EE452: Network Machine Learning - Introduction

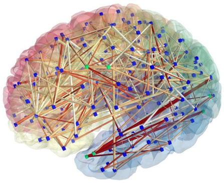
Dr Dorina Thanou
Prof. Pascal Frossard

Outline

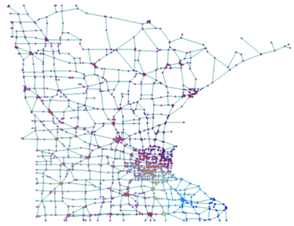
- Why studying networks?
- Graphs as flexible tools for modelling networks
- Network / Graph machine learning
- Overview of EE-452

Network data is everywhere

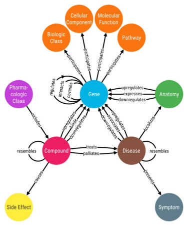
Social networks



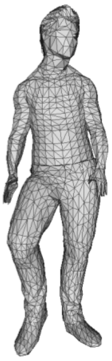
Biological networks



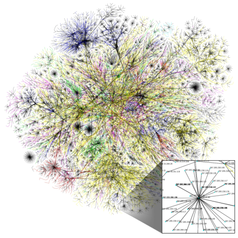
Transportation networks



Knowledge graphs



3D shapes

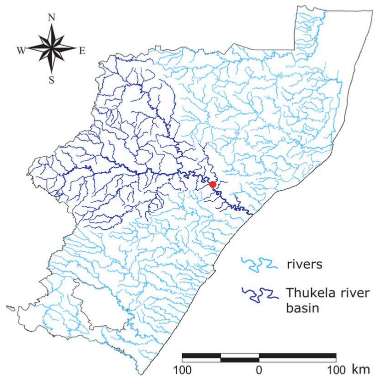


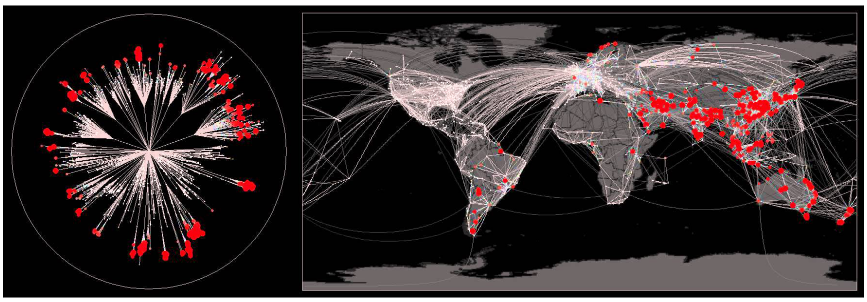
Communication networks

n-body system

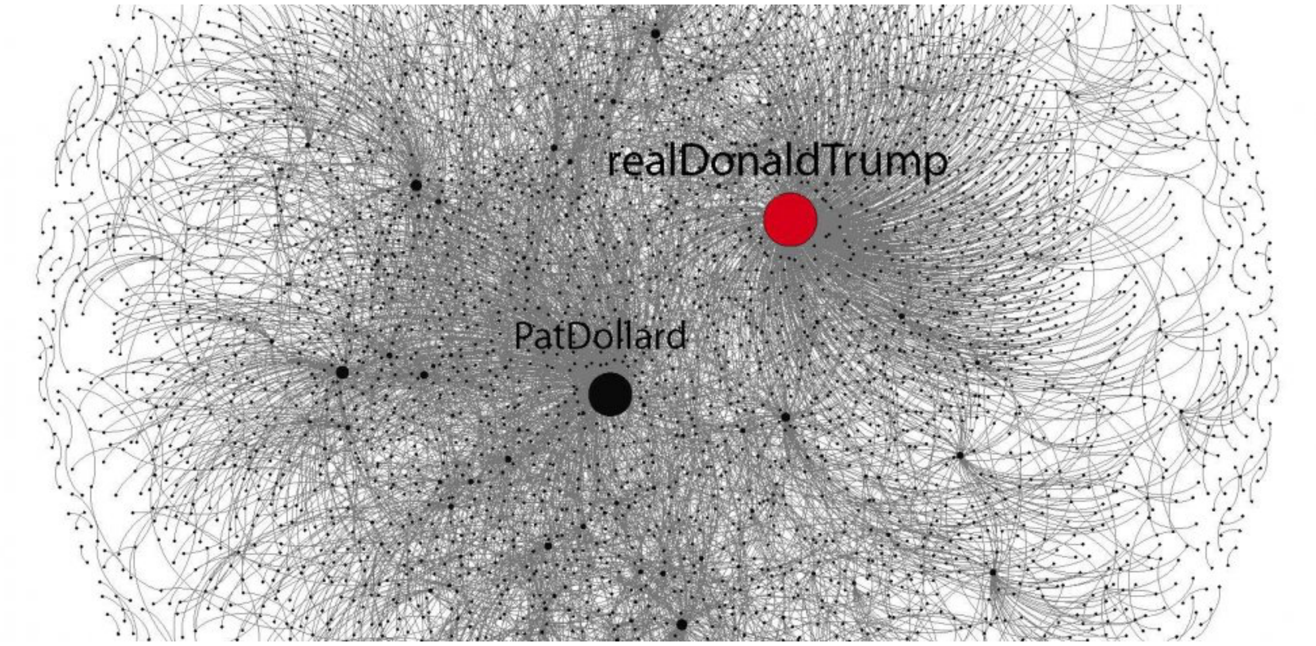
molecule

Diseases spread, then and now



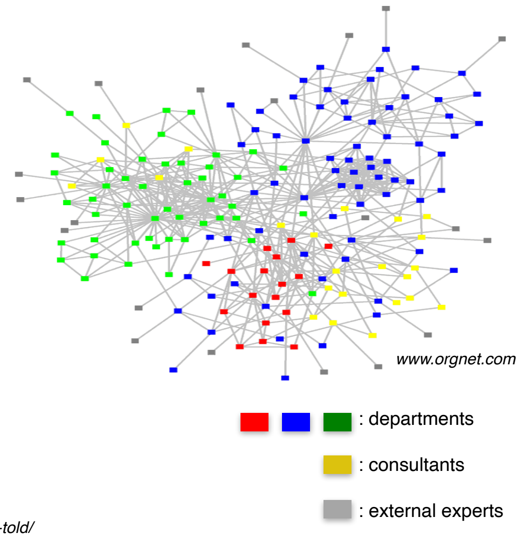


Rumors spread, too



Complex systems and networks

- Behind complex systems there is often a network, which defines the interactions between the components!

