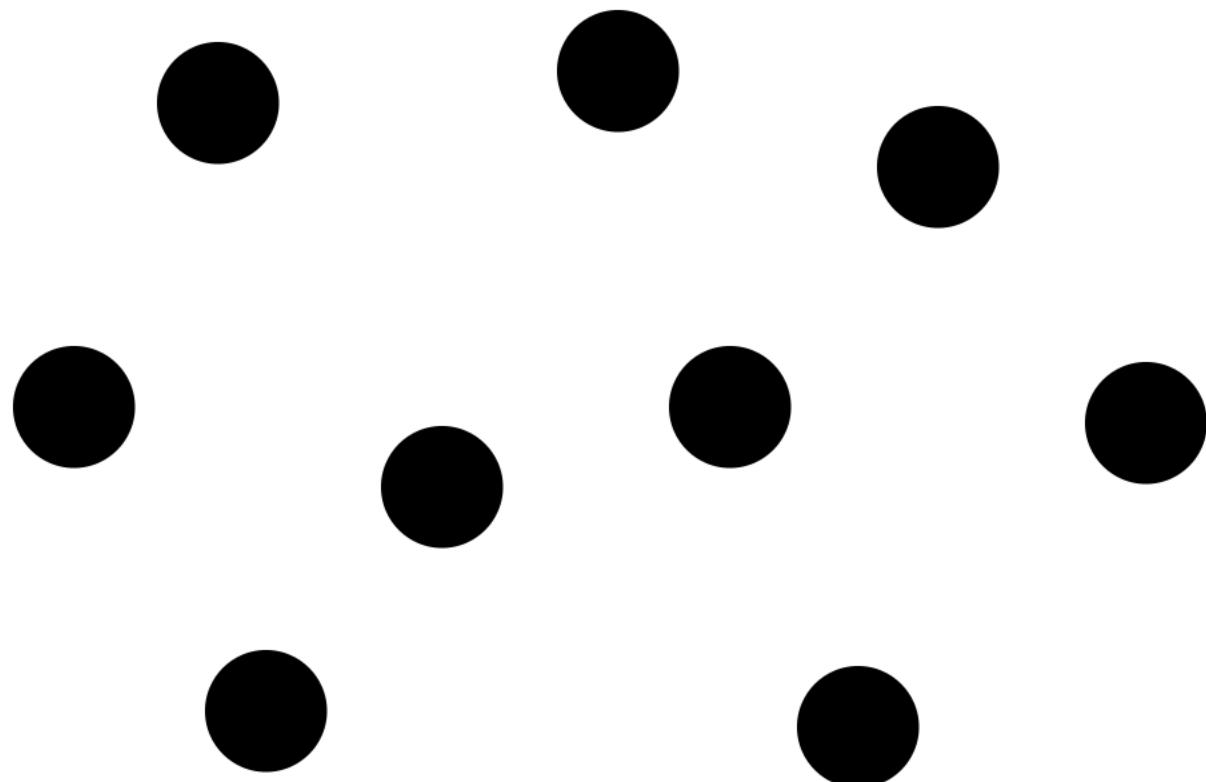
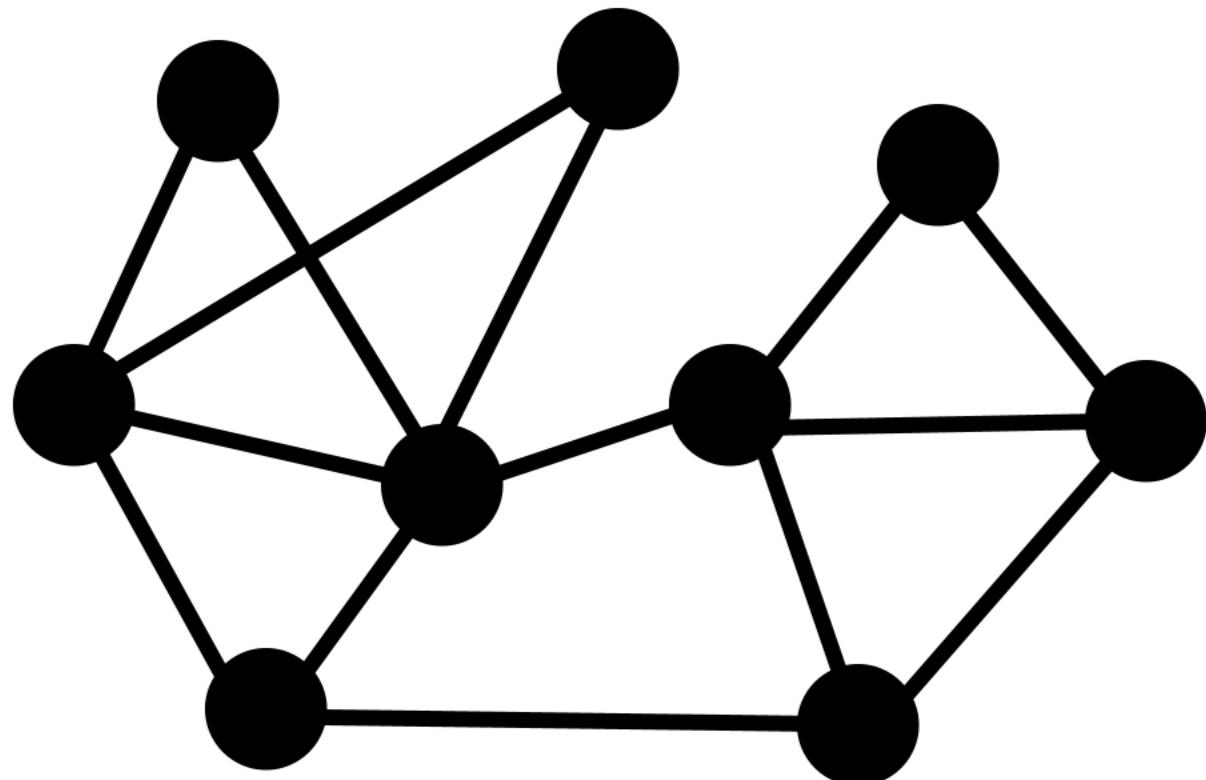


Machine Learning with Graphs

Why Graphs?

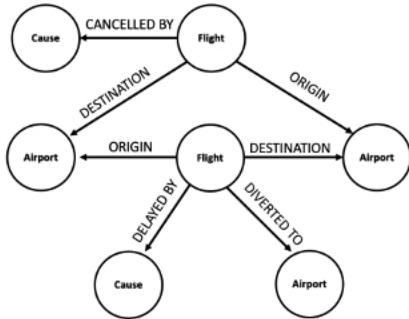
Graphs are a general language for describing and analyzing entities with relations/interactions





Graph

Many Types of Data are Graphs (1)

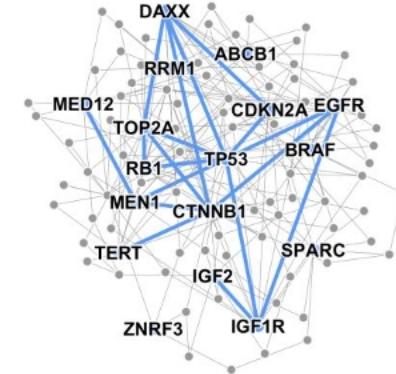


Event Graphs



Image credit: [SalientNetworks](#)

Computer Networks



Disease Pathways

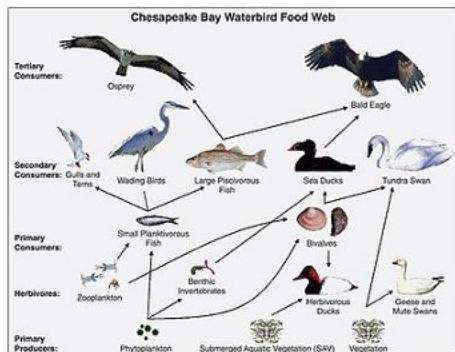


Image credit: [Wikipedia](#)

Food Webs



Image credit: [Pinterest](#)

Particle Networks



Image credit: [visitlondon.com](#)

Underground Networks

Many Types of Data are Graphs (2)



Image credit: [Medium](#)

Social Networks

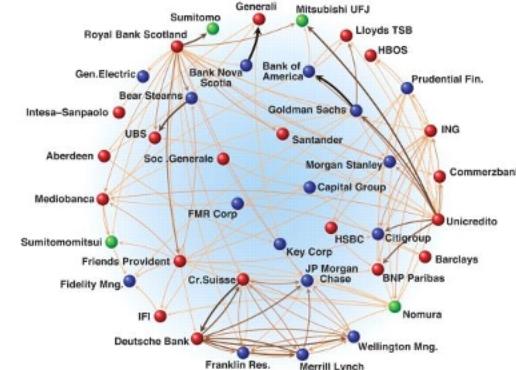


Image credit: [Science](#)



Image credit: [Lumen Learning](#)

Communication Networks



Image credit: [Missoula Current News](#)

