EE626: Graph representations for biology and medicine - Introduction

Dr Dorina Thanou

11.09.2024

Technological advancements are transforming biomedicine

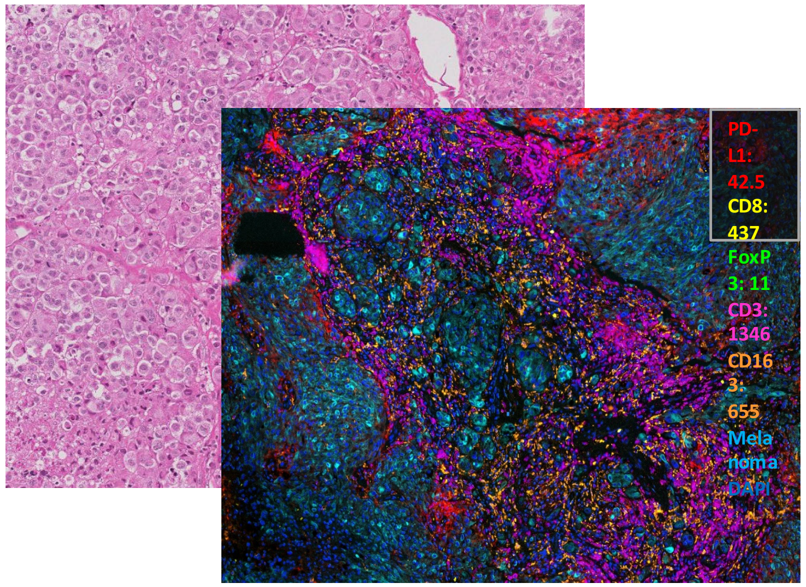
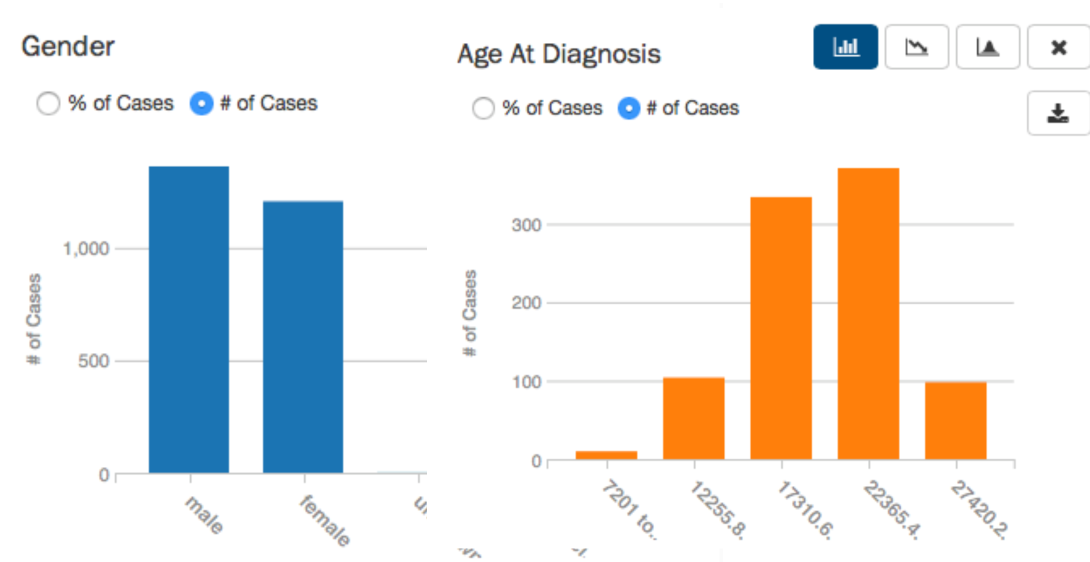
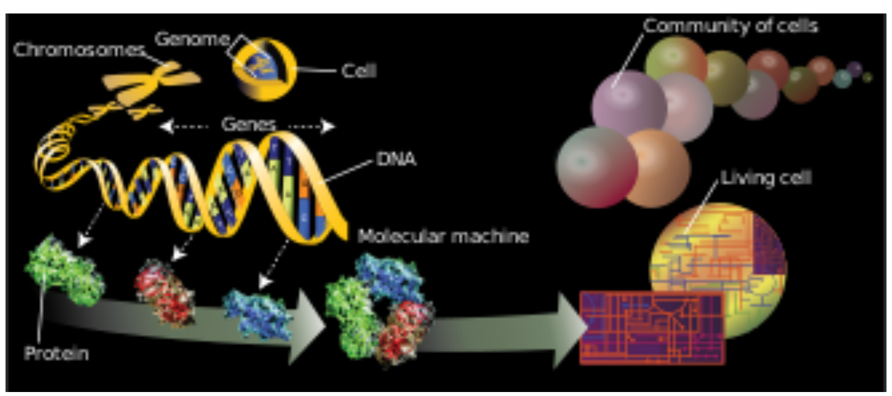


Image modalities



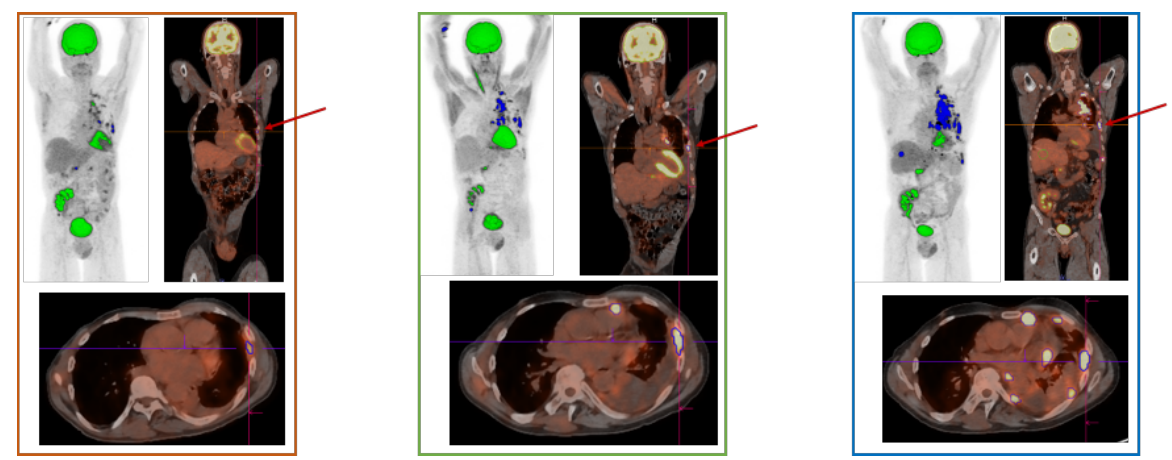
https://portal.gdc.cancer.gov

Clinical data



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

OMICS

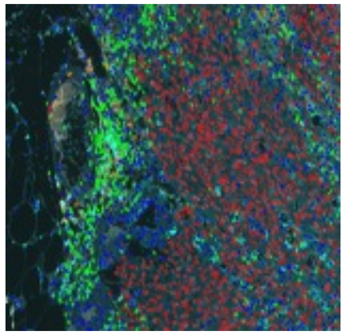


Time varying data

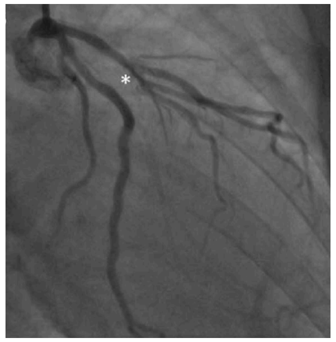
Unprecedented opportunities for novel biomedical discoveries!

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?

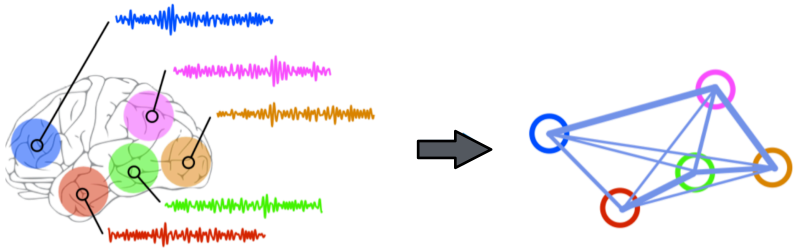


Respondent or non respondent?



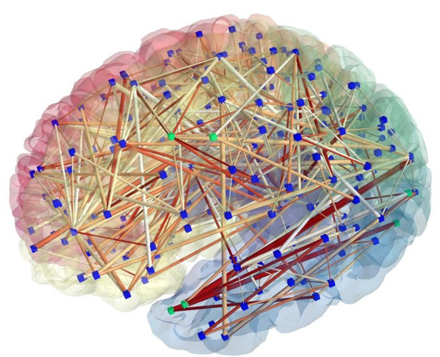
Healthy or not?

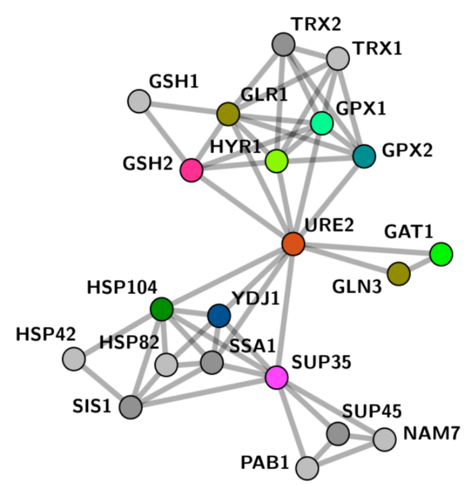
- Knowledge discovery (usually unsupervised):
  - What are predictive biomarkers of a disease?
  - What is the underlying biological mechanism?
  - Can we discover new therapeutics?

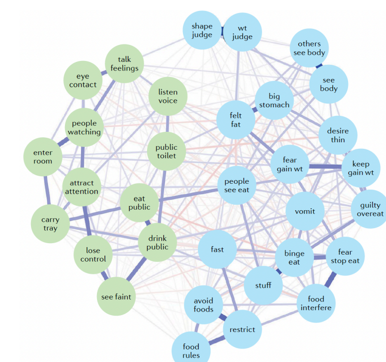


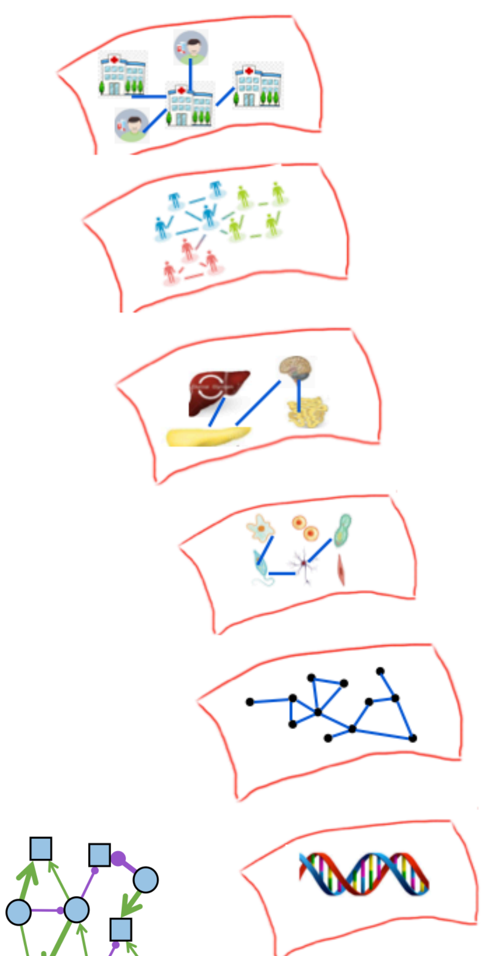
Networks for biology and medicine

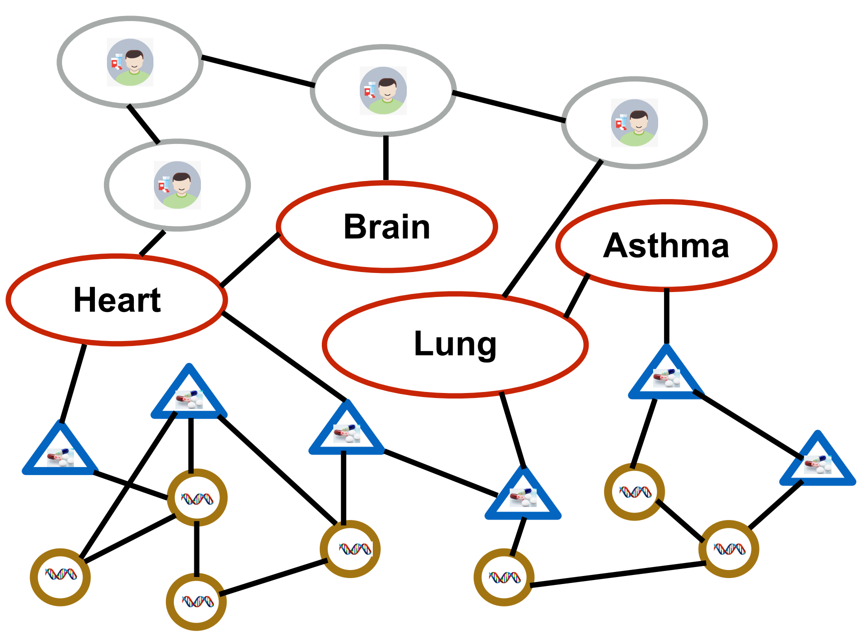
- Many biomedical data are represented by some networks
  - Spatial information, functional interactions, anatomical structures

