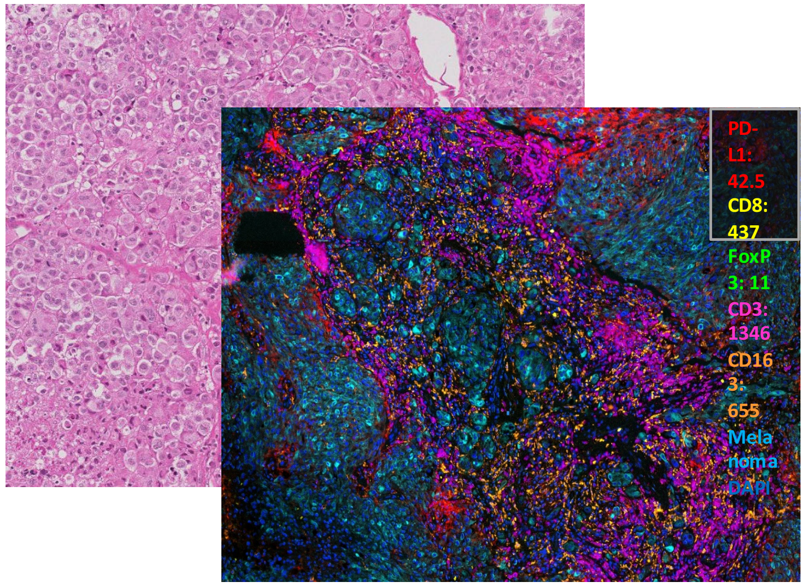
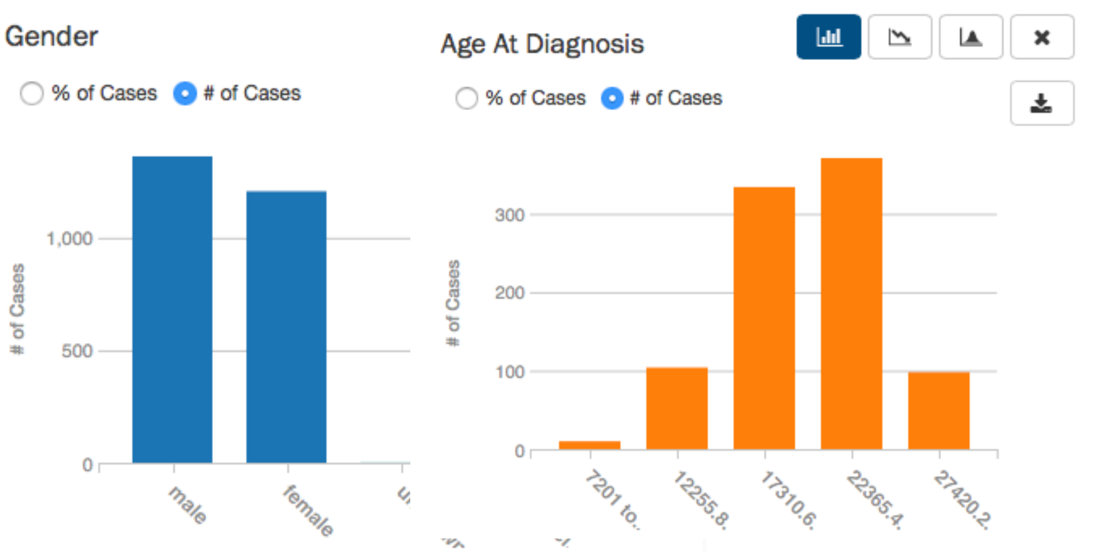
EE626: Graph representations for biology and medicine - Introduction

Dr Dorina Thanou
10.09.2025

Technological advancements are transforming biomedicine

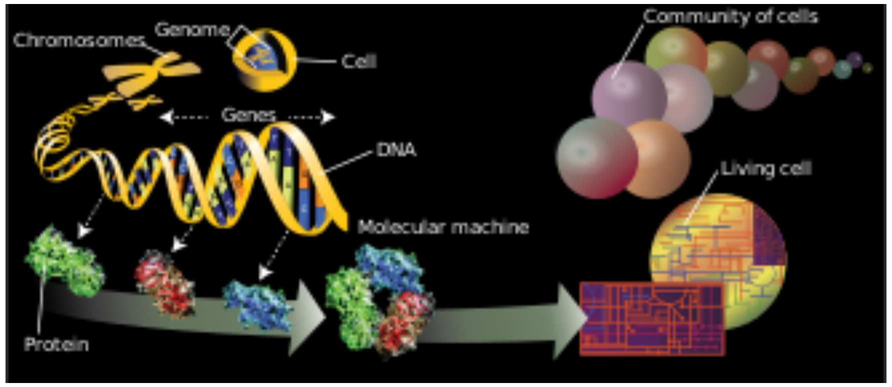




https://portal.gdc.cancer.gov

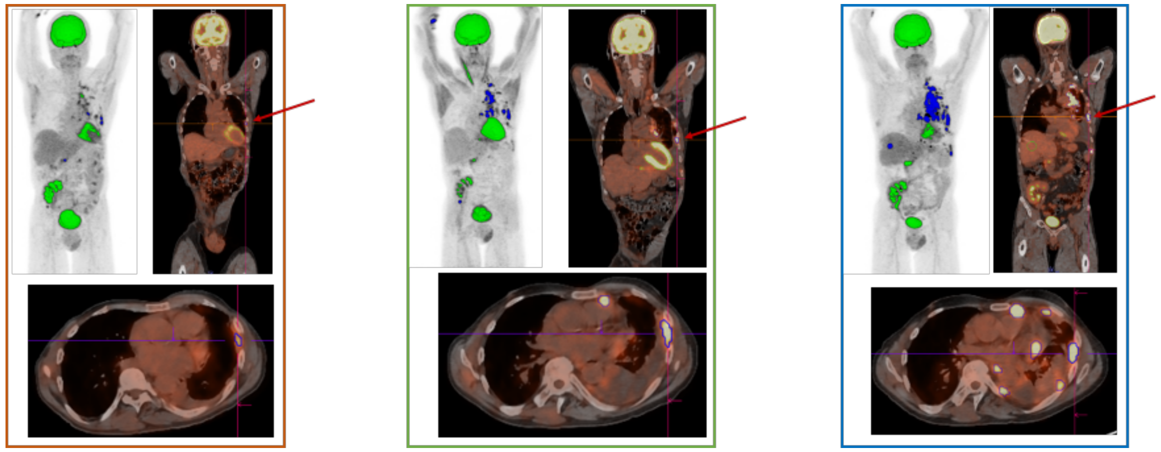
Clinical data

Image modalities



https://en.wikipedia.org/wiki/Omics

OMICS

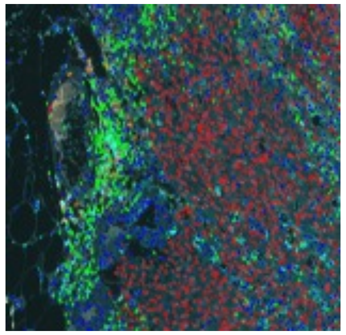


Time varying data

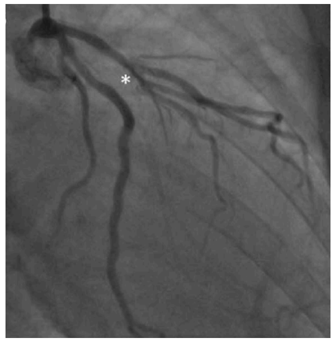
Unprecedented opportunities for novel biomedical discoveries!

Typical challenges in biomedical applications

- Diagnostic or predictive tasks (usually supervised):
  - Is a patient diagnosed with a disease?
  - Is a patient going to respond to a therapy?

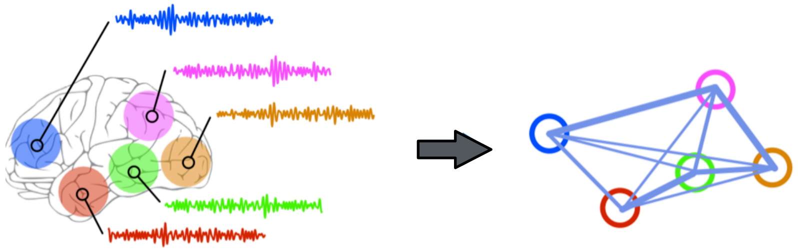


Respondent or non respondent?



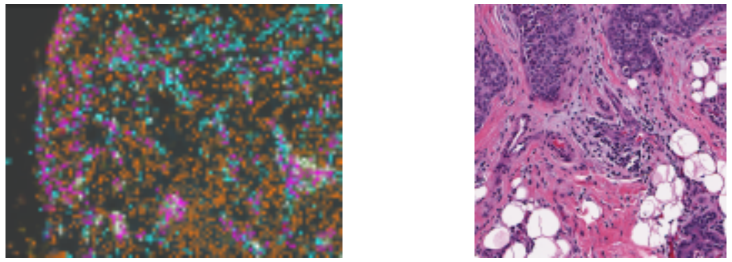
Healthy or not?

- Knowledge discovery (usually unsupervised):
  - What are predictive biomarkers of a disease?
  - What is the underlying biological mechanism?
  - Can we discover new therapeutics?



Why inference from biomedical data is challenging?

- **Data quality/availability**
  - Typically noisy, incomplete, heterogenous, captured from different technologies

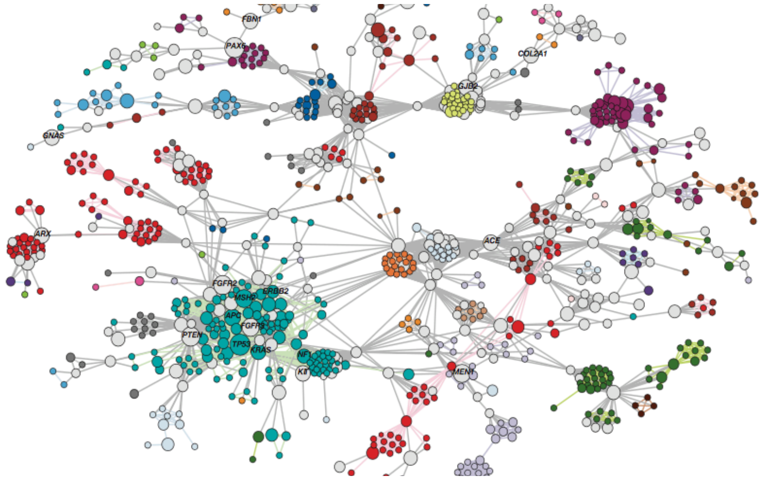


Why inference from biomedical data is challenging?

- **Data quality/availability**
  - Typically noisy, incomplete, heterogenous, captured from different technologies

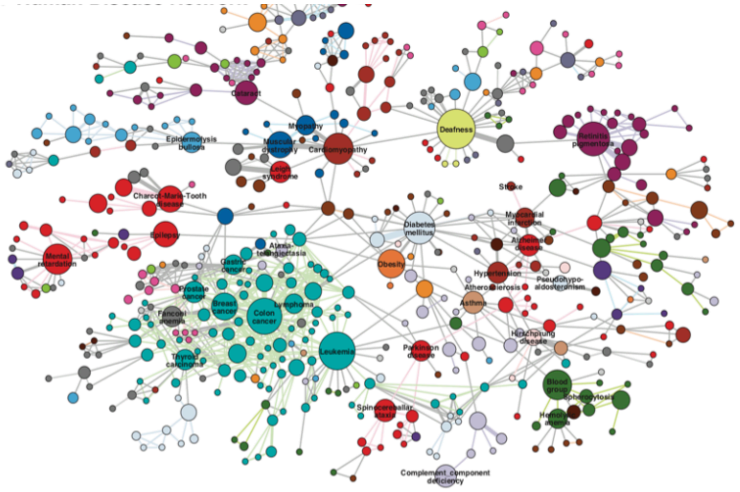
- **Complex interactions at different levels**
  - High complexity due to multiple, and heterogenous interactions at different levels

(from molecules to society)

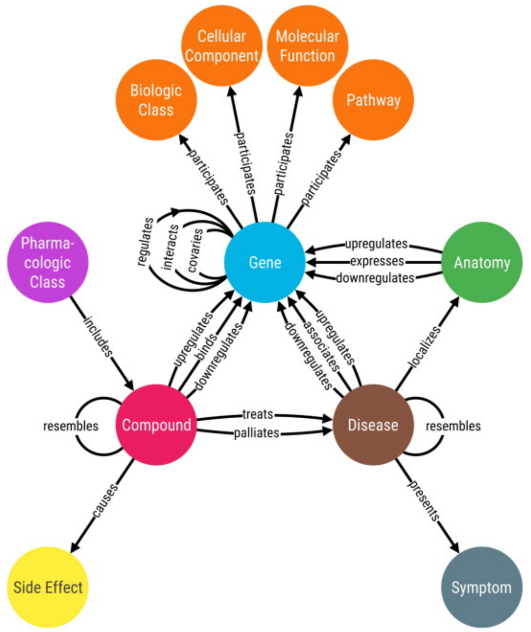


Disease gene network

[Goh et al 2007]



Human disease network



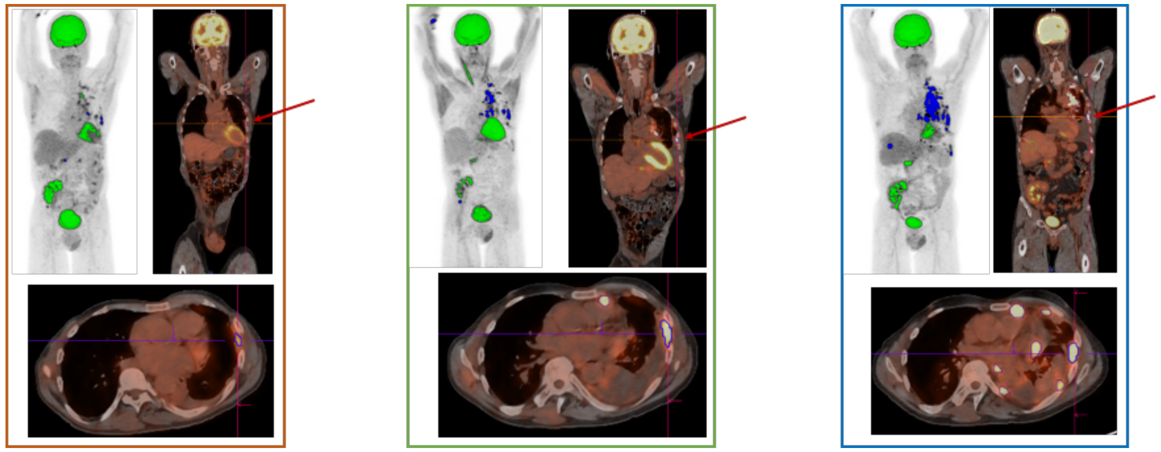
Biomedical knowledge graph

Why inference from biomedical data is challenging?