Citation Networks

Internet

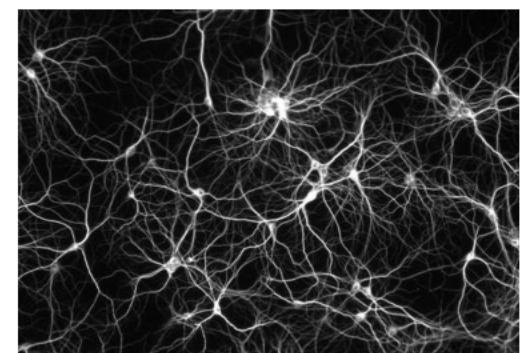


Image credit: [The Conversation](#)

Networks of Neurons

Many Types of Data are Graphs (3)

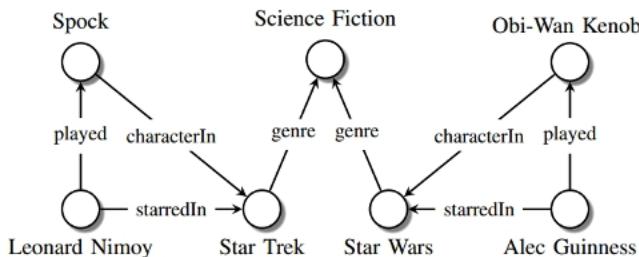


Image credit: [Maximilian Nickel et al](#)

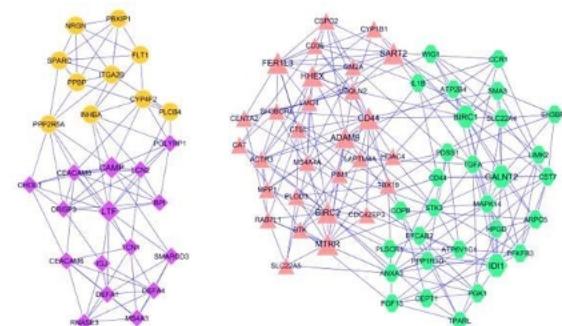


Image credit: [ese.wustl.edu](#)

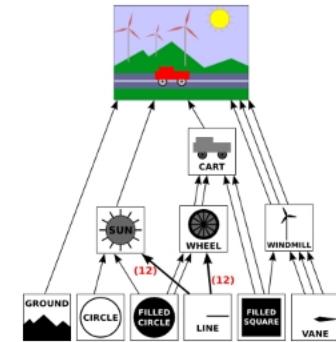


Image credit: [math.hws.edu](#)

Knowledge Graphs

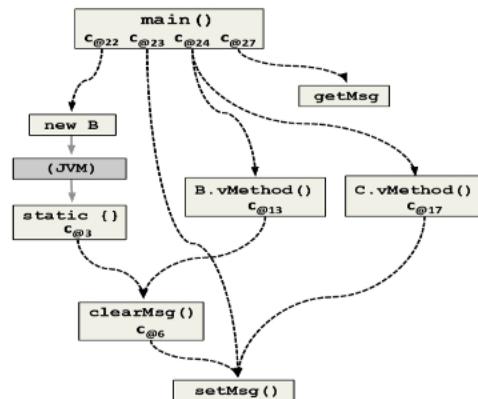


Image credit: [ResearchGate](#)

Code Graphs

Regulatory Networks

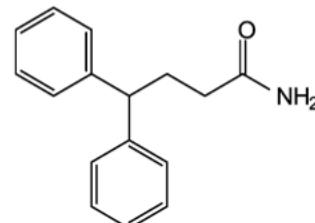


Image credit: [MDPI](#)

Molecules

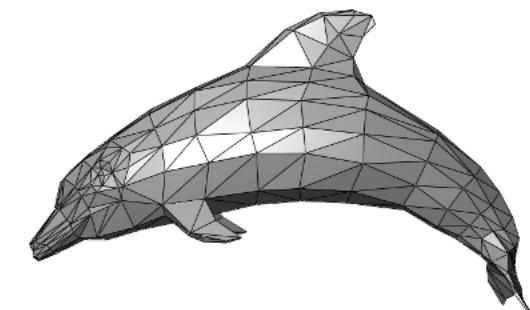


Image credit: [Wikipedia](#)

3D Shapes

Graphs and Relational Data

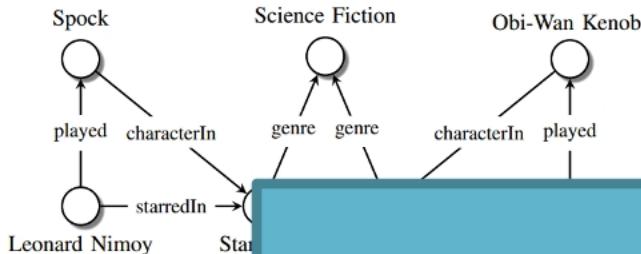


Image credit:

Known

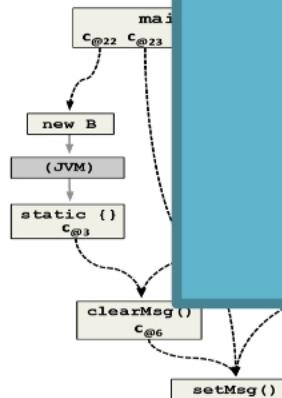


Image credit: ResearchGate

Code Graphs

Main question:

How do we take advantage of relational structure for better prediction?

Image credit: MDPI

Molecules

Image credit: MDPI

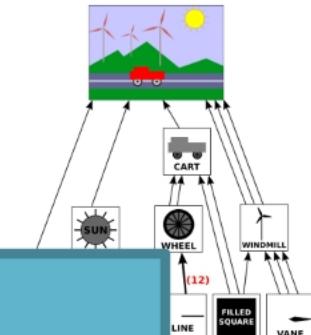


Image credit: math.hws.edu

Graphs

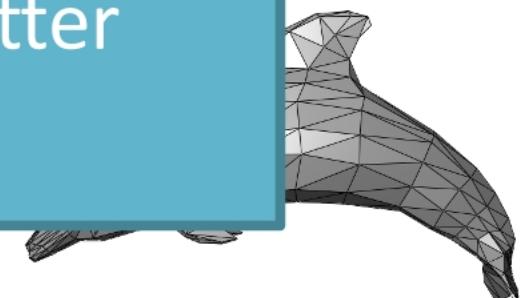


Image credit: Wikipedia

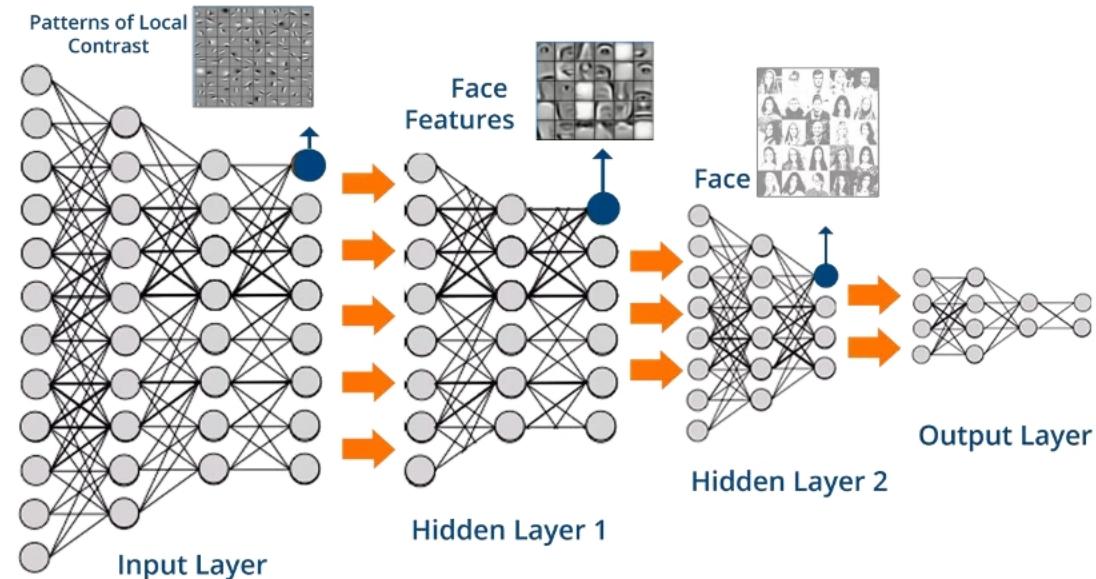
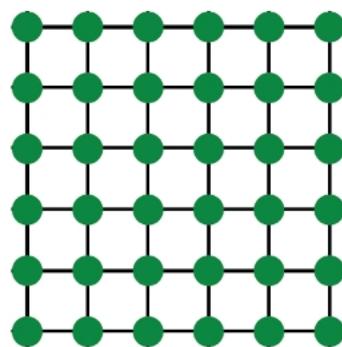
3D Shapes

Graphs: Machine Learning

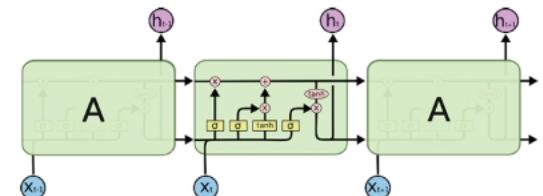
Complex domains have a rich relational structure, which can be represented as a **relational graph**

By explicitly modeling relationships we achieve better performance!