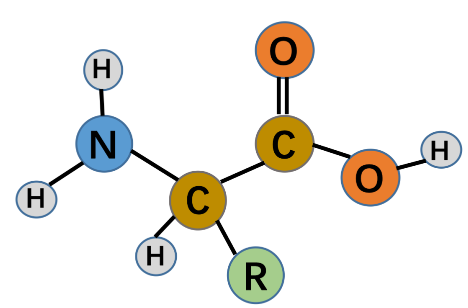


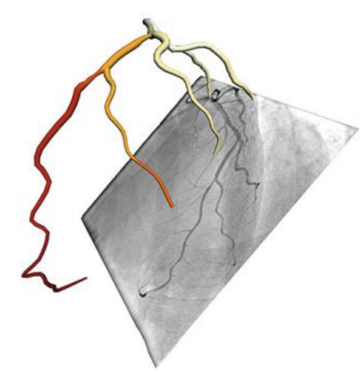


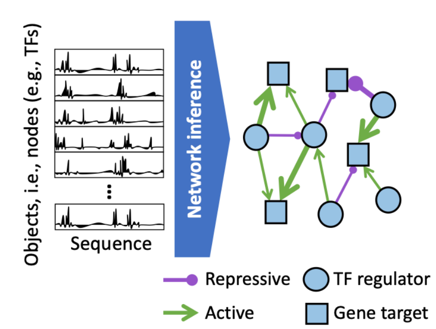






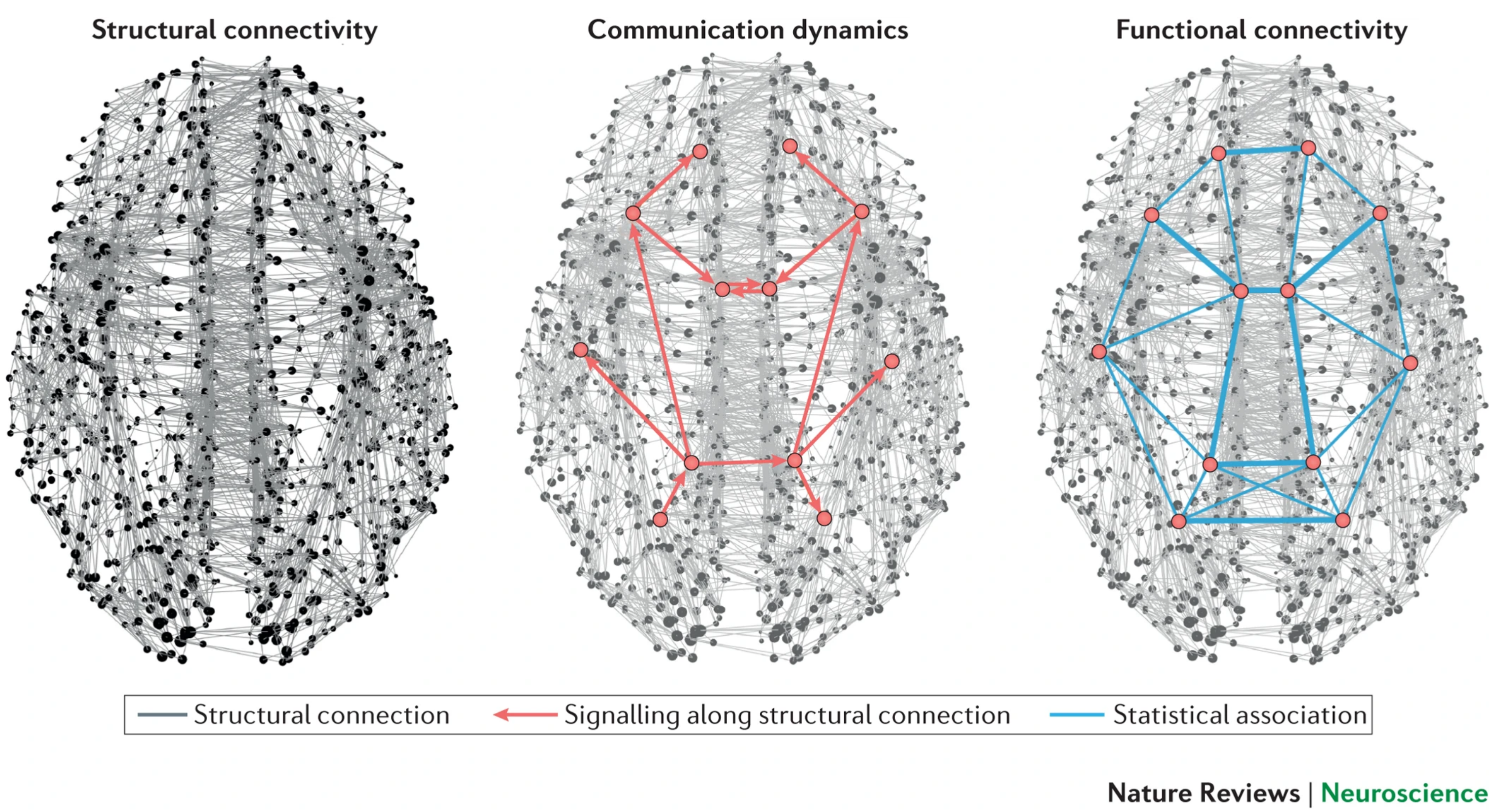






Capturing structure-function relationships

- 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

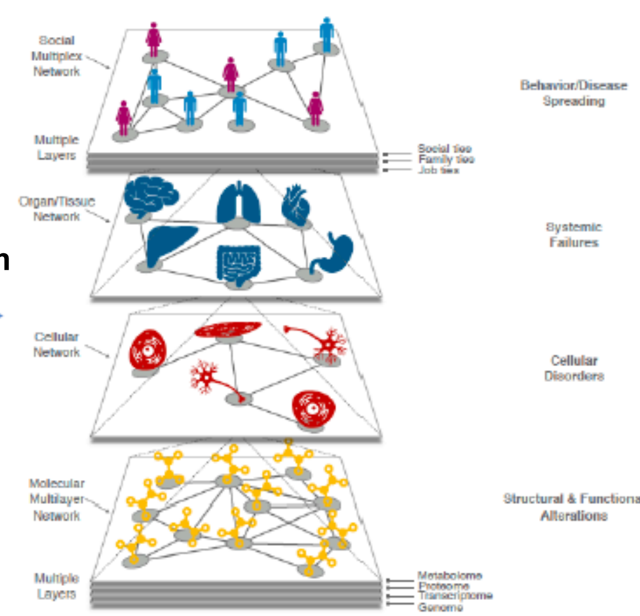


[Avena-Koenigsberger et al., Communication dynamics in complex brain networks, Nature Reviews Neuroscience, 2017]

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]

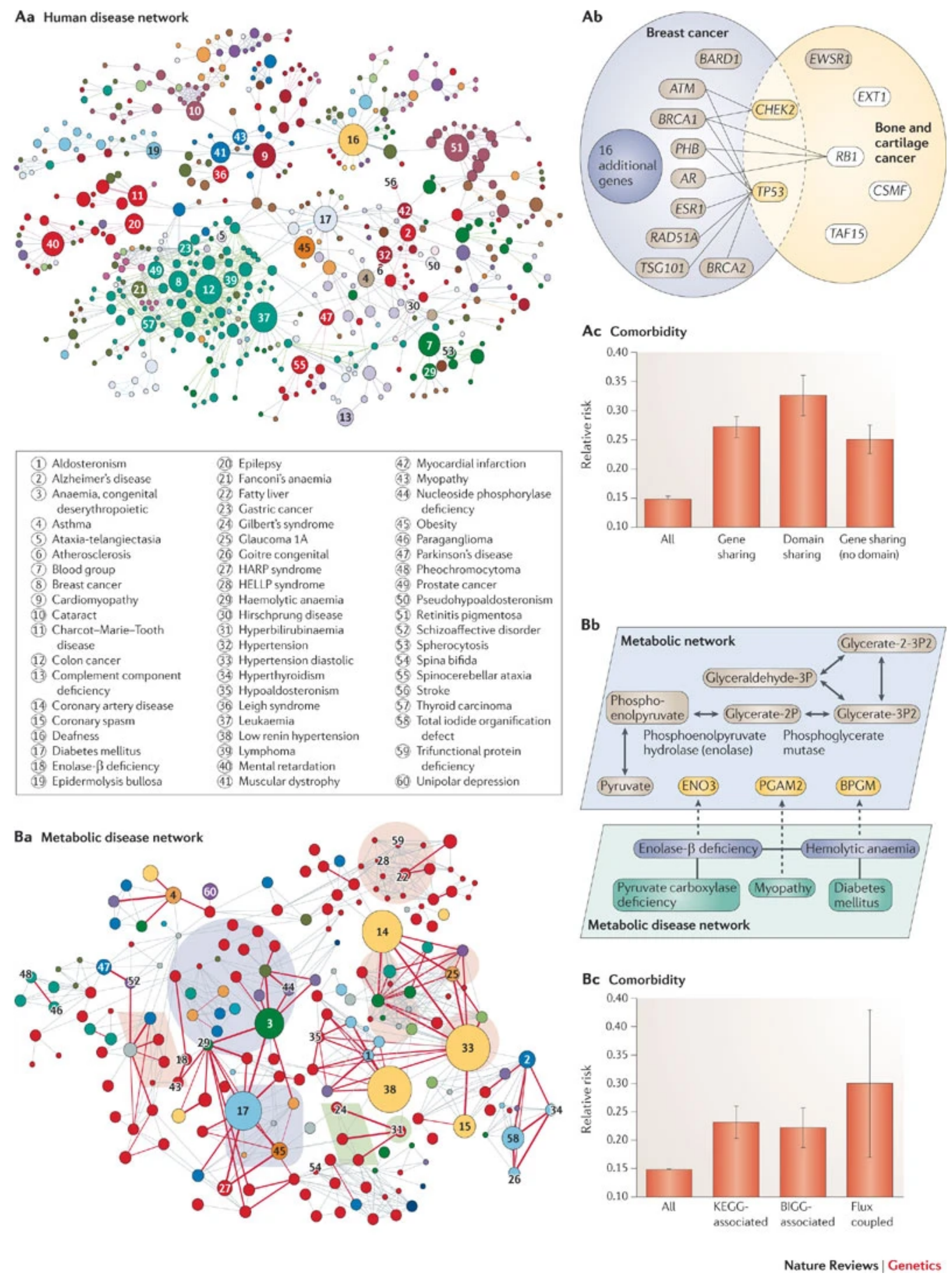
Network medicine: a network-based approach to human disease