- **Data quality/availability**
    - Typically noisy, incomplete, heterogenous, captured from different technologies

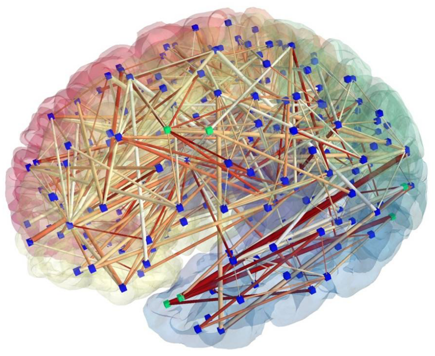
- **Complex interactions at different levels**
    - High complexity due to multiple, and heterogenous interactions at different levels (from molecules to society)

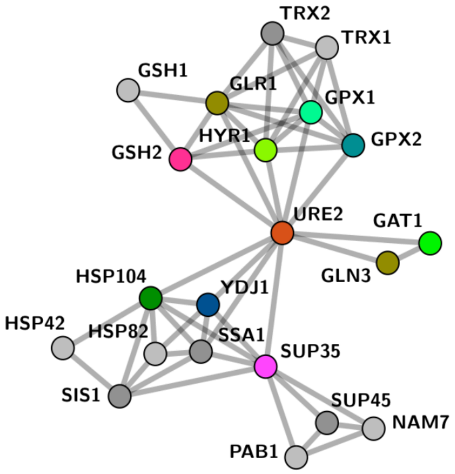
- **Many biological systems exhibit time-varying patterns**

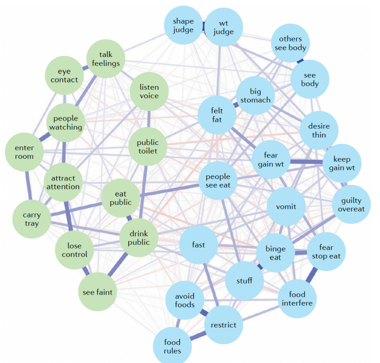


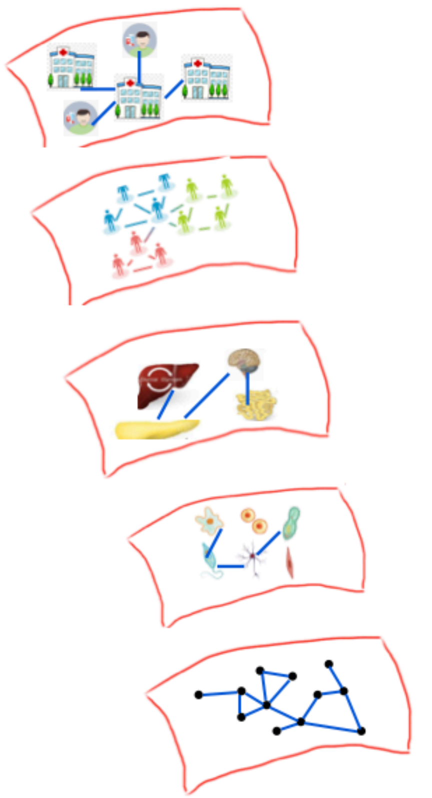
Networks for biology and medicine

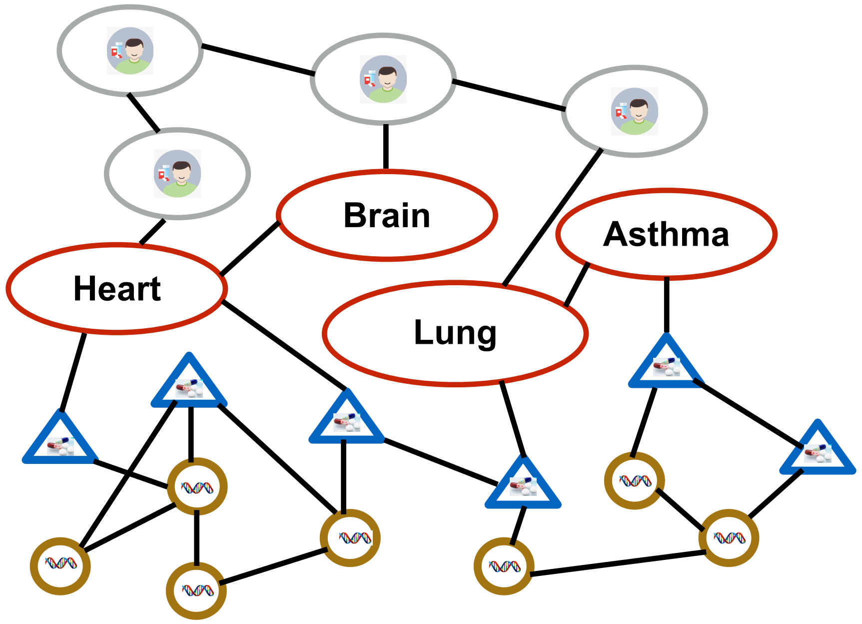
- Many biomedical data are represented by some networks
  - Spatial information, functional interactions, anatomical structures

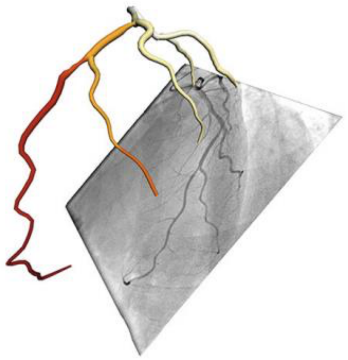


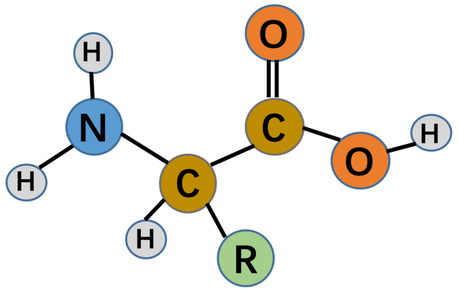


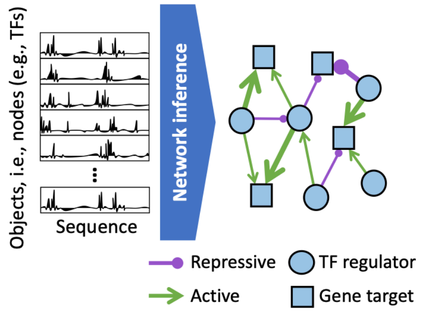






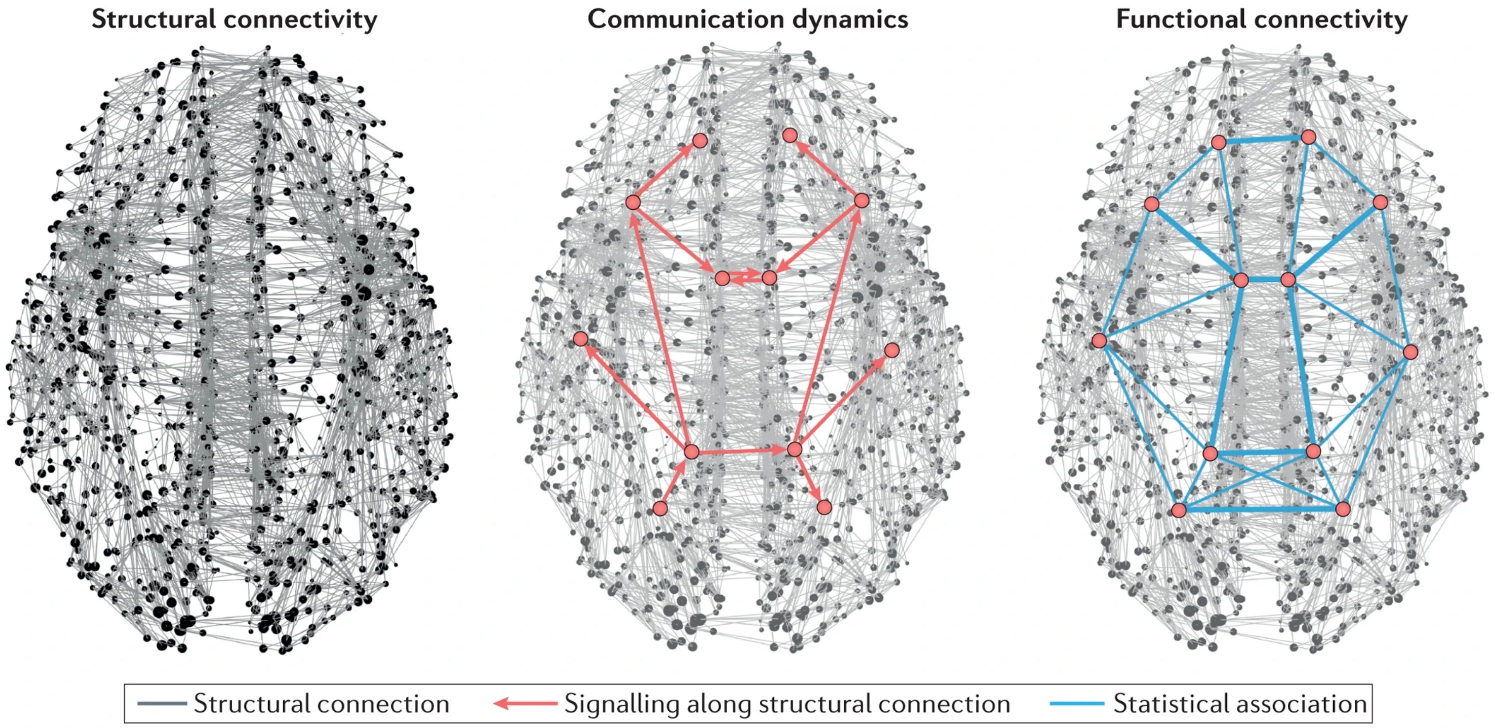






Capturing structure-function relationships

- Very often complex, biological systems generate communication events that temporally evolve on top of a structural network
- There is a strong interplay between anatomical structure and communication dynamics
- A network approach enables studying the underlying mechanism

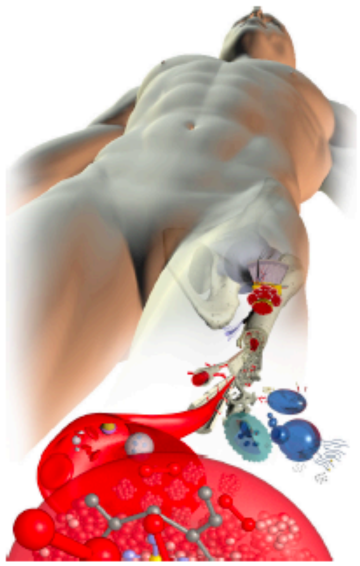


Nature Reviews | Neuroscience

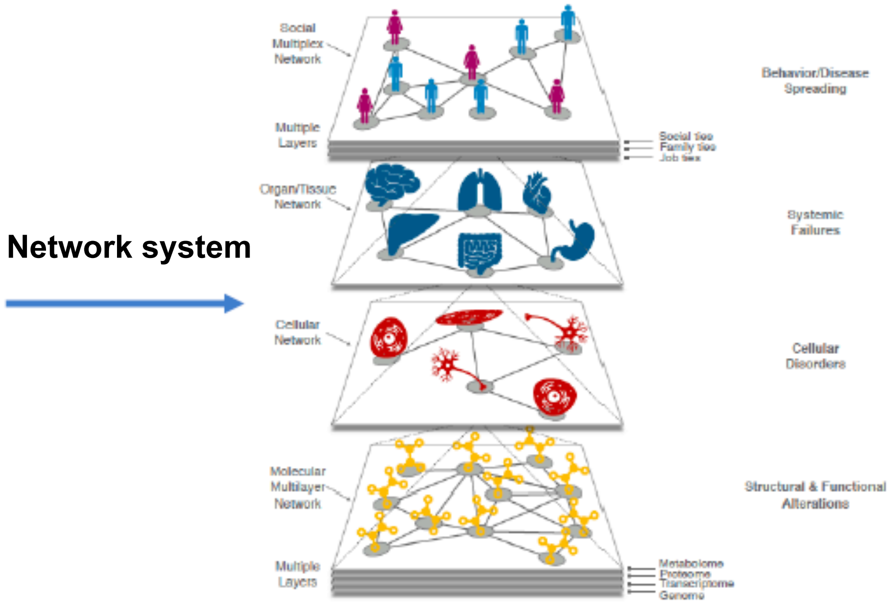
[Avena-Koenigsberger et al., Communication dynamics in complex brain networks, Nature Reviews Neuroscience, 2017]

A systemic view: biology is interconnected

- Nowadays, we can observe measurements at almost all scales
- Understanding complex mechanisms requires synthesizing information across different scales
- The human body can be seen as a **network system**
- A systemic, multi-scale, multi-organ approach could provide new insights by capturing processes that are mechanistically related



Network system



[Illustration adapted from P. Lio]

Network medicine: a network-based approach to human disease

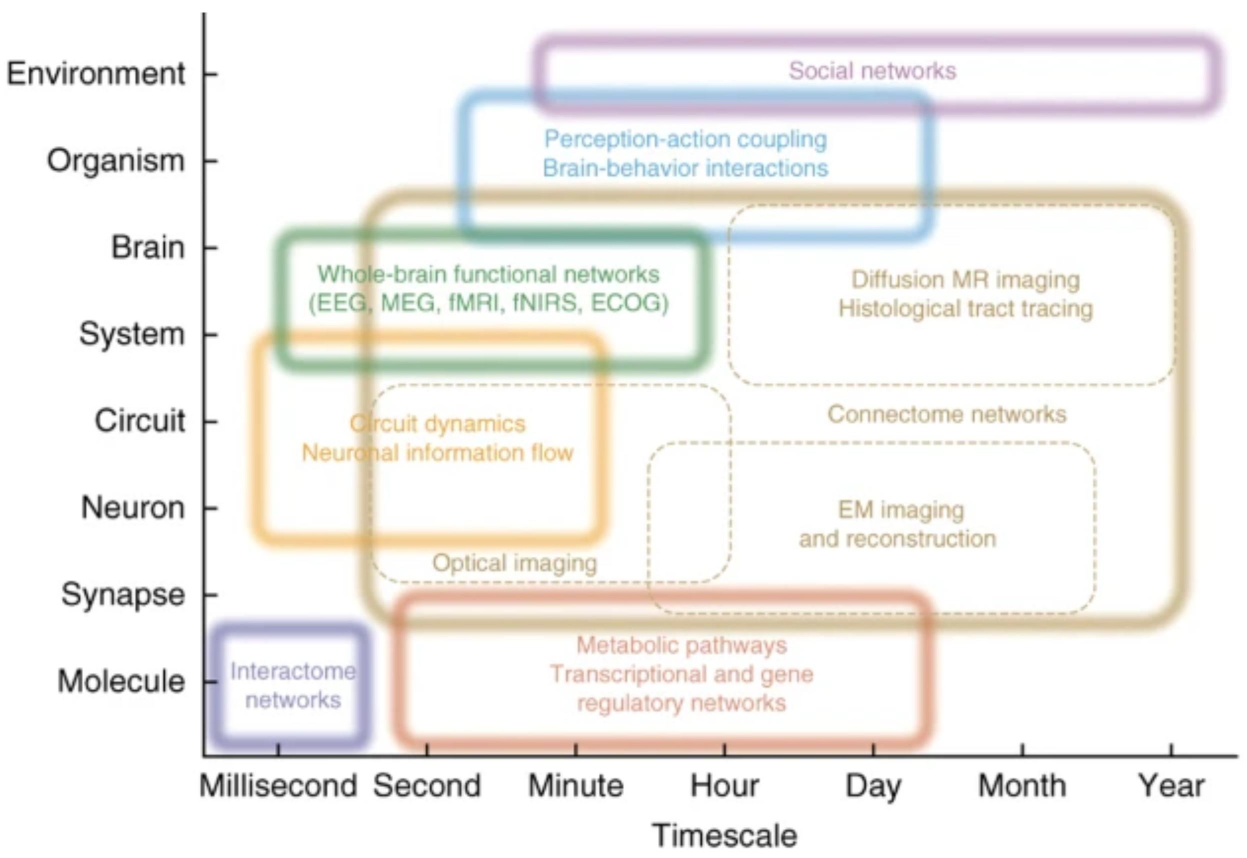
- A disease is rarely an abnormality in a single gene
- It reflects the perturbation of the complex intercellular network
- An integrated understanding of the interactions among the genome, the proteome, the environment mediated by the underlying cellular network, offers a basis for future advances

Think globally, act locally!