Today: Modern ML Toolbox



Text/Speech



Modern deep learning toolbox is designed
for simple sequences & grids

Doubt thou the stars are fire,
Doubt that the sun doth move;
Doubt truth to be a liar;
But never doubt I love...

Text



Audio signals



Images

Modern
deep learning toolbox
is designed for
sequences & grids

Not everything
can be represented as
a sequence or a grid

**How can we develop neural
networks that are much more
broadly applicable?**

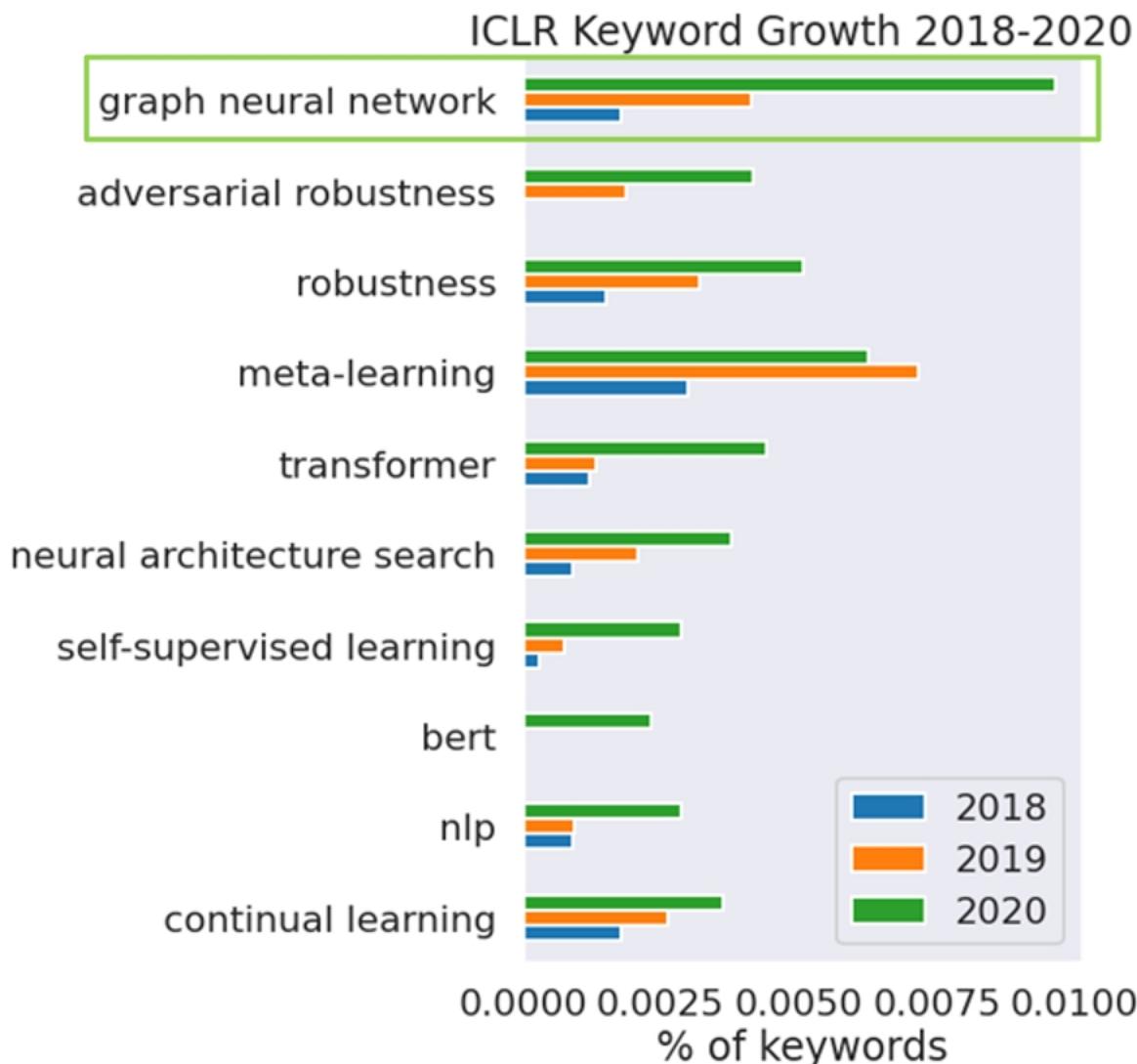
New frontiers beyond classic neural
networks that only learn on images
and sequences

This Class

Graphs are the new frontier
of deep learning

Graphs connect things.

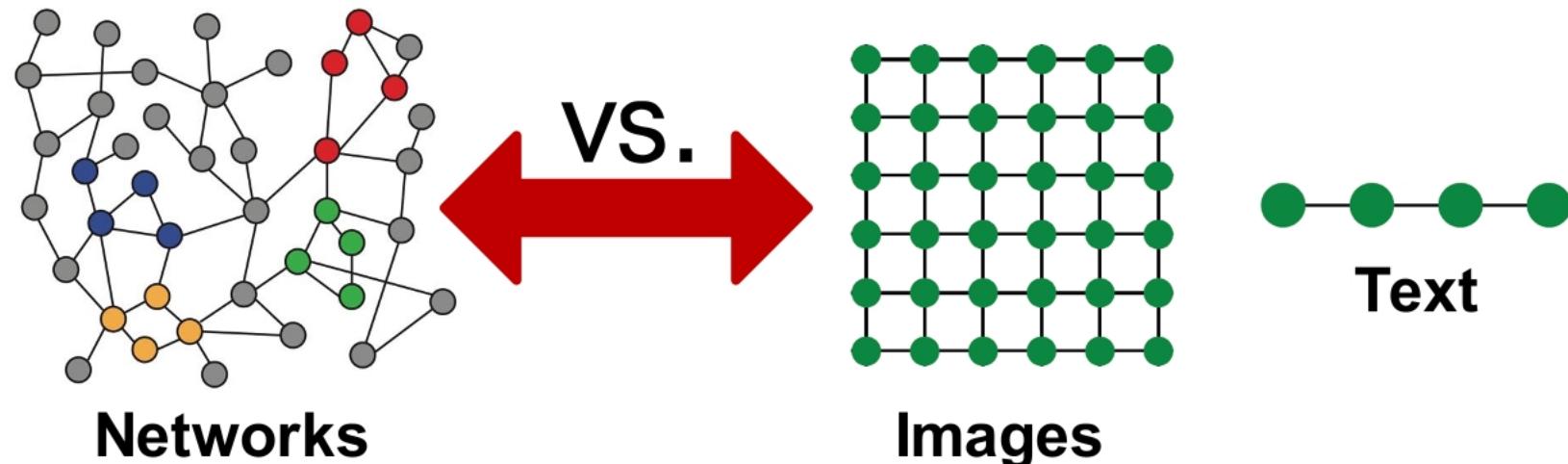
The hottest subfield in ML



Why is Graph Deep Learning Hard?

Networks are complex.

- Arbitrary size and complex topological structure (*i.e.*, no spatial locality like grids)



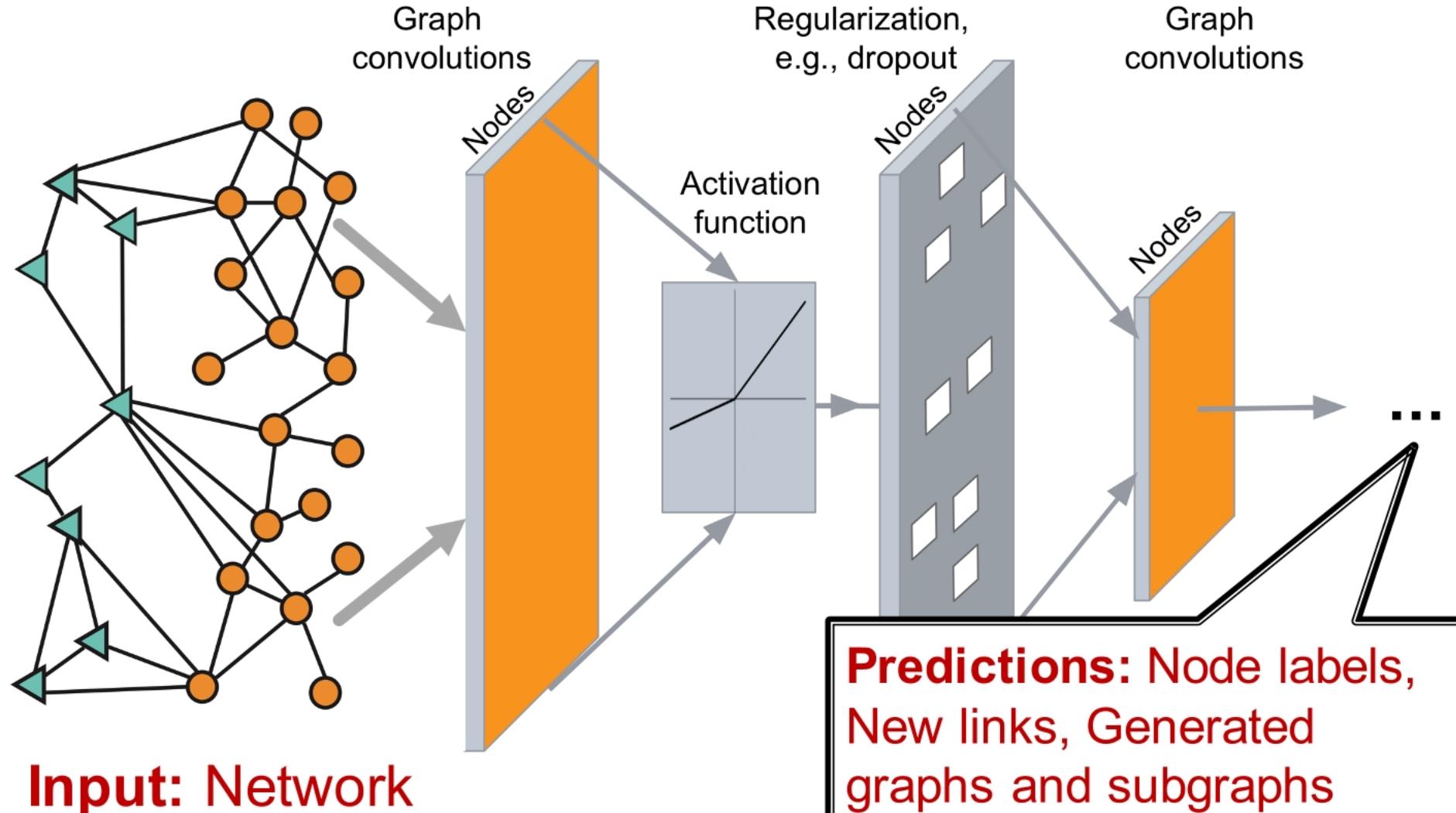
- No fixed node ordering or reference point
- Often dynamic and have multimodal features