- 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]

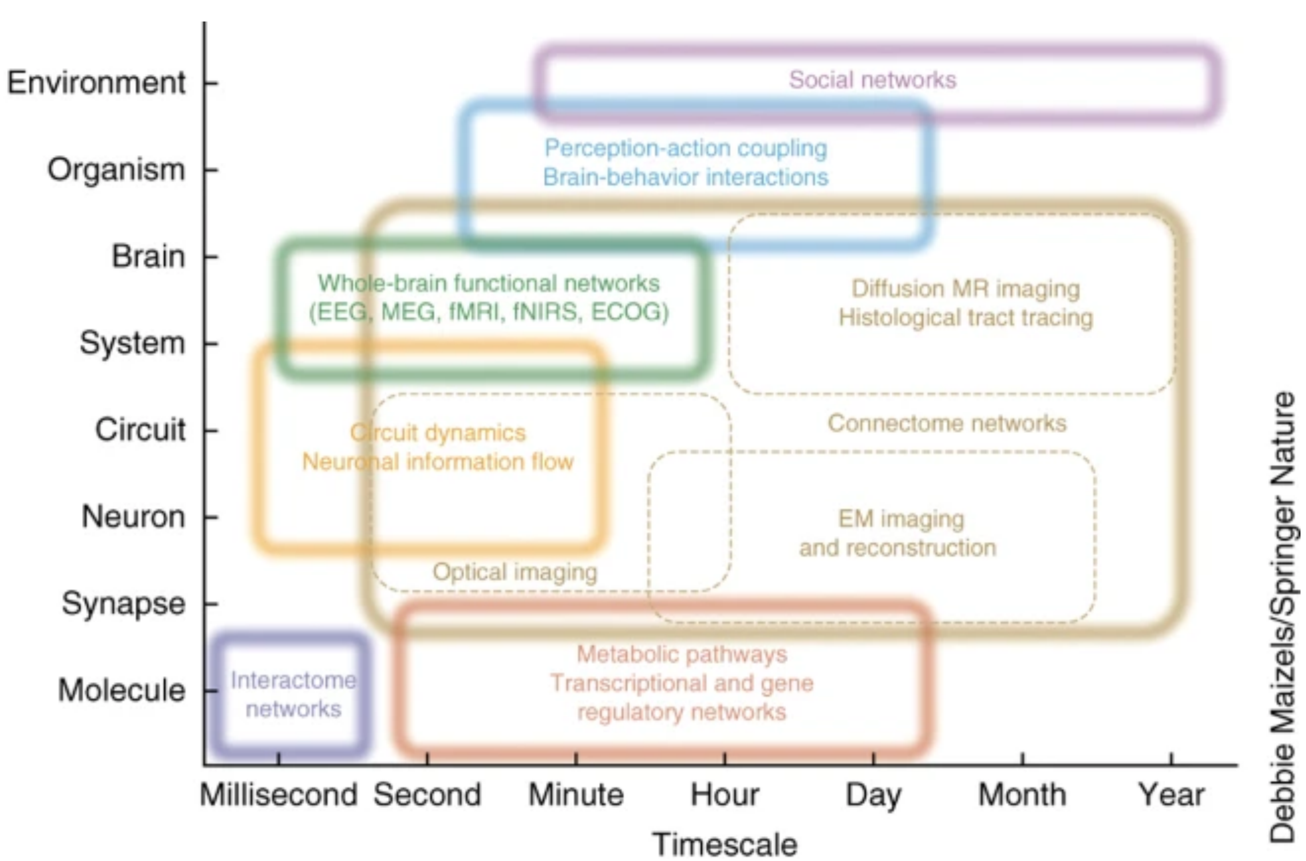
The human disease network

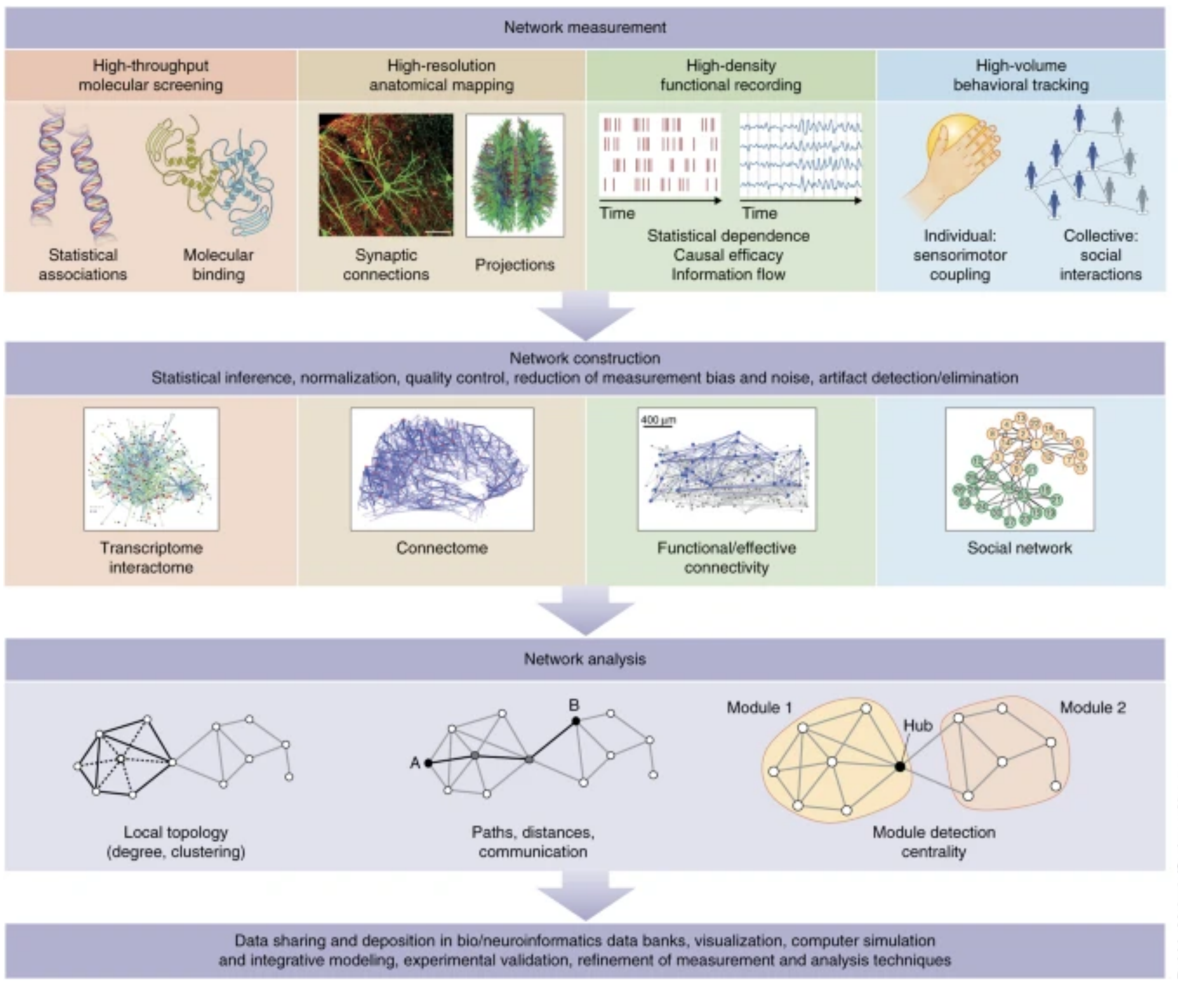


[A. Barabasi et al, Network Medicine: A network-based approach to human disease, Nature Reviews Genetics, 2011]

Network neuroscience

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





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

Network biomedicine: a growing field

2023

Current and future directions in network biology

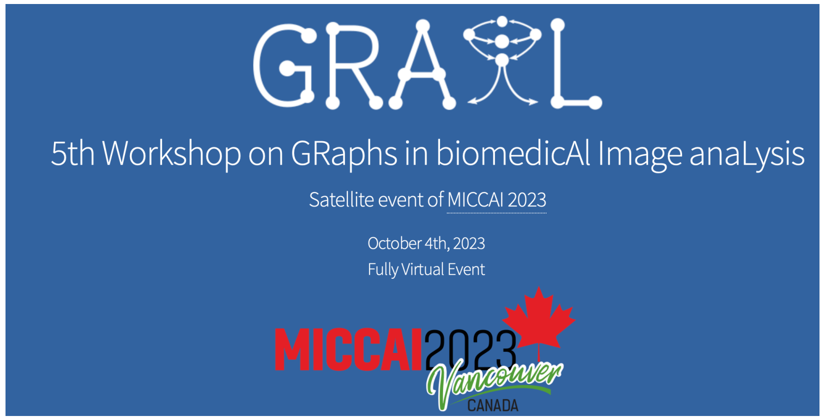
Marinka Zitnik[1,†], Michelle M. Li[1,†], Aydin Wells[2,3,4,†], Kimberly Glass[5,#], Deisy Morselli Gysi[5,6,7,#], Arjun Krishnan[8,#], T. M. Murali[9,#], Predrag Radivojac[10,#], Sushmita Roy[11,12,#], Anaïs Baudot[13], Serdar Bozdag[14,15], Danny Z. Chen[2], Lenore Cowen[16], Kapil Devkota[16], Anthony 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,*]

ANNUAL REVIEW OF BIOMEDICAL DATA SCIENCE Volume 7, 2024

Review Article | Open Access

Graph Artificial Intelligence in Medicine

Ruth Johnson[1,2], Michelle M. Li[1,3], Ayush Noori[1,4], Owen Queen[1], and Marinka Zitnik[1,5,6,7]

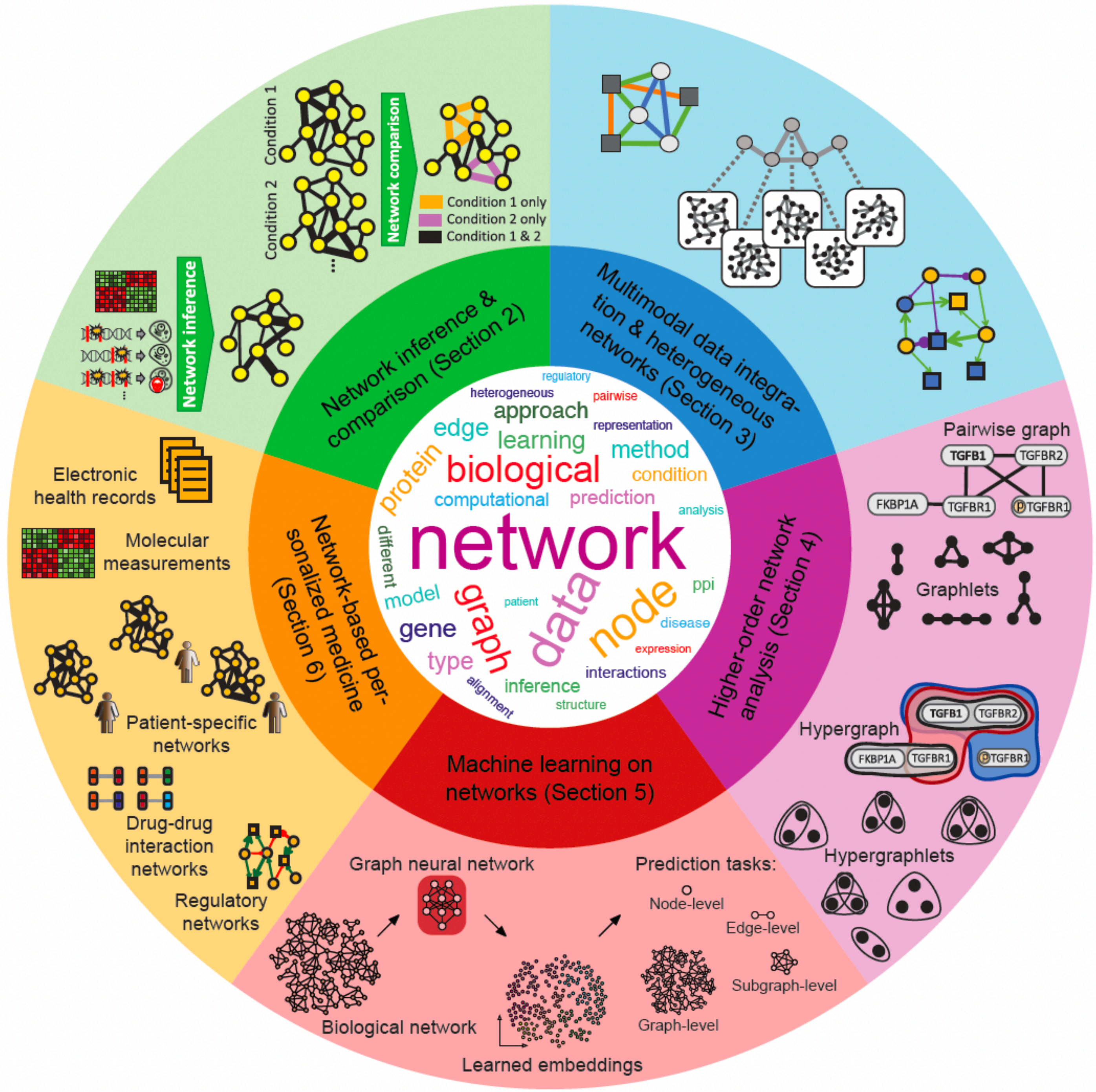


NetBioMed 2023: Networks in Biology and Medicine

What is it? The NetBioMed satellite is part of the NetSci onference 2023 in Vienna **[https://netsci2023.wixsite.com/netsci2023]** and will take place in the morning of July 10.