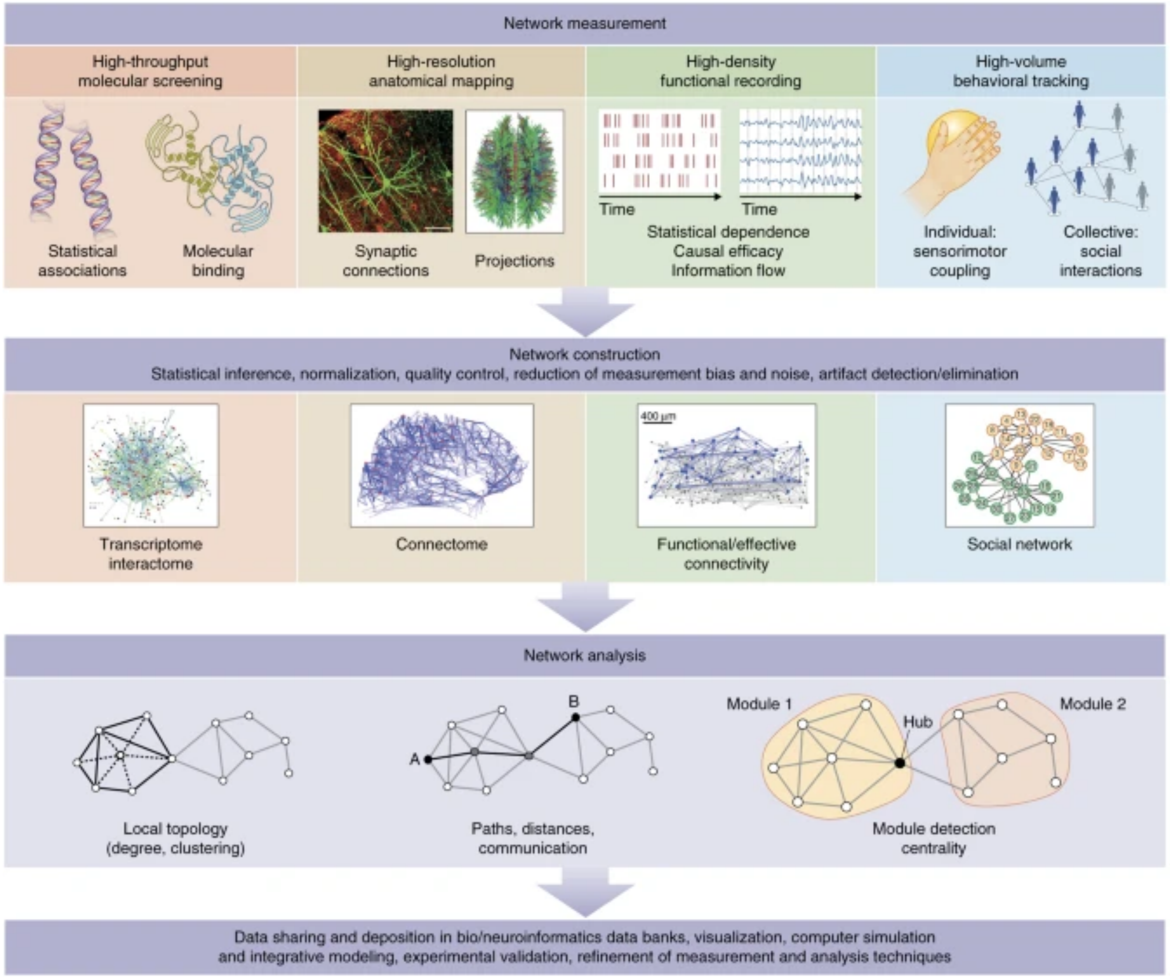
[A. Barabasi et al,.,Network Medicine: A network-based approach to human disease, Nature Reviews Genetics, 2011]

Network neuroscience

- Understand the principles and mechanisms underlying complex brain function and cognition from an integrative perspective, i.e., as a multiscale networked system



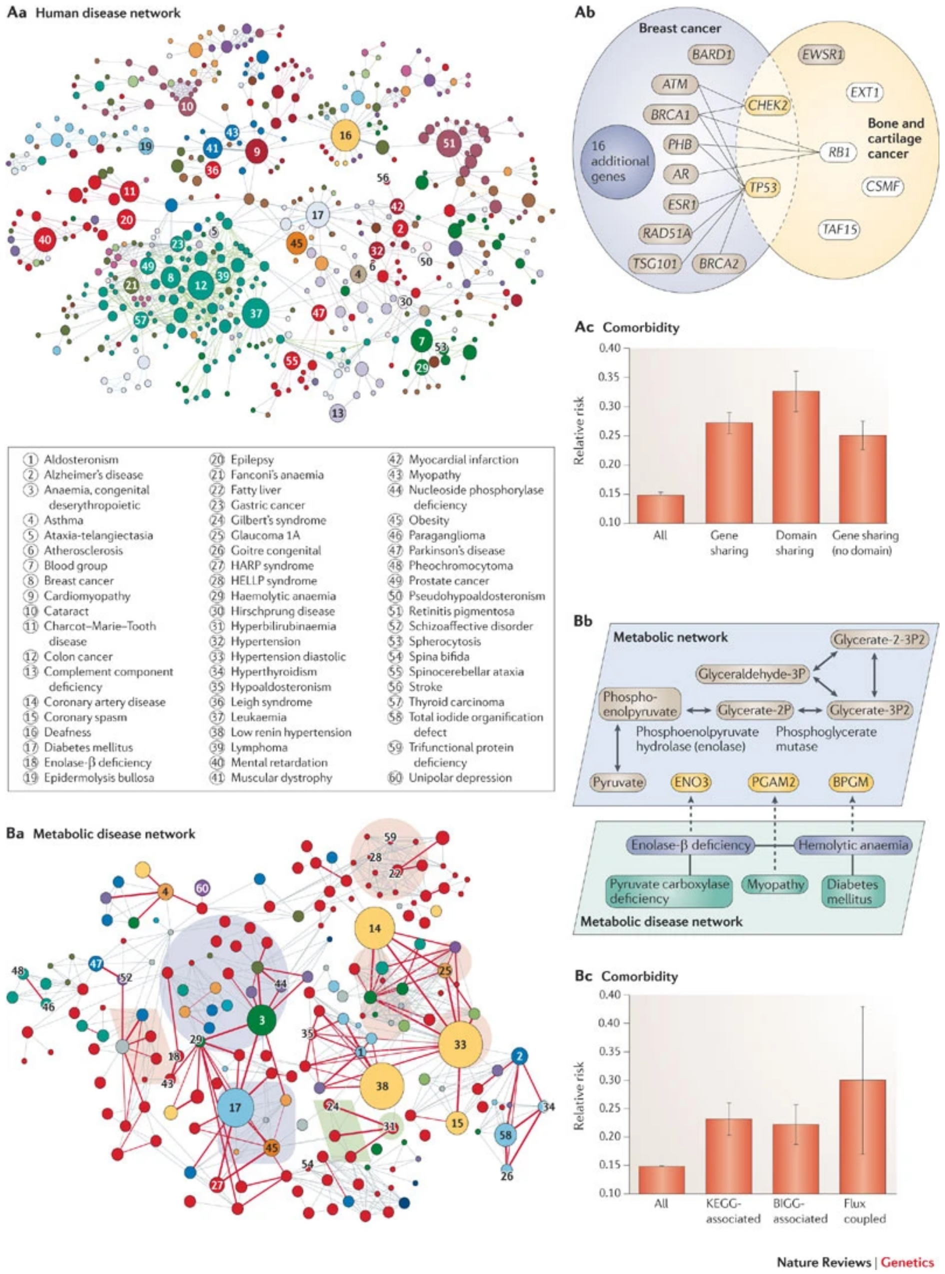
Debbie Maizels/Springer Nature



Debbie Maizels/Springer Nature

[A. Bassett et al., Network Neuroscience: A network-based approach to human disease, Nature Neuroscience, 2017]

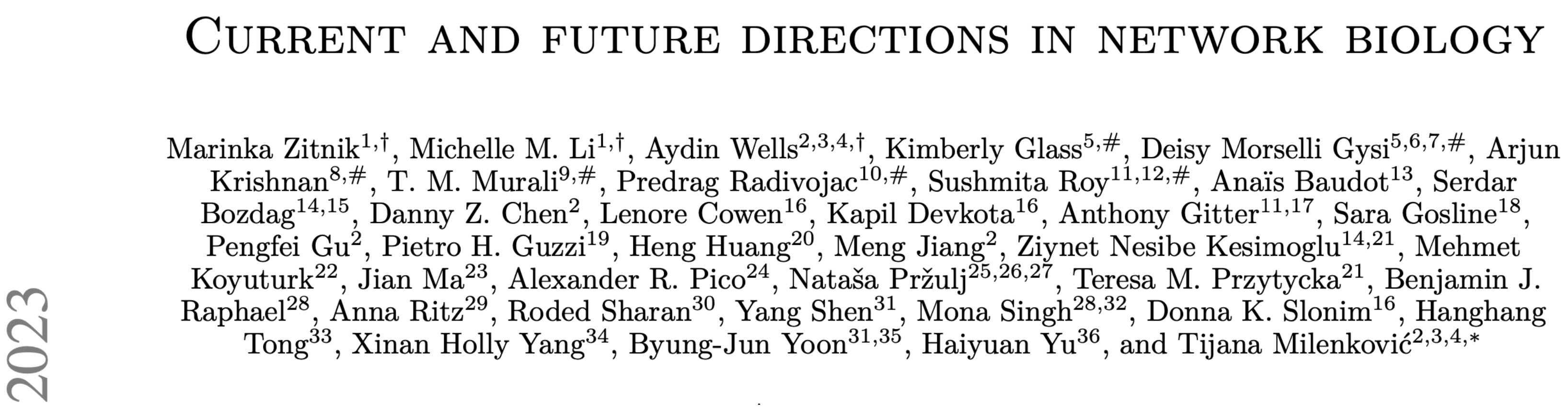
The human disease network



Nature Reviews | Genetics

[A. Barabasi et al, Network Medicine: A network-based approach to human disease, Nature Reviews Genetics, 2011]

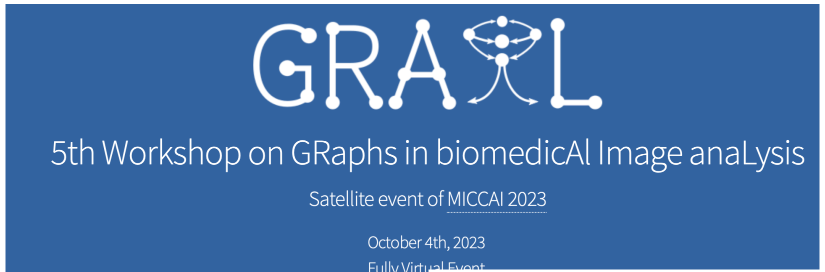
Network biomedicine: a growing field



2023

Current and future directions in network biology

Marinka Zitnik[1,†], Michelle M. Li[1,†], Aydin Wells[2,3,4,†], Kimberly Glass[5,#], Deisy Morselli Gysi[5,6,7,#], Arjun Krishnan[8,#], T. M. Murali[9,#], Predrag Radivojac[10,#], Sushmita Roy[11,12,#], Anaïs Baudot[13], Serdar Bozdag[14,15], Danny Z. Chen[2], Lenore Cowen[16], Kapil Devkota[16], Anthony Gitter[11,17], Sara Gosline[18], Pengfei Gu[2], Pietro H. Guzzi[19], Heng Huang[20], Meng Jiang[2], Ziynet Nesibe Kesimoglu[14,21], Mehmet Koyuturk[22], Jian Ma[23], Alexander R. Pico[24], Nataša Pržulj[25,26,27], Teresa M. Przytycka[21], Benjamin J. Raphael[28], Anna Ritz[29], Roded Sharan[30], Yang Shen[31], Mona Singh[28,32], Donna K. Slonim[16], Hanghang Tong[33], Xinan Holly Yang[34], Byung-Jun Yoon[31,35], Haiyuan Yu[36], and Tijana Milenković[2,3,4,*]