This Course

How can we develop neural networks
that are much more broadly
applicable?

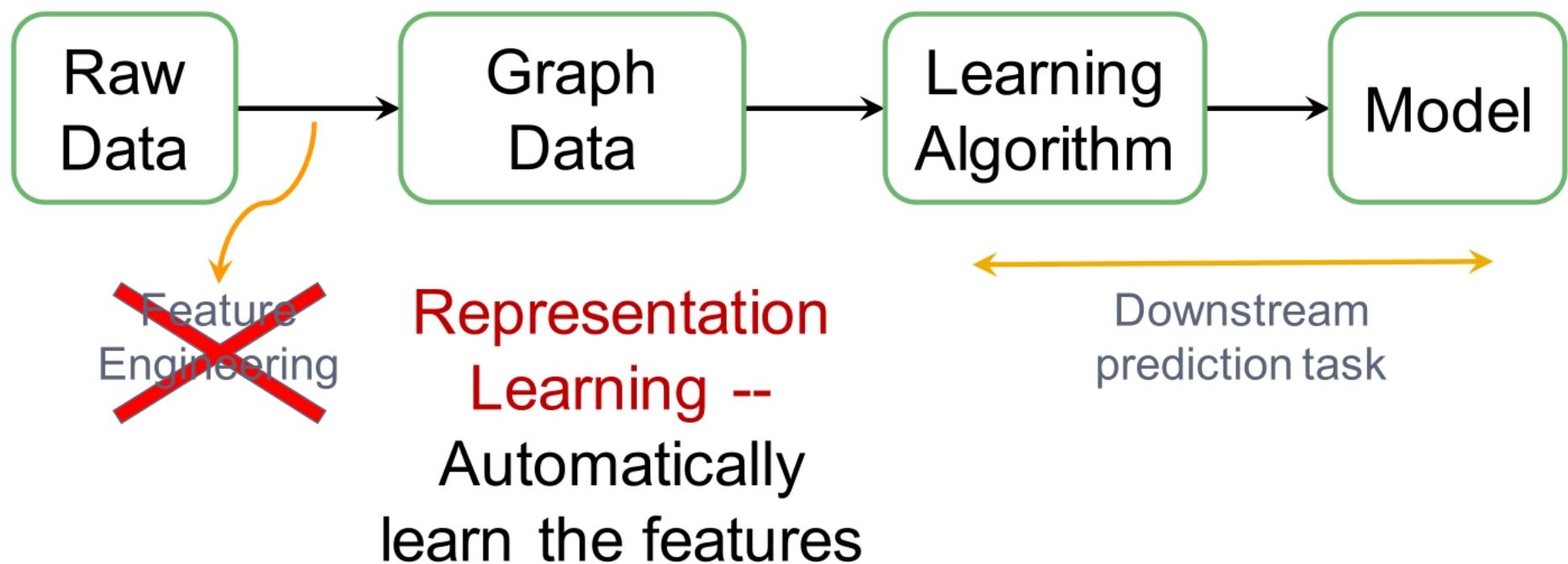
Graphs are the new frontier
of deep learning

CS224W: Deep Learning in Graphs



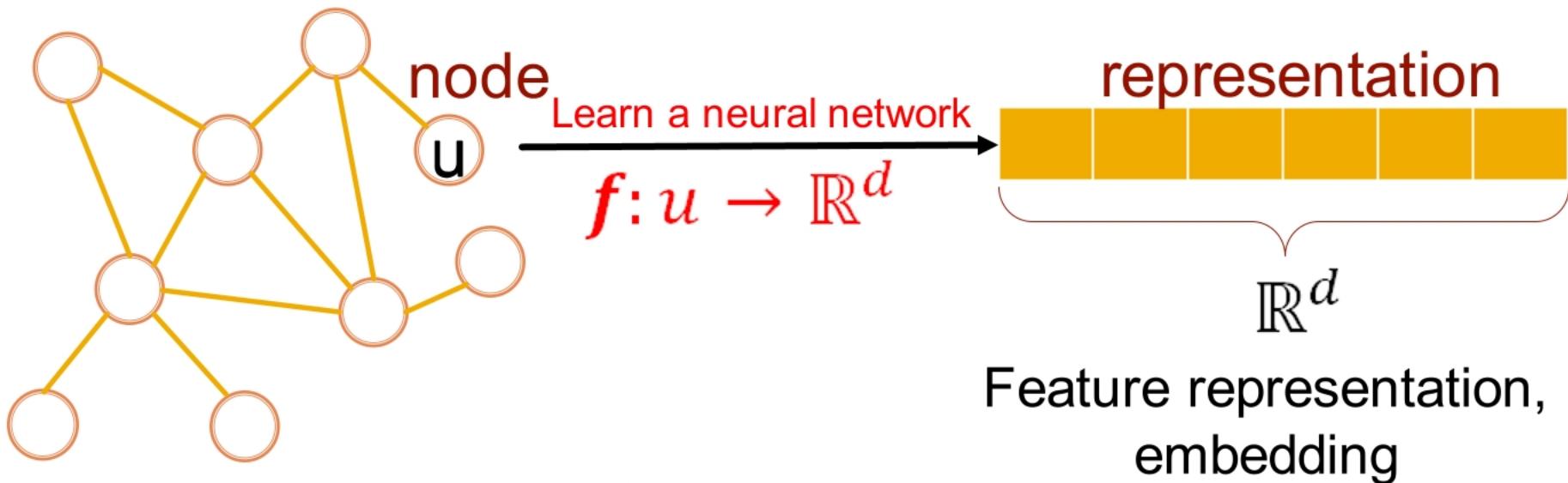
CS224W & Representation Learning

(Supervised) Machine Learning Lifecycle:
This feature, that feature. **Every single time!**



CS224W & Representation Learning

Map nodes to d-dimensional **embeddings** such that similar nodes in the network are **embedded close together**