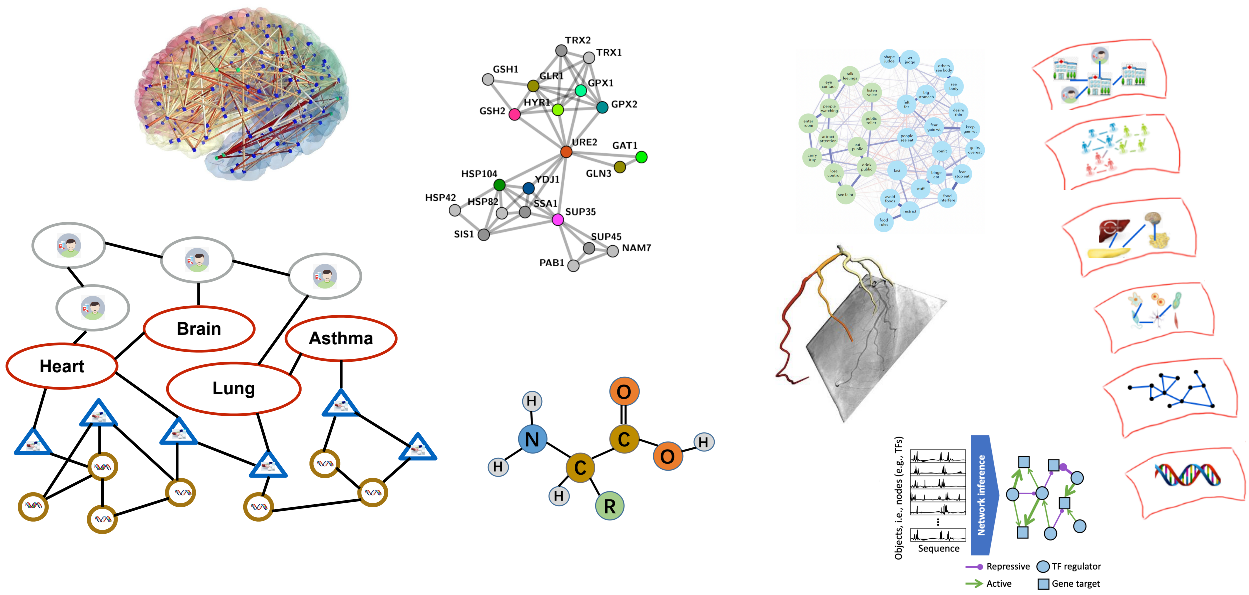


Keith Shepherd's "Sunday Best". http://baseballart.com/2010/07/shades-of-greatness-a-story-that-needed-to-be-told/

We will never understand complex systems unless we understand and properly account for the networks behind them!

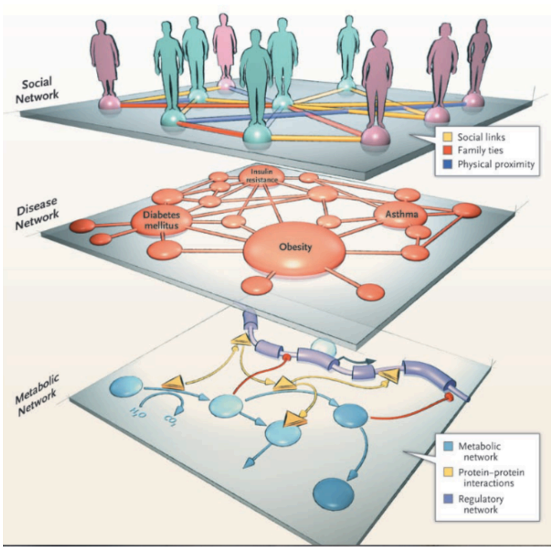
Networks for biology and medicine

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



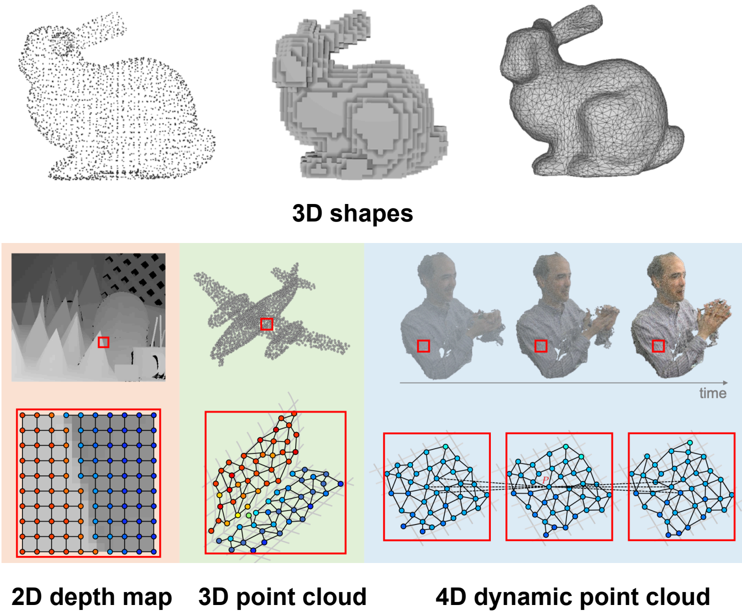
Complex interactions at different scales