5th Workshop on GRaphs in biomedicAl Image anaLysis

Satellite event of MICCAI 2023

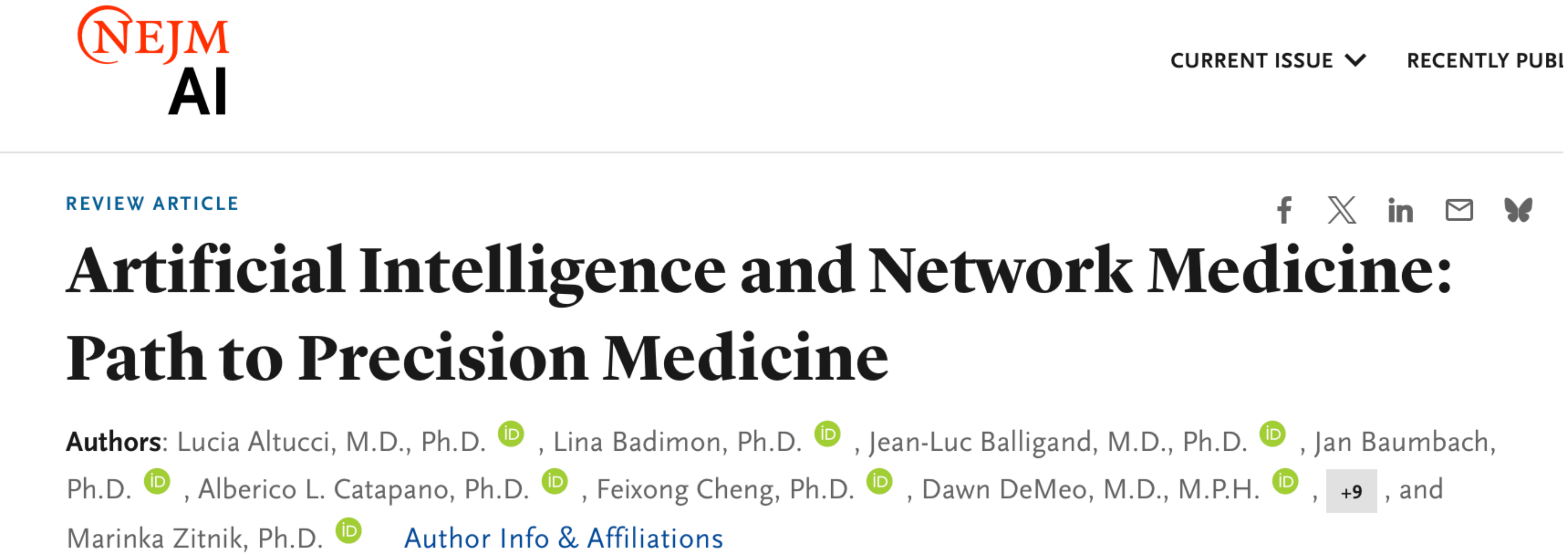
October 4th, 2023

Fully Virtual Event

ANNUAL REVIEW OF BIOMEDICAL DATA SCIENCE Volume 7, 2024

Review Article | Open Access

Graph Artificial Intelligence in Medicine

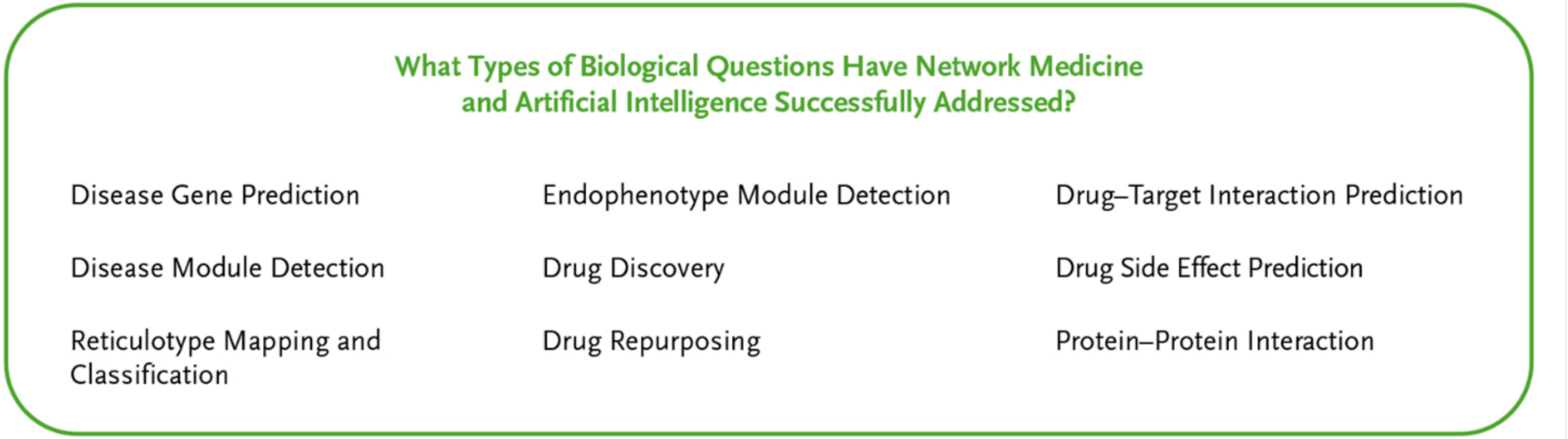


REVIEW ARTICLE

Artificial Intelligence and Network Medicine: Path to Precision Medicine

Authors: Lucia Altucci, M.D., Ph.D., Lina Badimon, Ph.D., Jean-Luc Balligand, M.D., Ph.D., Jan Baumbach, Ph.D., Alberico L. Catapano, Ph.D., Feixong Cheng, Ph.D., Dawn DeMeo, M.D., M.P.H., +9, and Marinka Zitnik, Ph.D. Author Info & Affiliations

Success stories

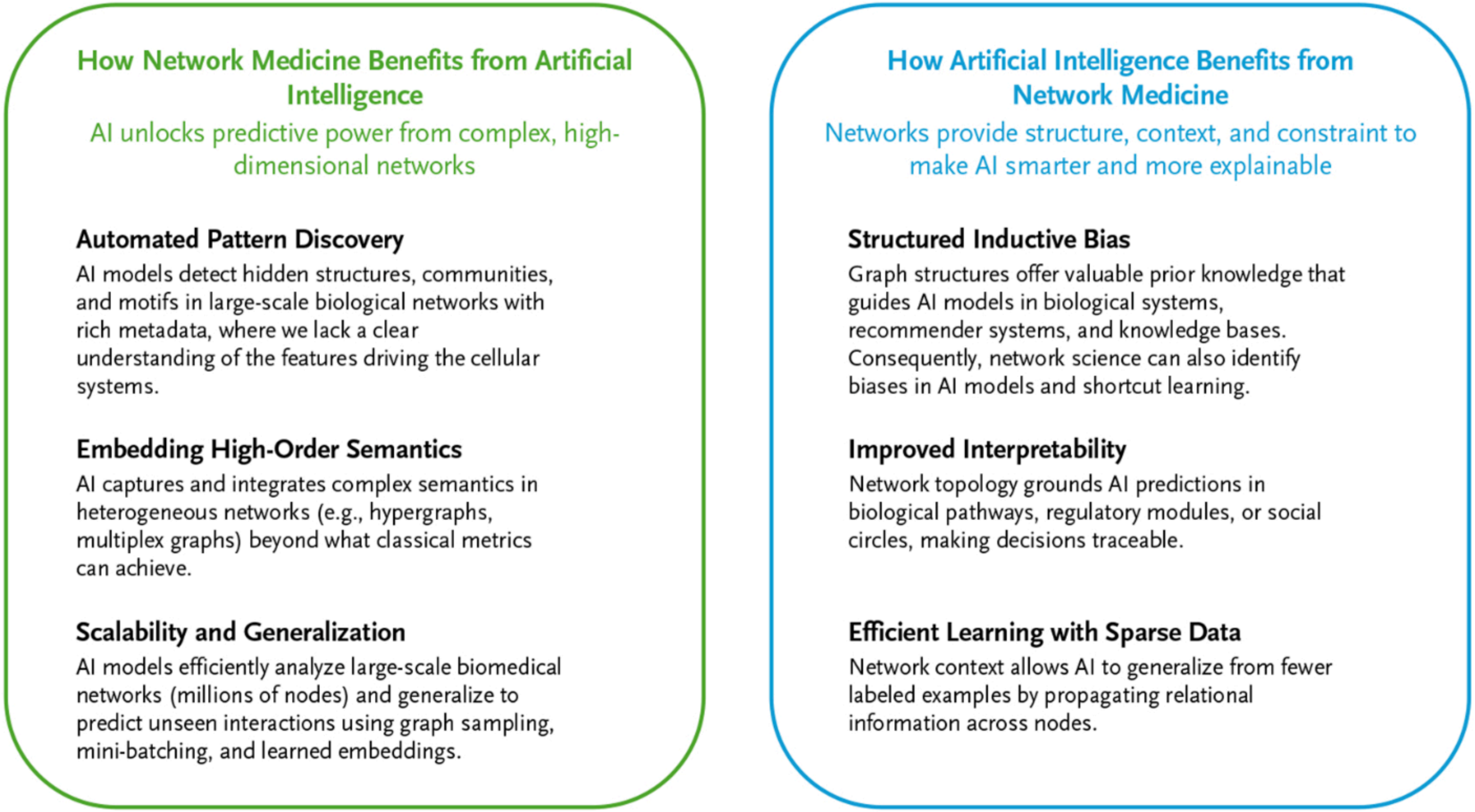


What Types of Biological Questions Have Network Medicine and Artificial Intelligence Successfully Addressed?

- Disease Gene Prediction
- Disease Module Detection
- Reticulotype Mapping and Classification
- Endophenotype Module Detection
- Drug Discovery
- Drug Repurposing
- Drug–Target Interaction Prediction
- Drug Side Effect Prediction
- Protein–Protein Interaction

[Altucci et al, Artificial Intelligence and Network Medicine: Path to Precision Medicine, NEMJ AI, 2025]

AI for Network Medicine & Network Medicine for AI



How Network Medicine Benefits from Artificial Intelligence

AI unlocks predictive power from complex, high-dimensional networks

Automated Pattern Discovery

AI models detect hidden structures, communities, and motifs in large-scale biological networks with rich metadata, where we lack a clear understanding of the features driving the cellular systems.

Embedding High-Order Semantics

AI captures and integrates complex semantics in heterogeneous networks (e.g., hypergraphs, multiplex graphs) beyond what classical metrics can achieve.

Scalability and Generalization

AI models efficiently analyze large-scale biomedical networks (millions of nodes) and generalize to predict unseen interactions using graph sampling, mini-batching, and learned embeddings.

How Artificial Intelligence Benefits from Network Medicine

Networks provide structure, context, and constraint to make AI smarter and more explainable

Structured Inductive Bias

Graph structures offer valuable prior knowledge that guides AI models in biological systems, recommender systems, and knowledge bases. Consequently, network science can also identify biases in AI models and shortcut learning.

Improved Interpretability

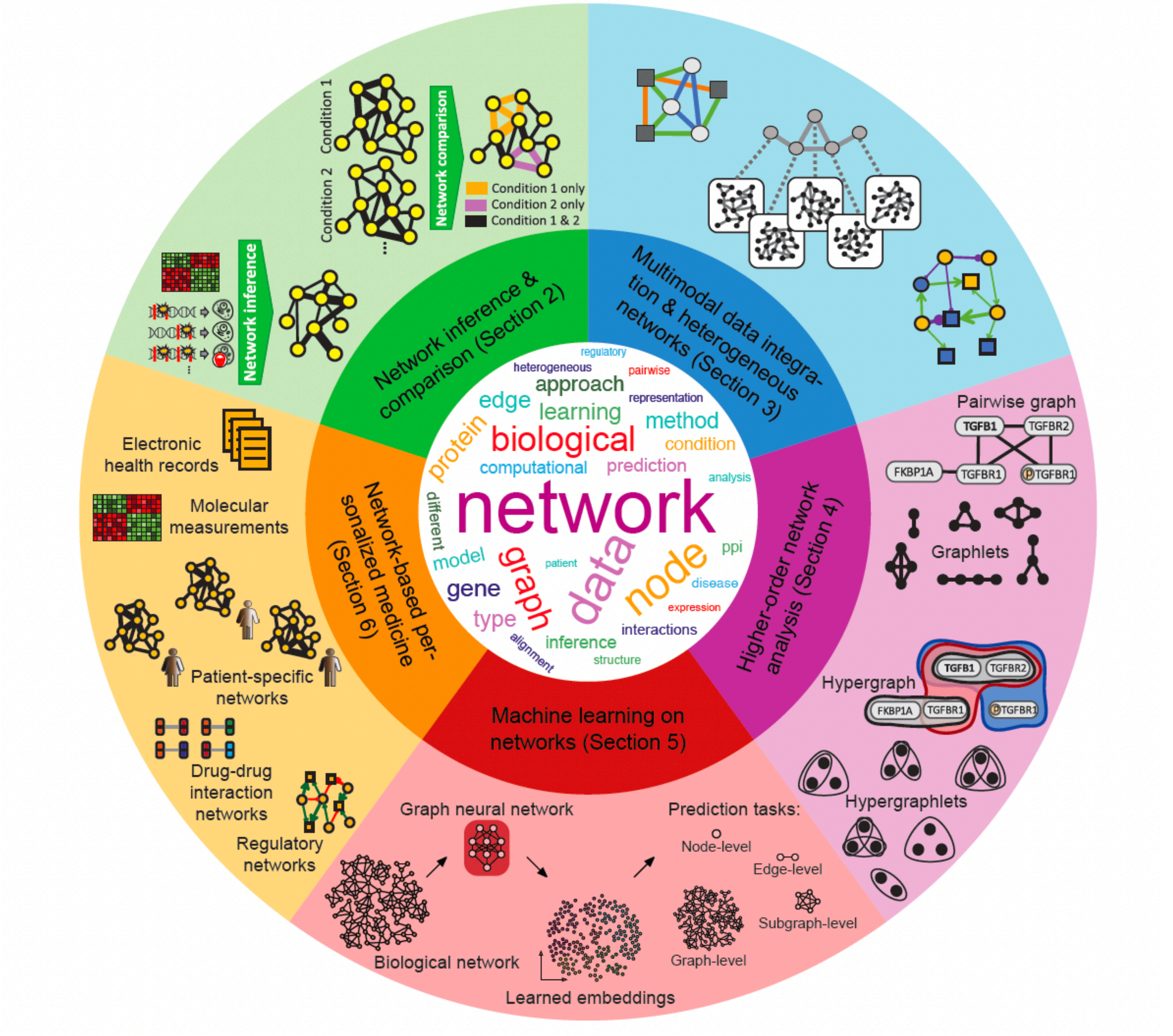
Network topology grounds AI predictions in biological pathways, regulatory modules, or social circles, making decisions traceable.

Efficient Learning with Sparse Data

Network context allows AI to generalize from fewer labeled examples by propagating relational information across nodes.