NetBioMed is dedicated to network science in biology and medicine. It was installed in 2020, when the two long-running satellites NetMed and NetSciReg were combined into a single, integrated satellite.

Some of the current challenges



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

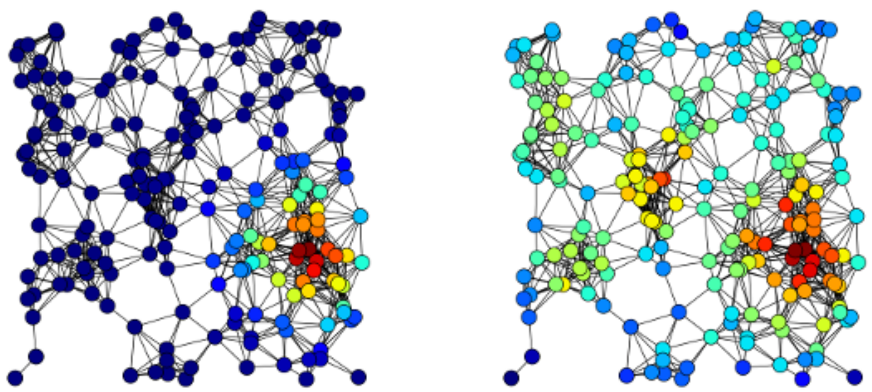
Networks as graphs

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

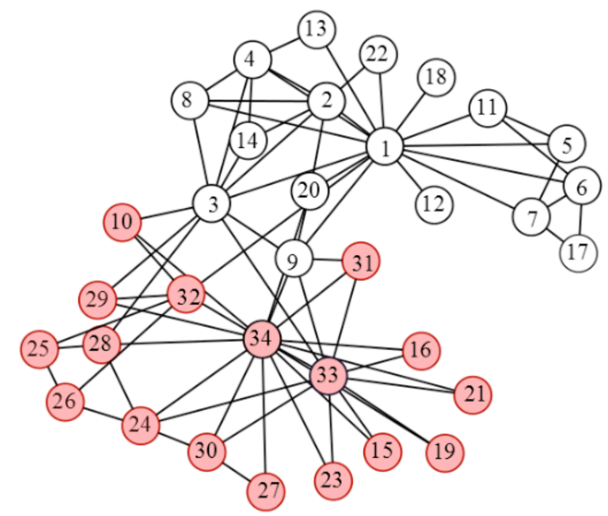
The Königsberg Bridge Problem
[Leonhard Euler, 1736]

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

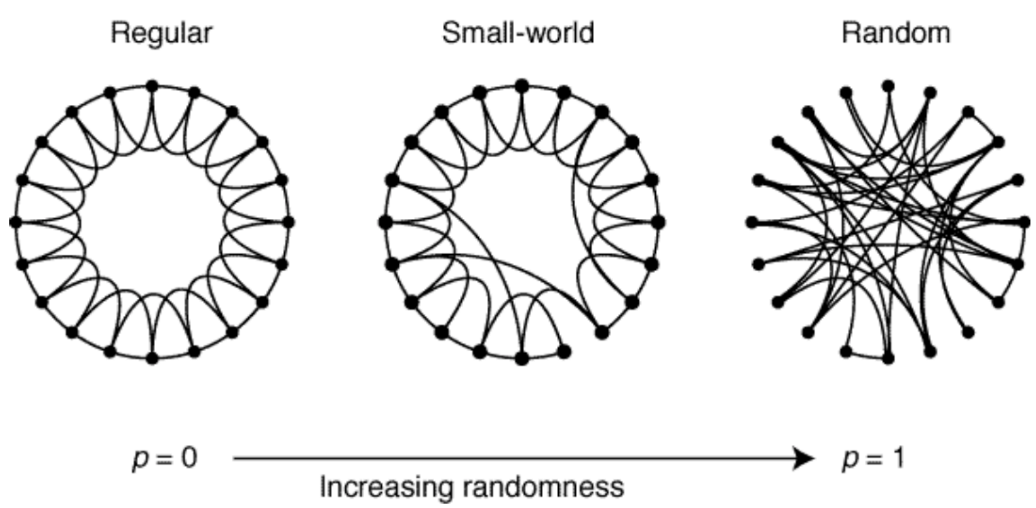
Graph topology analysis



Node centrality



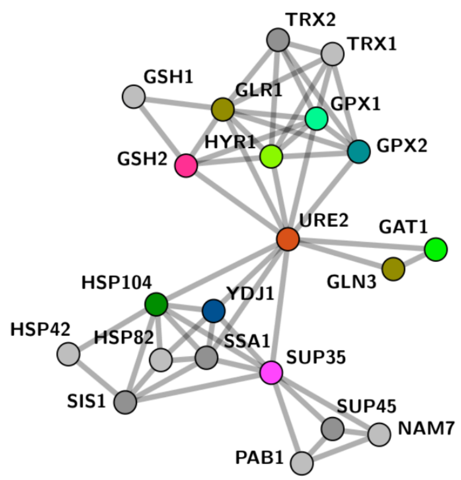
Community detection

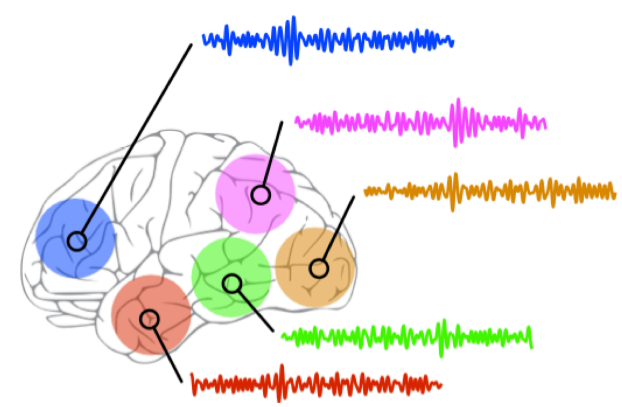


Random graph models

Graph structured data

- From edges to node attributes





Network/graph structured data

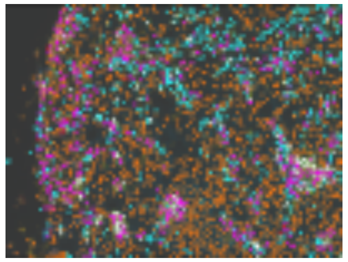
- Need to take into account both structure (i.e., edges), and data (i.e., information on the nodes of the network)

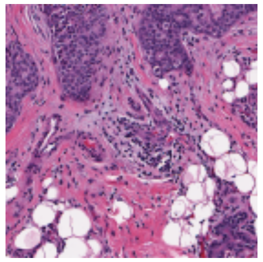
Why learning from graphs is hard?

- 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

Why inference for biomedical data is even harder?

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

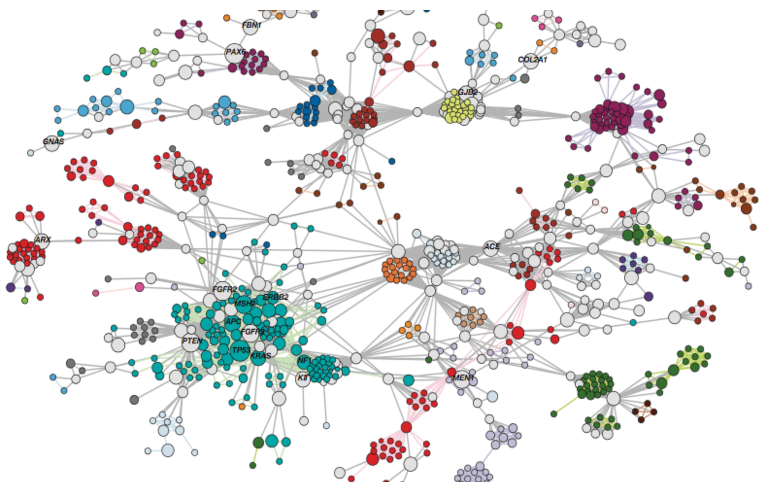




Why inference for biomedical data is even harder?

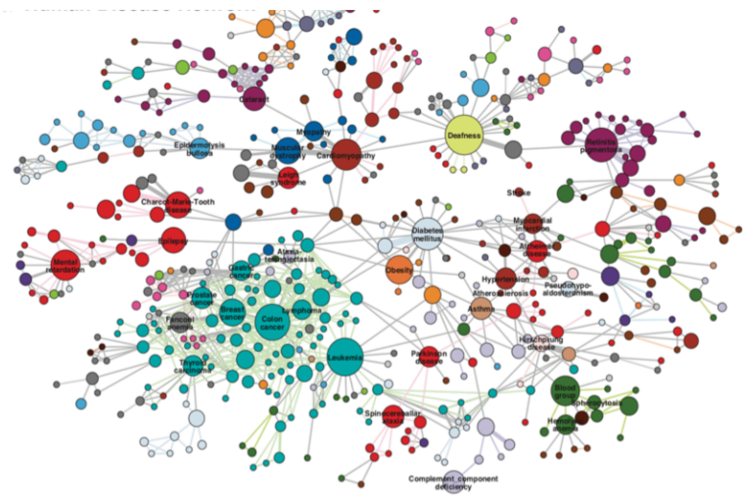
- **Data quality/availability**
  - Typically noisy, incomplete, heterogenous, captured from different technologies

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

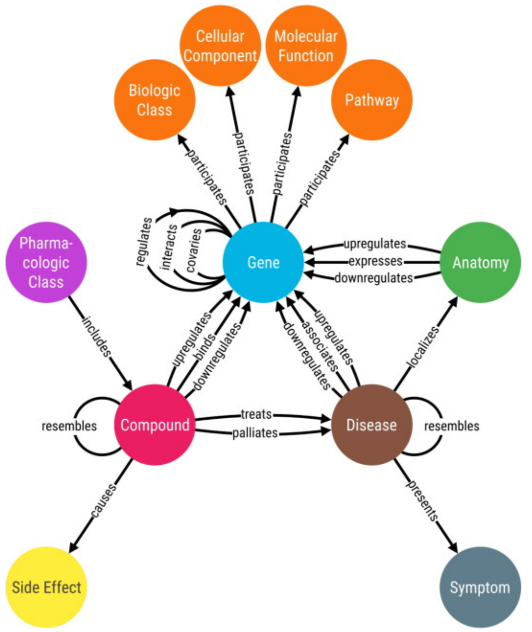


Disease gene network

[Goh et al 2007]