- The human body can be seen as a network system



Networks can have completely different scales, yet share some common properties

Networks as a proxy for geometry



Networks can capture functional and semantic relationships in complex objects

Outline

- Why studying networks?
- **Graphs as flexible tools for modeling networks**
- Network / Graph machine learning
- Overview of EE-452

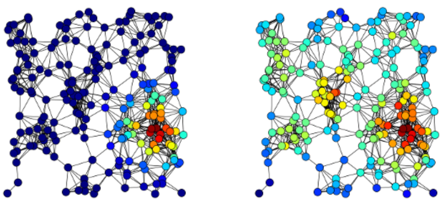
Networks as graphs

- Graphs provide a mathematical representation for describing and modeling complex systems

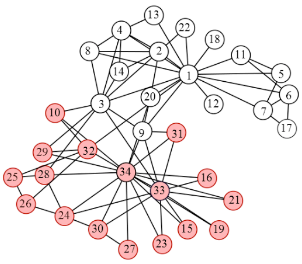
The Königsberg Bridge Problem
[Leonhard Euler, 1736]

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

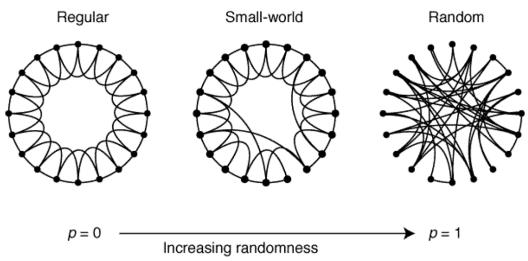
Graph topology analysis



Node centrality



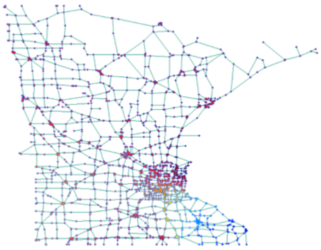
Community detection



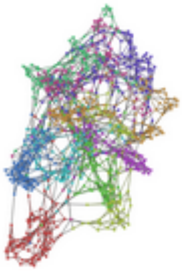
Random graph models

Graph structured data

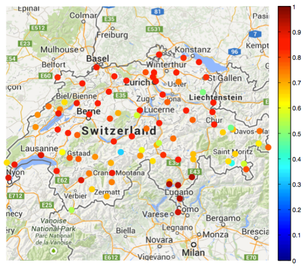
- From edges to node attributes



Transportation networks



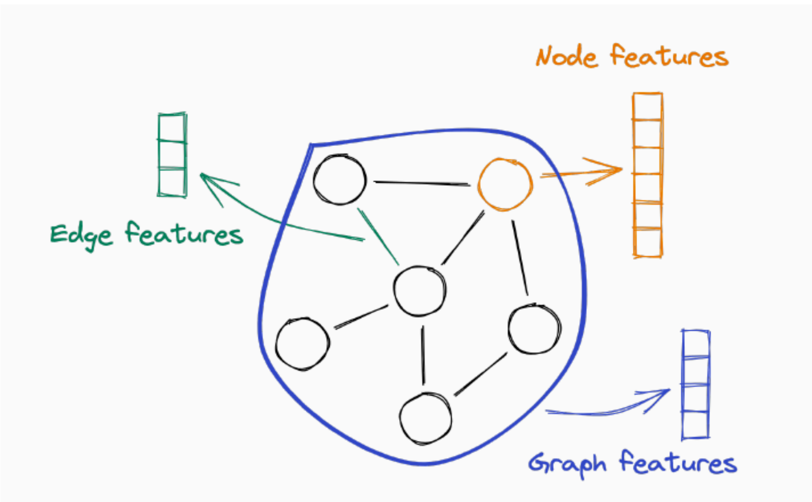
Network/graph structured data



Weather networks

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

Graph features



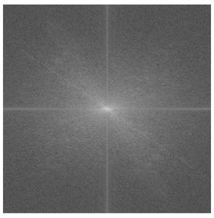
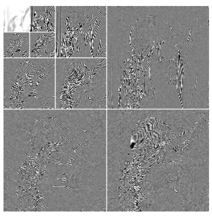
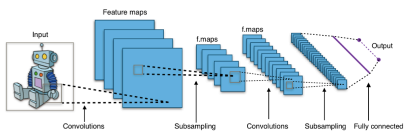
- Example
  - Node features: atom type
  - Edge features: bond type
  - Graph features: molecule energy

Why learning from graphs is hard?

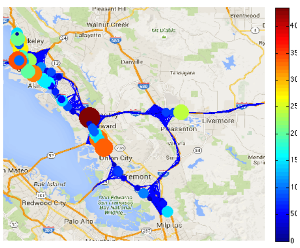
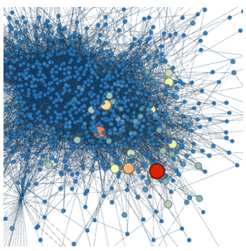
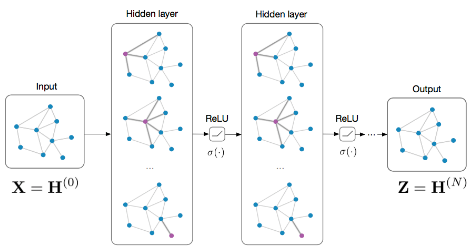
- Irregular domain: complex interactions between nodes of graph
- The size of graphs varies and the number of neighbors changes
- No specific node ordering, leading to different symmetries
  - Permutation equivariance/invariance
- Topologies can change over time: node and edge can appear and disappear