[Altucci et al, Artificial Intelligence and Network Medicine: Path to Precision Medicine, NEMJ AI, 2025]

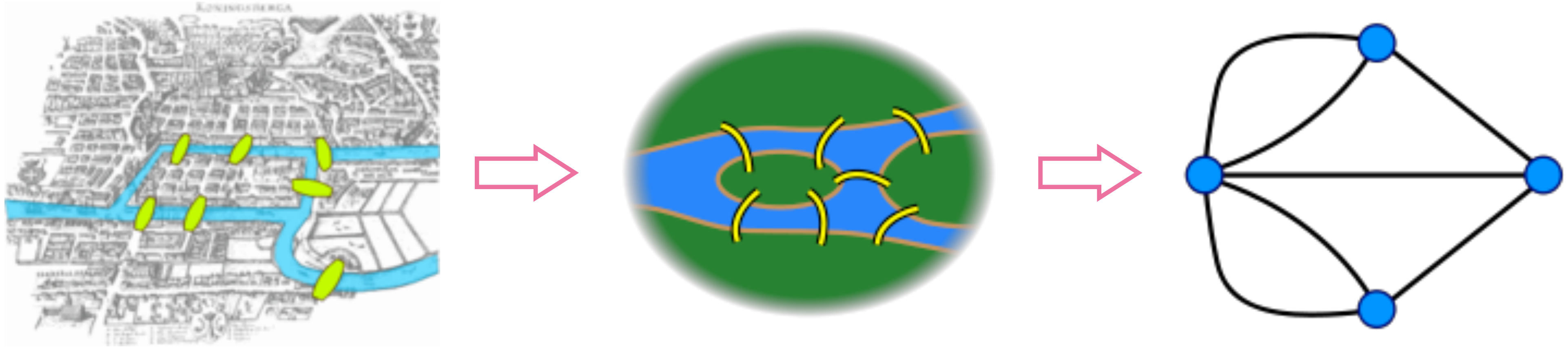
Some of the current challenges



[Fig. from M. Zitnik et al., Current and future directions in network biology, 2023]

Networks as graphs

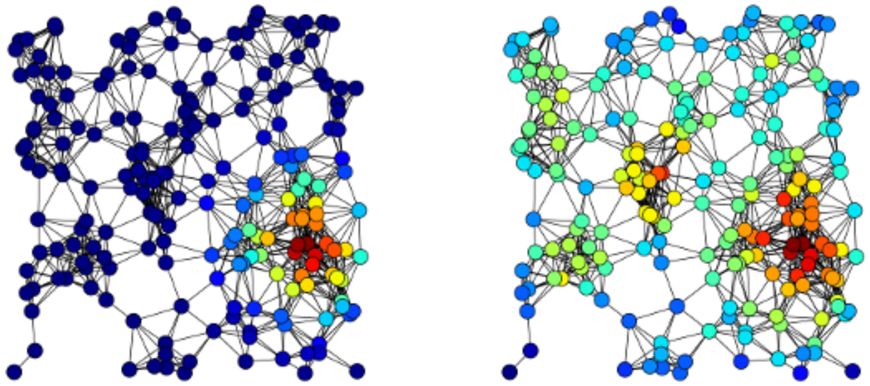
- Graphs provide a mathematical representation for describing and modeling complex structure, geometry and knowledge of complex systems



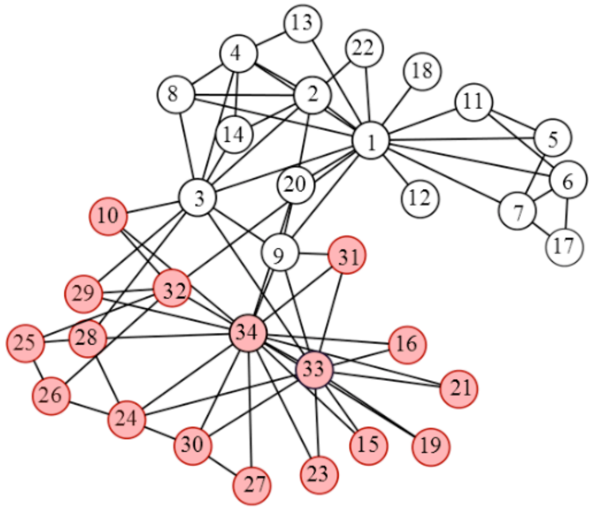
The Königsberg Bridge Problem
[Leonhard Euler, 1736]

- “Graphs are the most important discrete models in the world!” - G. Strang (MIT)

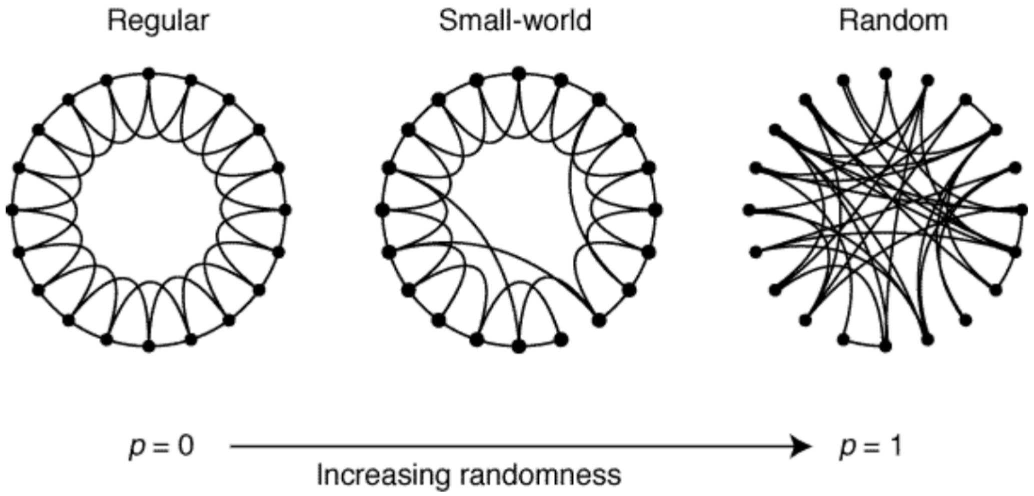
Graph topology analysis



Node centrality



Community detection



Random graph models

Graph structured data

- From edges to node attributes

Network/graph structured data

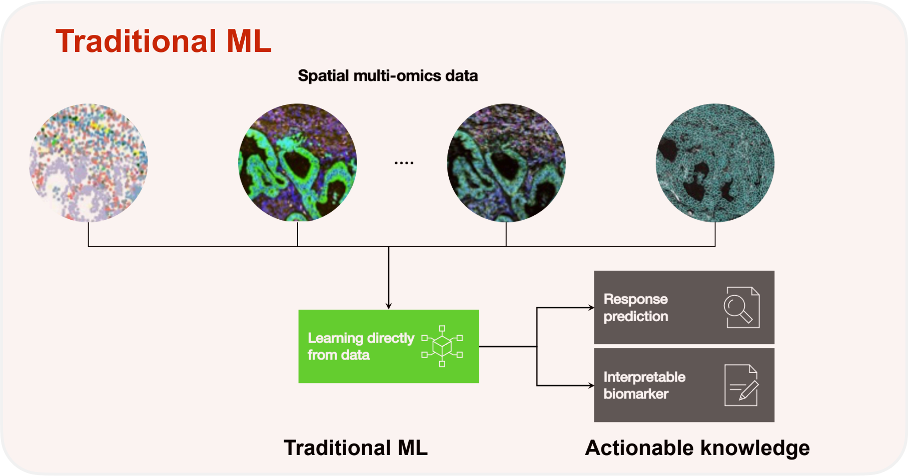
- Need to take into account both structure (i.e., edges), and data (i.e., information on the nodes of the network)

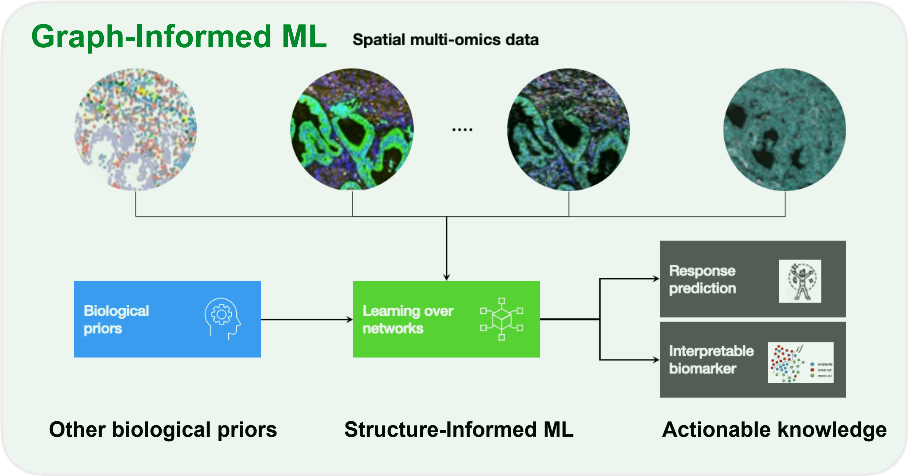
Why learning from graphs is hard?

- Irregular domain: complex interactions between nodes of graph
- The size of graphs varies and the number of neighbors changes
- No specific node ordering, leading to different symmetries
  - Permutation equivariance/invariance
- Topologies can change over time: node and edge can appear and disappear

Representation of graph-structured data

- Traditional signal processing and machine learning approaches: Harmonic analysis on Euclidean domain (e.g., Fourier, wavelets), (deep) representation learning

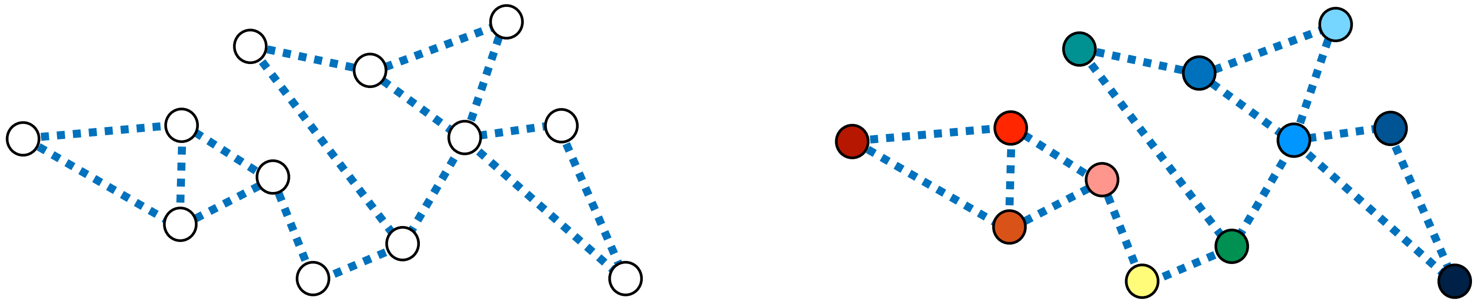




How can we build principled frameworks for graph-structured data?

A signal processing approach

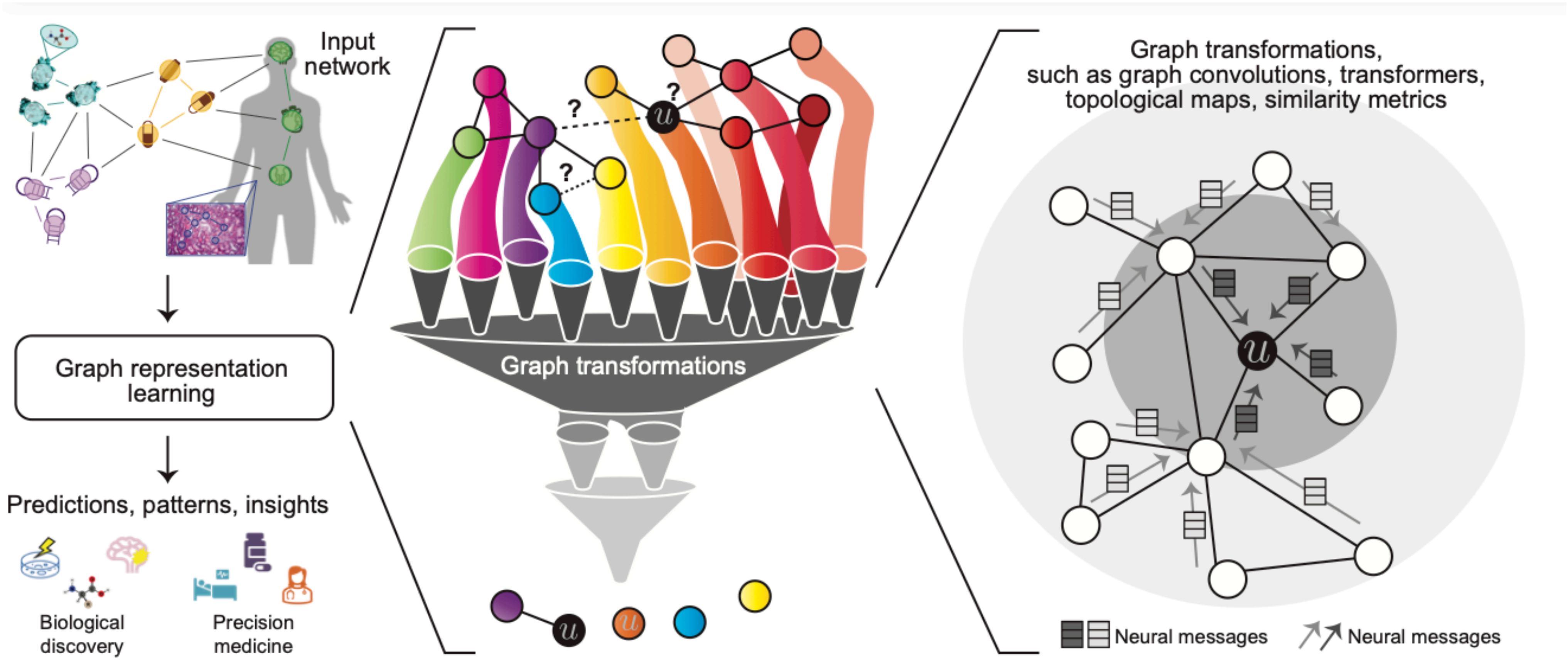
- Exploit the interplay between
  - **graph:** flexible tools to represent a discrete and irregular domain
  - **signal on graph:** function value (attribute) at each vertex



- **Challenges:**
  - Generalize classical signal processing notions such as sampling, convolution, frequency analysis, Fourier, wavelet transforms
  - Use these notions for inference and learning from structured data

