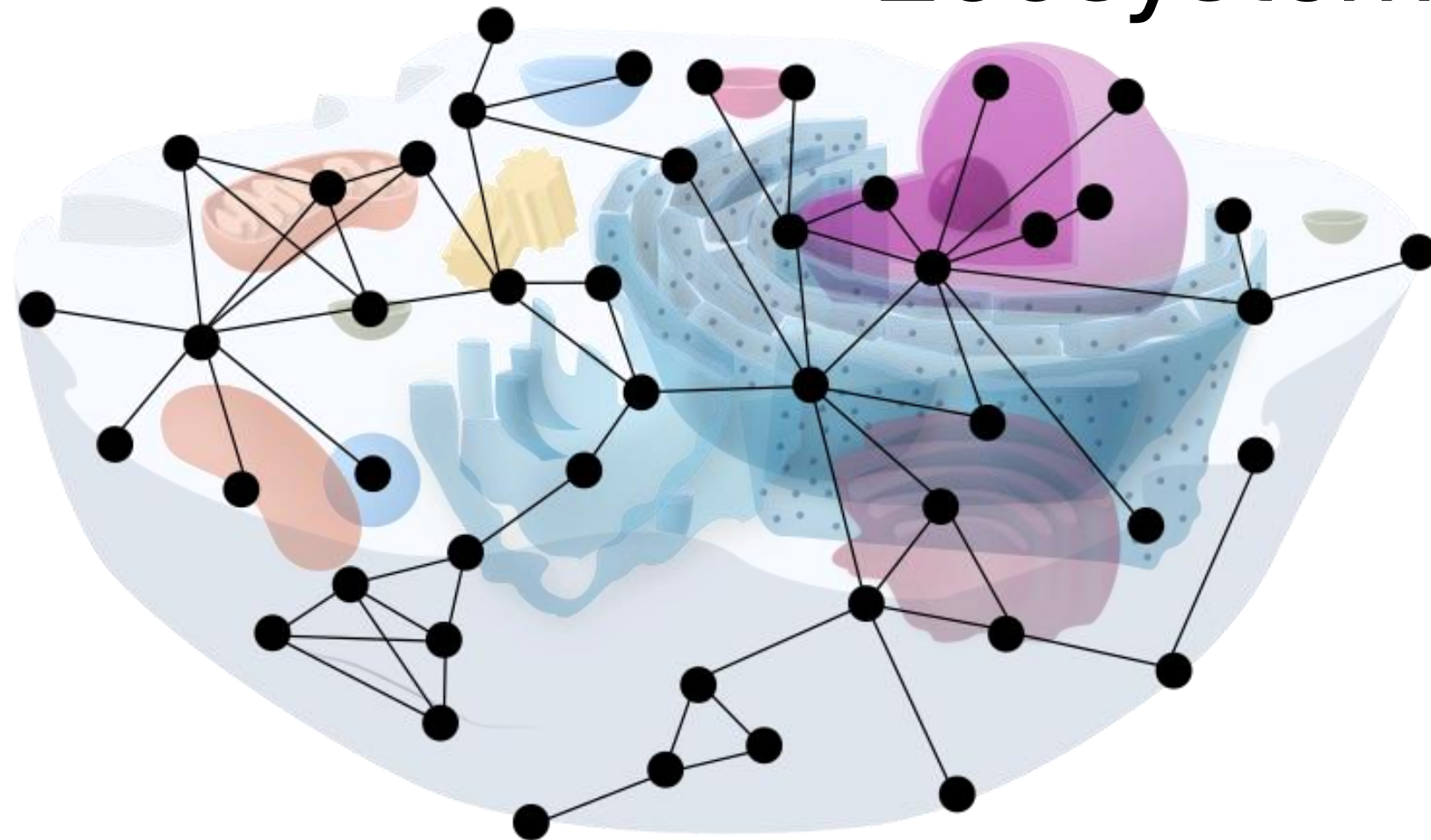


Networks Across Biological Scales: From molecules to Ecosystems

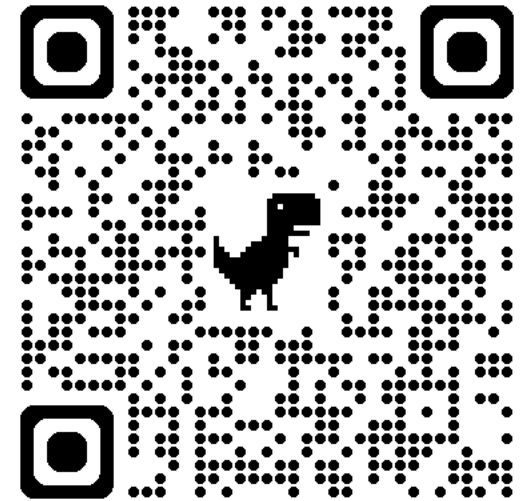


Ankit Agrawal
University of Würzburg
Germany

Outline agenda

- Biological networks are ubiquitous
- Basic component of a network
- Constructing and modeling networks
- Network motifs
- Multilayer networks
- Introduction of Knowledge graph
- Thoughtful experiment of networks
- Application of knowledge graph

Scan me to download the presentation and the codes



Social networks

Prompt

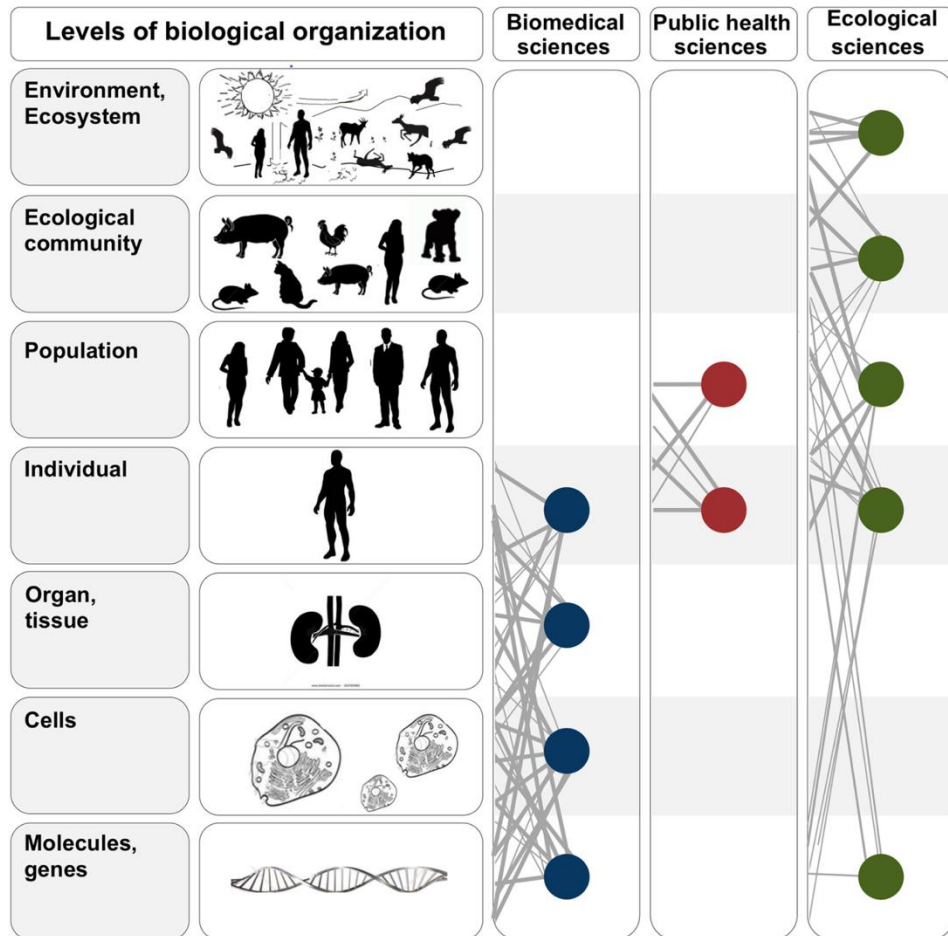
Create a cartoon-style illustration showing the evolution of a person's social network from infancy to adulthood. Start with a baby in the center who first connects with their parents and siblings. As the child grows, add connections to extended family (grandparents, aunts, uncles), followed by friends 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



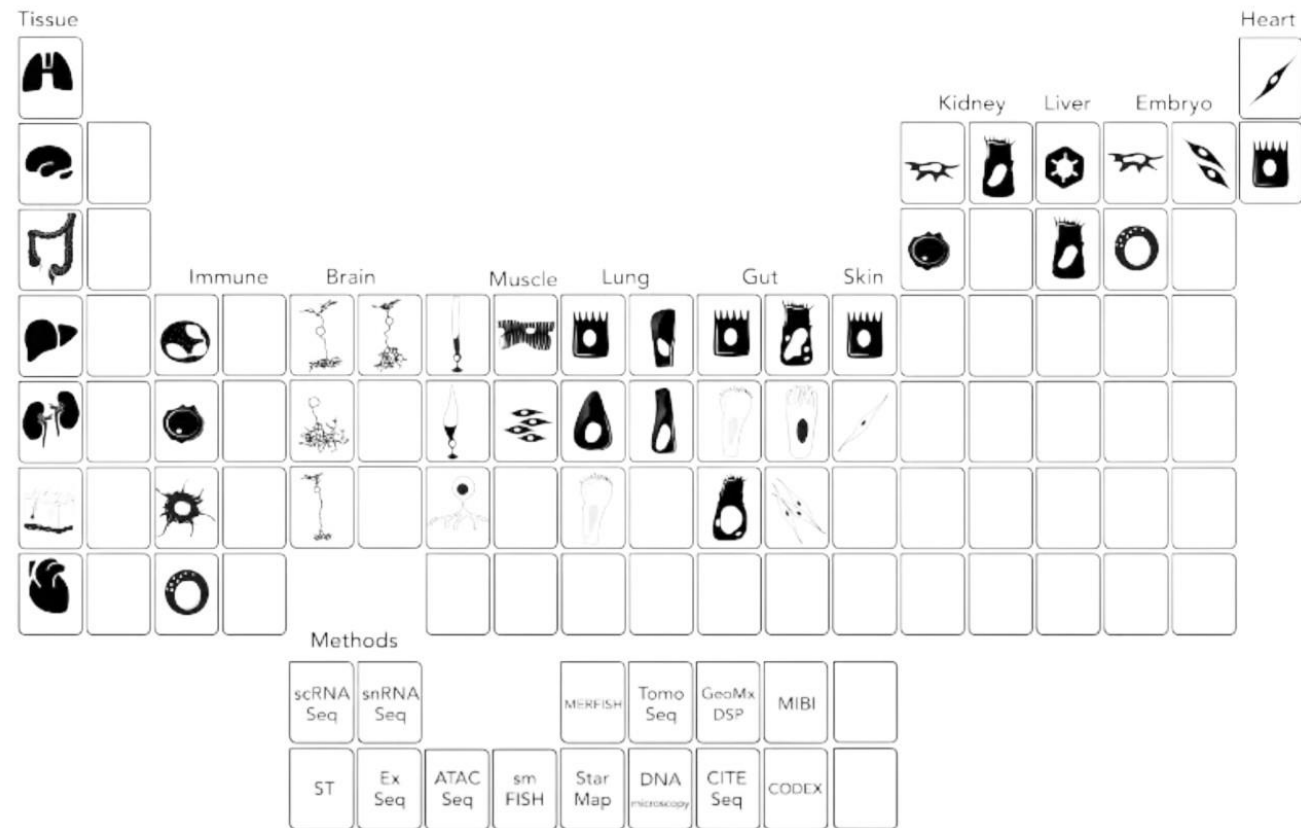
<https://deepai.org/machine-learning-model/text2img>

Hierarchy of biological organization, from molecules and genes to ecosystems

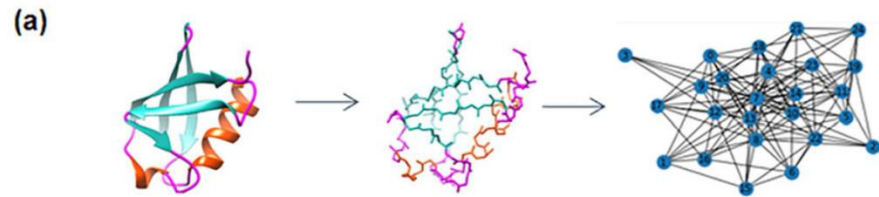
Concentrates on the structural and topological organization of biological systems