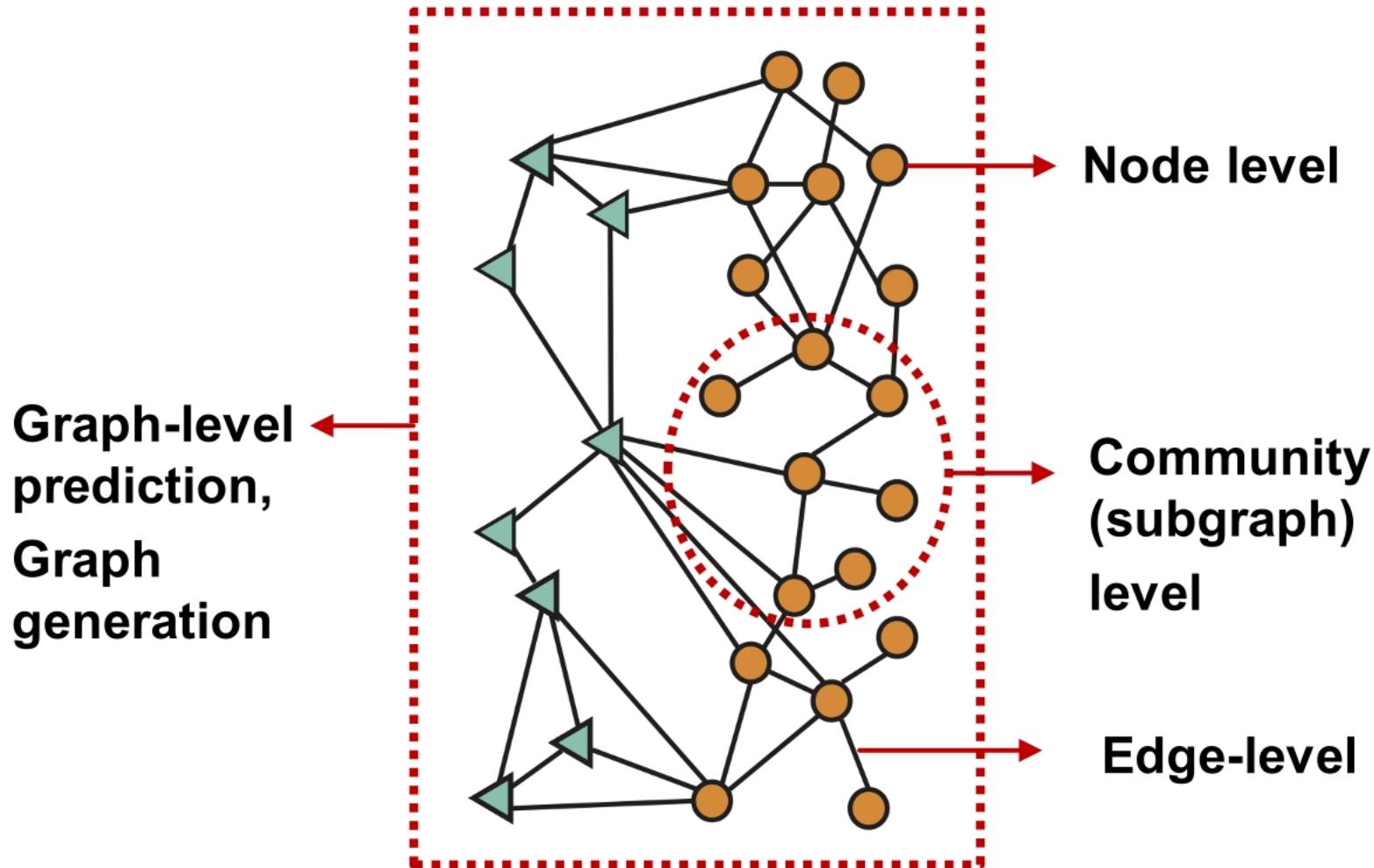
CS224W Course Outline

We are going to cover various topics in Machine Learning and Representation Learning for graph structured data:

- Traditional methods: Graphlets, Graph Kernels
- Methods for node embeddings: DeepWalk, Node2Vec
- Graph Neural Networks: GCN, GraphSAGE, GAT, Theory of GNNs
- Knowledge graphs and reasoning: TransE, BetaE
- Deep generative models for graphs: GraphRNN
- Applications to Biomedicine, Science, Industry

Applications of Graph ML

Different Types of Tasks



Classic Graph ML Tasks

- **Node classification**: Predict a property of a node
 - **Example**: Categorize online users / items
- **Link prediction**: Predict whether there are missing links between two nodes
 - **Example**: Knowledge graph completion
- **Graph classification**: Categorize different graphs
 - **Example**: Molecule property prediction
- **Clustering**: Detect if nodes form a community
 - **Example**: Social circle detection
- **Other tasks**:
 - **Graph generation**: Drug discovery
 - **Graph evolution**: Physical simulation

Classic Graph ML Tasks

- **Node classification:** Predict a property of a node
 - Example: Categorize online users / items
- **Link prediction:** Predict whether there are missing links
 - Example
- **Graph classification:** Predict a property of a graph
 - Example
- **Clustering:** Group nodes into clusters
 - Example
- **Others:**
 - **Graph generation:** Drug discovery
 - **Graph evolution:** Physical simulation

These Graph ML tasks lead to high-impact applications!

Example of Node-level ML Tasks

Example (1): Protein Folding

A protein chain acquires its native 3D structure

Every protein is made up of a sequence of amino acids bonded together

These amino acids interact locally to form shapes like helices and sheets

These shapes fold up on larger scales to form the full three-dimensional protein structure

Proteins can interact with other proteins, performing functions such as signalling and transcribing DNA

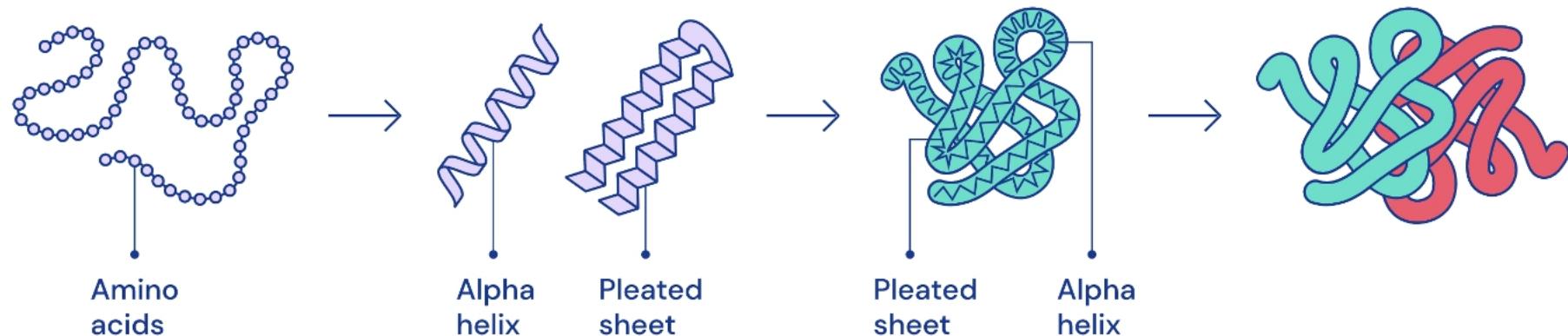
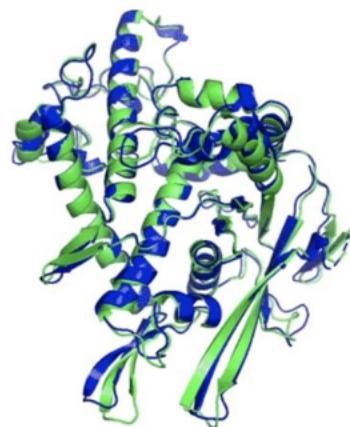


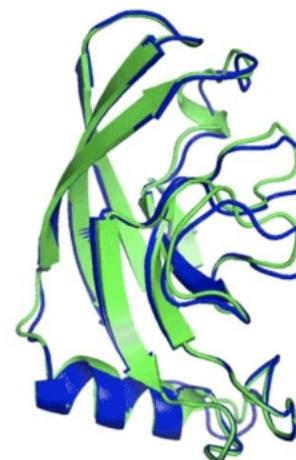
Image credit: [DeepMind](#)

The Protein Folding Problem

Computationally predict a protein's 3D structure based solely on its amino acid sequence



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

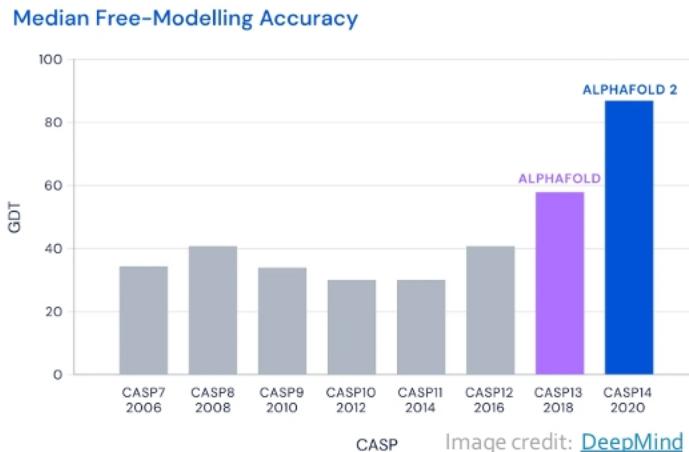


T1049 / 6y4f
93.3 GDT
(adhesin tip)

- Experimental result
- Computational prediction

Image credit: [DeepMind](#)