Representation of structured data

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

- Irregular structures: complicated interconnected network structures

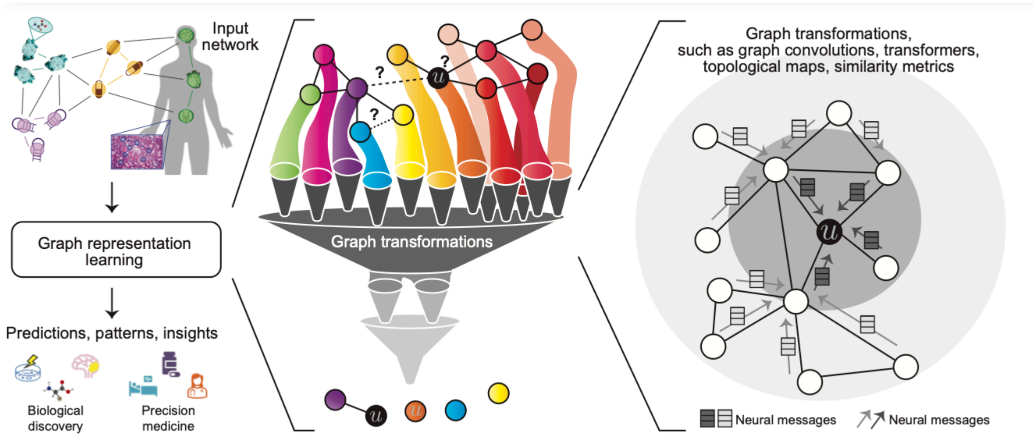
How can we build principled frameworks for graph-structured data?

Outline

- Why studying networks?
- Graphs as flexible tools for modeling networks
- **Network / Graph machine learning**
- Overview of EE-452

Network/Graph machine learning

- A recent research topic, with many practical applications

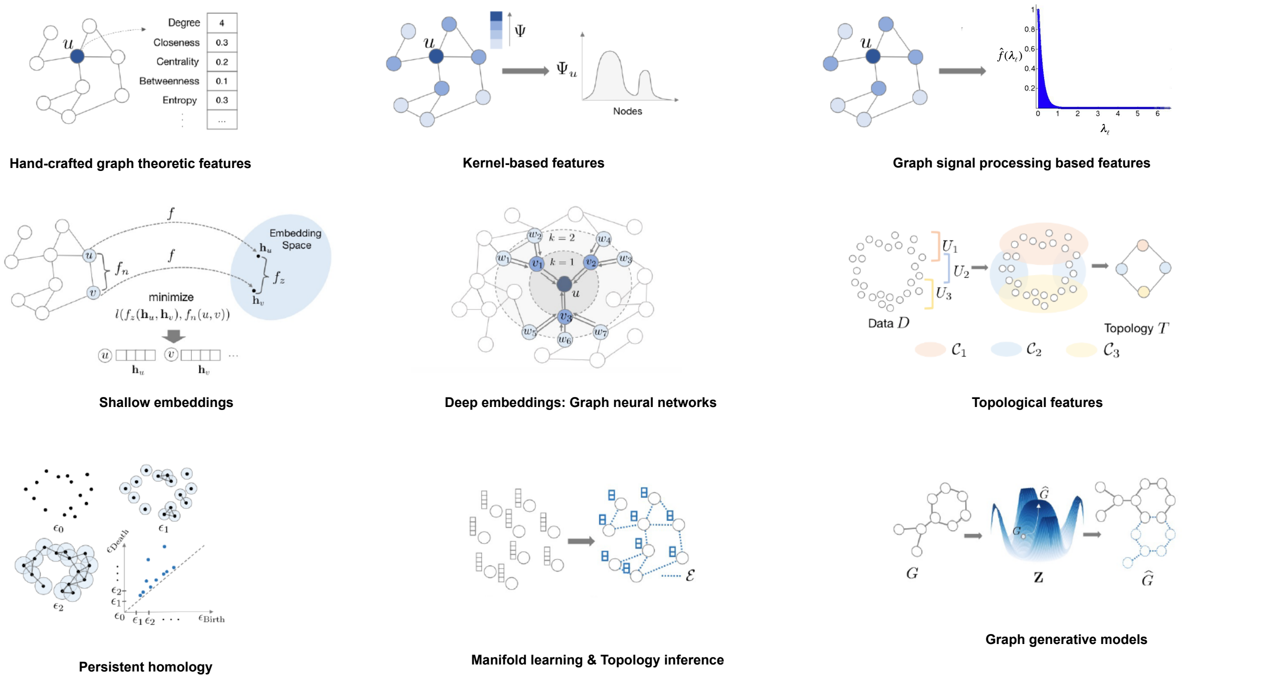


[Fig. from Li'2020]

Generate actionable knowledge by learning directly from network data

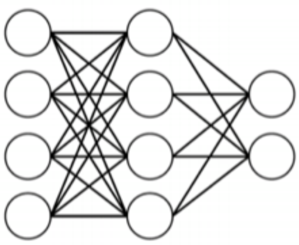
[M. Li, K. Hunag, and M. Zitinik.,Graph Representation Learning in Biomedicine and Healthcare, Nature Biomedical Engineering, 2022]

Predominant graph representation learning paradigms

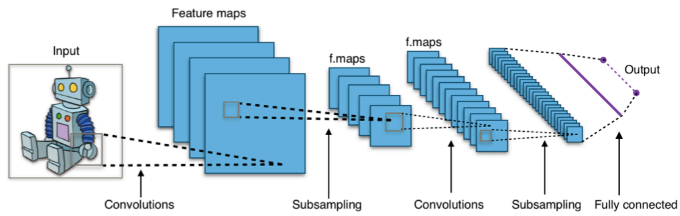


[Fig modified from M. Li, K. Hunag, and M. Zitinik.,Graph Representation Learning in Biomedicine and Healthcare, Nature Biomedical Engineering, 2022]