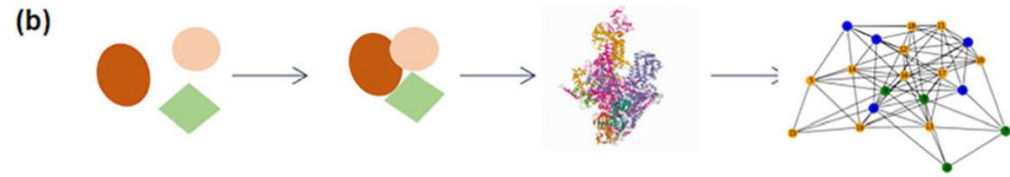
Focuses on the holistic understanding of biological systems, emphasizing how components interact to produce emergent behaviors



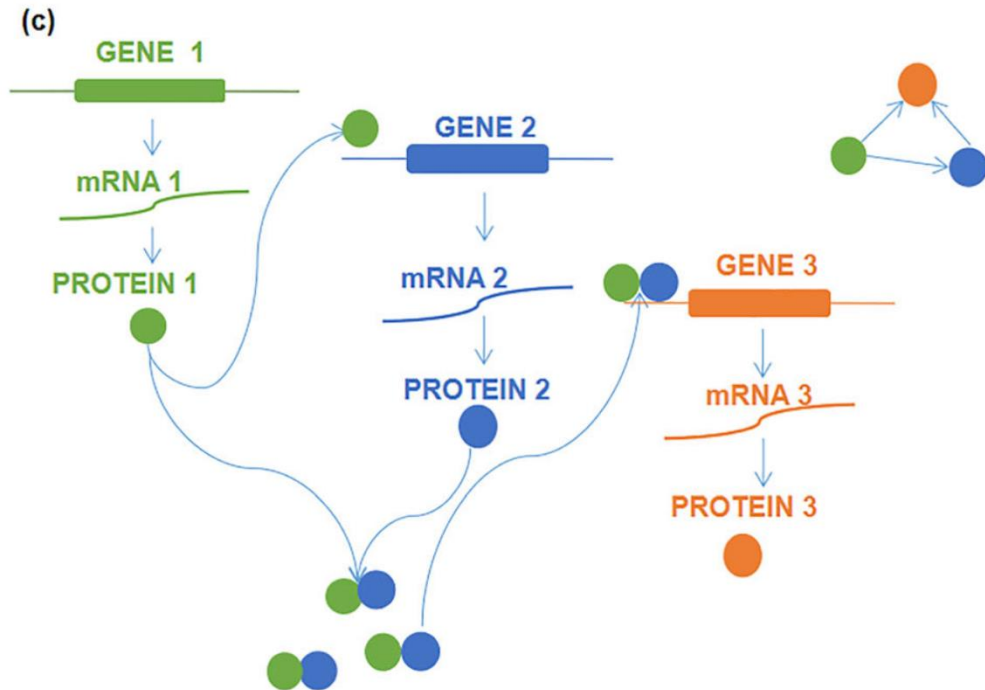
Types of Biological Network



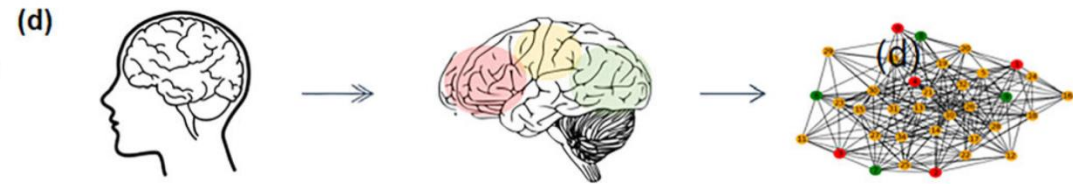
Protein Contact Network



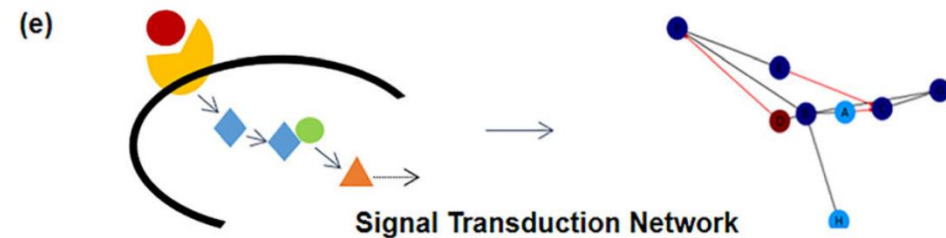
Protein-Protein Interaction Network



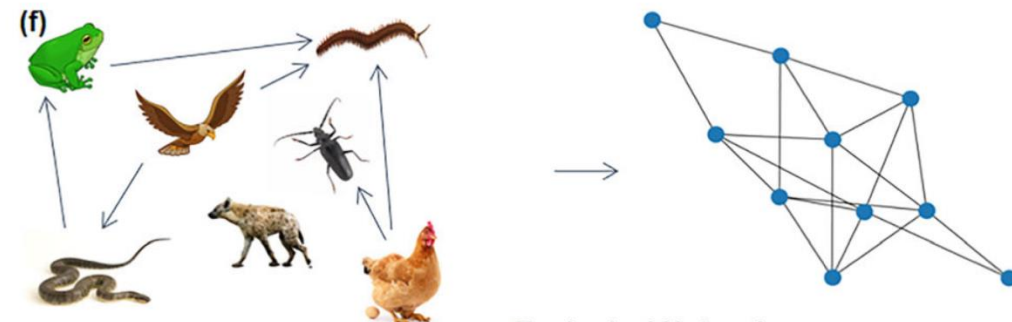
Gene Regulatory Network



Neuronal Network



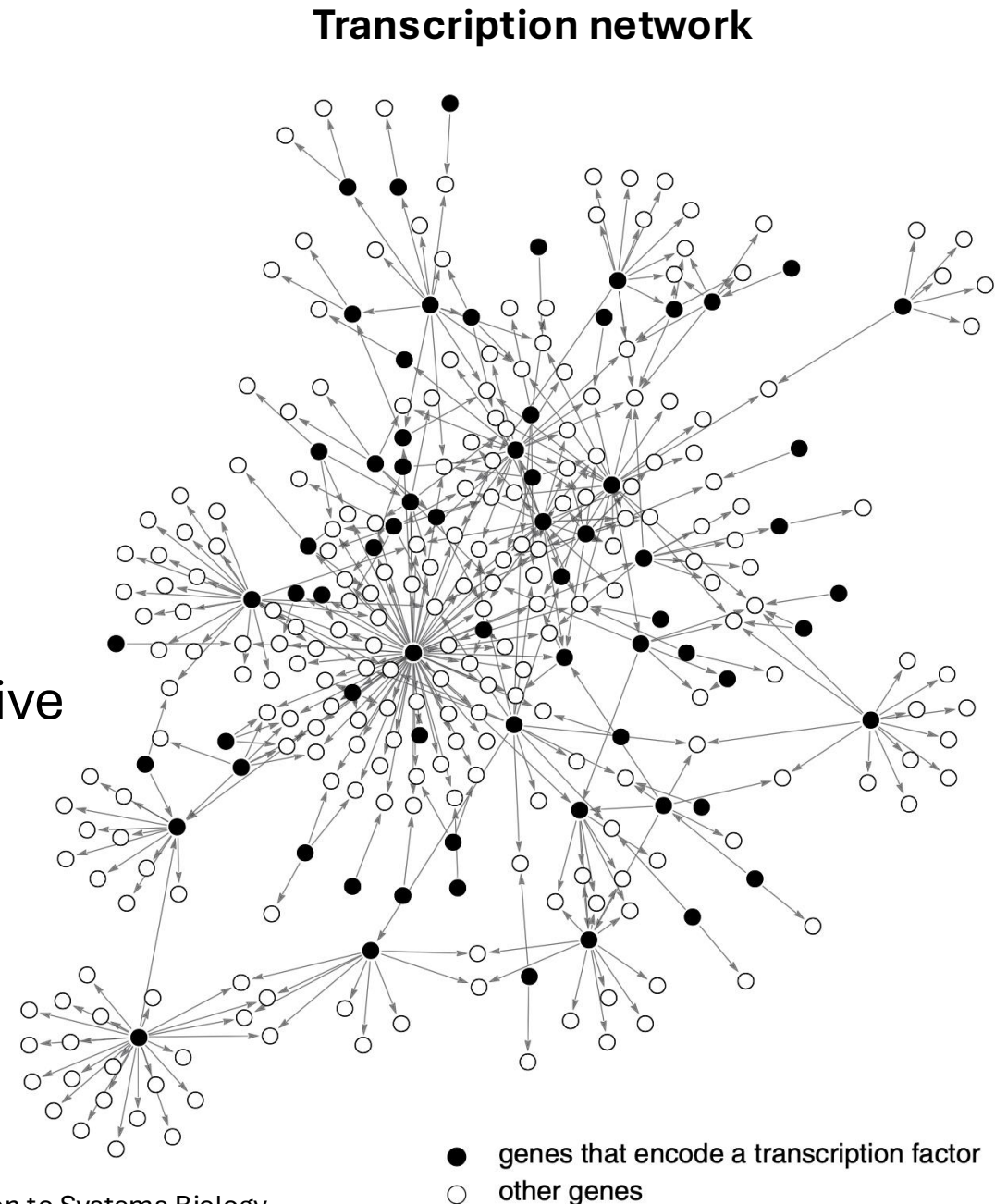
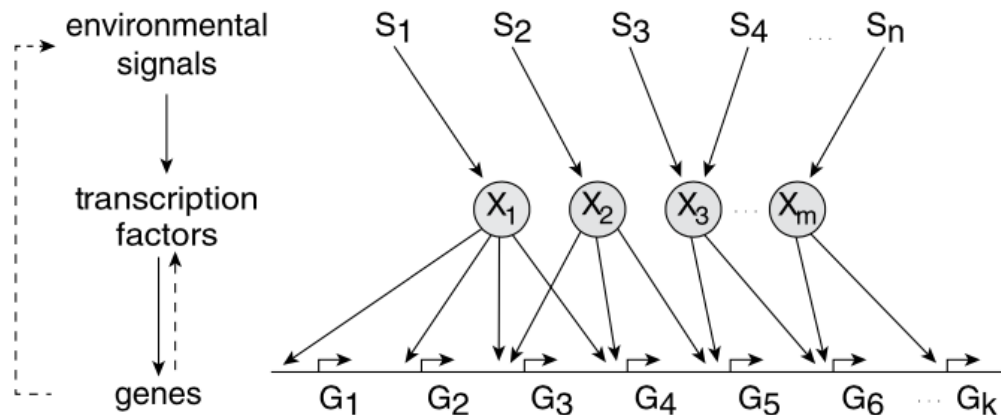
Signal Transduction Network