Human disease network



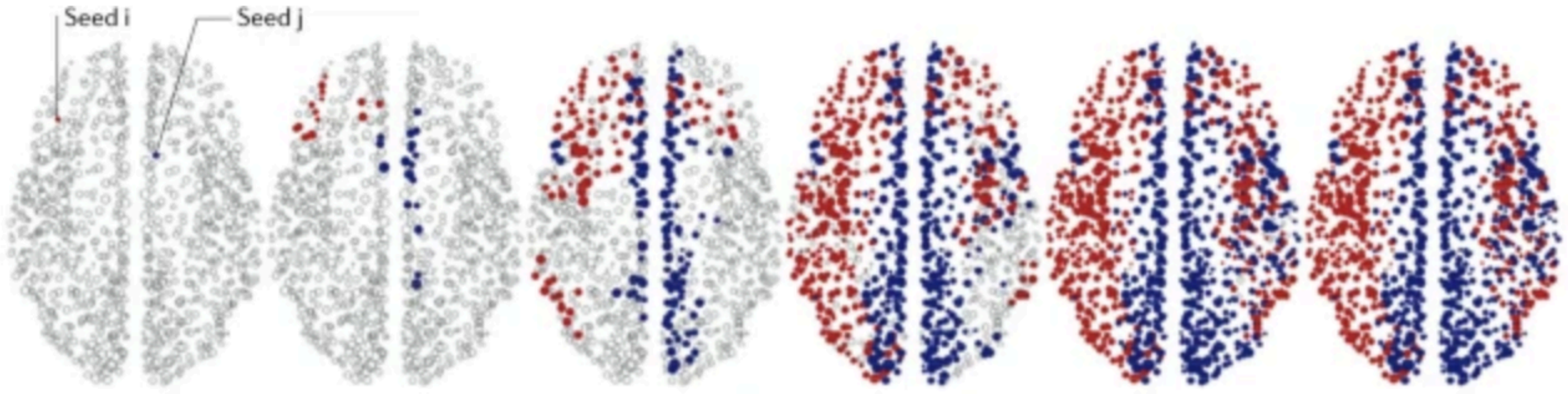
Biomedical knowledge graph

Why inference for biomedical data is even harder?

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

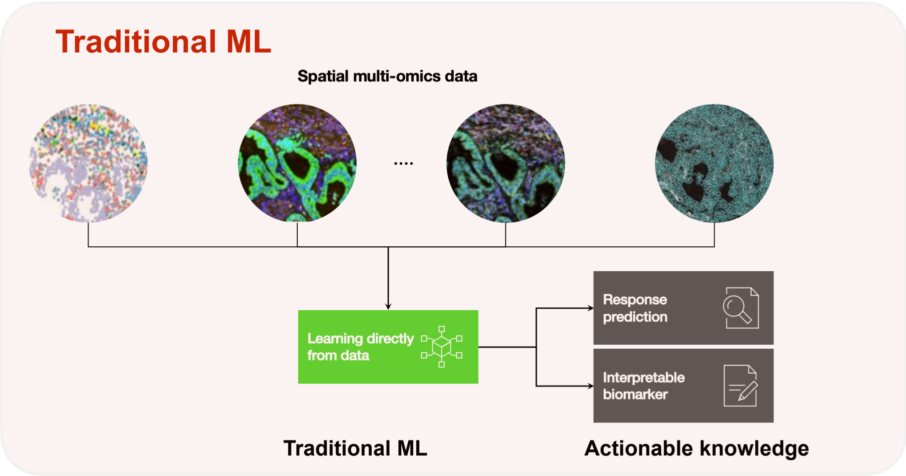
- **Complex interactions at different levels**
  - High complexity due to multiple, and heterogenous interactions at different levels (from molecules to society)

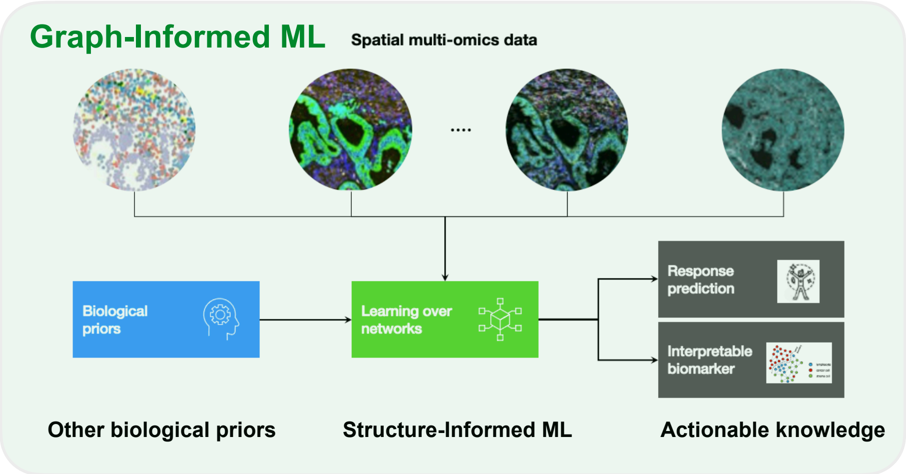
- **Many biological systems exhibit time-varying patterns**



Representation of graph-structured data

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

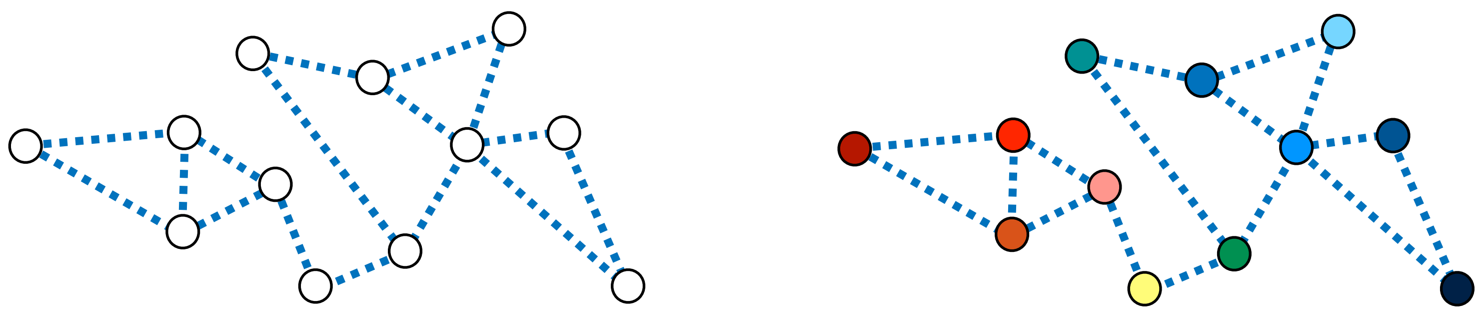




How can we build principled frameworks for graph-structured data?

A signal processing approach

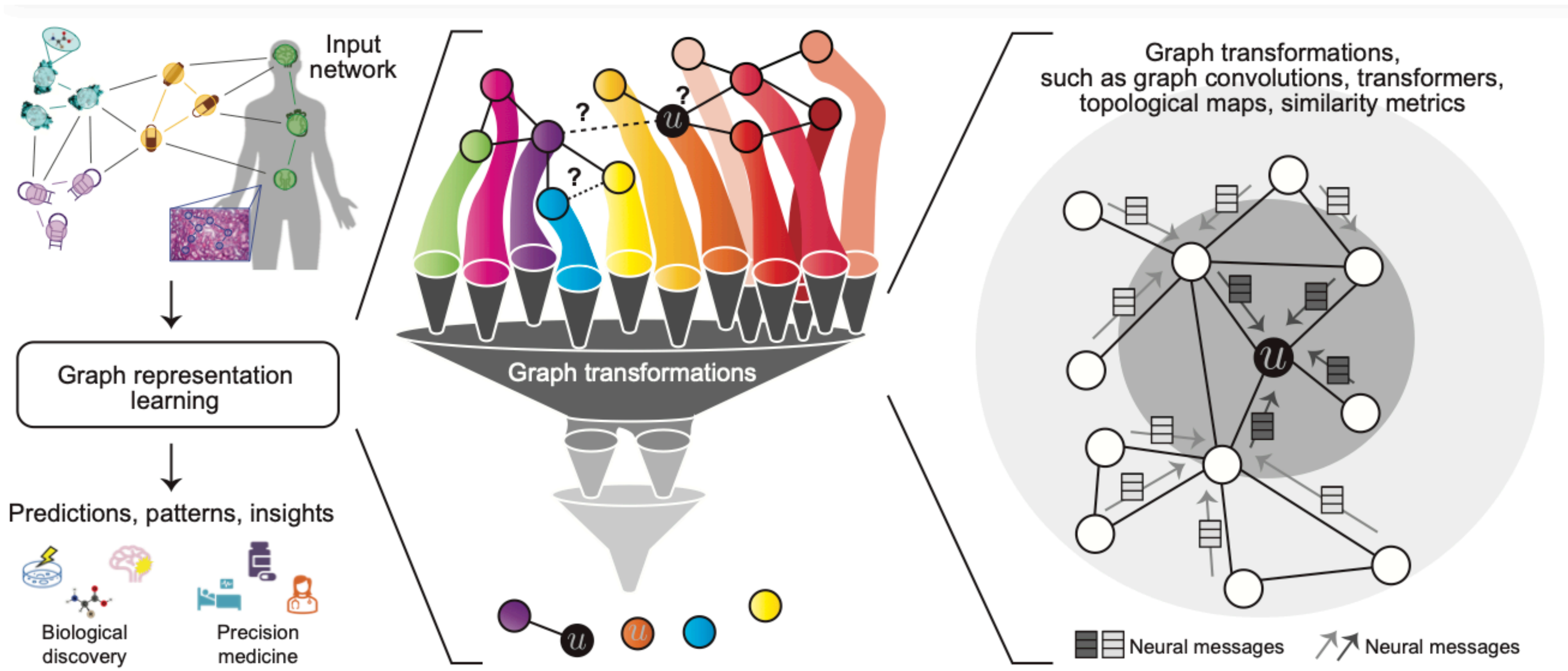
- Exploit the interplay between
  - **graph:** flexible tools to represent a discrete and irregular domain
  - **signal on graph:** function value (attribute) at each vertex



- **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

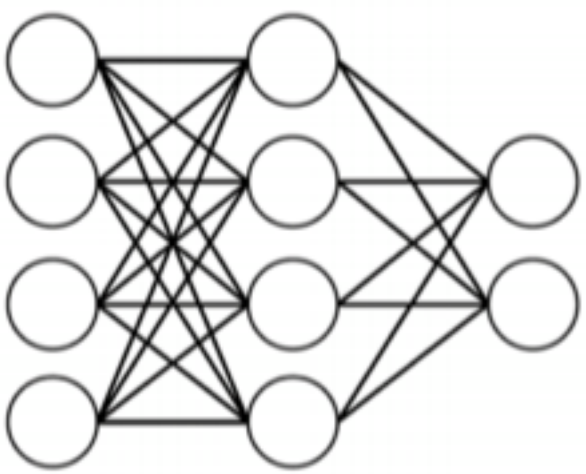
- An emerging research topic, with many practical applications



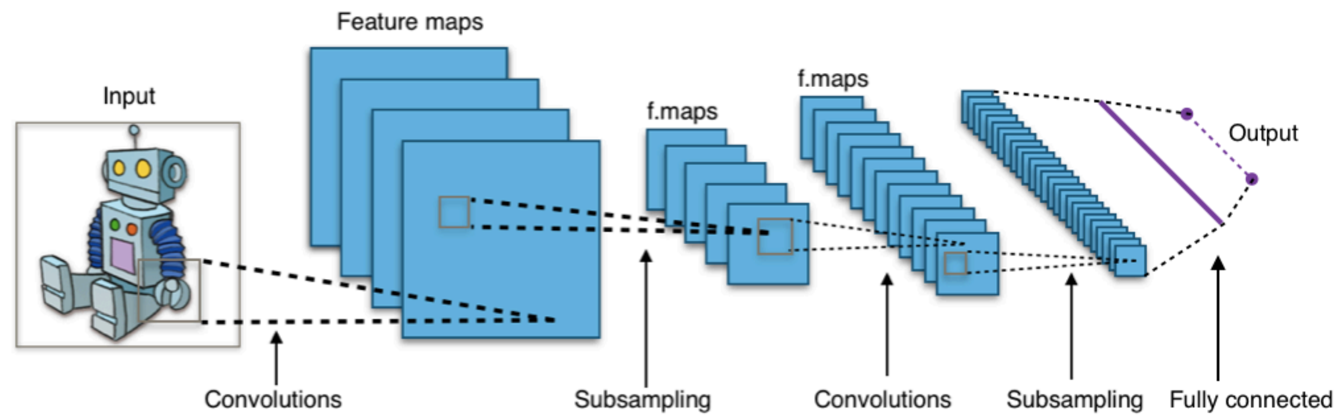
Generate actionable knowledge by learning directly from network data