AlphaFold: Impact



AlphaFold's AI could change the world of biological science as we know it

DeepMind's latest AI breakthrough can accurately predict the way proteins fold

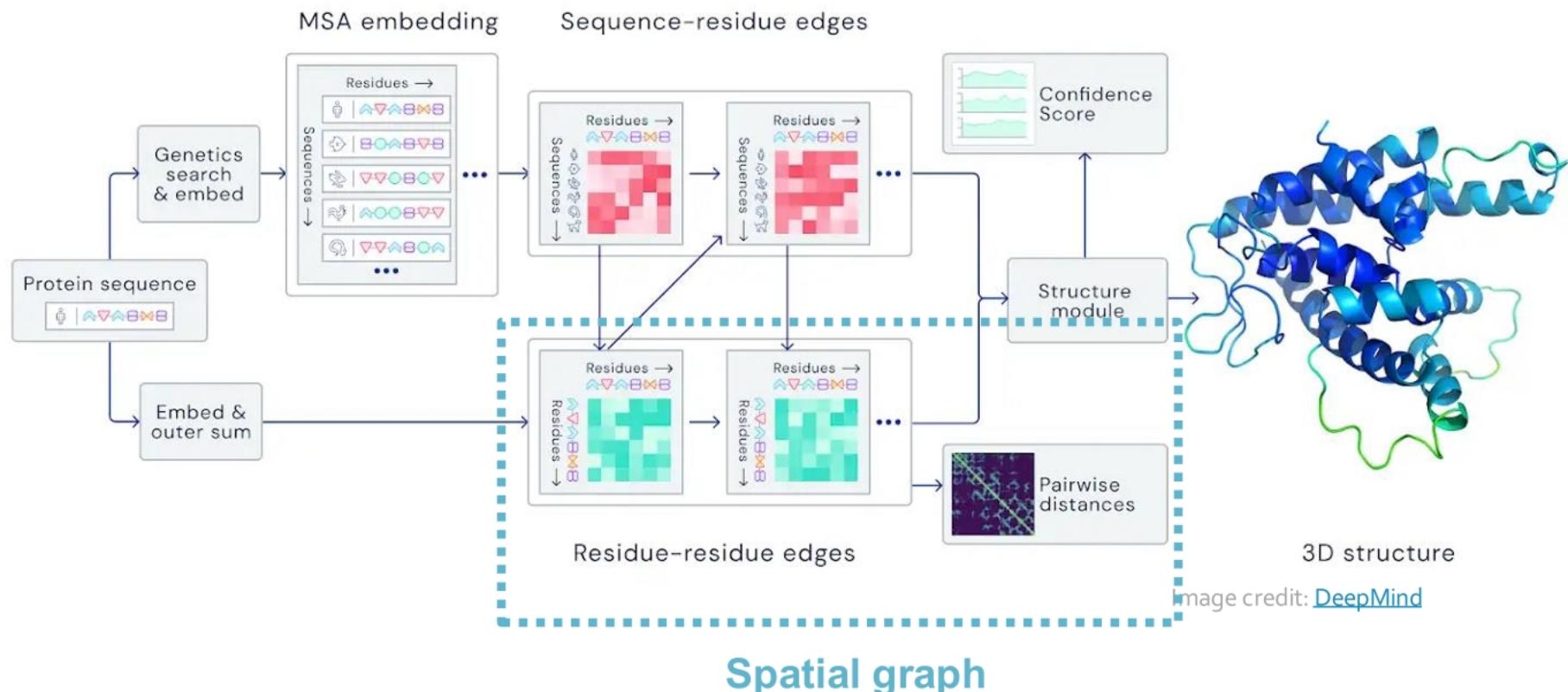
Has Artificial Intelligence 'Solved' Biology's Protein-Folding Problem?

12-14-20

DeepMind's latest AI breakthrough could turbocharge drug discovery

AlphaFold: Solving Protein Folding

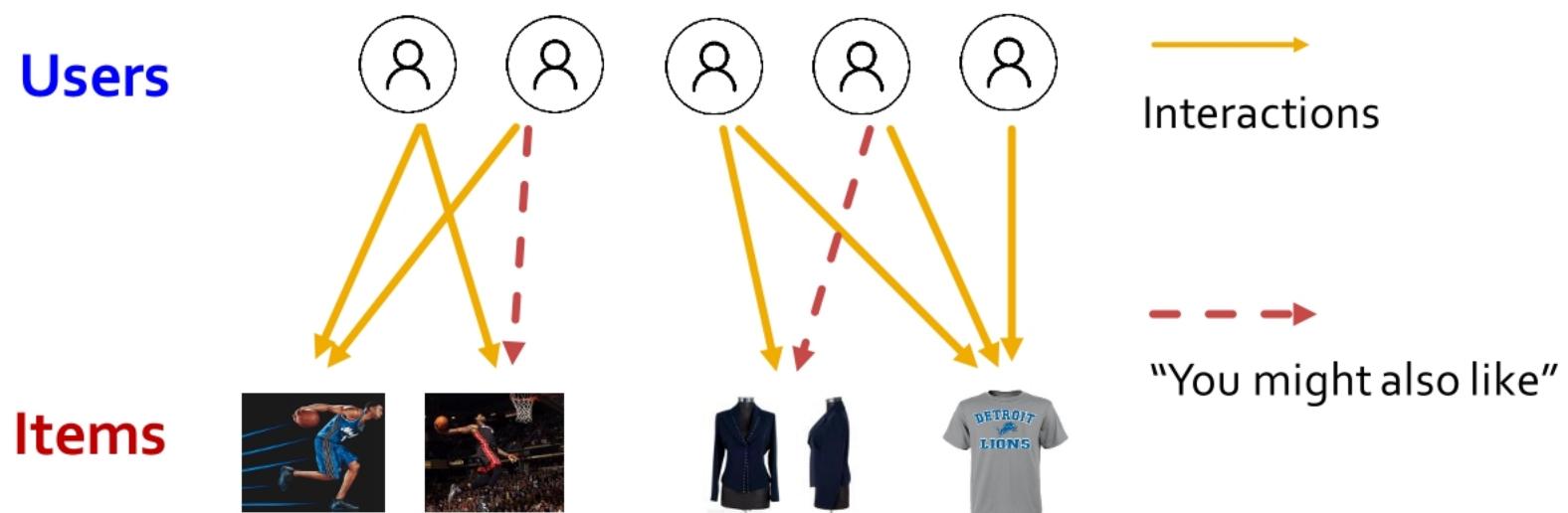
- **Key idea:** “Spatial graph”
 - **Nodes:** Amino acids in a protein sequence
 - **Edges:** Proximity between amino acids (residues)



Examples of Edge-level ML Tasks

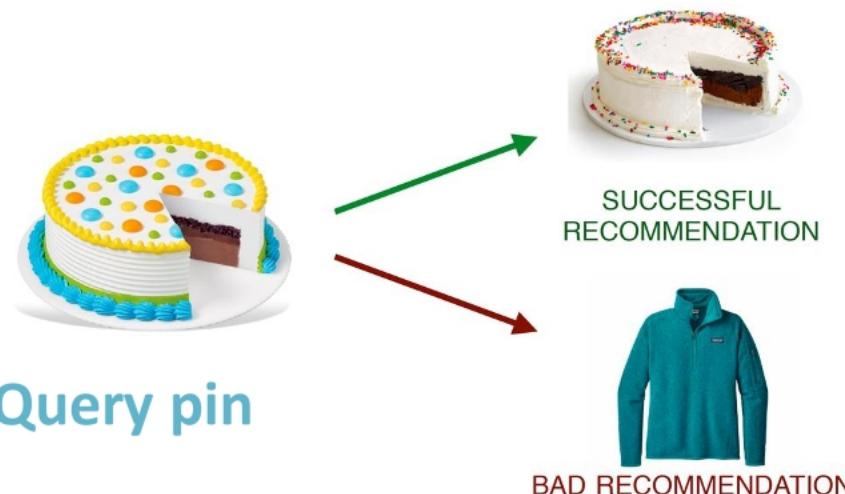
Example (2): Recommender Systems

- **Users interacts with items**
 - Watch movies, buy merchandise, listen to music
 - **Nodes:** Users and items
 - **Edges:** User-item interactions
- **Goal: Recommend items users might like**



PinSage: Graph-based Recommender

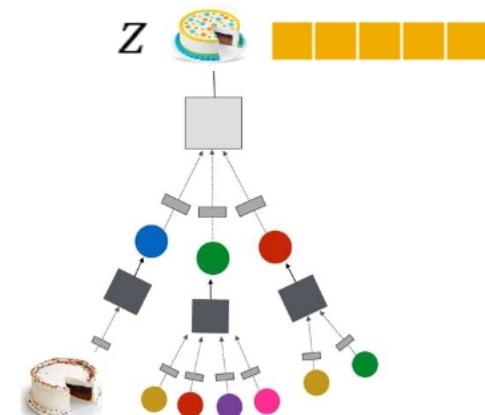
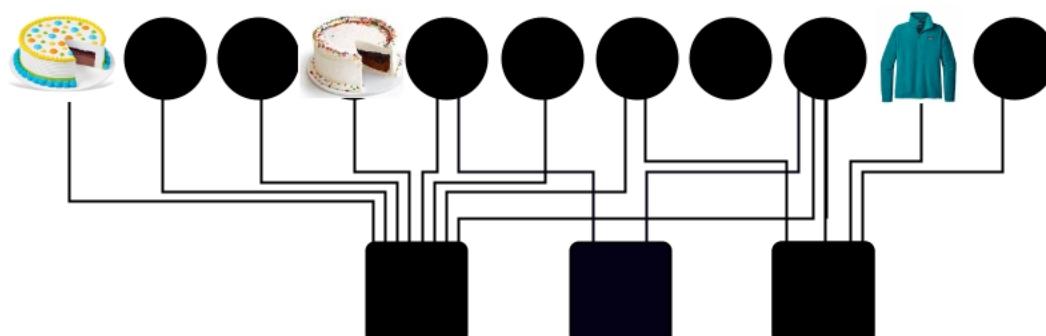
Task: Recommend related pins to users