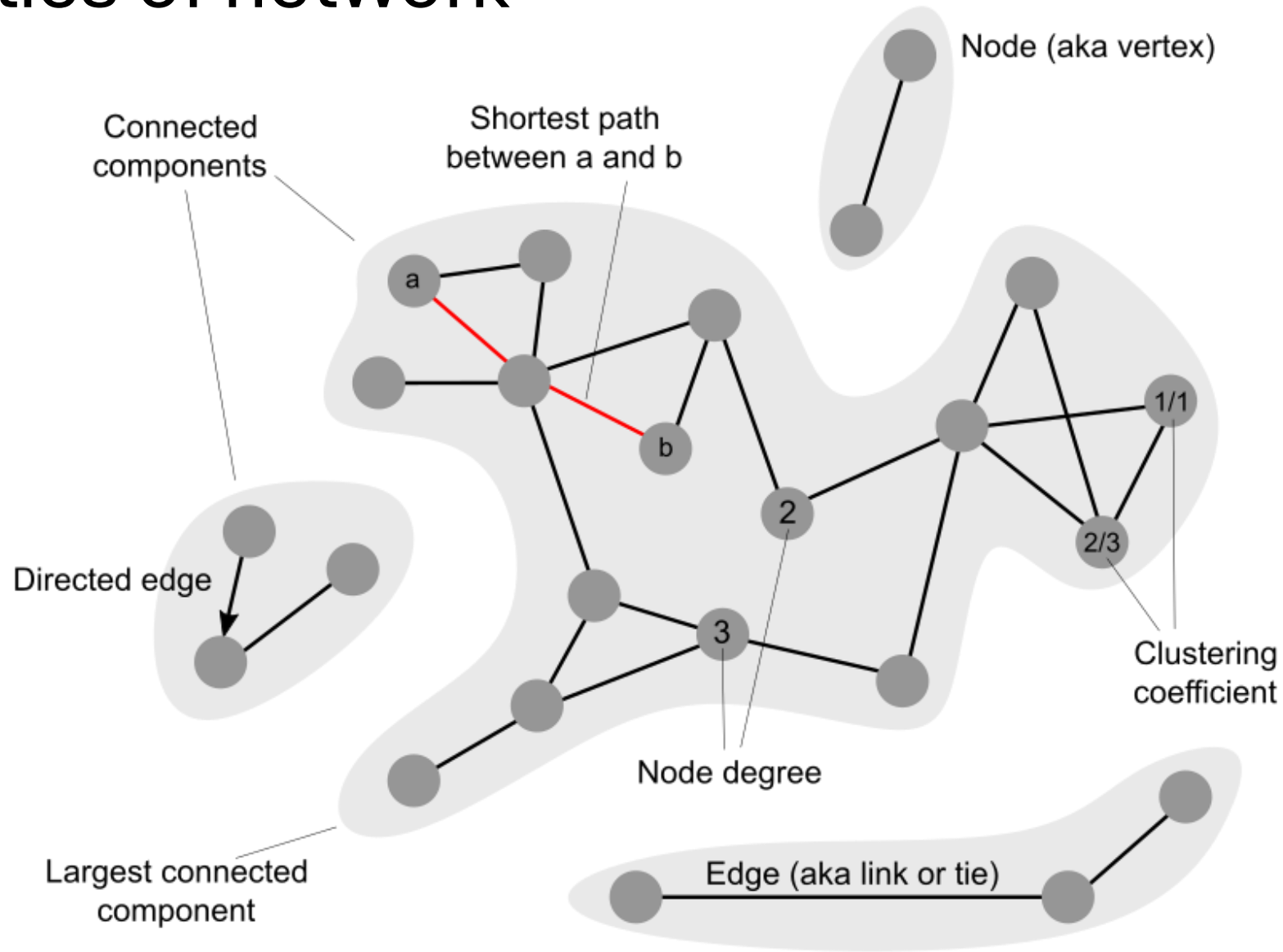
Ecological Network

Cells are living machines

- Cells sense many different signals
 - Temperature and pressure
 - Damage
 - Signal from other cell
 - Nutrients and chemicals
- Cells respond to these signals by producing appropriate proteins
- These special protein called TF
- TF are designed to transit rapidly between active and inactive states

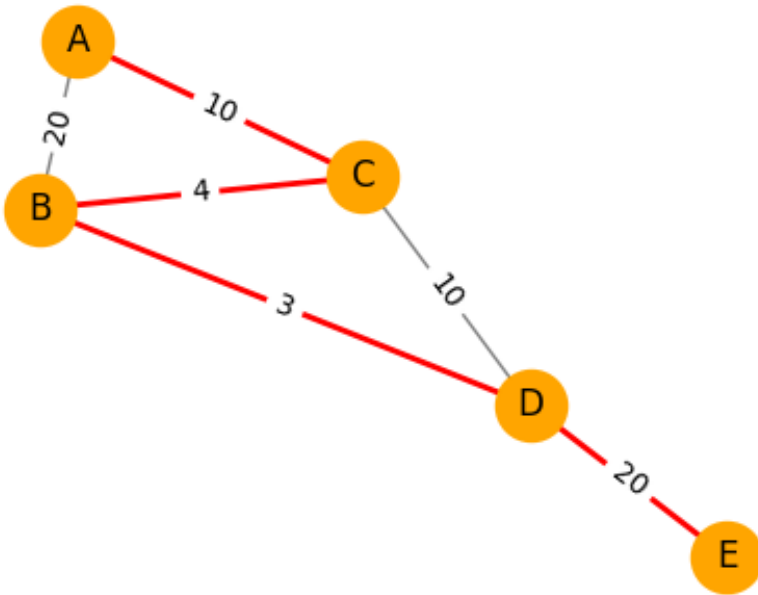
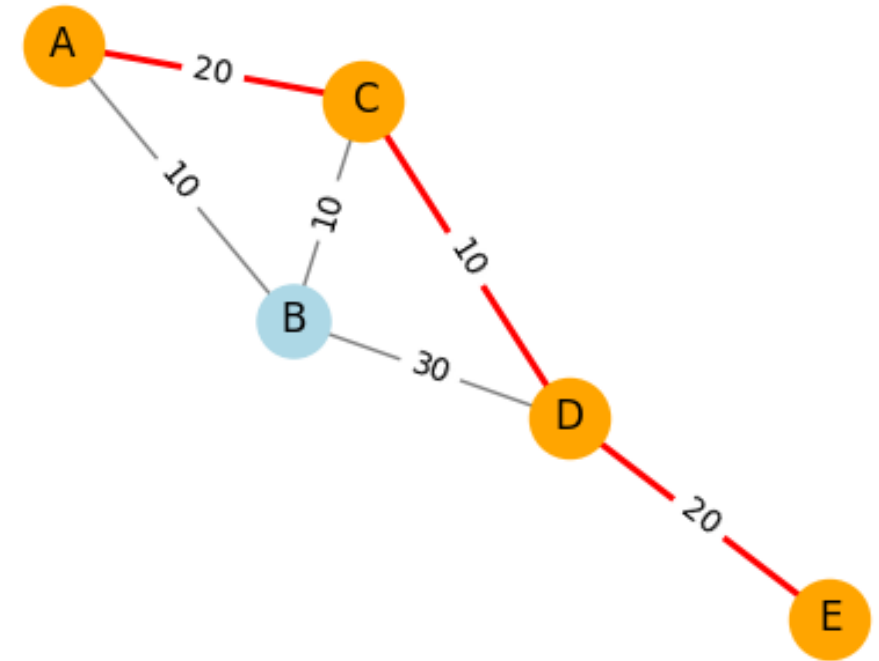


Properties of network



Shortest path in weighted graph

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



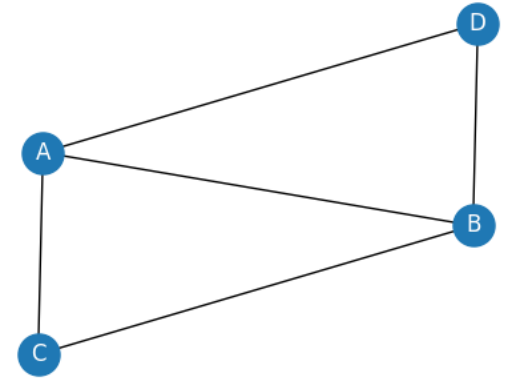
- Shortest Path:
['A', 'C', 'B', 'D', 'E']
- Shortest Path Length: 37

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.

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

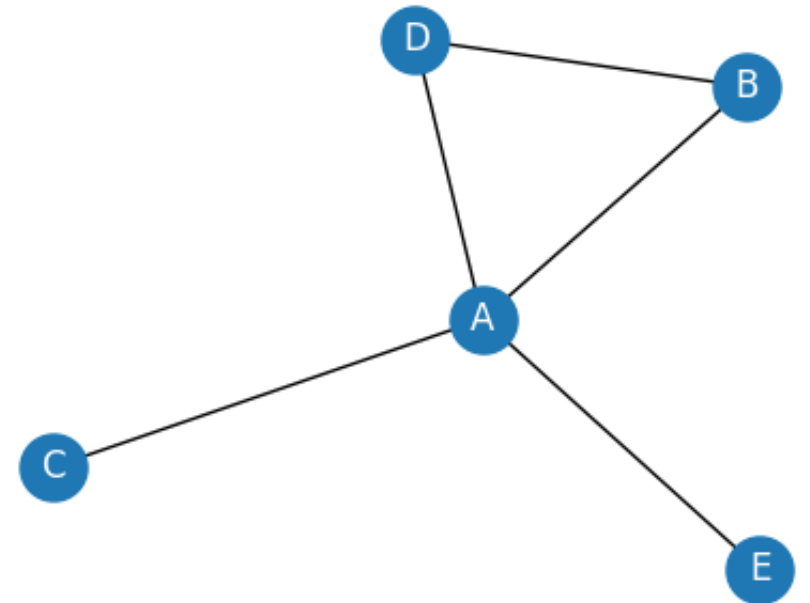
$$A = 1.0 = 4 / (1+1+1+1) = 4/4$$

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

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

$$D = 0.66 \text{ same as B}$$

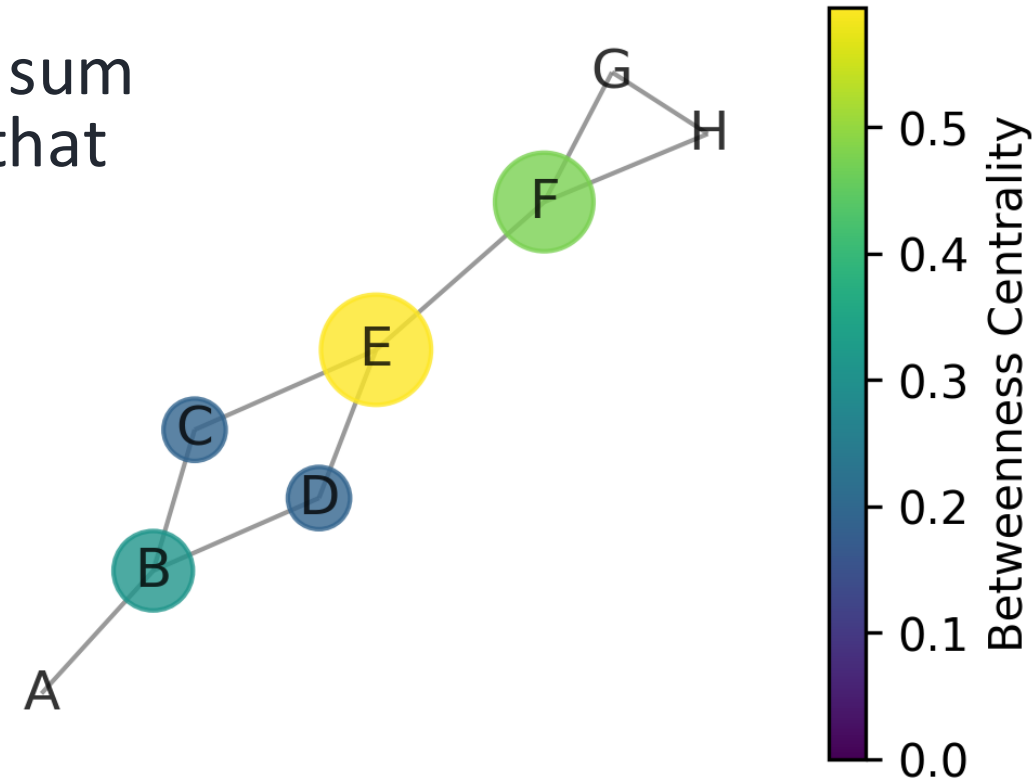
$$E = 0.57 \text{ same as C}$$



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

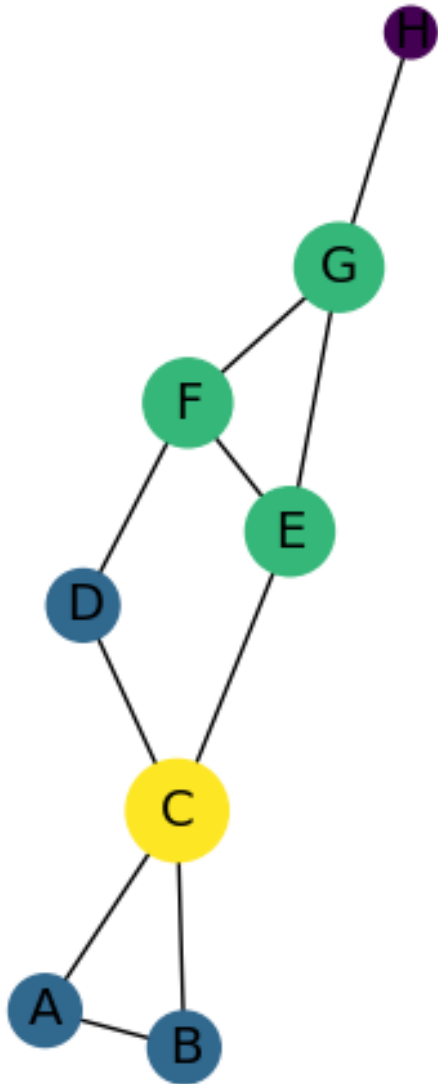
F = 10

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.