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

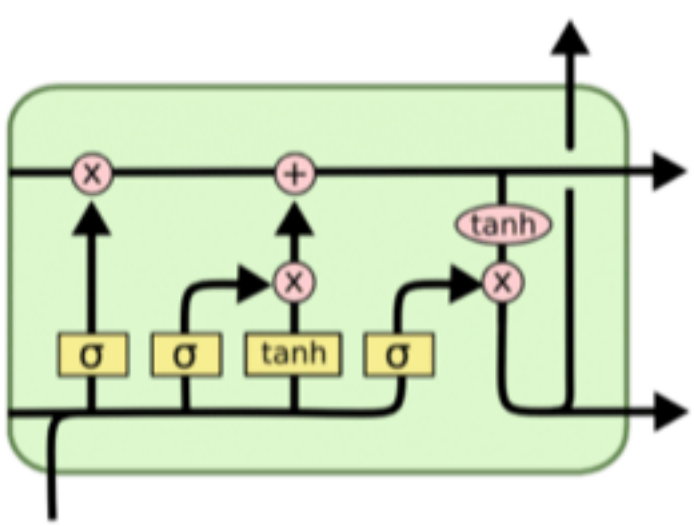
From classical to graph machine learning



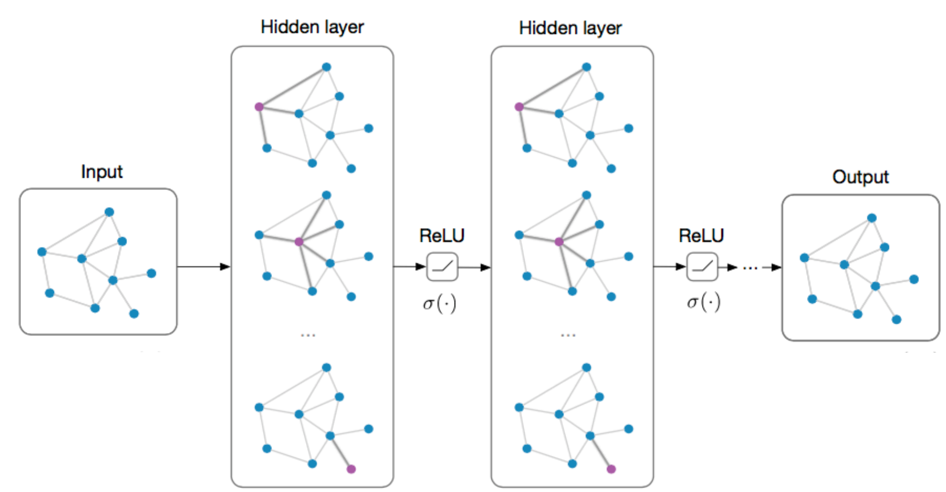
Perceptrons:
Function regularity



CNNs:
Translation invariance

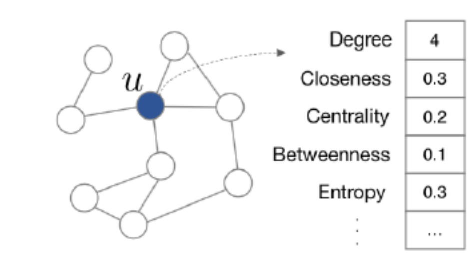


RNNs:
Time warping

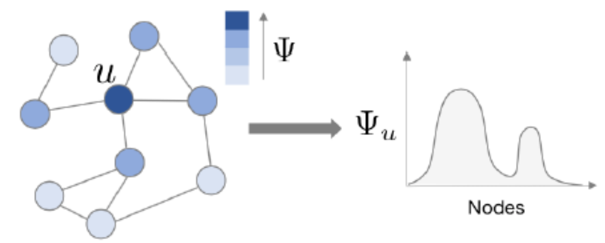


GNNs:
Permutation invariance

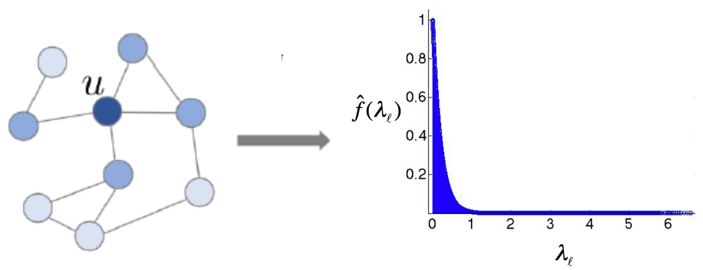
Predominant graph representation learning paradigms



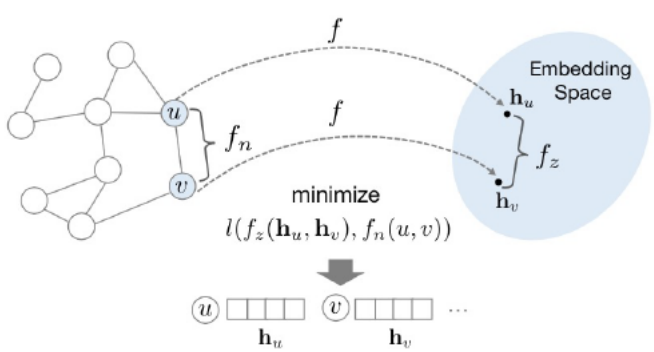
Hand-crafted graph theoretic features



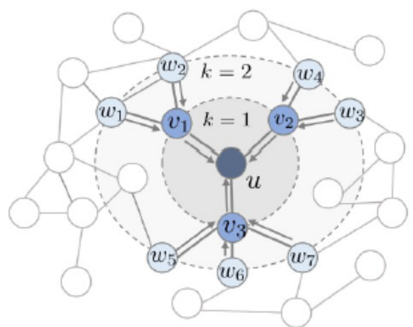
Kernel-based features



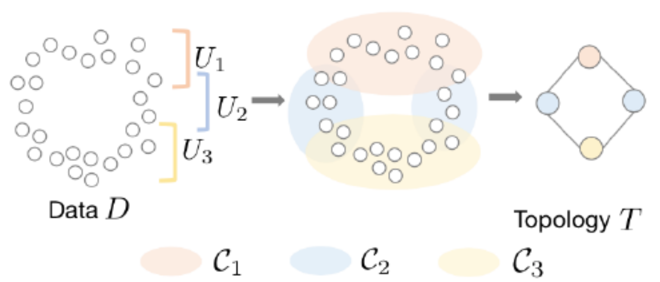
Graph signal processing based features



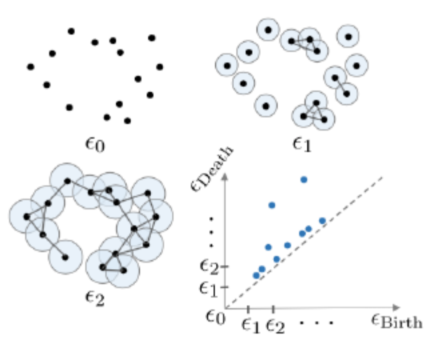
Shallow embeddings



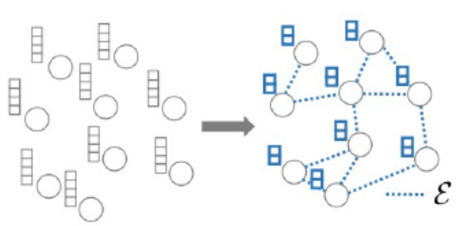
Deep embeddings: Graph neural networks



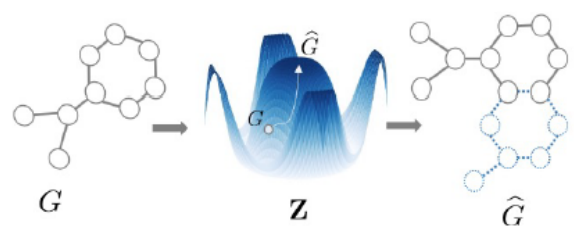
Topological features



Persistent homology



Manifold learning & Topology inference



Graph generative models

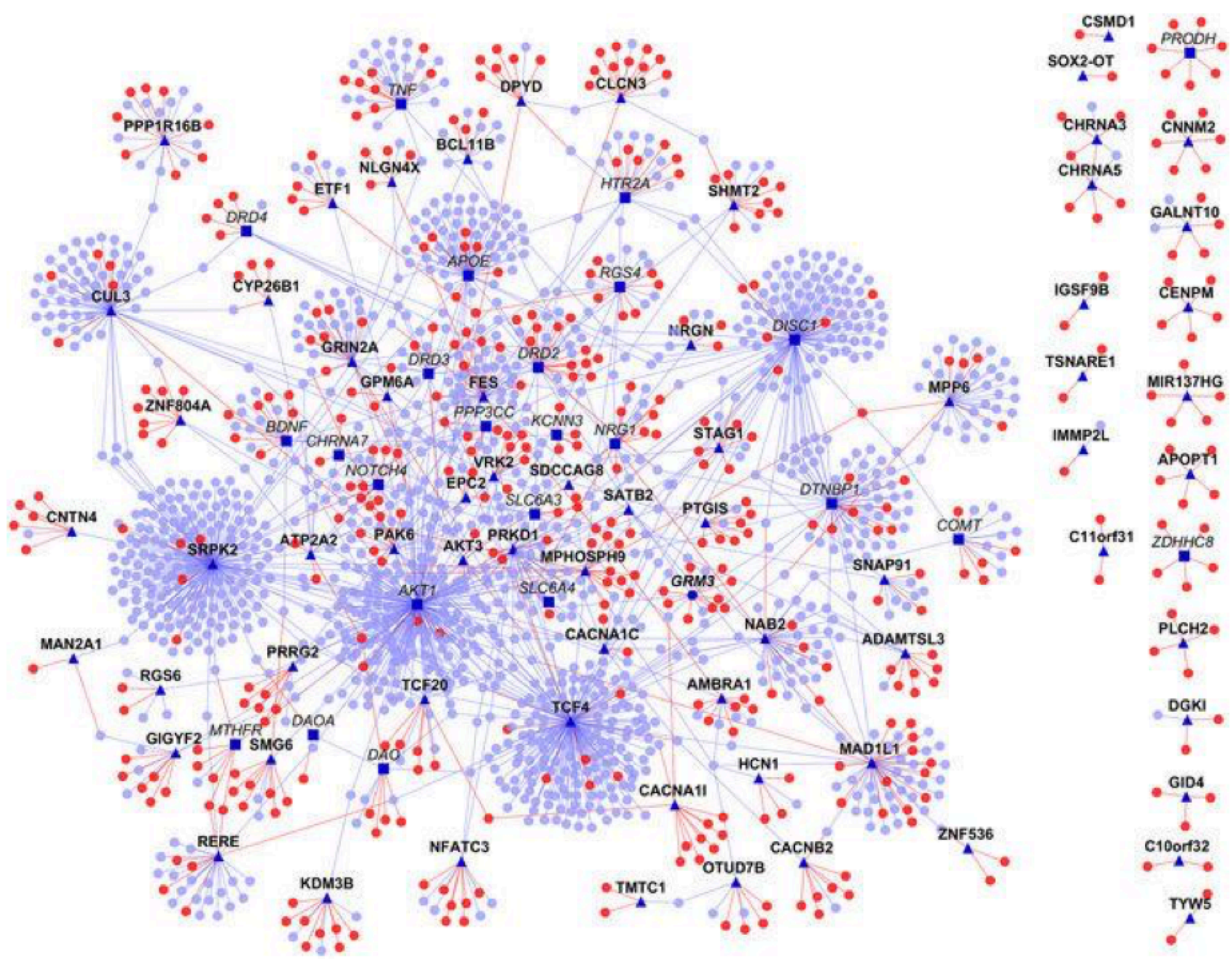
[Fig modified from M. Li, K. Hunag, and M. Zitinik.,Graph Representation Learning in Biomedicine and Healthcare, Nature Biomedical Engineering, 2022]