Graph machine learning

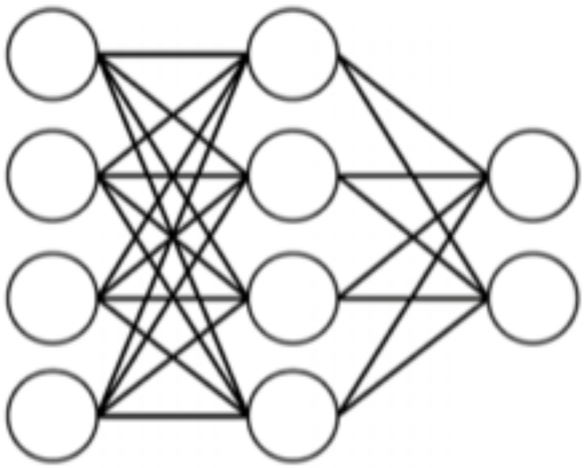
- An emerging research topic, with many practical applications



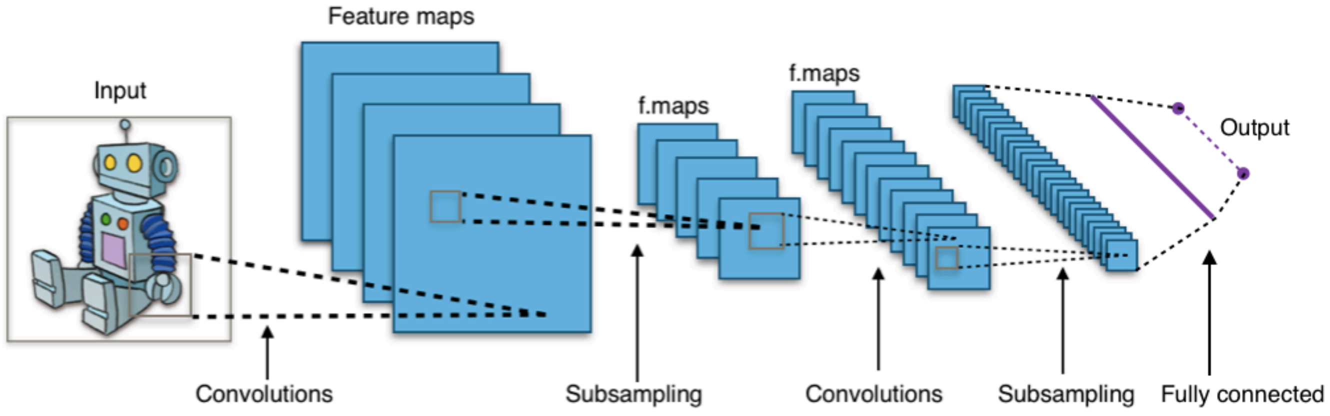
Generate actionable knowledge by learning directly from network data

[M. Li, K. Hunag, and M. Zitinik.,Graph Representation Learning in Biomedicine and Healthcare, Nature Biomedical Engineering, 2022]

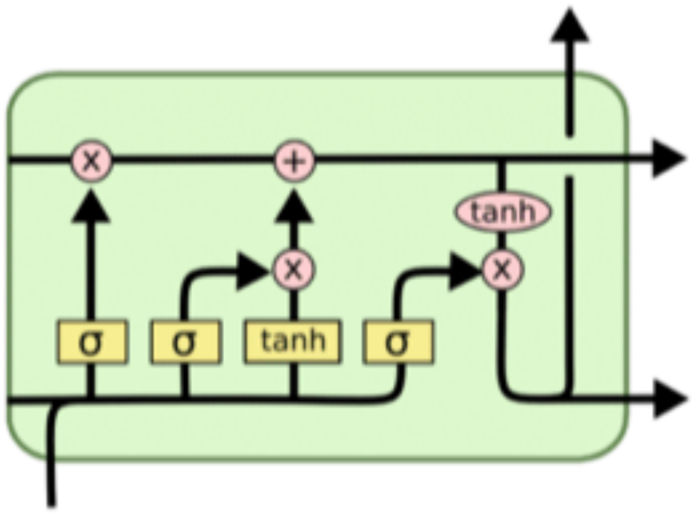
From classical to graph machine learning



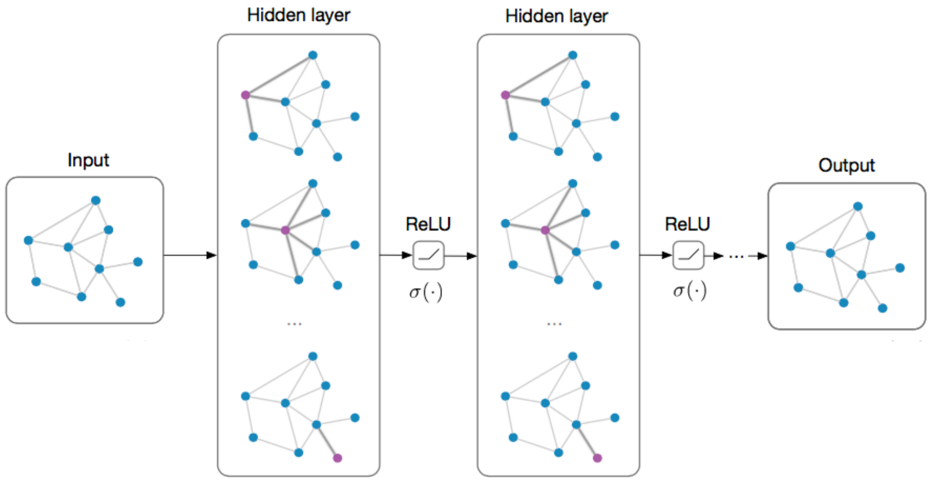
Perceptrons:
Function regularity



CNNs:
Translation invariance

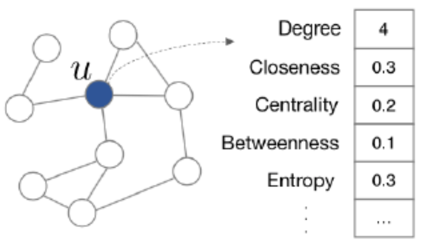


RNNs:
Time warping

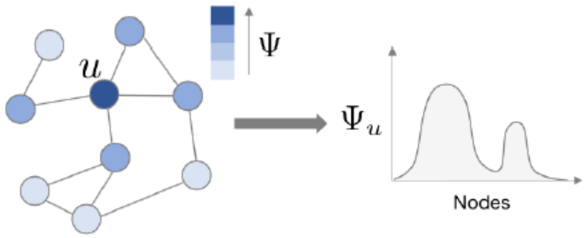


GNNs:
Permutation invariance

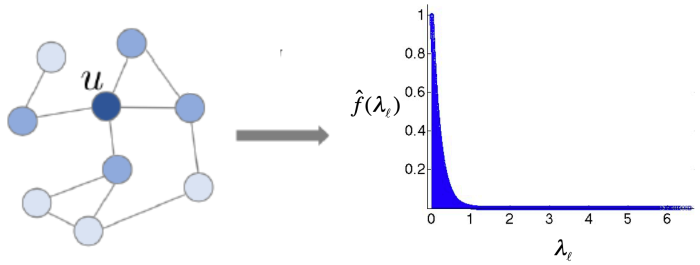
Predominant graph representation learning paradigms



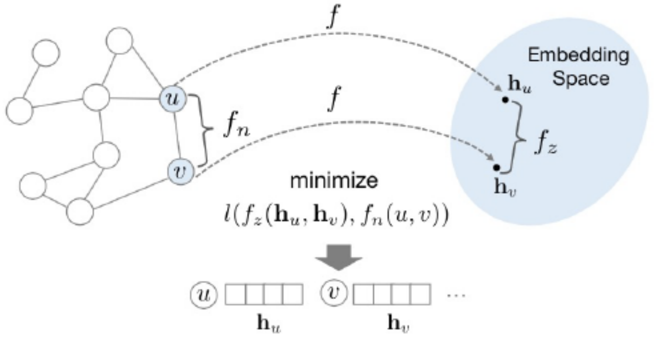
Hand-crafted graph theoretic features



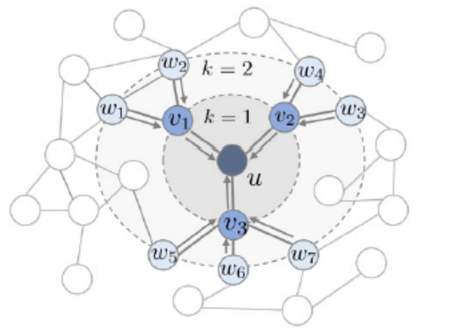
Kernel-based features



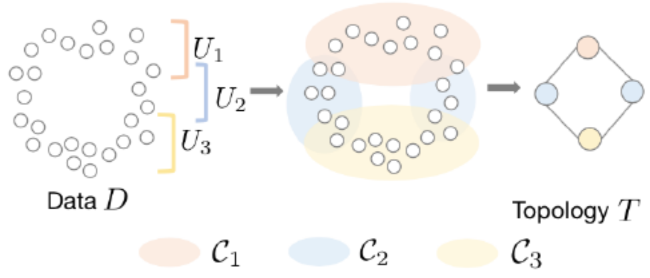
Graph signal processing based features



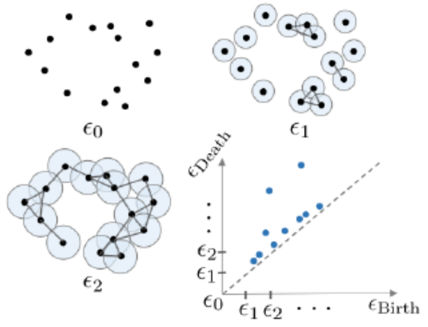
Shallow embeddings



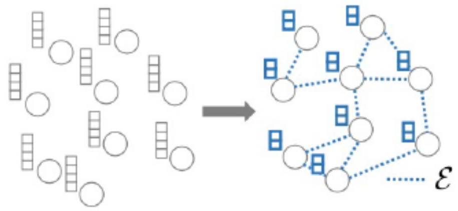
Deep embeddings: Graph neural networks



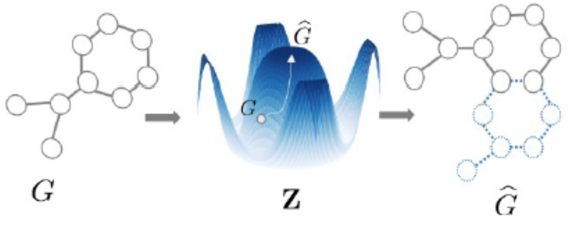
Topological features



Persistent homology



Manifold learning & Topology inference



Graph generative models

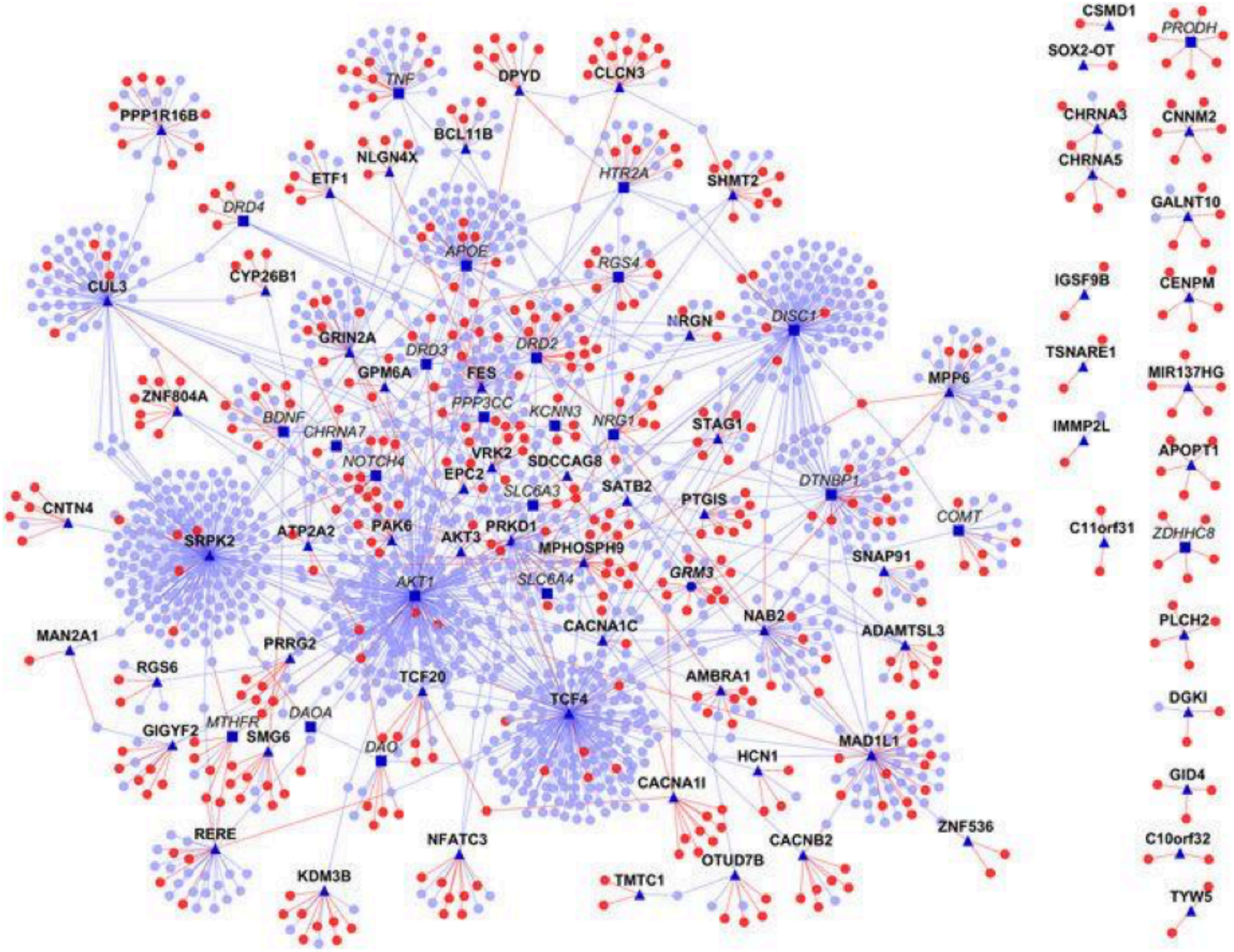
[Fig modified from M. Li, K. Hunag, and M. Zitinik.,Graph Representation Learning in Biomedicine and Healthcare, Nature Biomedical Engineering, 2022]

Common tasks in graph structured data

- Predict a type of a given node: node classification/clustering
- Predict whether two nodes are linked: link prediction
- Identify densely linked clusters of nodes: clustering/community detection
- How similar are two nodes/networks: graph classification
- Design graphs with desirable properties: graph generation