Task: Learn node embeddings z_i such that

$$d(z_{cake1}, z_{cake2}) < d(z_{cake1}, z_{sweater})$$

Predict whether two nodes in a graph are related

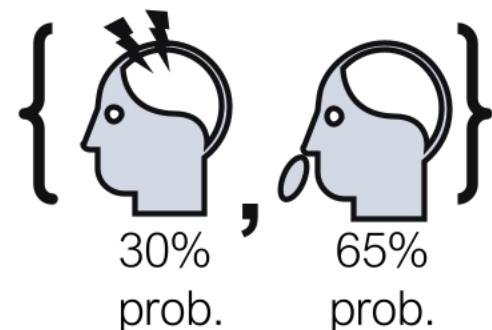
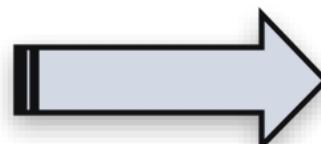


Example (3): Drug Side Effects

Many patients **take multiple drugs** to treat
complex or co-existing diseases:

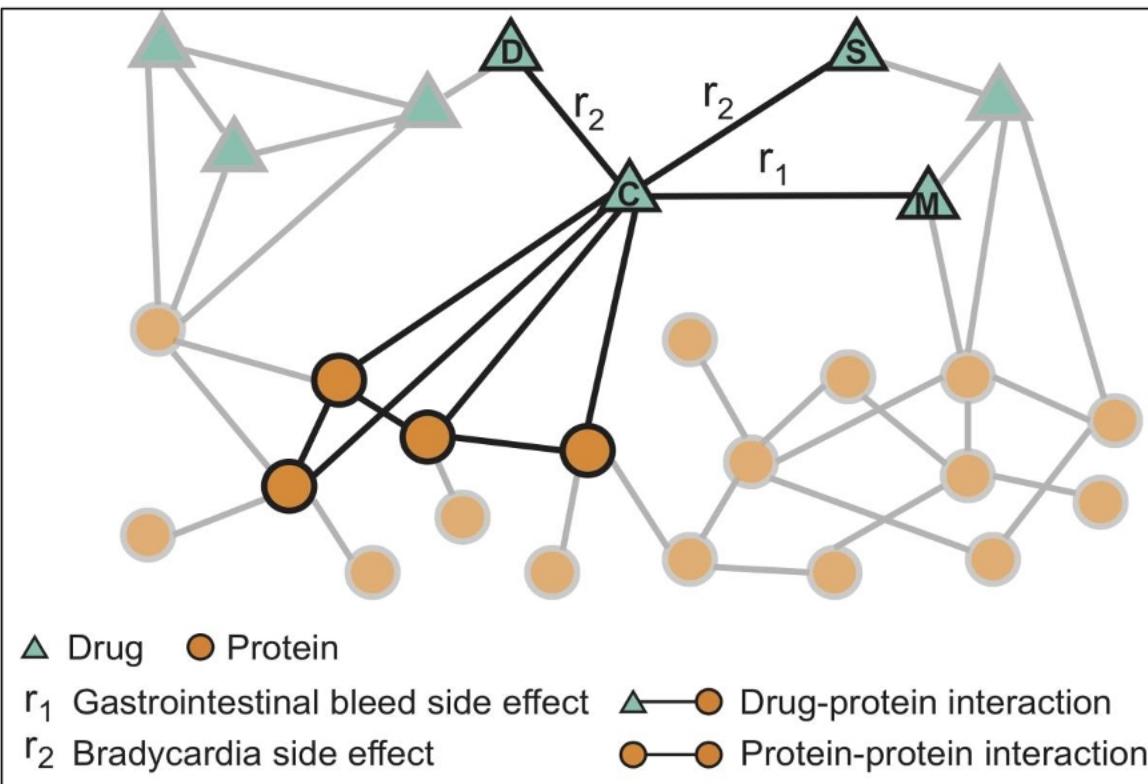
- 46% of people ages 70-79 take more than 5 drugs
- Many patients take more than 20 drugs to treat heart disease, depression, insomnia, etc.

**Task: Given a pair of drugs predict
adverse side effects**

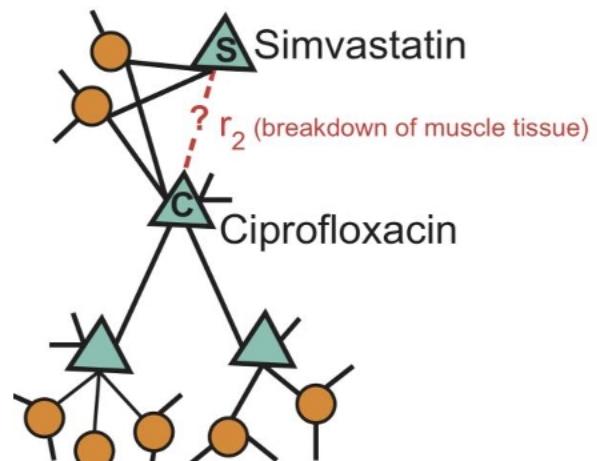


Biomedical Graph Link Prediction

- **Nodes:** Drugs & Proteins
- **Edges:** Interactions



Query: How likely will Simvastatin and Ciprofloxacin, when taken together, break down muscle tissue?



Results: *De novo* Predictions

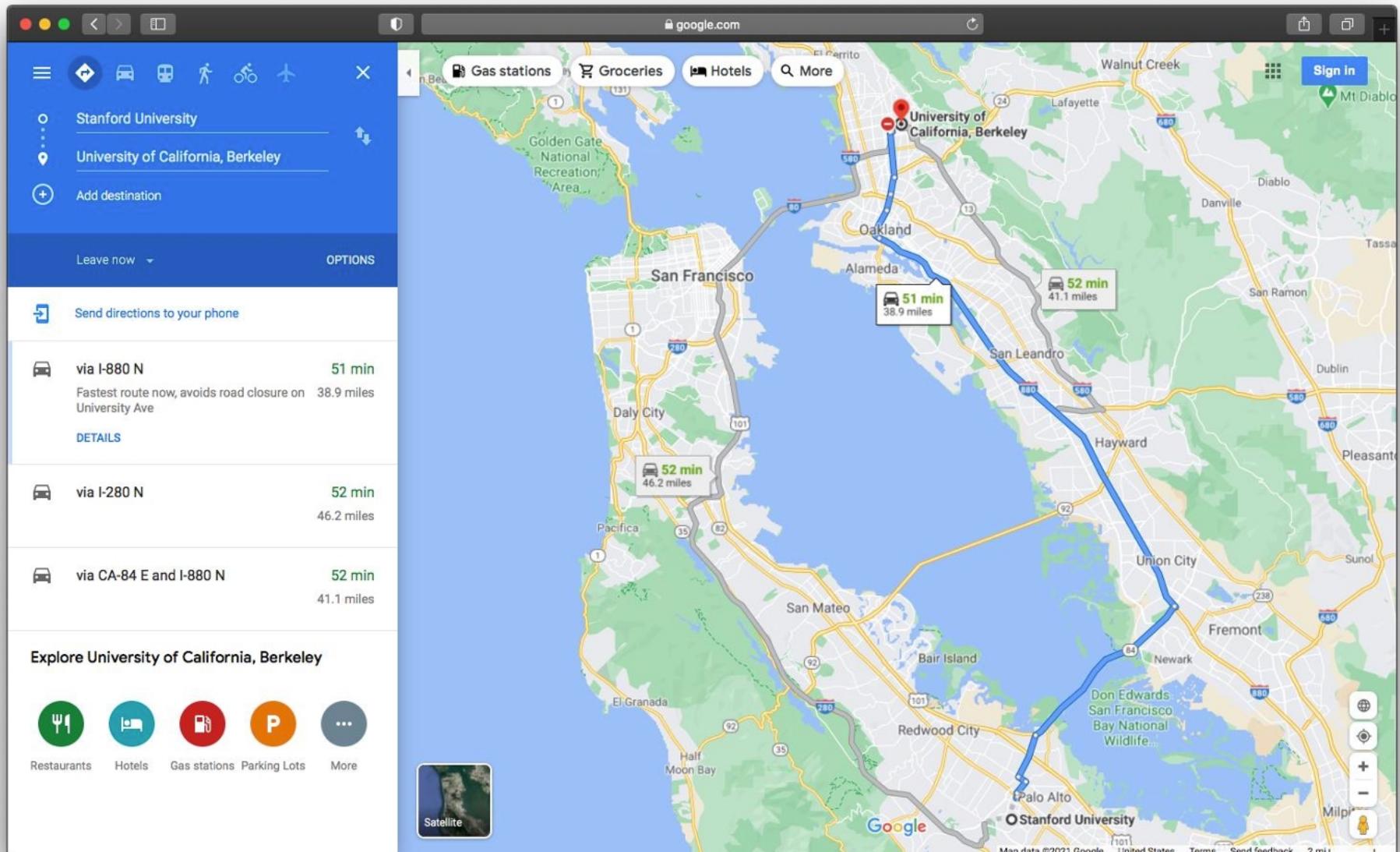
Rank	Drug c	Drug d	Side effect r	Evidence found
1	Pyrimethamine	Aliskiren	Sarcoma	Stage et al. 2015
2	Tigecycline	Bimatoprost	Autonomic neuropathy	
3	Omeprazole	Dacarbazine	Telangiectases	
4	Tolcapone	Pyrimethamine	Breast disorder	Bicker et al. 2017
5	Minoxidil	Paricalcitol	Cluster headache	
6	Omeprazole	Amoxicillin	Renal tubular acidosis	Russo et al. 2016
7	Anagrelide	Azelaic acid	Cerebral thrombosis	
8	Atorvastatin	Amlodipine	Muscle inflammation	Banakh et al. 2017
9	Aliskiren	Tioconazole	Breast inflammation	Parving et al. 2012
10	Estradiol	Nadolol	Endometriosis	