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

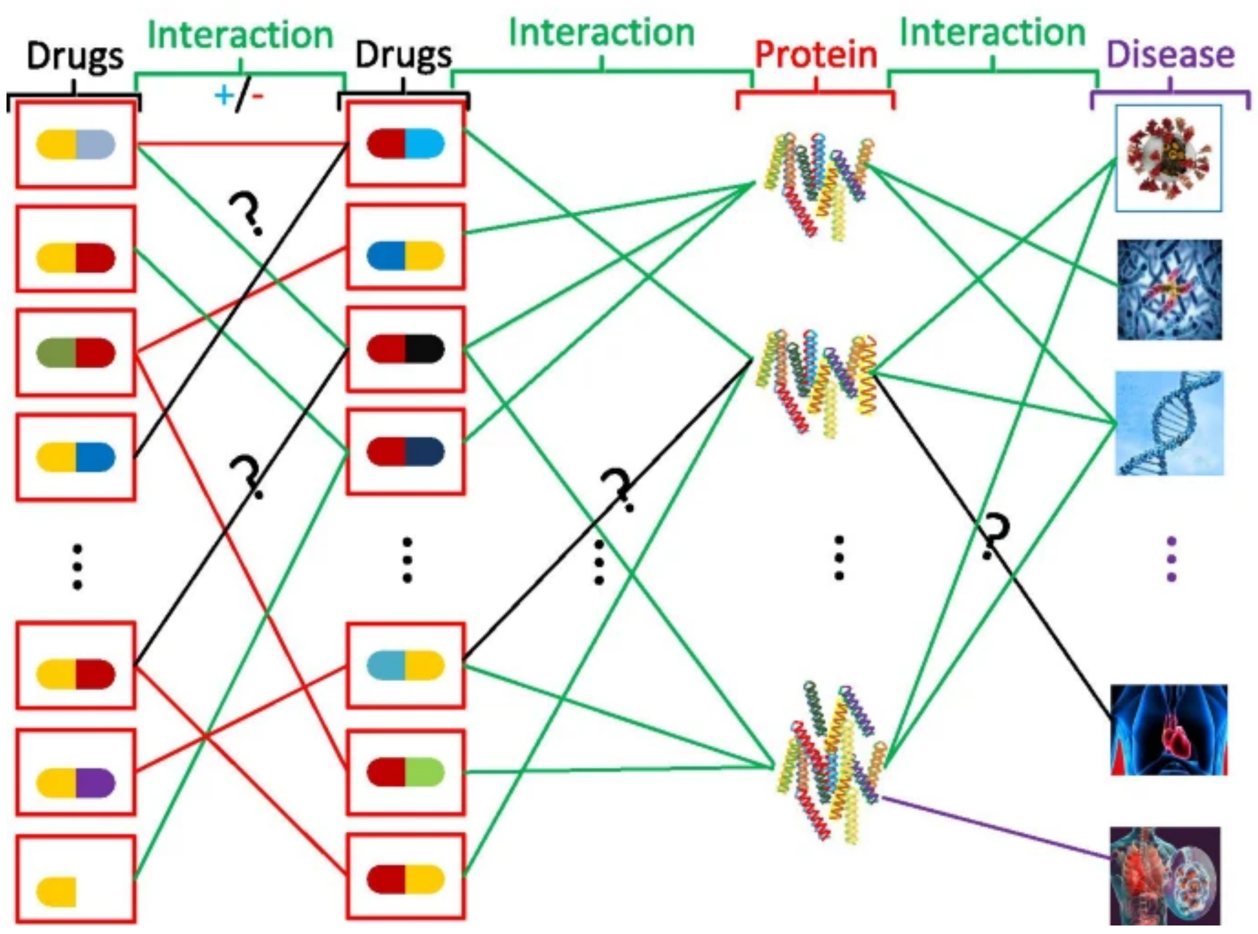
- Classifying the function of proteins in the interactome



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

Link prediction example

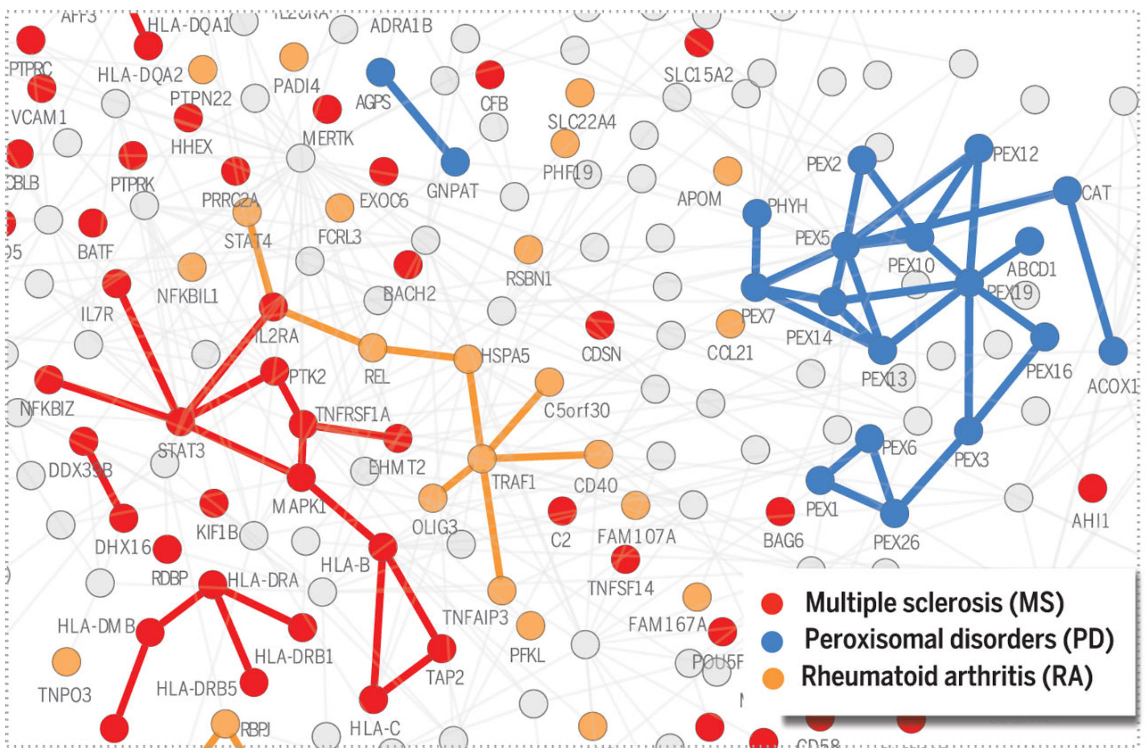
- Predicting drug-target and drug-drug interaction links



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

Cluster identification example

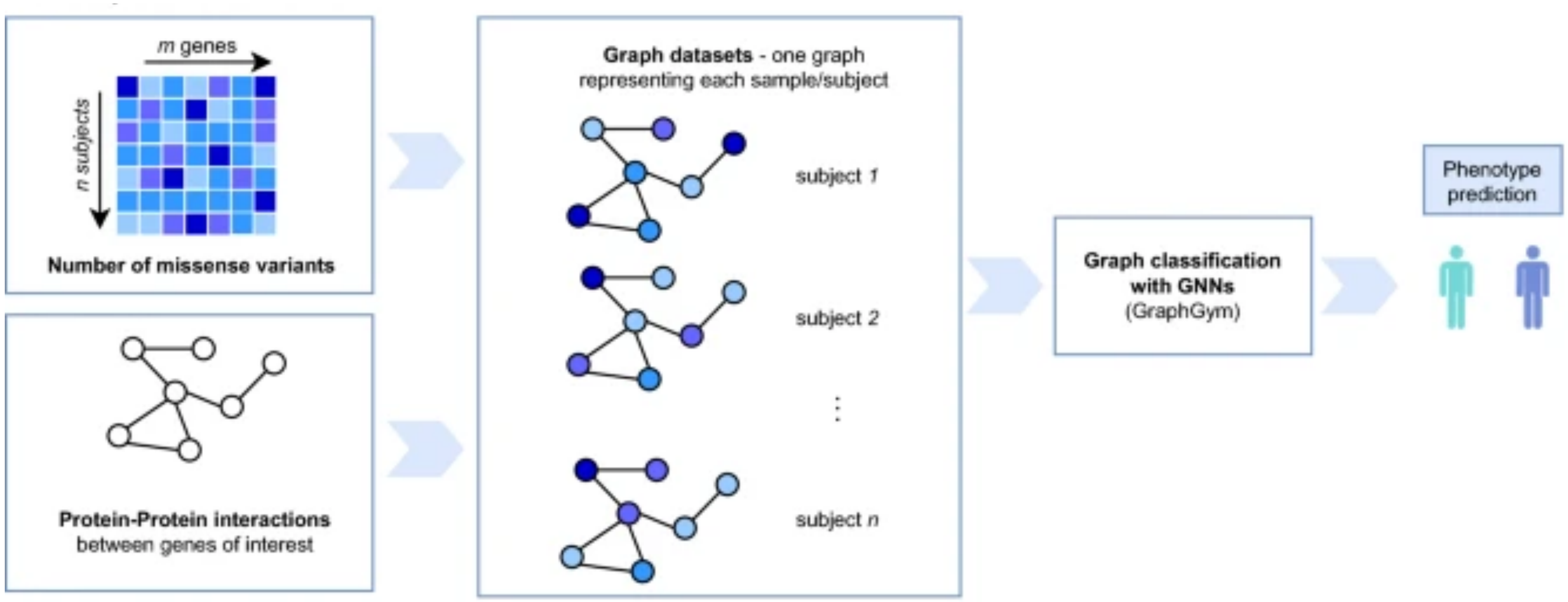
- Identifying proteins associated with the same disease from connected subgraphs



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

Graph classification example

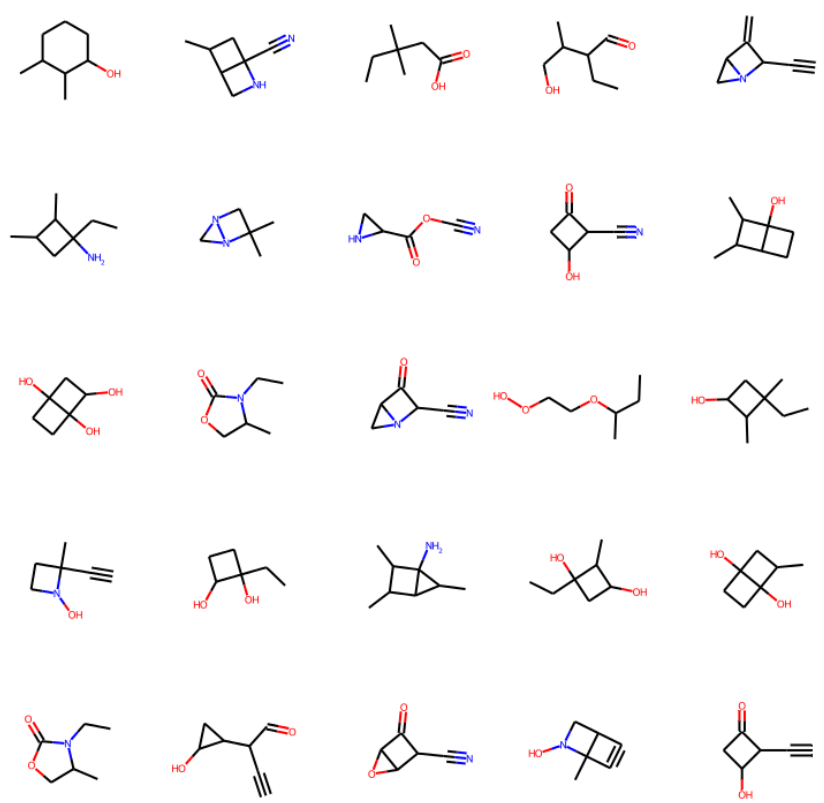
- Predicting patients' phenotype for easy diagnosis of Alzheimer's disease