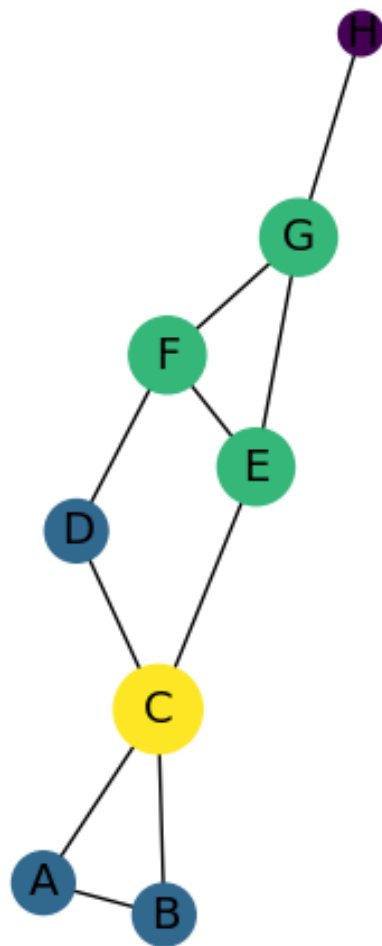
Neighbors of a given node are connected to each other.

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

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

Adjacency matrix



	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	

Network modularity

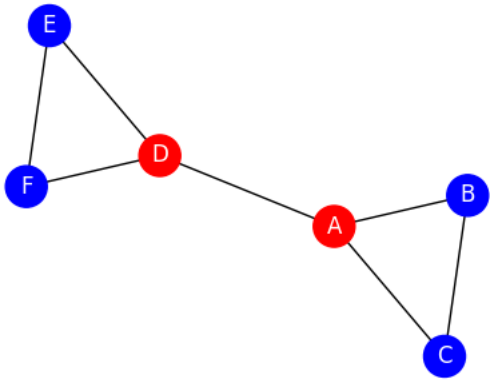
- Modularity measures the strength of division of a network into modules (also called clusters or communities).
- Values
 - closer to 1 indicate strong community structures.
 - near 0 indicate no significant community structure.
 - Negative values means worse-than-random divisions.

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

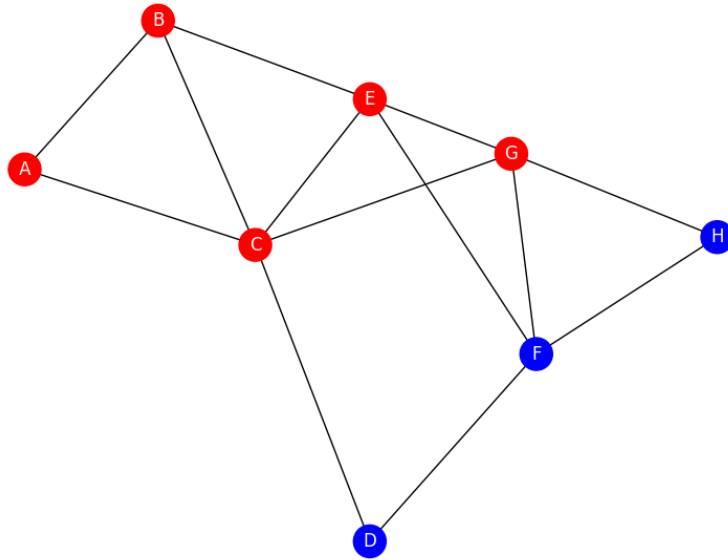
- m is the number of edges (or sum of all edge weights)
- A is adjacency matrix of graph G
- k_i is the (weighted) degree of node i
- γ is the resolution parameter
- $\delta(c_i, c_j)$ is 1 if i and j are in same cluster else 0.

Examples

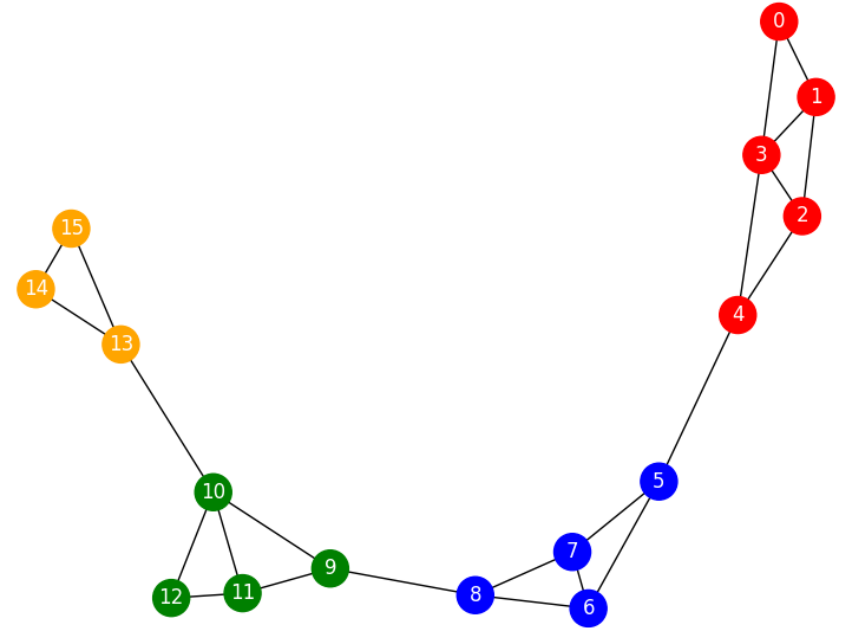
```
mod = nx.community.modularity(G,  
nx.community.label_propagation_communities(G))
```



Mod=-0.08



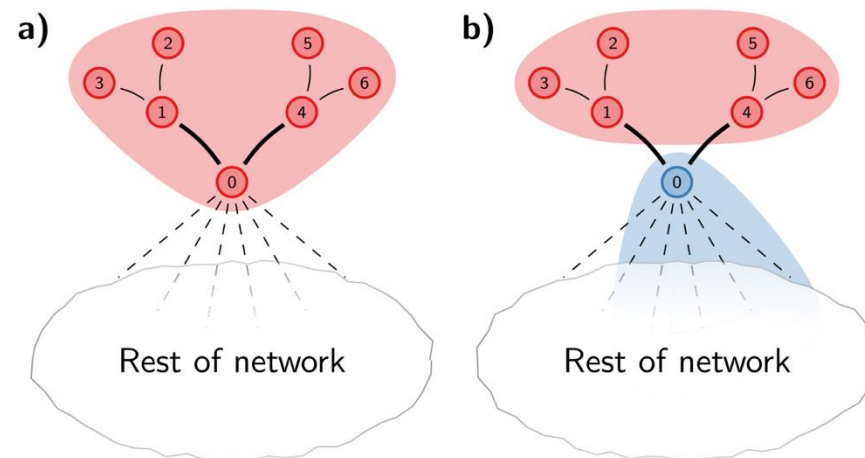
Mod=0.12



Mod=0.60

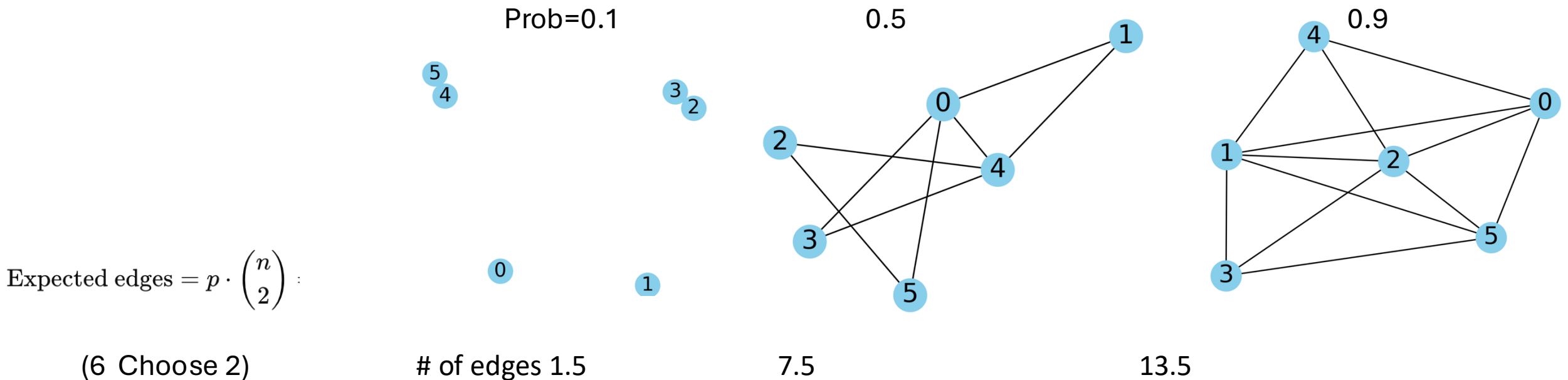
Application of modularity

- 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



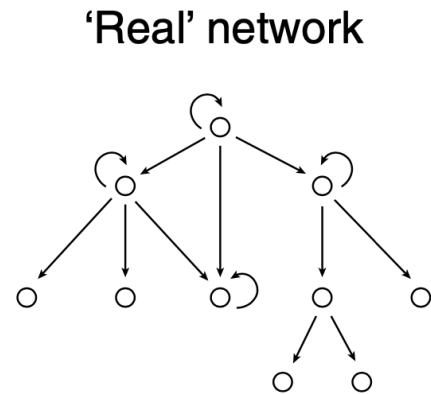
Random graph

- A graph is constructed by connected nodes randomly.
- Each edge is included in the graph with probability p , independently from every other edge.
- Equivalently, the probability for generating each graph that n nodes and M edges is

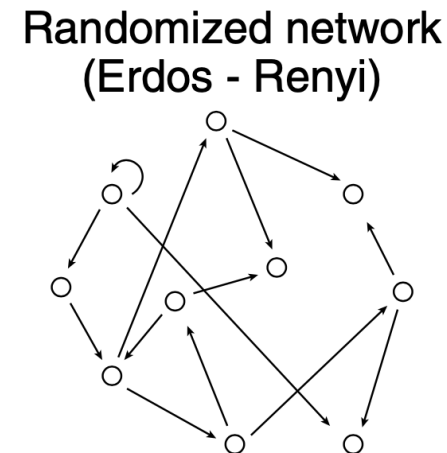


Network motifs

- 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

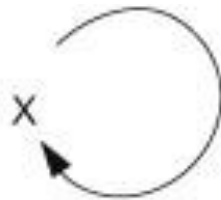
Autoregulation is a network motif

- Regulation of a gene by its own gene product is known as autogenous control, or autoregulation.
- Is autoregulation significantly more frequent in the real network than at random?
- Probability of connecting to itself: $p_{\text{self}} = 1/N$
- If E edges are placed at random to form the random network, the average number of self-arrows is

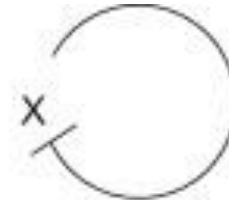
$$\langle N_{\text{self}} \rangle_{\text{rand}} = E p_{\text{self}} = E / N$$