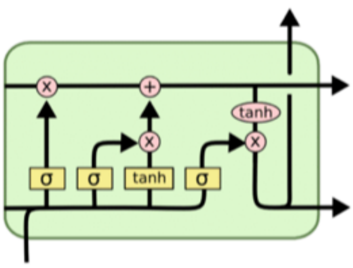
From classical to graph machine learning



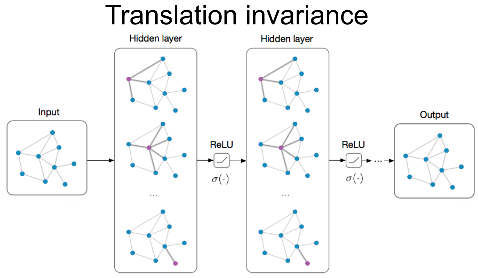
Perceptrons:
Function regularity



CNNs:
Translation invariance



RNNs:
Time warping



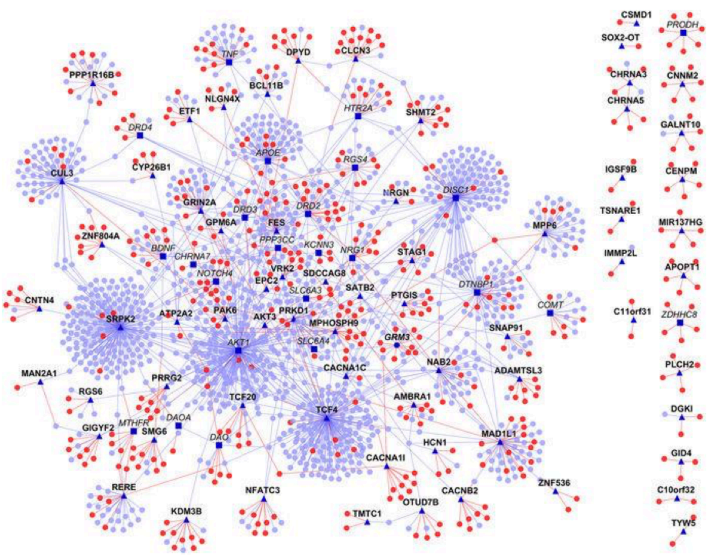
GNNs:
Permutation invariance

Common tasks in network machine learning

- 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

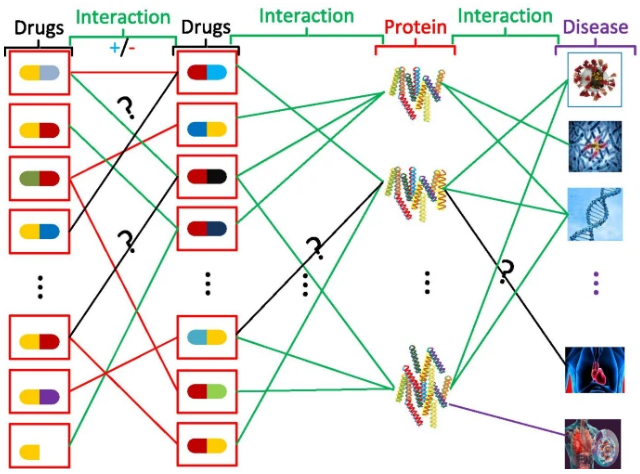
- Classifying the function of proteins in the interactome



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

Link prediction example

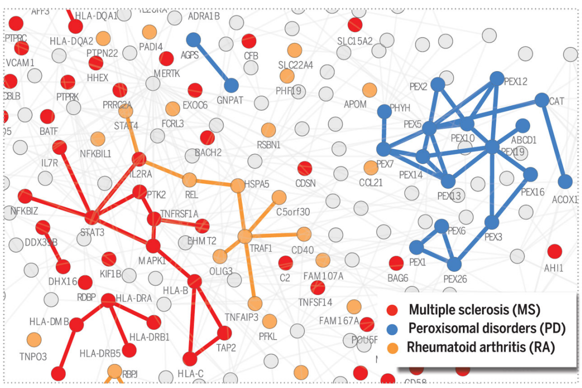
- Predicting drug-target and drug-drug interaction links



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

Cluster identification example

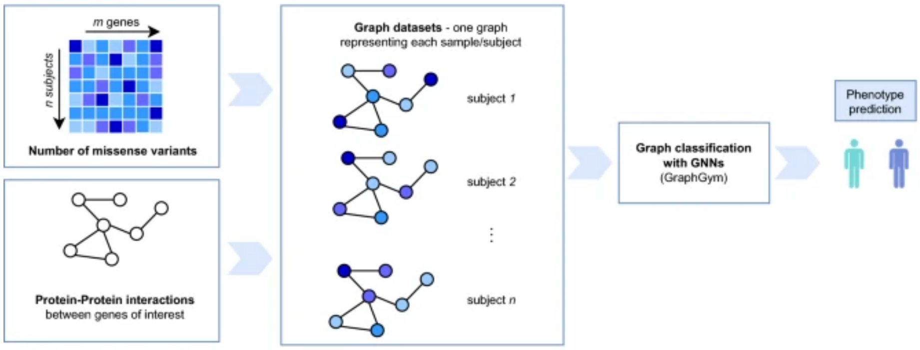
- Identifying proteins associated with the same disease from connected subgraphs