Node classification example

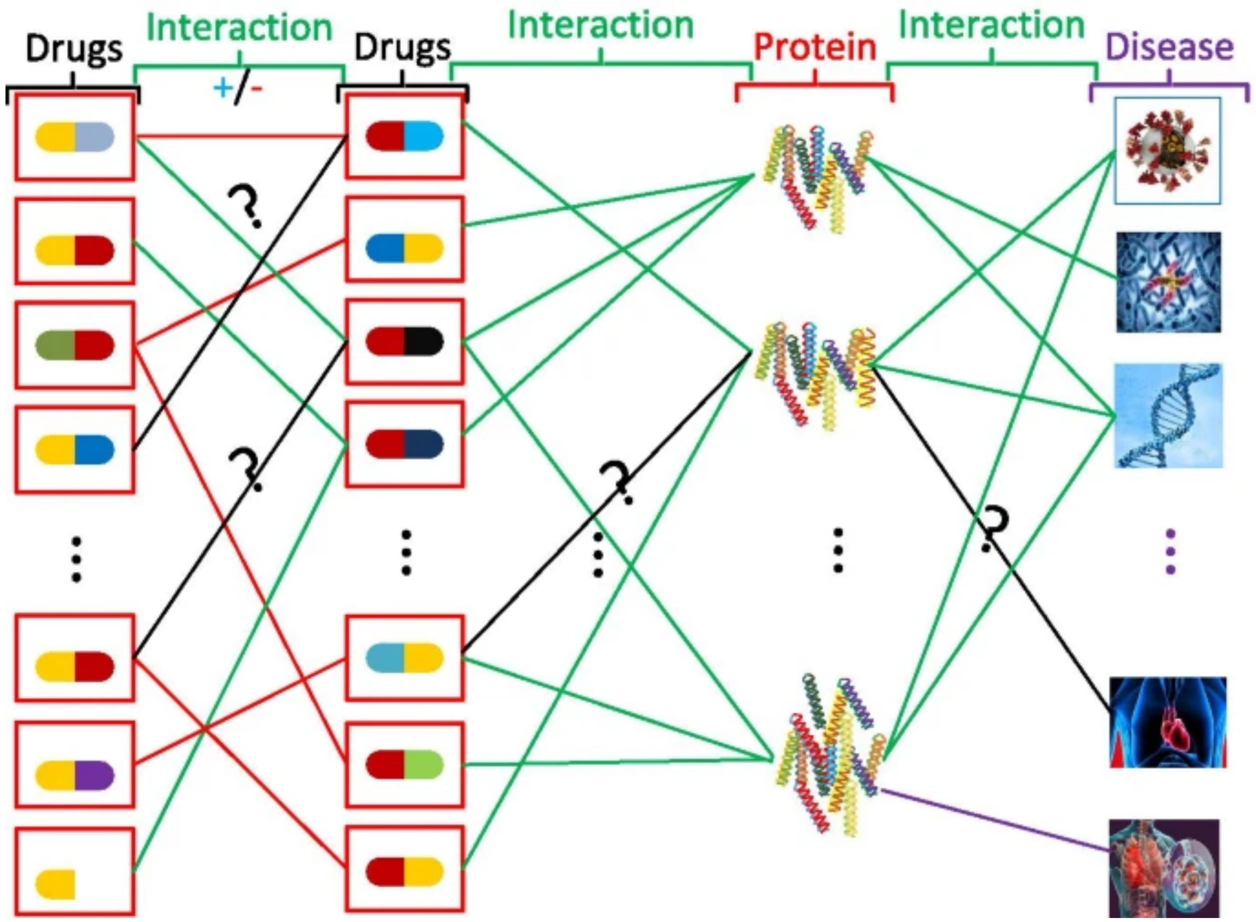
- Classifying the function of proteins in the interactome



[Ganapathiraju et al. 2016. Schizophrenia interactome with 504 novel protein–protein interactions. Nature]

Link prediction example

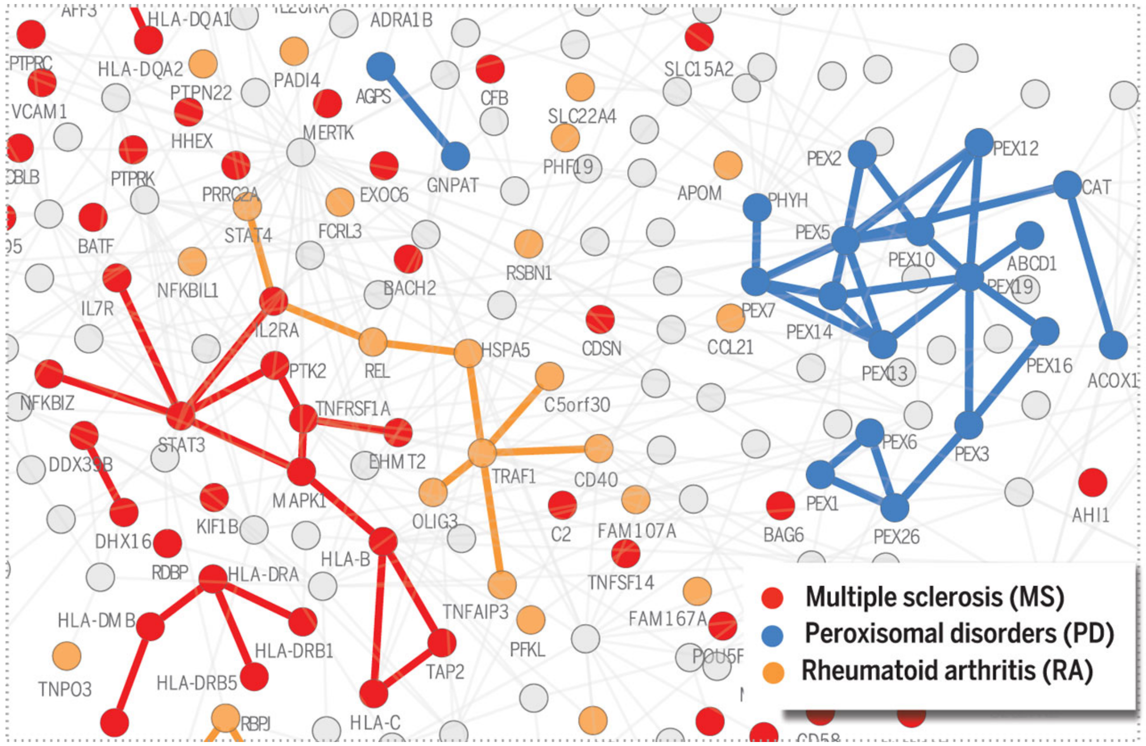
- Predicting drug-target and drug-drug interaction links



[Abbas et al., 2021. Application of network link prediction in drug discovery, BMC Bioinformatics]

Cluster identification example

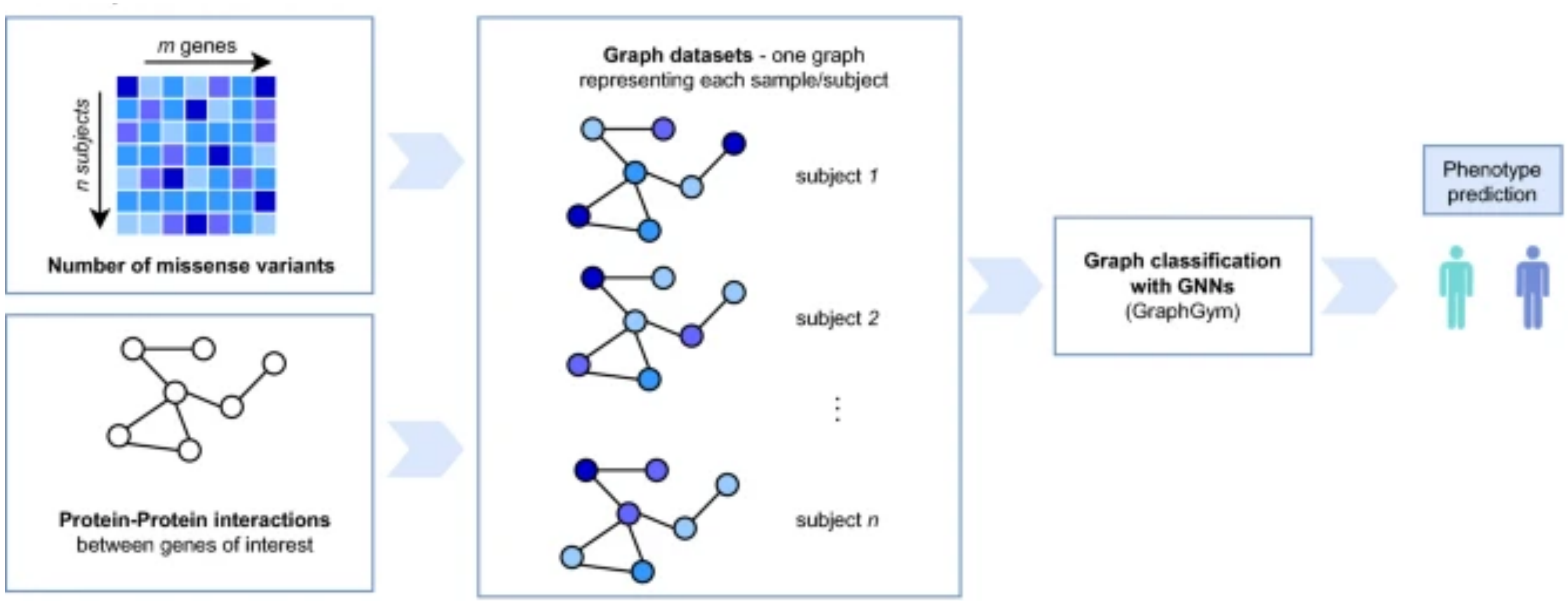
- Identifying proteins associated with the same disease from connected subgraphs



[Menche et al., 2015. Uncovering disease-disease relationships through the incomplete interactome, Science]

Graph classification example

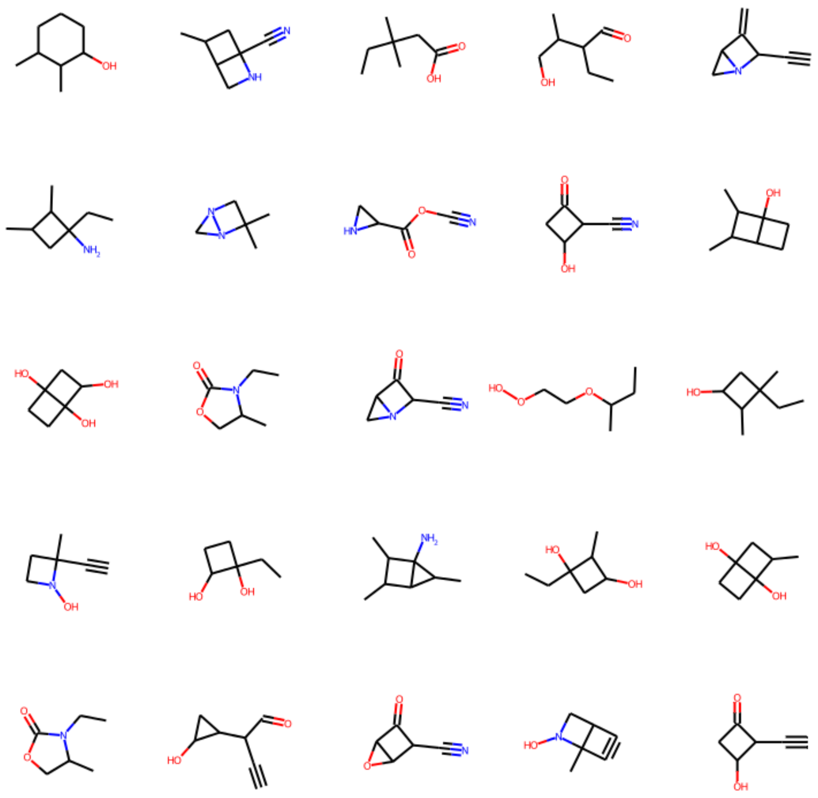
- Predicting patients' phenotype for easy diagnosis of Alzheimer's disease



[Hernandez-Lorenzo et al., 2022. On the limits of graph neural networks for the early diagnosis of Alzheimer's disease, Nature Scien. Rep.]