X → Y

A Simple regulation



B Positive
autoregulation



C Negative
autoregulation

Autoregulation example

Real Path length

	A	B	C
A	0	1	2
B	2	0	1
C	1	2	0

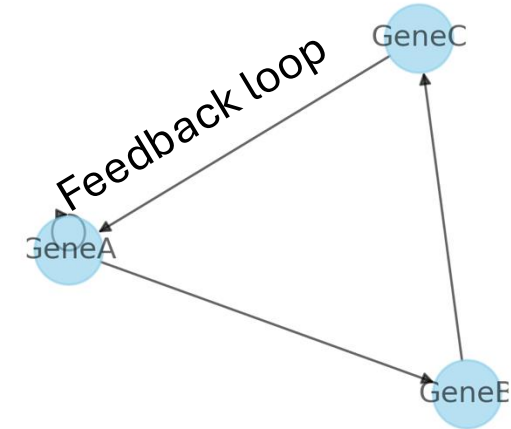
Radom Path length

	A	B	C
A	0	1	2
B	1	0	1
C	1	2	0

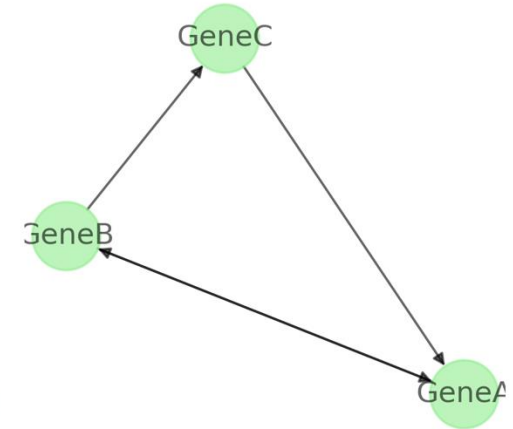
- In the E. coli transcription network
N = 424 and E = 519
- In random network with the same N and E would give the number of self-arrows

$$\langle N_{\text{self}} \rangle_{\text{rand}} = E / N = 1.22$$
- The real network has 40 self-arrows, which exceeds the random networks

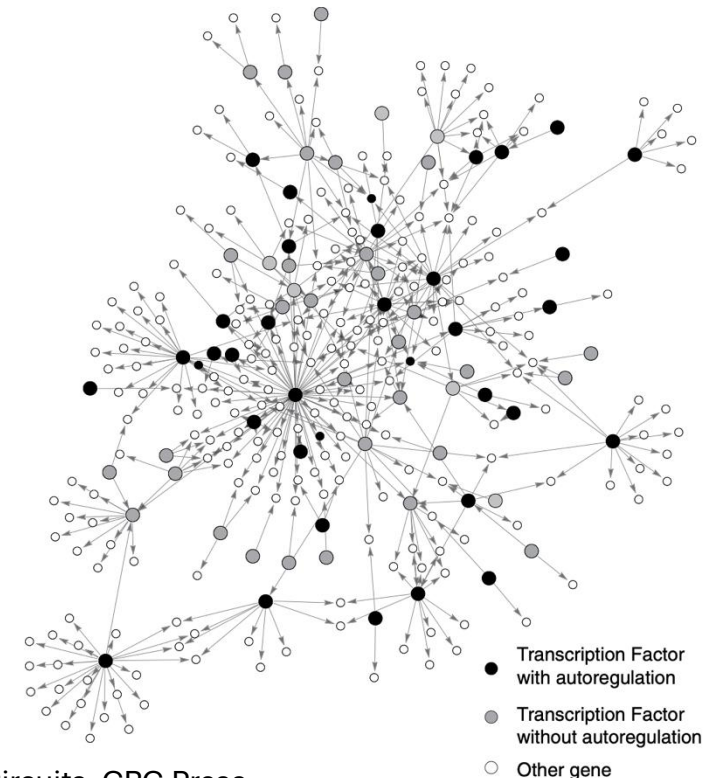
Real Network with Autoregulation



Random Network

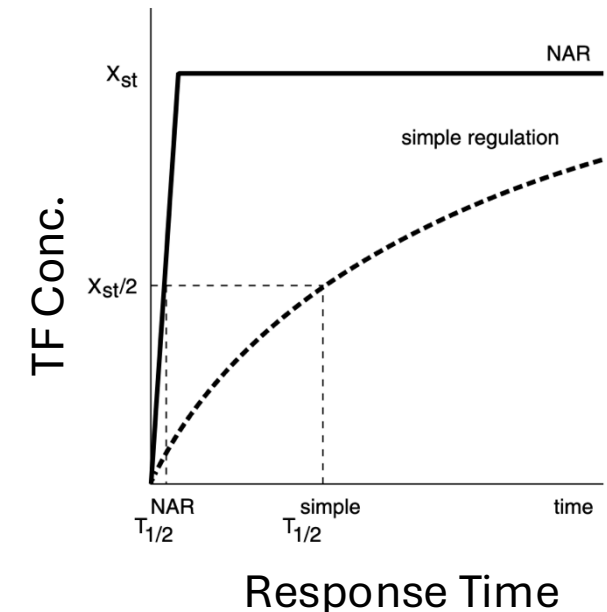
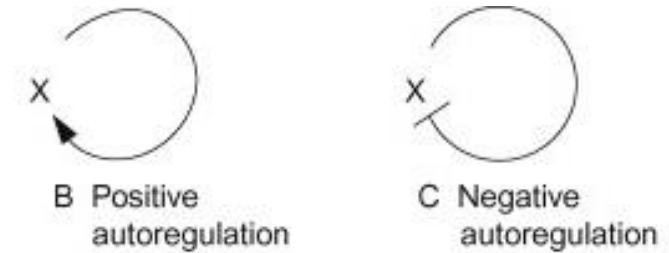


Clustering Coefficient: 1.0

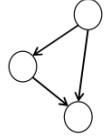
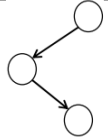
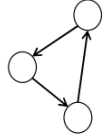
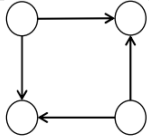
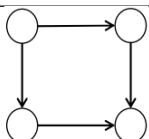
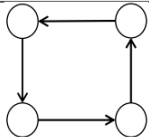


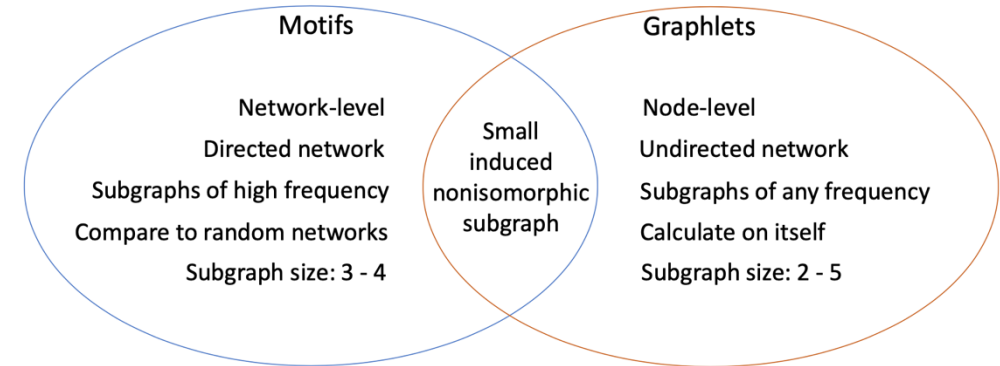
Regulation slows/speeds the response

- Positive regulation
 - **Slows response time**: Positive feedback often reinforces itself, leading to a delay before reaching a steady state.
 - **Bistability**: Positive feedback can create two stable states (e.g., ON or OFF), enabling decision-making processes in cells
- Negative Regulation
 - **Speeds response time**: Negative feedback counteracts changes quickly, stabilizing the system and making it more responsive to fluctuations.
 - **Reduces Variability**: Negative feedback ensures a steady-state concentration of molecules like proteins, reducing differences across cells in a population.

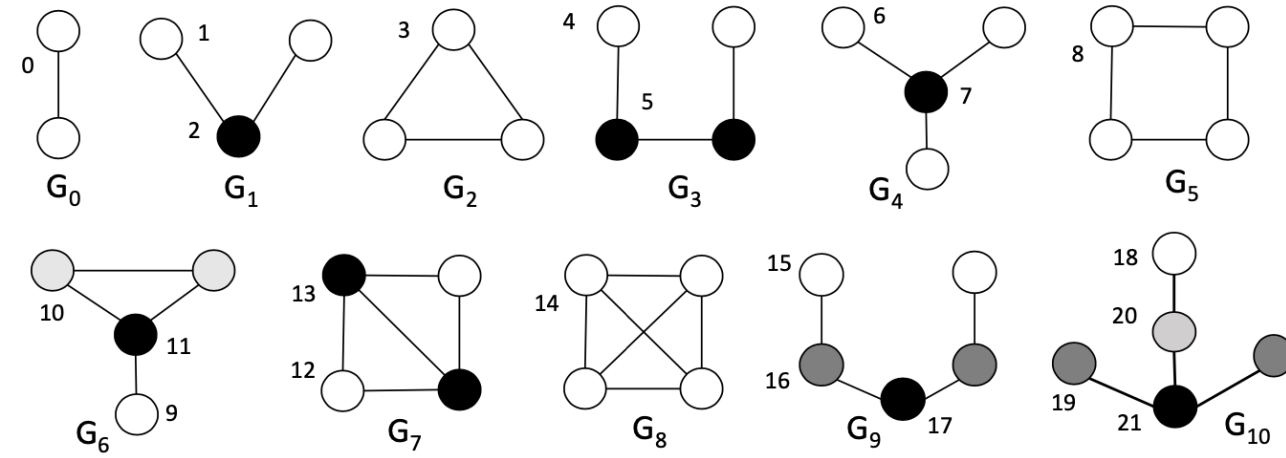


Occurrence of network motifs

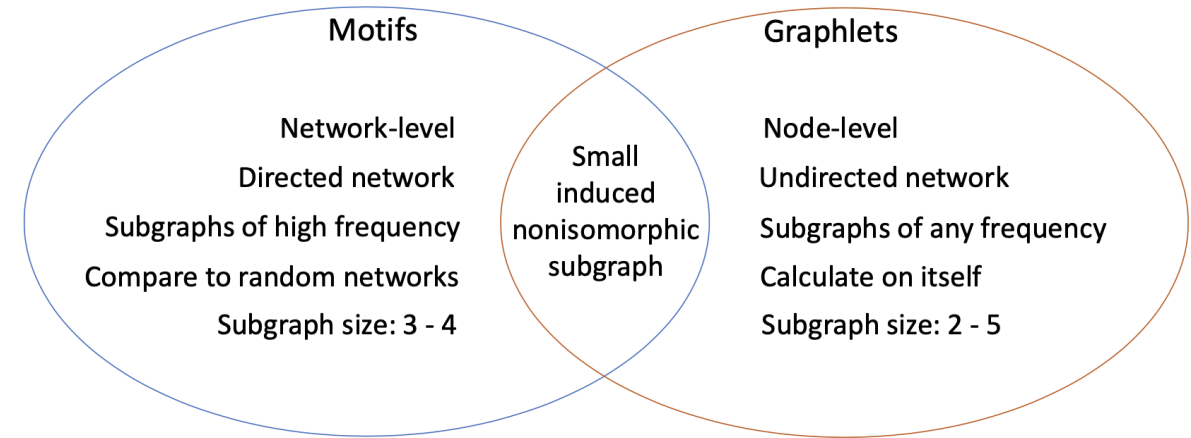
Motif	Designation	Type of network
	3-node feed-forward loop	Gene regulation network Neural network Electronic circuits (forward logic chips)
	3-chain	Food webs
	3-node feedback loop	Gene regulation network Neural network Electronic circuits (forward logic chips)
	Bi-fan	Gene regulation network Neural network Electronic circuits (forward logic chips) Electronic circuits II
	Bi-parallel	Neural network Food webs Electronic circuits (forward logic chips)
	4-node feedback loop	Electronic circuits II



Occurrence of graphlets



Orbits: Specific positions a node occupies within a graphlet, considering the symmetry of the graphlet. Nodes in symmetric positions are part of the same orbit.