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

Graph classification example

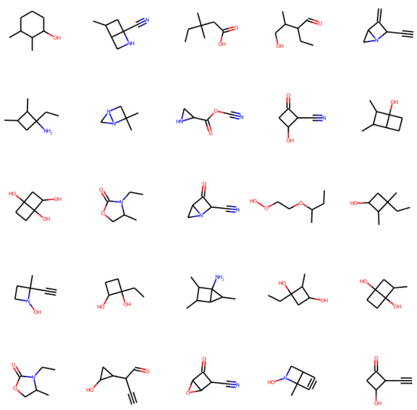
- Predicting patients' phenotype for easy diagnosis of Alzheimer's disease



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

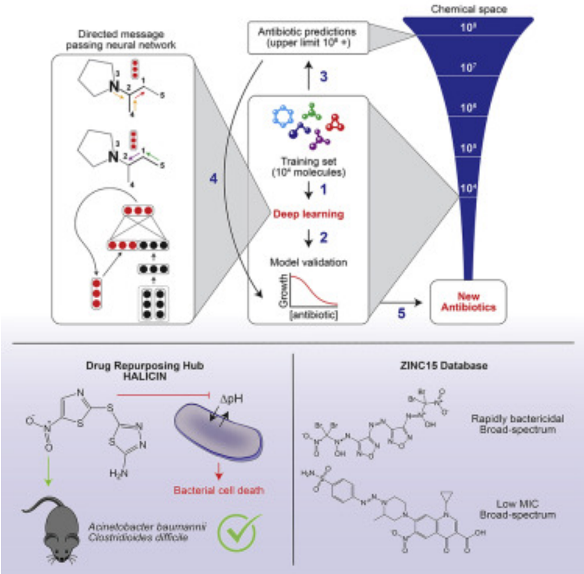
Graph generation example

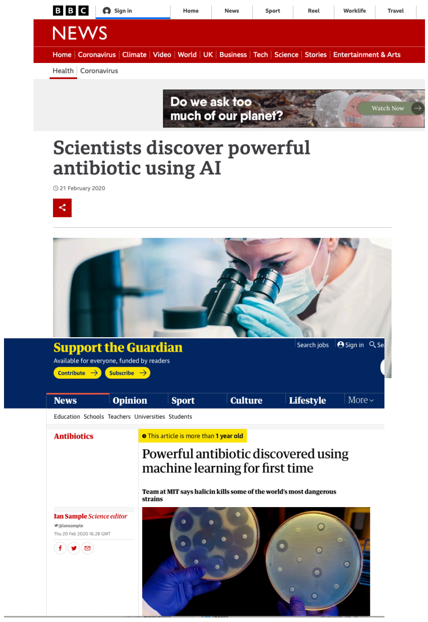
- Generating new molecules



[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




BBC NEWS

Scientists discover powerful antibiotic using AI

21 February 2020

Support the Guardian

Antibiotics

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

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

Recent success story: Protein folding

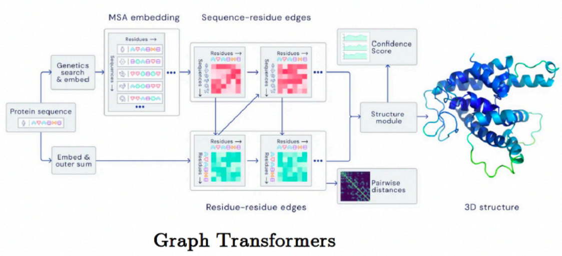

AlphaFold: a solution to a 50-year-old grand challenge in biology

SHARE

AUTHORS

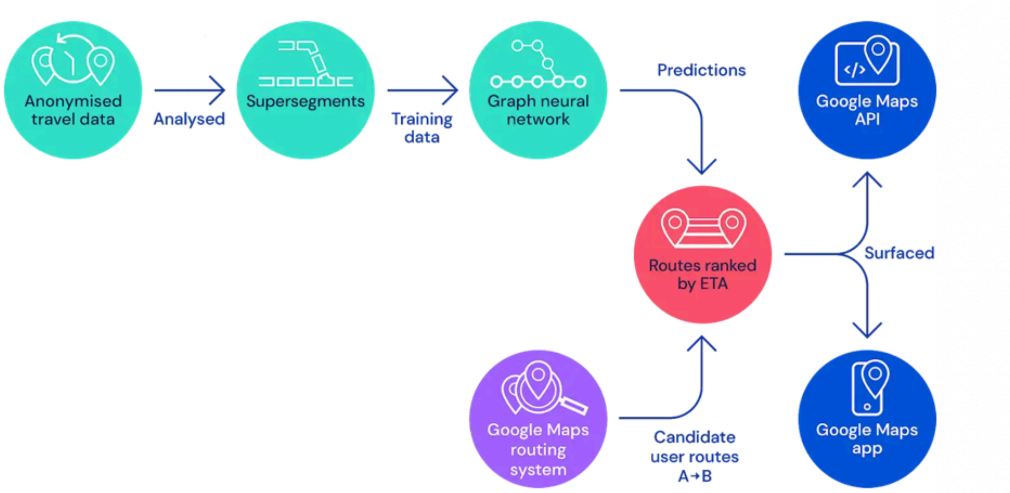
The AlphaFold team

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.



