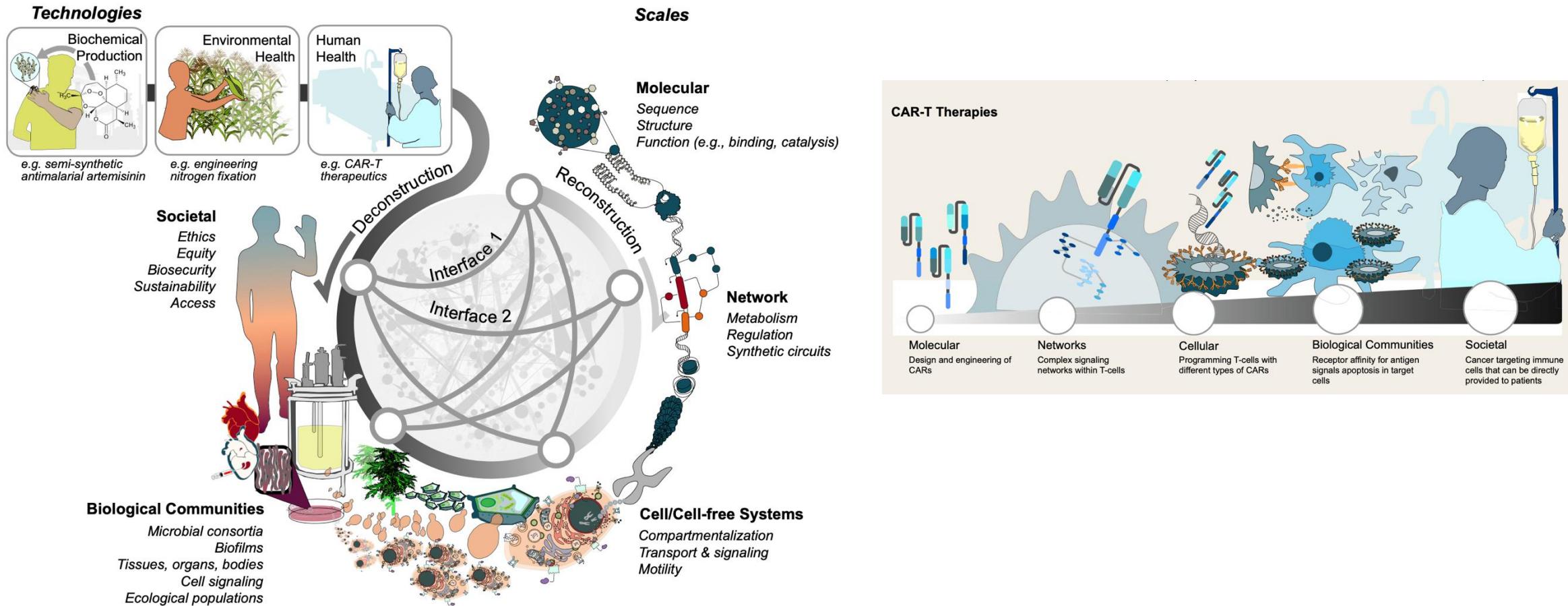
Famous algorithms related to graphs

- Dijkstra's (shortest path problem)
- PageRank
- Leiden community detection
- Kd-tree
- Random Forest
- Causal
- Markov clustering
- Search

Knowledge graph of a disease

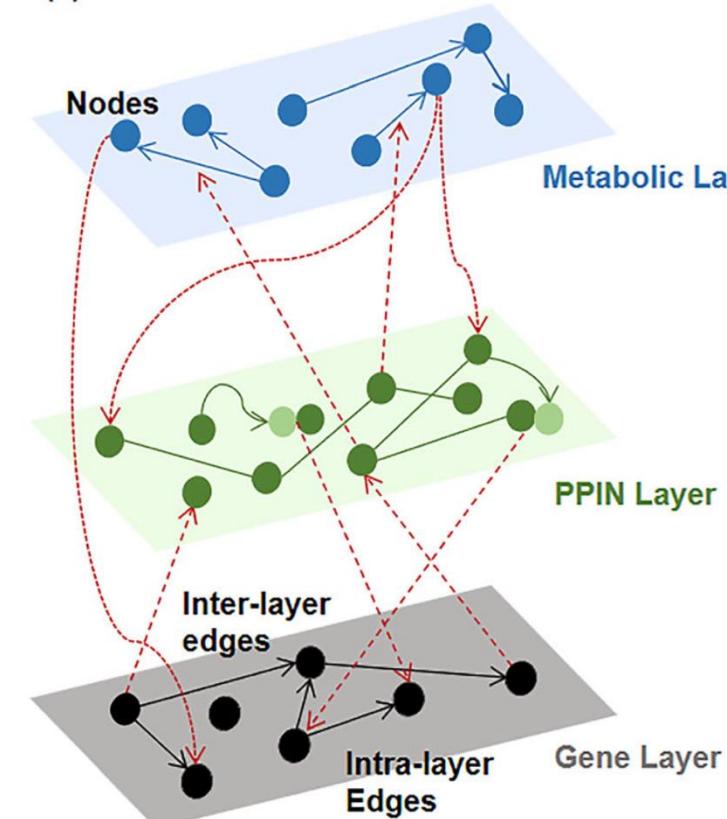


Deconstructing synthetic technologies across scales



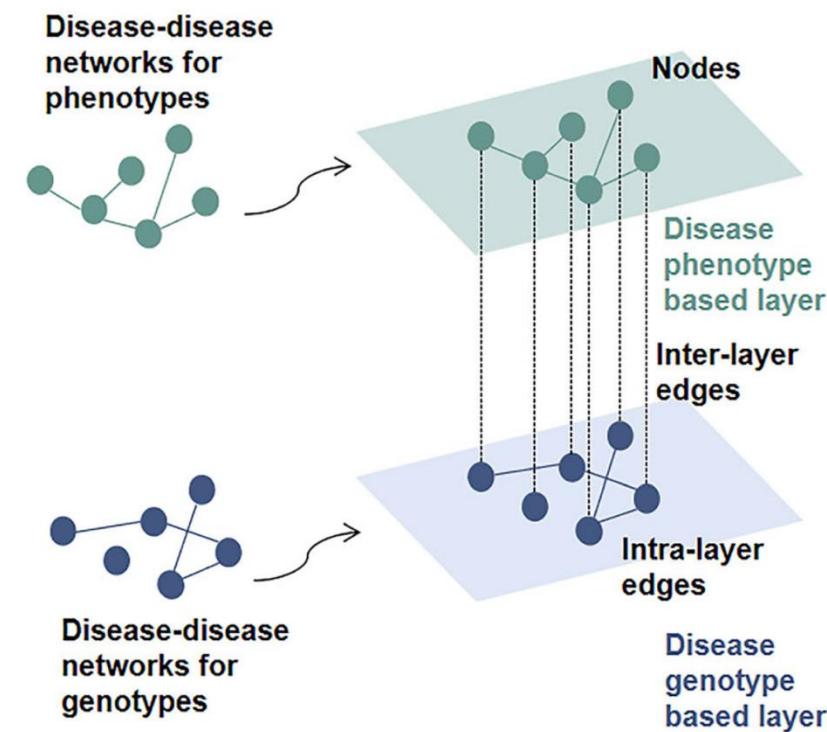
Multilayer network exa

An interdependent network of gene regulation and metabolism comprising three layers



Edges represent the inter-layer connections such as activation, repression, and catalysis.

Disease network



Disease connected by symptoms

Connections between diseases.

Diseases connected by common genes