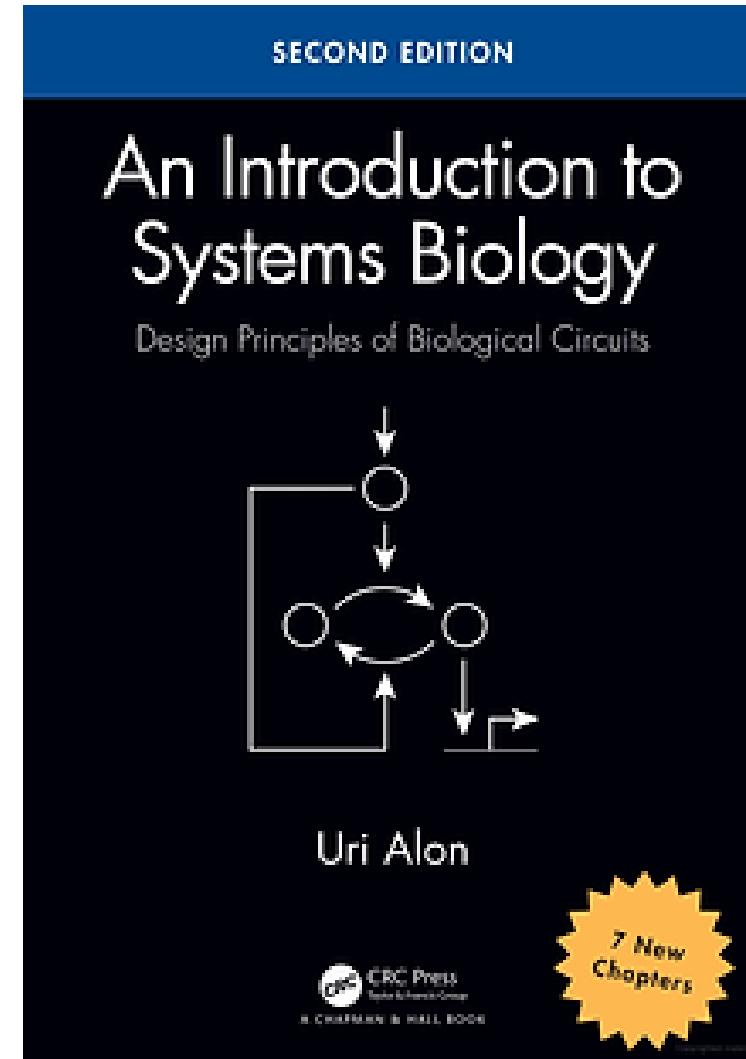
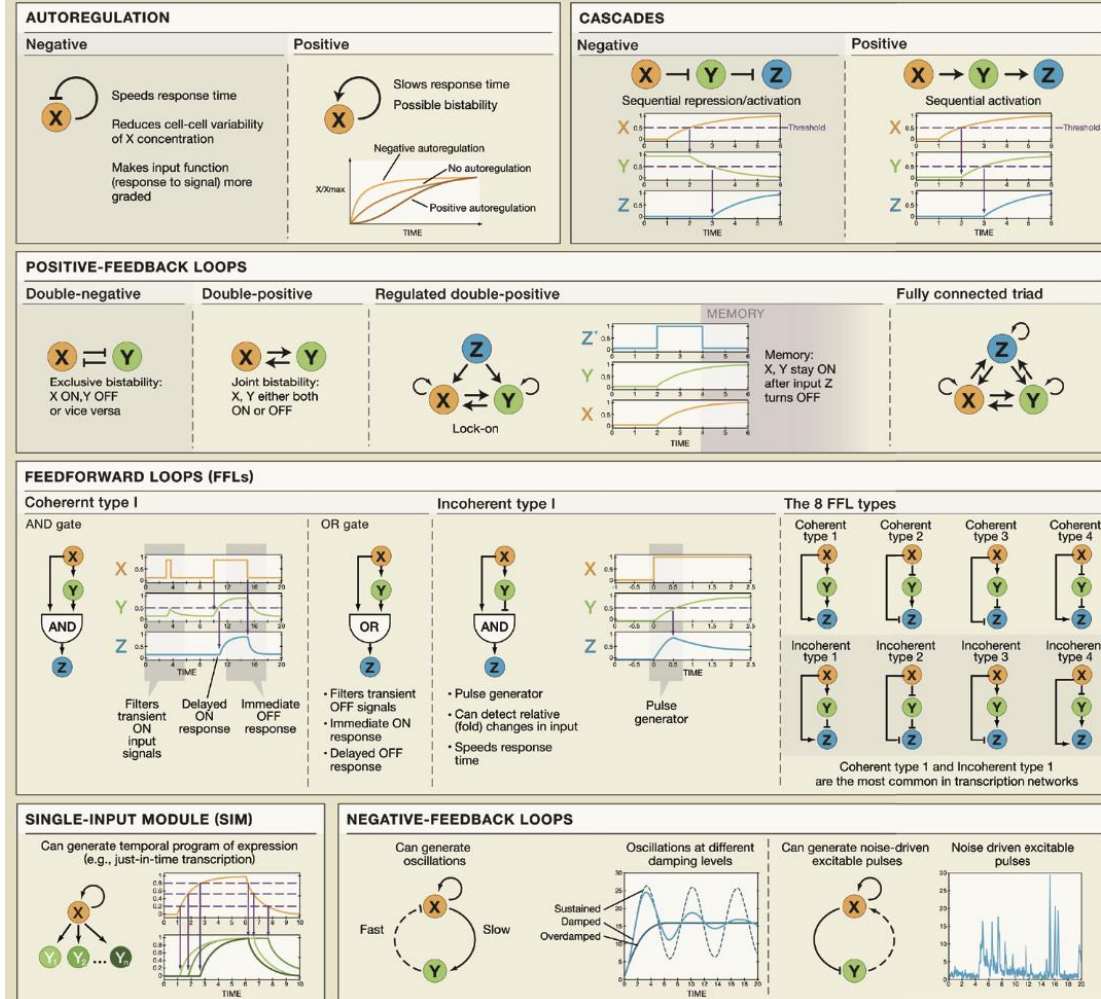


Resources: Motif connections to systems biology

SnapShot: Network Motifs

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

Cell

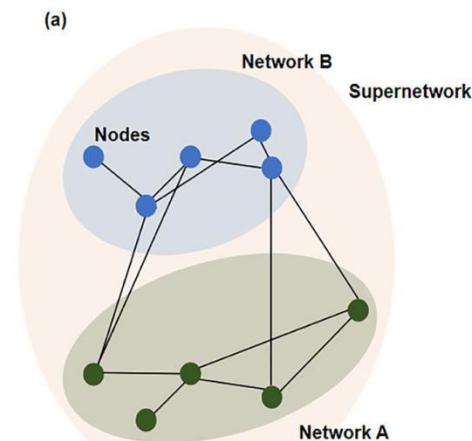


Oren Shoval and Uri Alon, Cell 2010

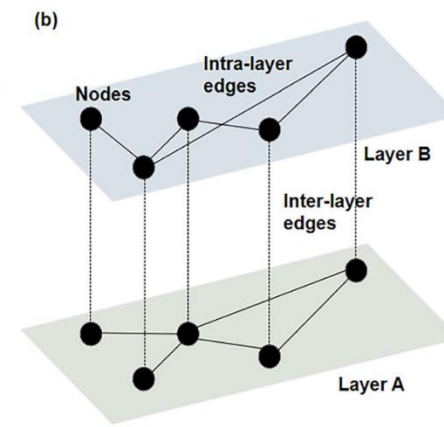
Let's have some fun with python
programming to find motifs

Multilayer networks

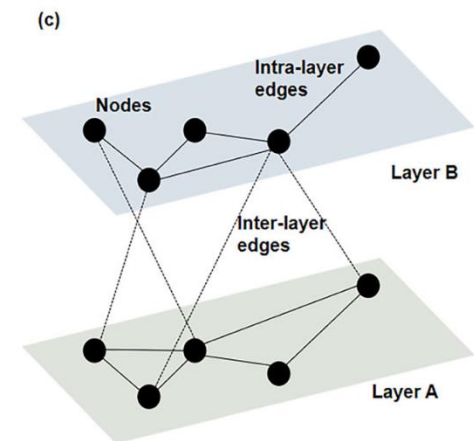
- Multilayer networks provide the framework to capture the complexity typical of systems of systems.
- It enable the analysis of biophysical, social and human-made networks from an integrated perspective.
- Layered structure of multilayer networks is responsible for phenomena that cannot be observed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



A network of two network systems A and B



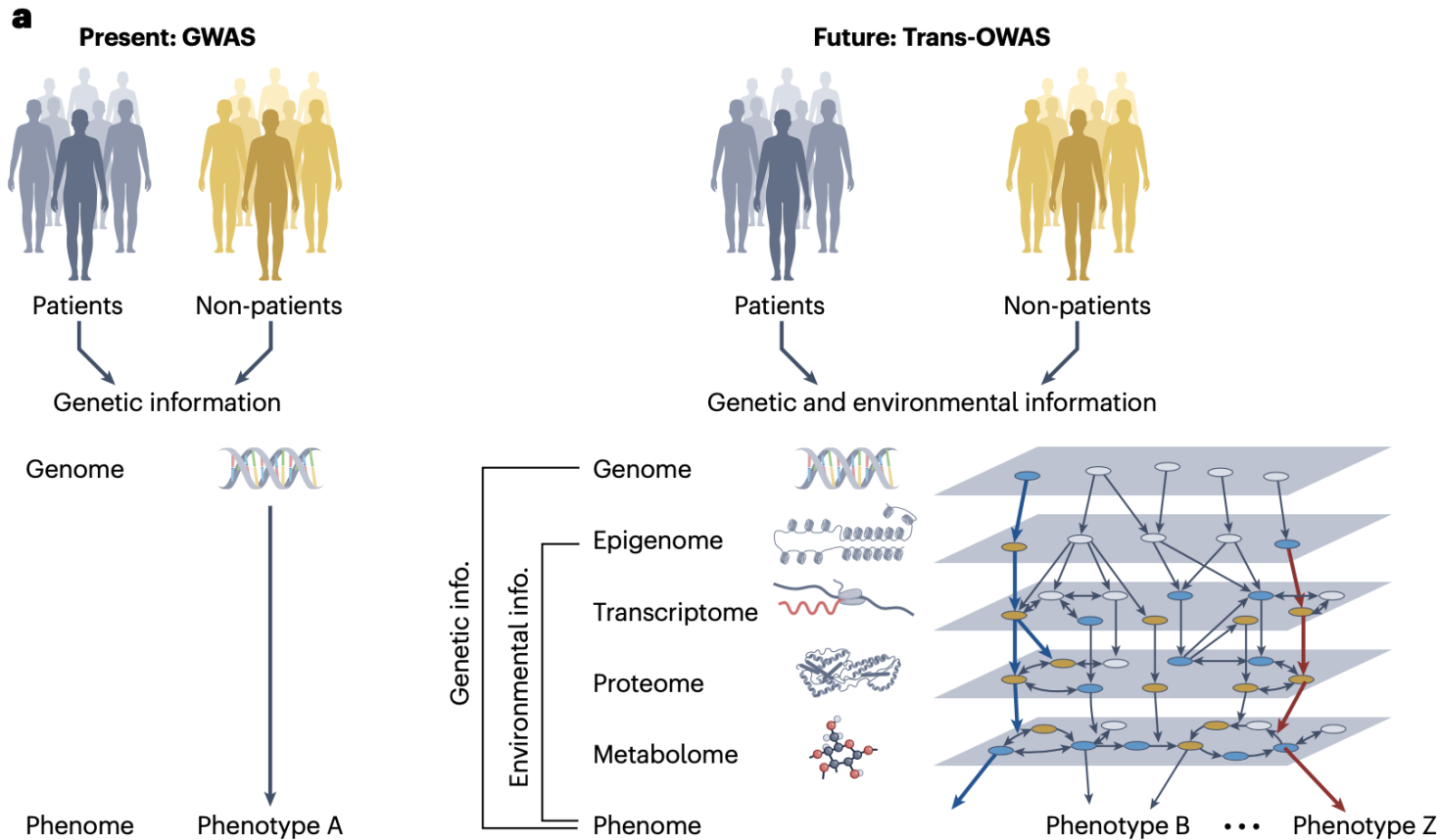
A multiplex network of two layers A and B



An interconnected network of two layers A and B

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

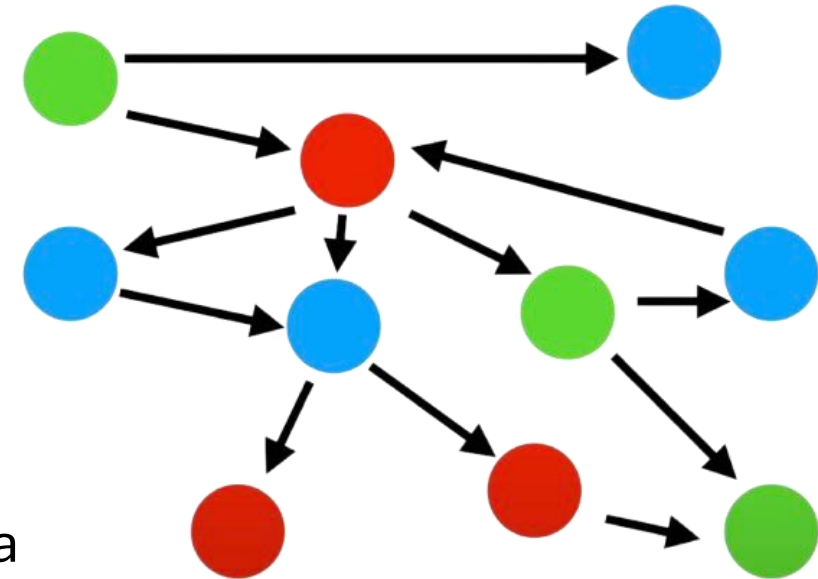
Multilayer networks



- 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.