[Jumper et al. 2021]

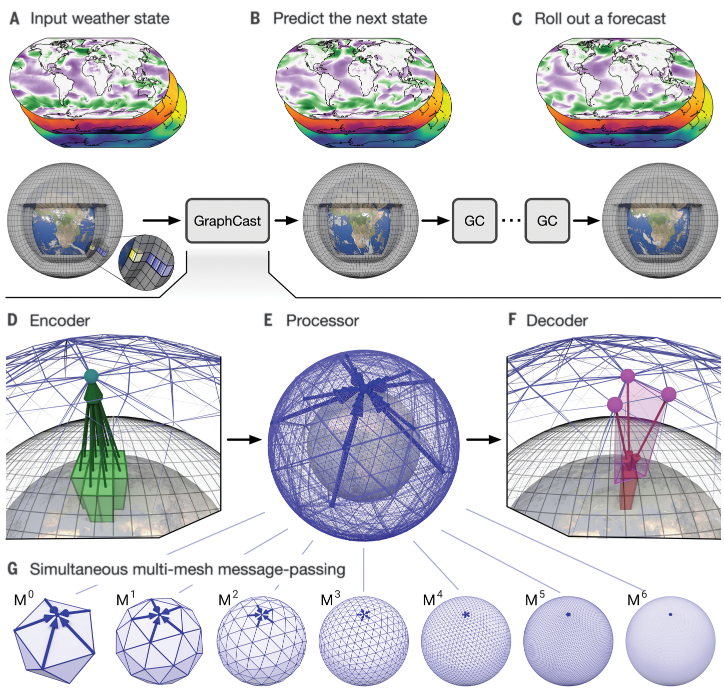
Recent success story: Traffic prediction



The model architecture for determining optimal routes and their travel time.

[Derrow-Pinion et al., 2021]

Recent success story: Weather forecasting



[Lam et al., 2023]

Outline

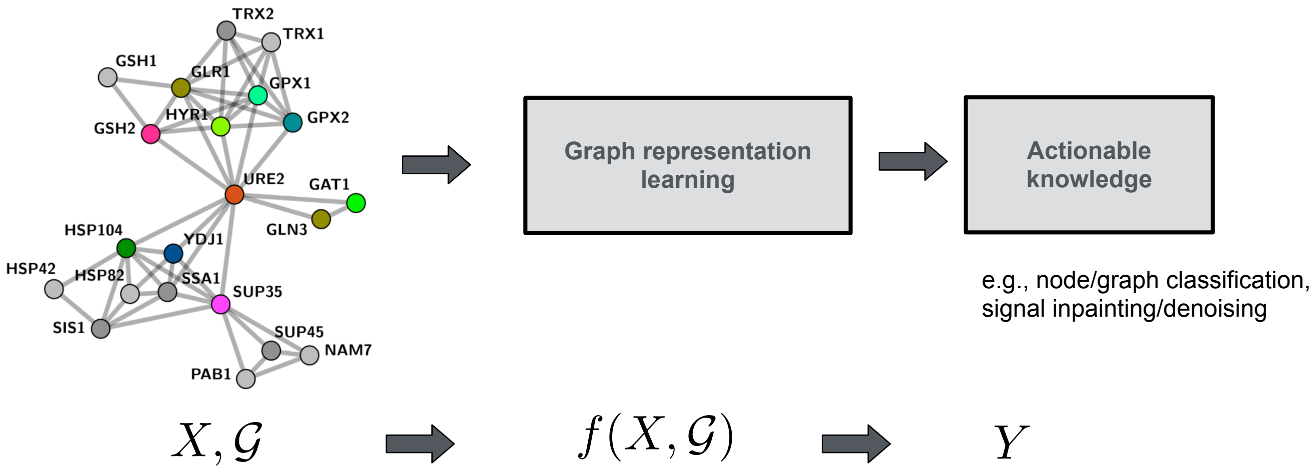
- Why studying networks?
- Graphs as flexible tools for modeling networks
- Network / Graph machine learning
- **Overview of EE-452**

Network Machine Learning - EE452

Dr Dorina Thanou
Prof. Pascal Frossard

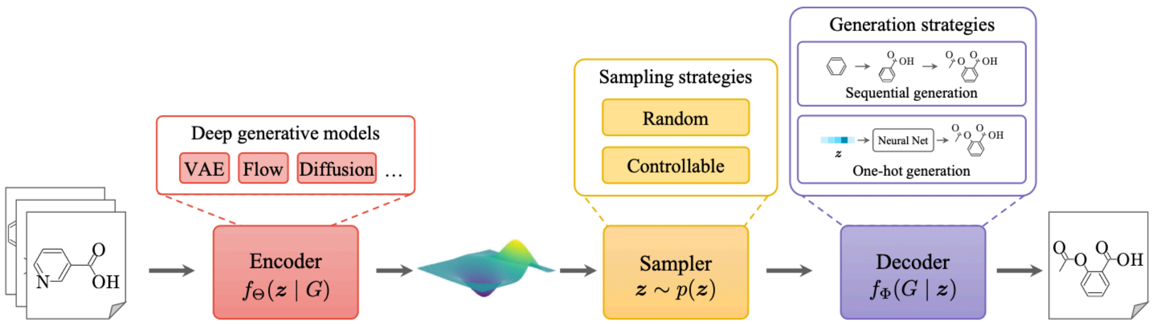
A snapshot of the class (I)

- How can we infer useful information from graph data?
  - Graphs are given
  - From structural graph features to (deep) learned features



A snapshot of the class (II)

- How are graphs generated?
  - How can we generate realistic graphs, using graph generative models (e.g., traditional network science models, deep graph generative models)



[Fig. from Zhu'2022]

Course Description

- **Objectives:** To provide an introduction to methods and algorithms in graph machine learning. A major goal is to understand, analyze and design network-based algorithms in the context of learning, and representation of structured data.

- **Prerequisites:**
  - linear algebra
  - statistics
  - calculus
  - digital signal processing or equivalent
  - machine learning
  - programming (python); familiarity with PyTorch is a plus

Goals of the course and learning outcomes

- Explore recent developments in network machine learning

- Apply these techniques to real world data, and get familiar with popular softwares/tools

- Synthesize arguments into scientific presentations

- Collaborate efficiently with other students

- Provide insights for further research

Organization

- Combination of lectures and lab sessions
  - Lectures: introduction to the theory and tools for the analysis and processing of networks and network data
  - Lab sessions: application of tools to real world data science problems - *please come with your laptops!*

- Sessions on Mon 16-18 (AAC 231) and Tue 13-15 (SG 0211)
  - Agenda available on Moodle
  - Lectures will take place on Tue
  - Lab sessions/tutorials will take place on Mon

- Grading: midterm (40%) and project (60%)

Communication

- All communication and material will be distributed via Moodle

- Slides will be posted before each lecture

- Feel free to post your questions in the forum or contact the team!