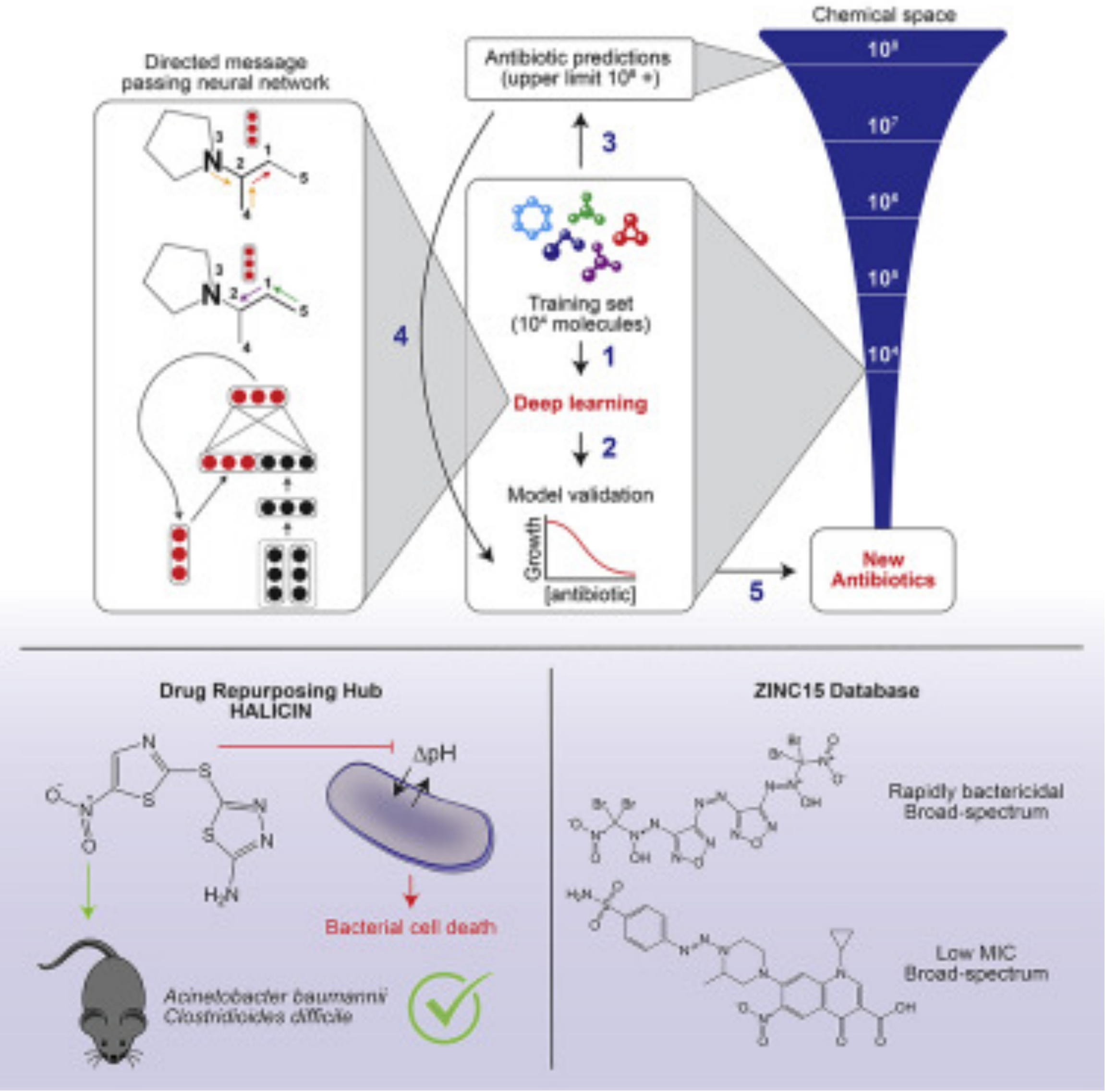
Graph generation example

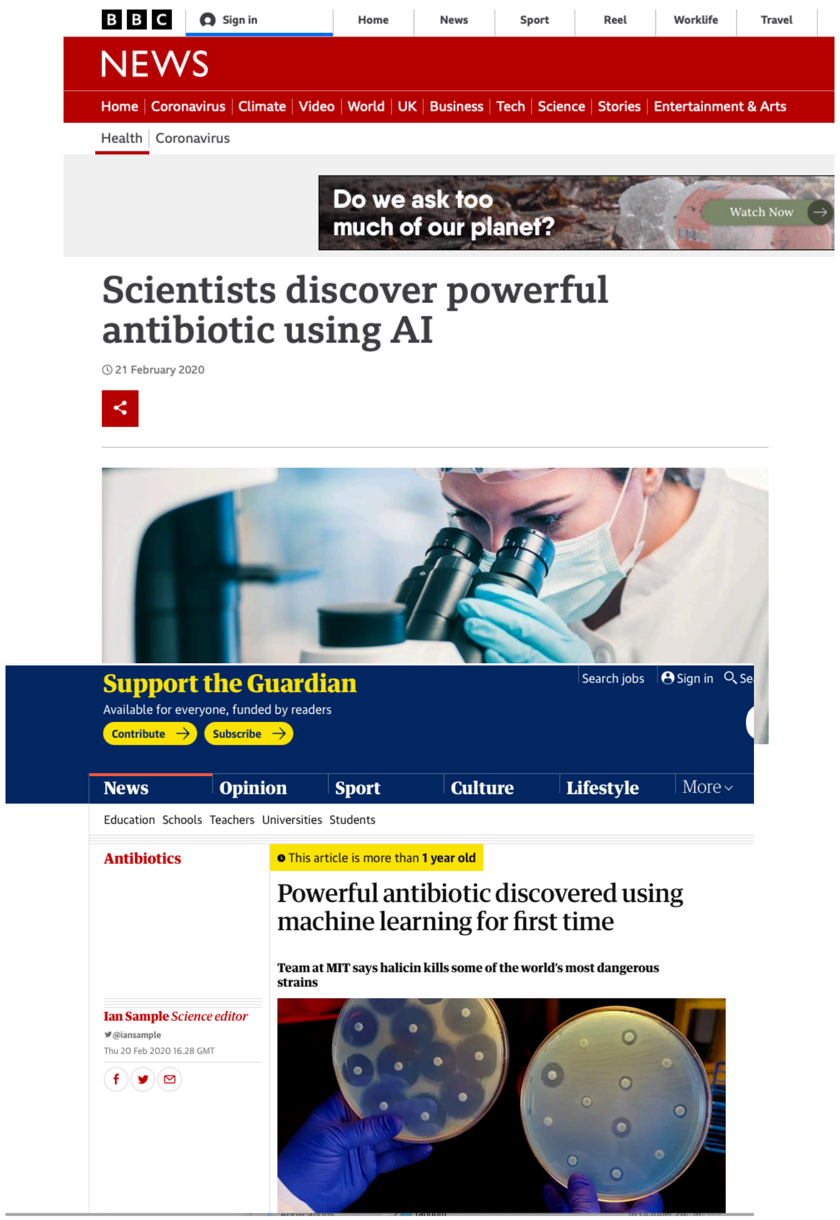
- Generating now molecules



[De Sao et al., 2022. MolGAN: An implicit generative model for small molecular graphs , ICML workshop on Theoretical Foundations and Applications of Deep Generative Models]

Recent success story: Antibiotic discovery





Scientists discover powerful antibiotic using AI

21 February 2020

This article is more than **1 year old**

Powerful antibiotic discovered using machine learning for first time

Team at MIT says halicin kills some of the world's most dangerous strains

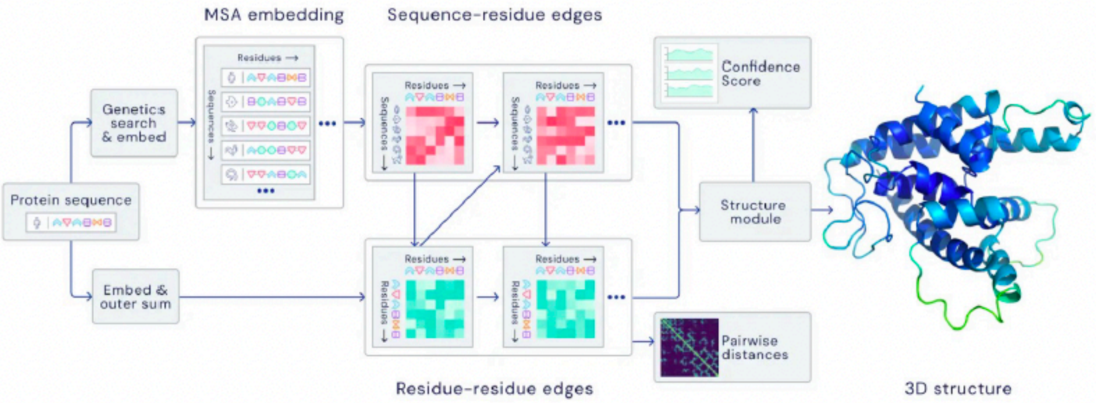
Ian Sample *Science editor*
@iansample
Thu 20 Feb 2020 16.28 GMT

[Simonovsky et al, 2017, De Cao et al 2018, Stokes et al 2020]

Recent success story: Protein folding

AlphaFold: a solution to a 50-year-old grand challenge in biology

Proteins are essential to life, supporting practically all its functions. They are large complex molecules, made up of chains of amino acids, and what a protein does largely depends on its unique 3D structure. Figuring out what shapes proteins fold into is known as the "protein folding problem", and has stood as a grand challenge in biology for the past 50 years. In a major scientific advance, the latest version of our AI system AlphaFold has been recognised as a solution to this grand challenge by the organisers of the biennial Critical Assessment of protein Structure Prediction (CASP). This breakthrough demonstrates the impact AI can have on scientific discovery and its potential to dramatically accelerate progress in some of the most fundamental fields that explain and shape our world.



Graph Transformers

[Jumper et al. 2021]

EE626: Graph representations for biology and medicine - Introduction

Every Wednesday 13:15-15:00
INF 019

Course description

- Discuss advanced topics in machine learning and signal processing on graphs, and showcase applications in biomedicine
- Every week will be an advanced seminar focused on a specific methodological aspect and its usefulness on various application domains
- The material will be based on research papers that will be presented and discussed in the class
- **Prerequisites:** good knowledge of machine learning, graph theory, graph ML; strong interest in AI for biology and medicine

Goals of the course and learning outcomes

- Explore recent developments in graph machine learning for biology and medicine
- Brainstorm on future developments of these tools in further medical or biological applications
- Provide insights for further research
- Analyze and summarize scientific, and interdisciplinary work
- Synthesize arguments into scientific presentations

Organization

- Join effort from all of us!
- Students are expected to present research topics and lead the discussion
- Research papers will be proposed, further suggestions are welcome!
- Use this class as an opportunity to interact, exchange, collaborate with other students on topics related to networks, biology and medicine

Communication

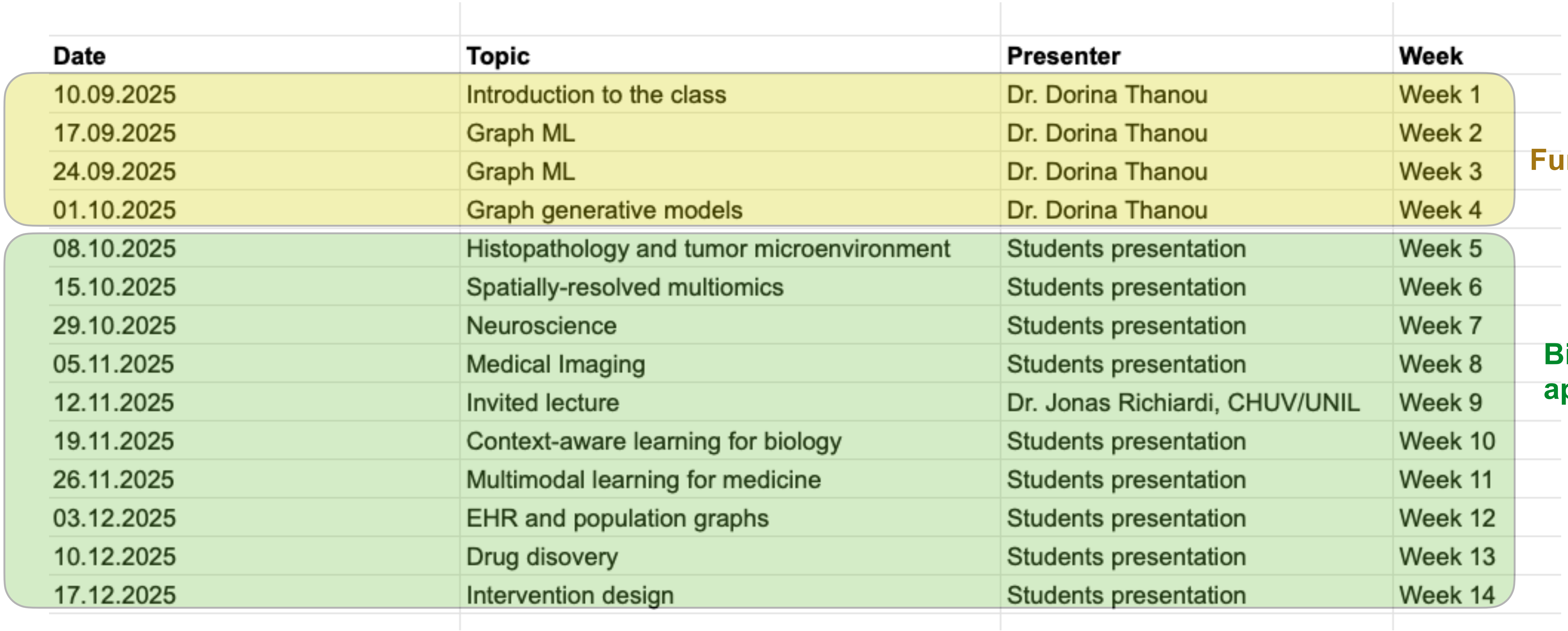
- All communication and material will be distributed via moodle
  - https://moodle.epfl.ch/course/view.php?id=18383

- Contact me anytime for further questions or discussions!

Content of the class

- This course has 14 lectures
  - Course overview and basic introduction into graphs and graph ML (4 lectures)
  - 9 lectures of paper presentations (~2-3 papers/week)
  - 1 invited lecture

Tentative agenda



Date	Topic	Presenter	Week
10.09.2025	Introduction to the class	Dr. Dorina Thanou	Week 1
17.09.2025	Graph ML	Dr. Dorina Thanou	Week 2
24.09.2025	Graph ML	Dr. Dorina Thanou	Week 3
01.10.2025	Graph generative models	Dr. Dorina Thanou	Week 4
08.10.2025	Histopathology and tumor microenvironment	Students presentation	Week 5
15.10.2025	Spatially-resolved multiomics	Students presentation	Week 6
29.10.2025	Neuroscience	Students presentation	Week 7
05.11.2025	Medical Imaging	Students presentation	Week 8
12.11.2025	Invited lecture	Dr. Jonas Richiardi, CHUV/UNIL	Week 9
19.11.2025	Context-aware learning for biology	Students presentation	Week 10
26.11.2025	Multimodal learning for medicine	Students presentation	Week 11
03.12.2025	EHR and population graphs	Students presentation	Week 12
10.12.2025	Drug disovery	Students presentation	Week 13
17.12.2025	Intervention design	Students presentation	Week 14

Fundamentals

Biomedical applications

Application domains

- Neuroscience
- Multi-omics
- Medical imaging
- Digital pathology
- Drug discovery
- EHR
- …

Material

- Mainly based on research articles

- Part A: Methodological components; Quick introduction into graph machine learning

- Part B: Biomedical applications
  - Each week, we will cover 2-3 research papers

- The final list of papers will be distributed by September 24; Please send me suggestions by then!

Class participation

- **Presenters (2 students ~45’):**
  - Introduction of themselves and their own research (~5’ each)
  - Summarize the papers’ main ideas, contributions and methods
  - Highlight why the paper is important
  - Prepare 2-3 open questions for discussions

- **Challengers (2 students ~15’):**
  - Critically analyse the paper; strengths, weaknesses, assumptions, open issues
  - Prepare 3-4 discussion points to challenge the presenters
  - Focus on constructive criticism

- **Audience (rest of the class/everyone ~20’):**
  - Engage in discussion
  - Connect the papers to your own research where possible

Weekly expectations

- Read the papers carefully and critically before the class
- Support arguments with evidence
- Brainstorm on possible connections with your own research
- Aim to learn from each role

- Guidelines for presentations will be communicated on Moodle!

- Each week final slides should be sent to me by Tuesday evening

Deadlines

- Action points before September 24:
  - Please send me your (un)availabilities for presenting
  - Preference to a specific topic can be expressed
  - Please choose your teammate

- An announcement on Moodle will follow

Questions and feedback?