A Knowledge Graph is a data set that is:

- **Structured** in the form of a specific data structure.
- **Normalised** consisting of small units, such as vertices and edges.
- **Connected** defined by the – possibly distant – connections between objects.
- **Explicit** created purposefully with an intended meaning.
- **Declarative** meaningful in itself, independent of a particular implementation or algorithm.
- **Annotated** enriched with contextual information to record additional details and meta-data.
- **Non-hierarchical** more than just a tree-structure.
- **Large** millions rather than hundreds of elements.

Biological molecule is also context dependent

“apple” is a **polysemic** word...

Google

🔍 grow an apple

🔍 buy an apple|

... whose **particular meaning** is resolved via **sentence context**

Google

🔍 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|

🔍 buy an apple **watch**

🔍 buy an apple **gift card**

buy an apple **tv**

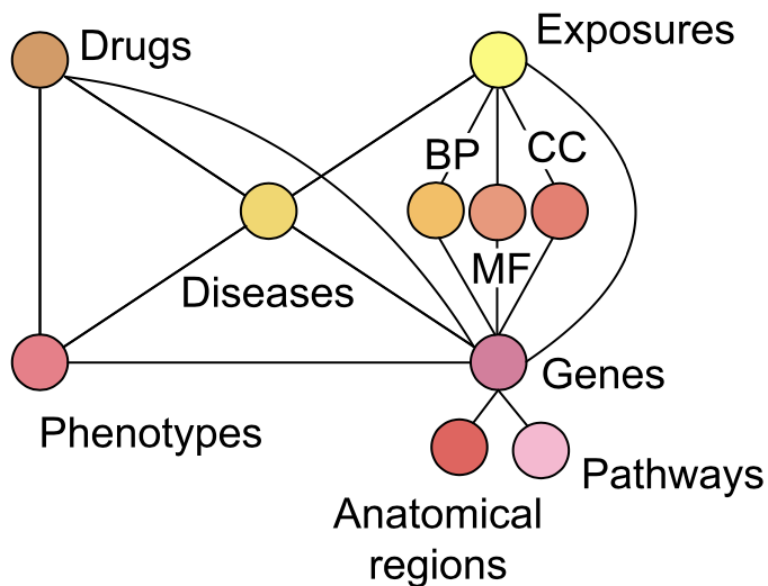


Knowledge graph in precision medicine

- 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

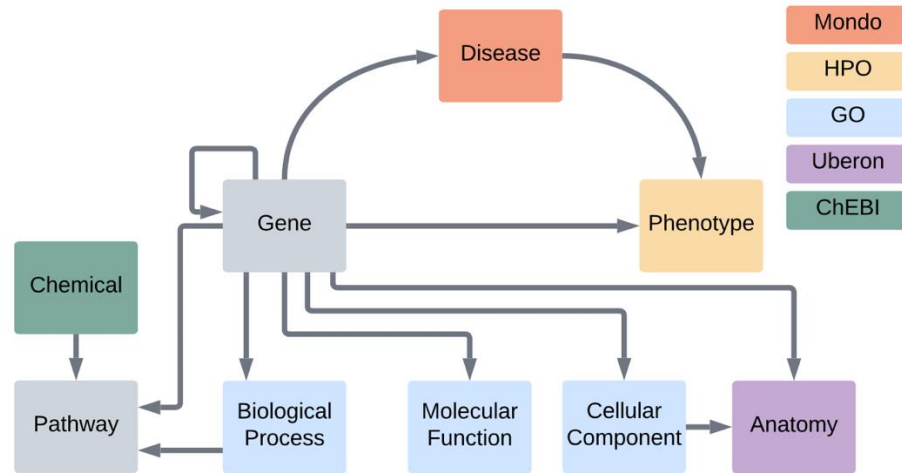
Existing knowledge graph in biomedical

It connects molecular, genetic, phenotypic, and drug-related information, enabling analyses like [drug-disease interactions](#), [disease mechanisms](#), and [therapeutic predictions](#).



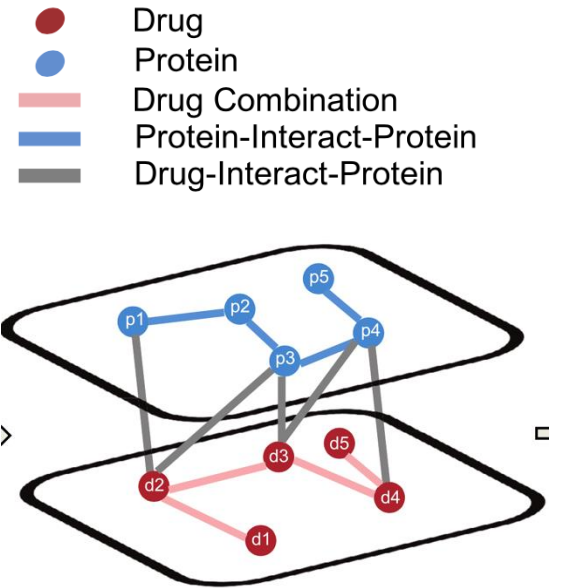
PrimeKG: Payal Chandak et al
Scientific Data 2023

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](#).



Monarch: Tim E. Putman et al
NAR 2024

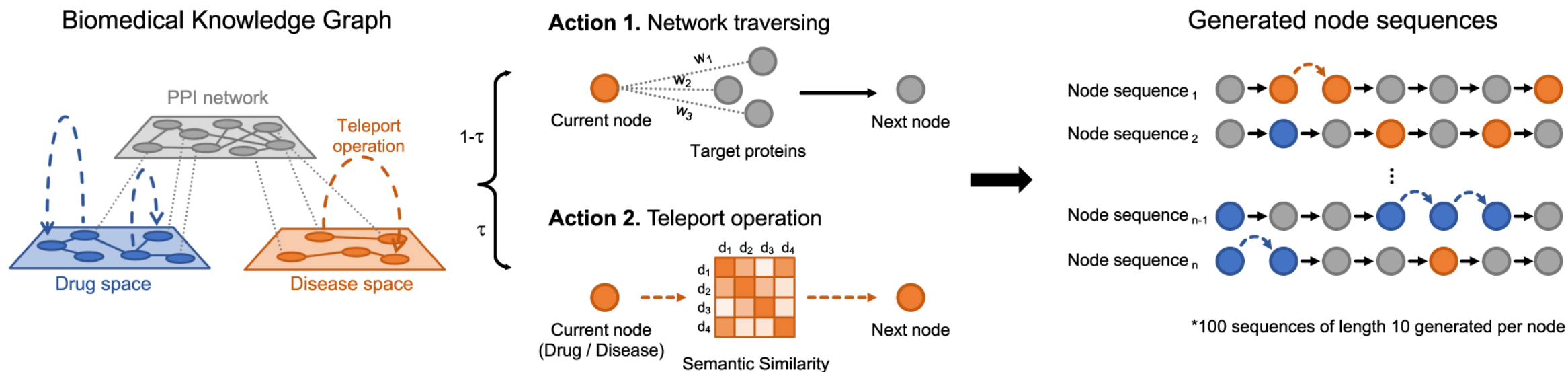
Predict Drug combination therapies (DCBs) for infectious diseases



Qing Ye et al Cell Reports 2023

Biomedical knowledge graph for drug repurposing

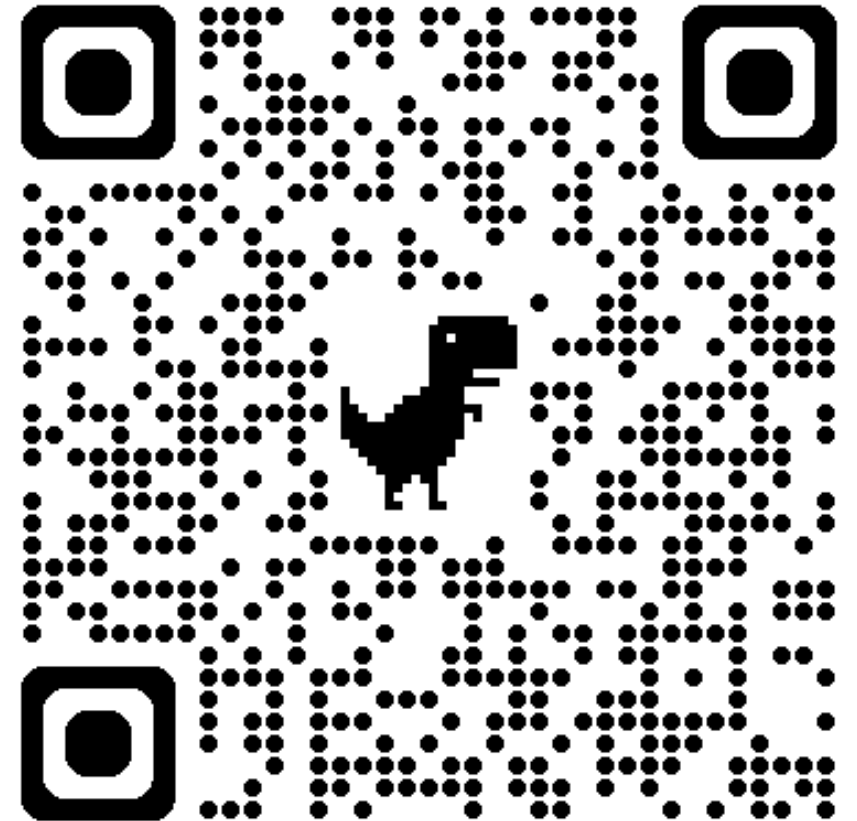
DREAMwalk map drugs and diseases into a unified embedding space, improving prediction of drug-disease associations.