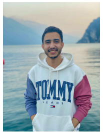
The team

- Main instructors

Dr Dorina Thanou

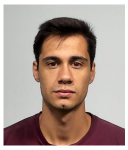
Prof. Pascal Frossard

- Teaching Assistants



Abdellah Rahmani

Jeremy Baffou



Manuel Madeira (lead TA)

Sevda Ogut

William Cappelletti

Yiming Qin

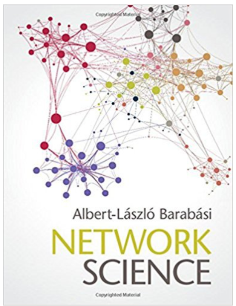
Content of the class

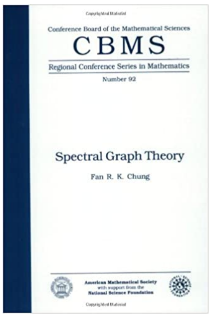
- It is not all about GNNs :)

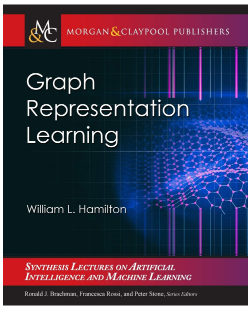
- Topics that will be covered
  - Basics of graph theory
  - Basic graph features & traditional ML
  - Learning (shallow) graph features (from structure)
  - Spectral graph theory and ML applications
  - Processing graphs with attributes
  - Learning (deep) graph features from attributed graphs (GNNs)
  - Graph transformers
  - Classical (random, scale free) and deep graph generative models (e.g., diffusion)
  - Applications

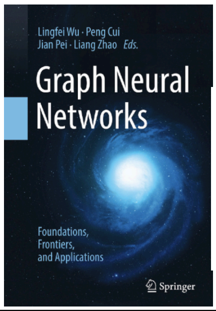
Material

- Course support
  - your notes
  - lecture slides (available on Moodle)
  - research papers
  - suggested textbooks:
    - Network Science, by Albert-László Barabási, 2016
    - Graph Representation Learning by William L. Hamilton, 2020
    - Geometric Deep Learning Grids, Groups, Graphs, Geodesics, and Gauges, by Bronstein et al. 2021









Geometric Deep Learning
Grids, Groups, Graphs,
Geodesics, and Gauges

Michael M. Bronstein[1], Joan Bruna[2], Taco Cohen[3], Petar Veličković[4]

May 4, 2021

Tentative agenda

Date	Lecture	Topic	Notebook	Description
2/17/2025	Introduction			
2/24/2025			Notebook 1	Python setup + Numpy, Pandas, SkLearn
2/25/2025	Lecture 1	Graph Theory basics		
3/3/2025			Notebook 1	NetworkX + graph stats + graph visualization
3/4/2025	Lecture 2	Spectral graph theory		
3/10/2025			Notebook 2 + Project Q&A	Spectral analysis/clustering
3/11/2025	Lecture 3	Structural network features + traditional ML on graphs		
3/17/2025			Notebook 3 + Project Q&A	Node classification + link prediction (Assignment 1 2024)
3/18/2025	Lecture 4	Node embeddings		
3/24/2025			Notebook 3 + Project Q&A	Node classification + link prediction (Assignment 1 2024)
3/25/2025	Lecture 5	Graph structured data		
3/31/2025			Notebook 4 + Project Q&A	GSP (Assignemt 3 2023 Part II - just confirm that is in lecture)
4/1/2025	Lecture 6	Graph Neural Networks Part I		
4/7/2025			Notebook 5 + Project Q&A	GNNs (Assignment 2 2024)
4/8/2025	Lecture 7	Graph Neural Networks Part II		
4/14/2025			Midterm Q&A	
4/15/2025		**Midterm**		
Easter break				
4/28/2025			Notebook 5 + Project Q&A	GNNs
4/29/2025	Lecture 9	Graph Neural Networks Part III - Limitations of GNNs		
5/5/2025			Notebook 6 + Project Q&A	Graph Transformer + Graph Diffusion Models
5/6/2025	Lecture 10	Graph transformers		
5/12/2025			Notebook 6 + Project Q&A	Graph Transformer + Graph Diffusion Models
5/13/2025	Lecture 11	Graph generative models		
5/19/2025			Notebook 6 + Project Q&A	Graph Transformer + Graph Diffusion Models
5/20/2025	Lecture 12	Graph generative models		
5/26/2025			Project Q&A	
5/27/2025			Project Q&A	

Grading: 40% midterm; 60% Project

Lab sessions

- The goal is to get hands-on experience dealing with the methods discussed in the class
- Tutorials
- 3 assignments in total (in Jupyter notebooks)
- Information on Moodle
- Assignments are not graded, but highly recommended!

Projects (60% of the grade)

- Projects by groups of 4 students
  - Addressing relevant questions on selected network datasets
  - Developed during the (2nd) part of the semester
- They will be based on topics covered during the semester
- Evaluation (precise instructions will be uploaded on Moodle)
  - Methodological approach
  - Achieved performance
  - Project report
  - Quality of the code
  - Presentation

Deadlines

- Project announcement: March 17
- Midterm: April 15
- Project report deadline: June 10 (to be confirmed)
- Project presentation: June, 16-17, 2025 (to be confirmed - let us know asap in case of hard constraints)

Questions?