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

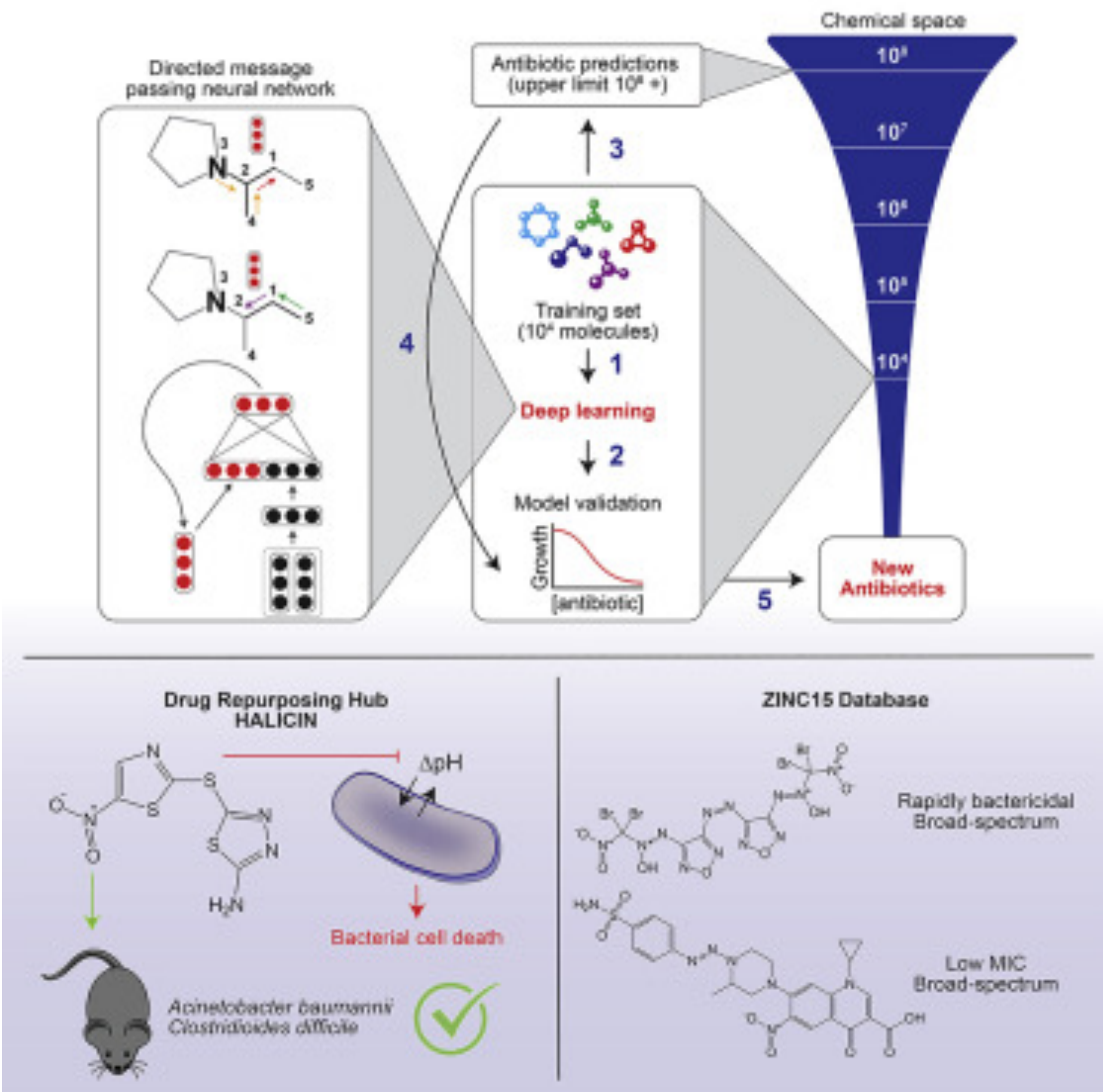
Graph generation example

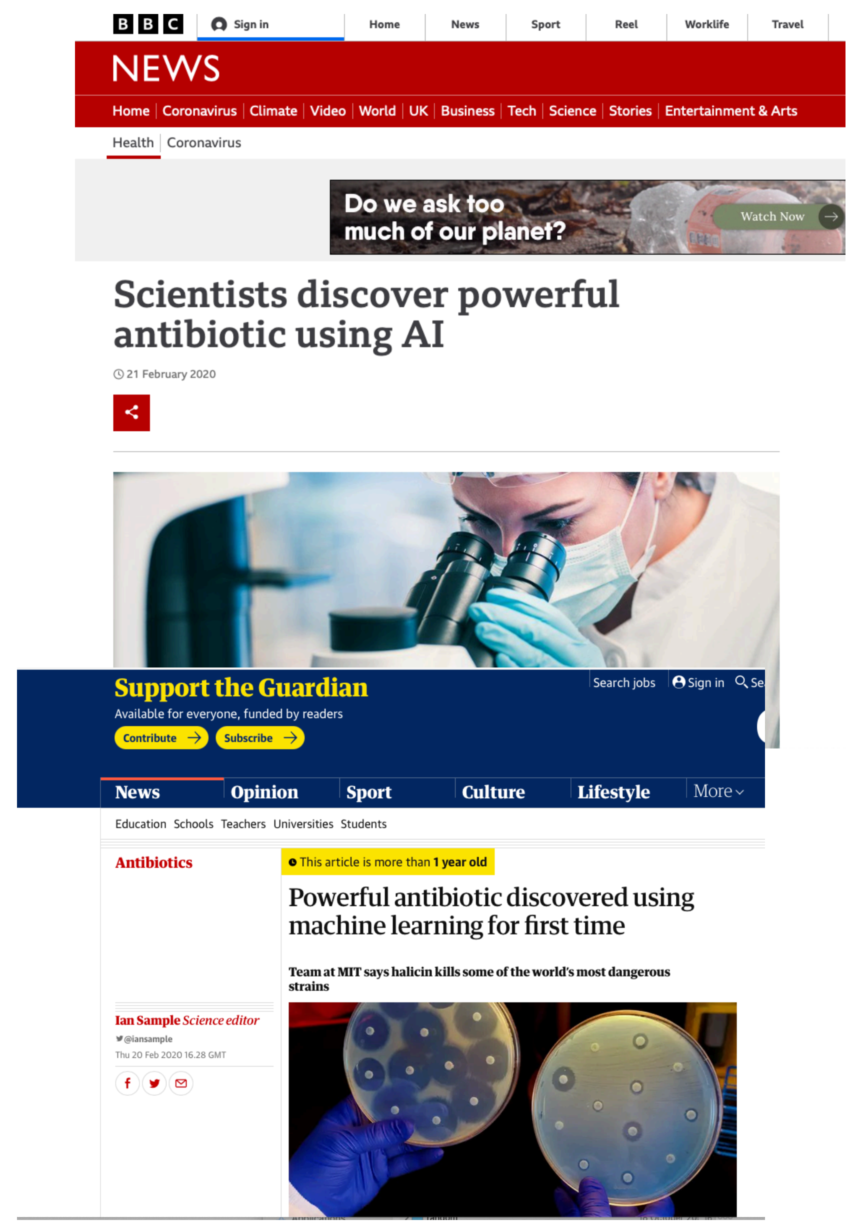
- Generating now molecules



[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





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

Recent success story: Protein folding

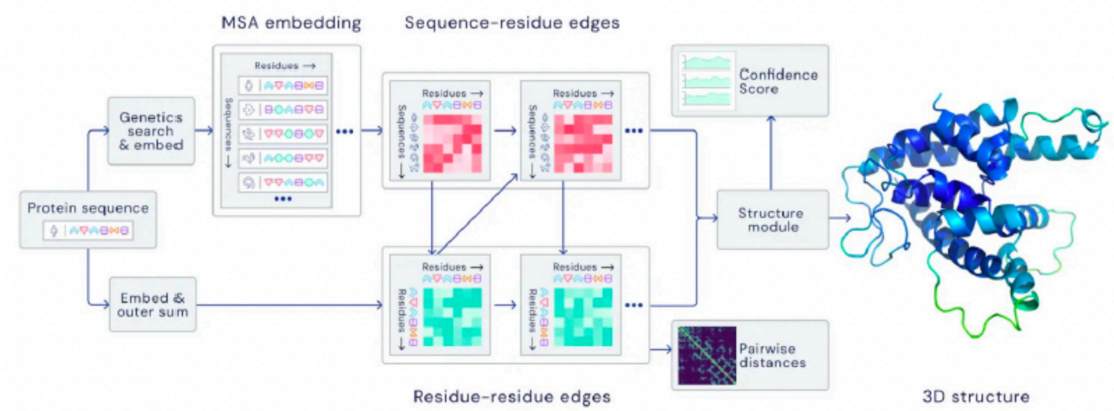
AlphaFold: a solution to a 50-year-old grand challenge in biology

SHARE

AUTHORS

TAt The AlphaFold team

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 10:15-12:00

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 (2 lectures)
  - 9 lectures of paper presentations (~3 papers/week)
  - 1 invited lecture on graphs for modelling the tumor microenvironment
  - 2 weeks of research project presentations

Tentative agenda

Date	Topic	Presenter
11.09.2024	Introduction to the class	Dr. Dorina Thanou
18.09.2024	Introduction into graph theory and graph ML	Dr. Dorina Thanou
25.09.2024	Graph theory and SP driven features	Students presentation
02.10.2024	Graph neural networks	Students presentation
09.10.2024	Graph transformers	Students presentation
16.10.2024	Learning on complex graphs: Higher order graph structures/ Hypergraph GNNs	Students presentation
30.10.2024	Learning on complex graphs: Subgraph GNNs	Students presentation
06.11.2024	Dynamic graphs	Students presentation
13.11.2024	Heterogeous graphs/Knowledge graphs	Students presentation
20.11.2024	Learning in low data regime: Self-supervised learning / Foundation models (?)	Students presentation
27.11.2024	Graph generative models	Students presentation
04.12.2024	Invited lecture	Prof. Marianna Rapsomaniki
11.12.2024	Project presentations	
18.12.2024	Project presentations	

Application domains

- Neuroscience
- Genomics/multi-omics
- Digital pathology
- Drug discovery
- …

Material

- Mainly based on research articles

- Each week, we will cover a few research papers
  - One (overview) paper on core methodological aspects, and two papers on applications

- A list of papers will be distributed next week; The list can be modified upon agreement; suggestions are welcome

Class participation

- **Each week:**
  - A (team of) student(s) will be responsible for presenting papers (T1)
  - A team of students will be driving the discussion and Q&A (T2)
  - All students should contribute to Q&A

- **Project presentation:** Why (or why not) graph-based modelling for your chosen application/research project
  - Discussion of advantages and limitations
  - Perspectives on how graph-based methods can contribute to the specific application field

- **Office hours:** Send me an email, and we will schedule a meeting!

Deadlines

- Please send me your (un)availabilities for presenting before **September 17th!**
- Preference to a specific topic can be expressed
- Please send at least two paper suggestions by **September 17th**
- Choose your teammate before **September 17th!**

Presentation instructions

- ~30 min on core methodological aspects
- ~30 min on the application aspects
- ~30 min of discussion
- You are welcome to add one slide for presenting yourself and your research

Grading

- The final grade will be defined based on:
  - The oral presentations and slides (T1)
  - The lead of the discussion (T2)
  - Participation in the class
  - Final presentation

Questions and feedback?