Last Tutorial

Thank you

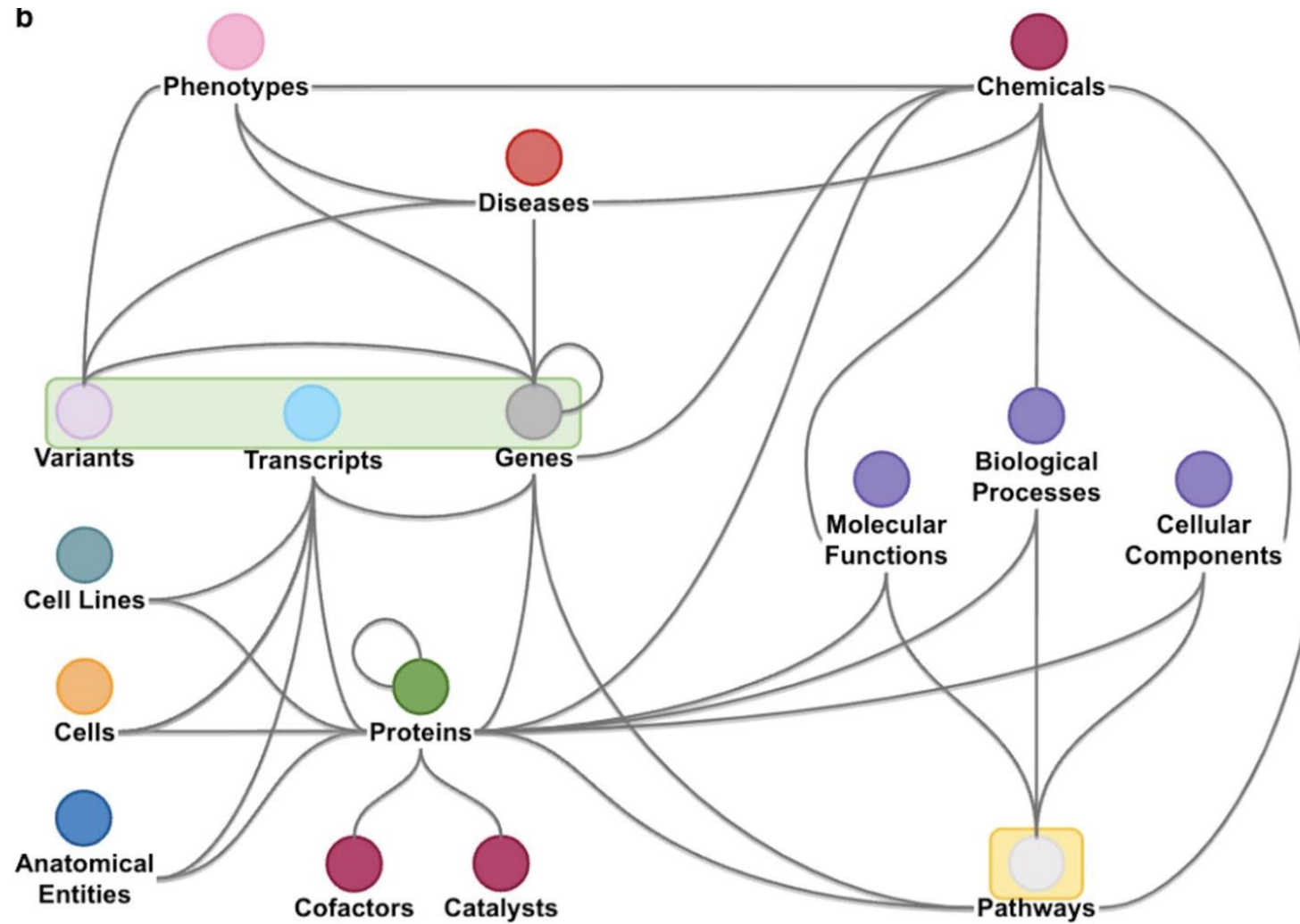
Scan me to download the
presentation and the codes



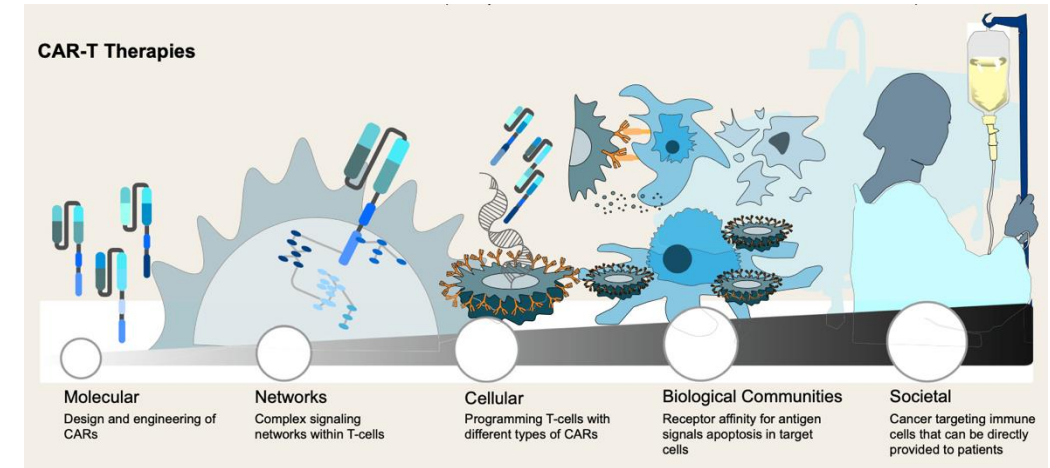
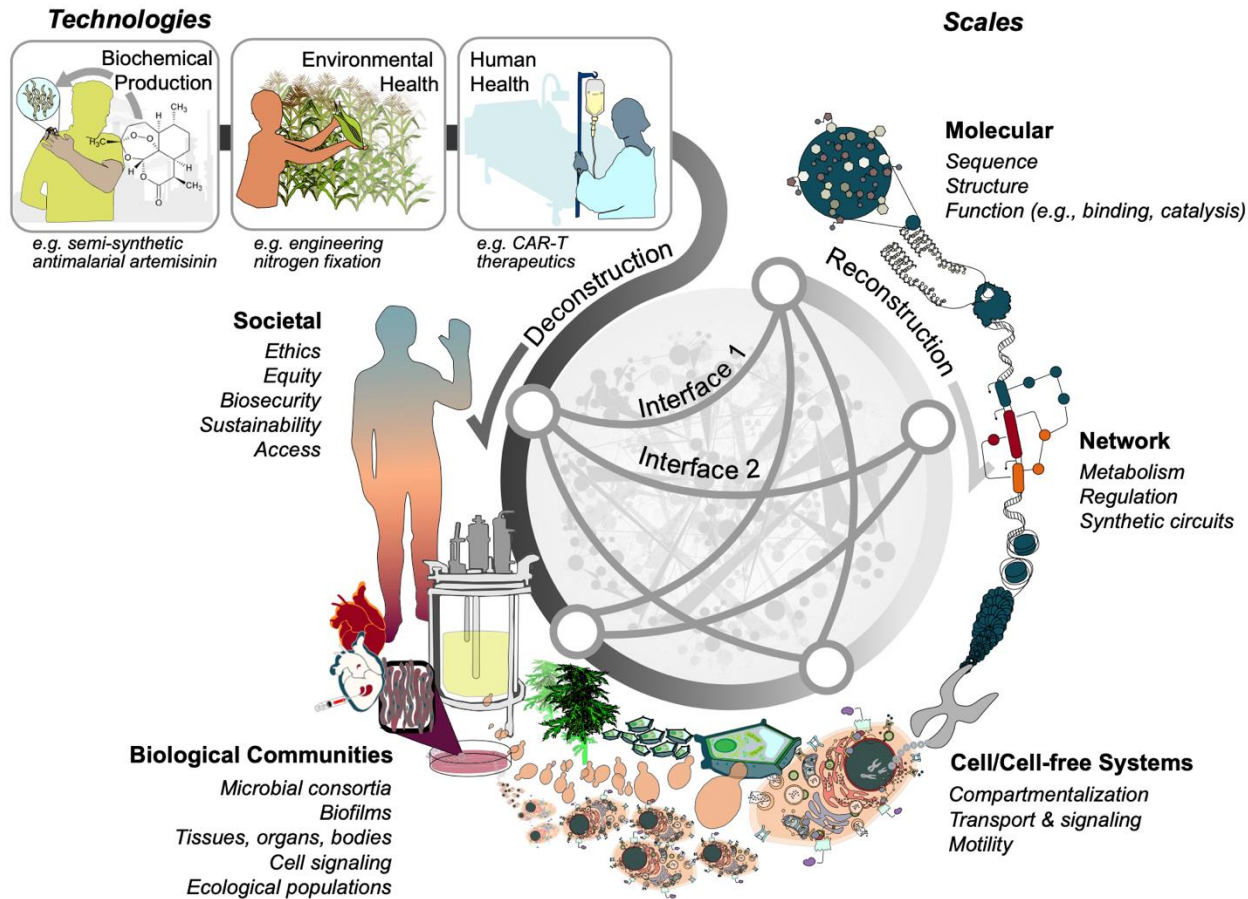
Famous algorithms related to graphs

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

Knowledge graph of a disease

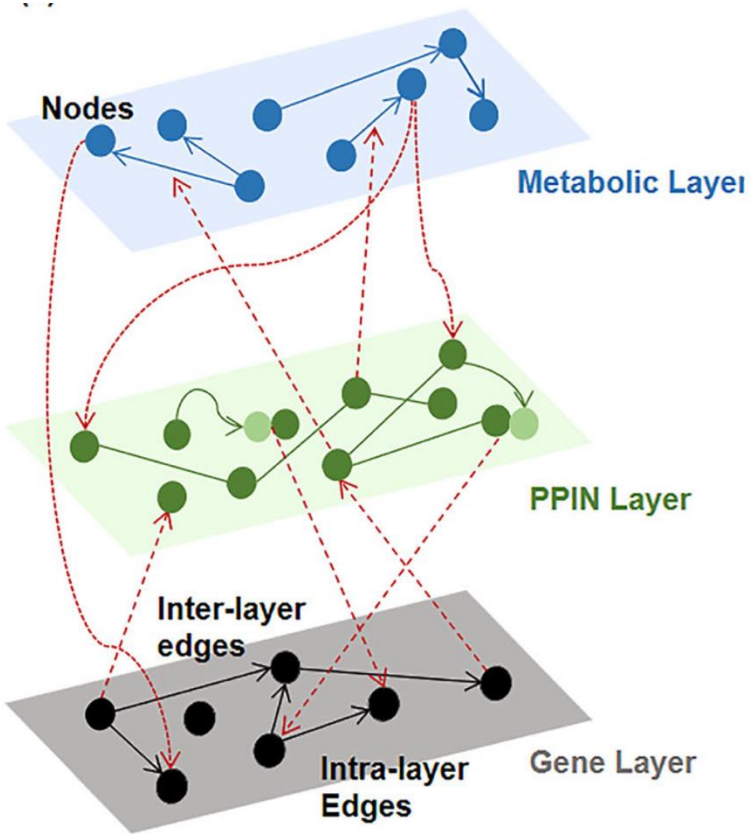


Deconstructing synthetic technologies across scales



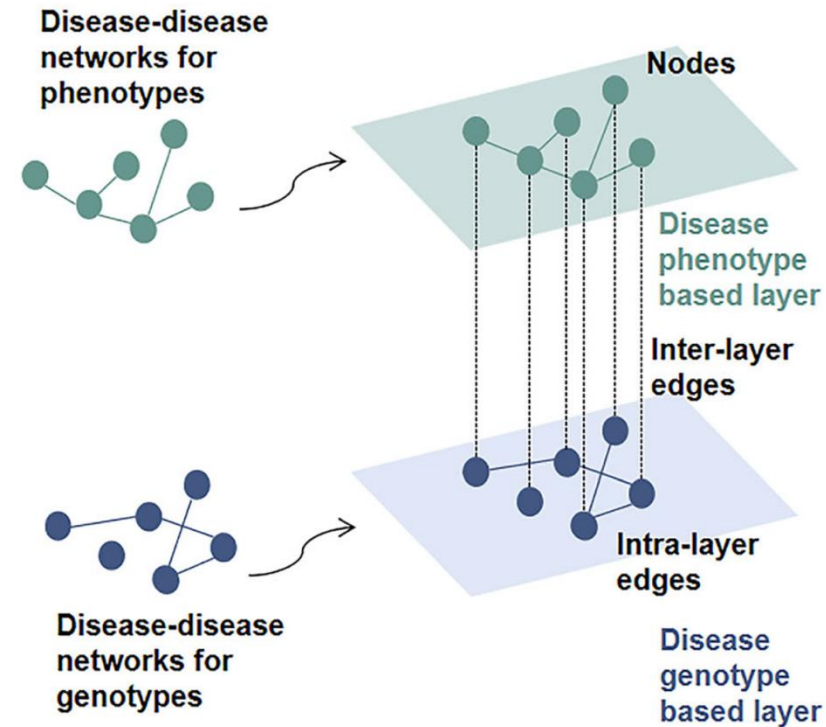
Multilayer network exa

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