Case Report

**Severe Rhabdomyolysis due to Presumed Drug Interactions
between Atorvastatin with Amlodipine and Ticagrelor**

Examples of Subgraph-level ML Tasks

Example (4): Traffic Prediction



Road Network as a Graph

- **Nodes:** Road segments
- **Edges:** Connectivity between road segments
- **Prediction:** Time of Arrival (ETA)

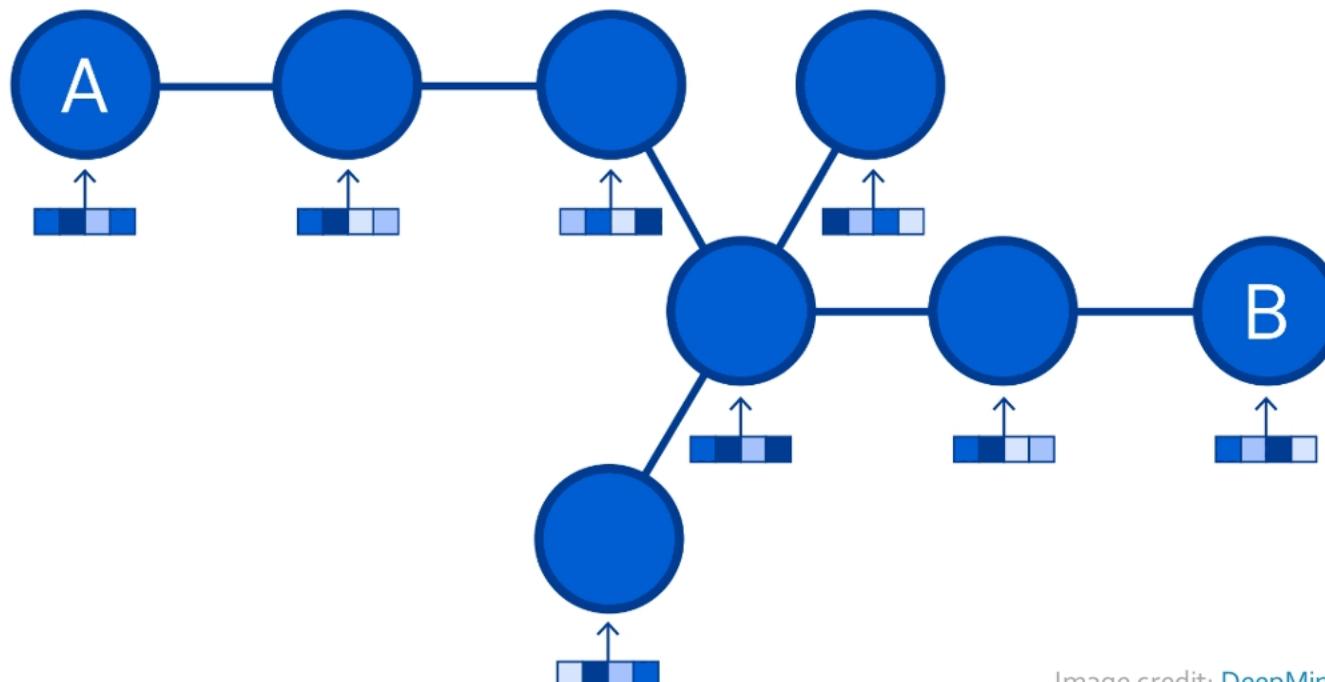
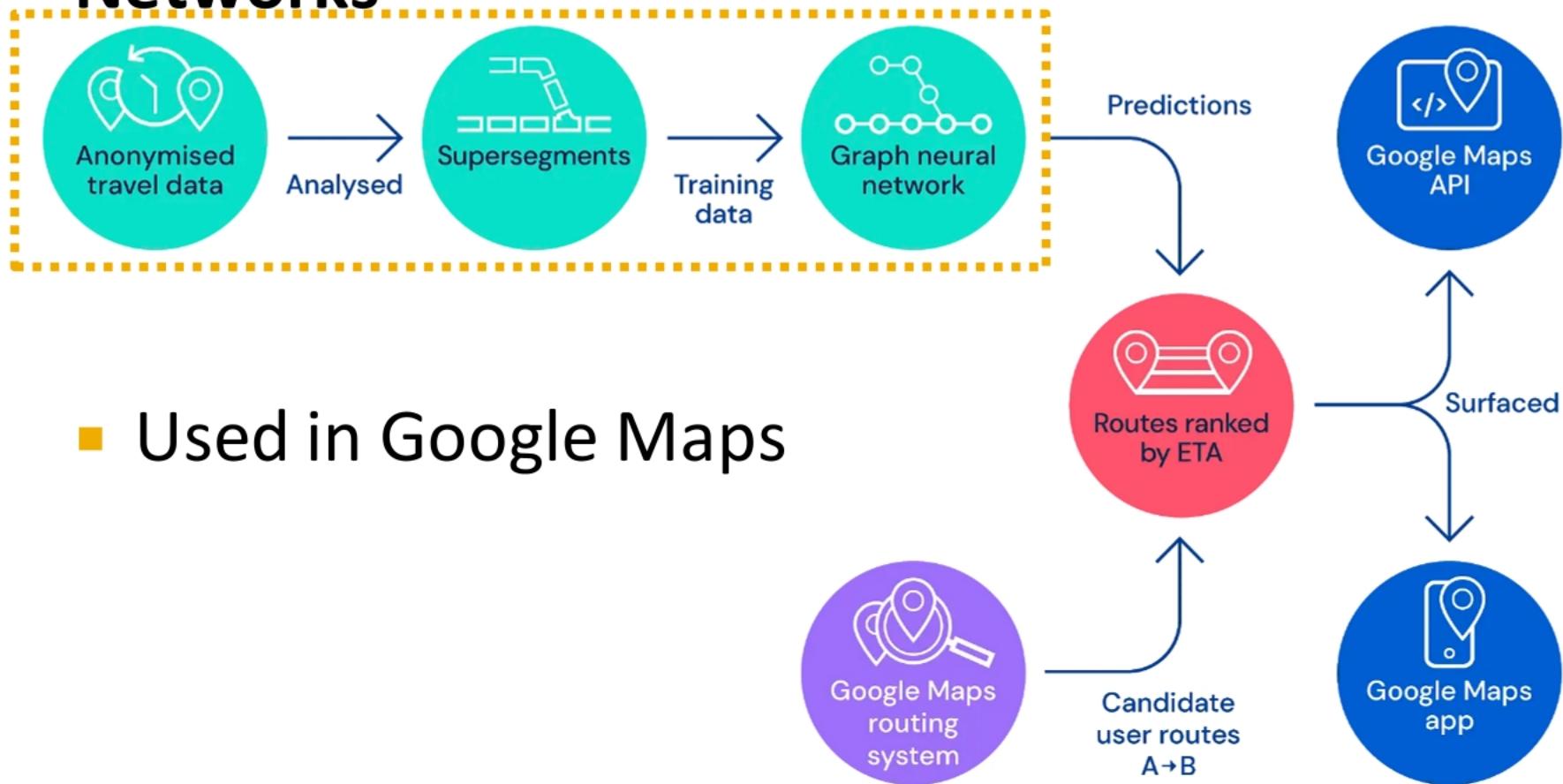


Image credit: [DeepMind](#)

Traffic Prediction via GNN

Predicting Time of Arrival with Graph Neural Networks



- Used in Google Maps

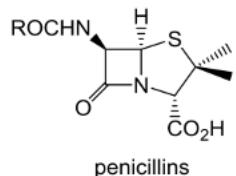
Examples of Graph-level ML Tasks

Example (5): Drug Discovery

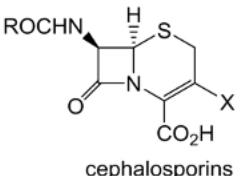
■ Antibiotics are small molecular graphs

- **Nodes:** Atoms

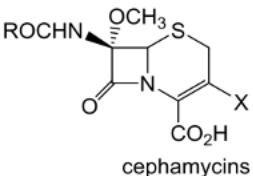
- **Edges:** Chemical bonds



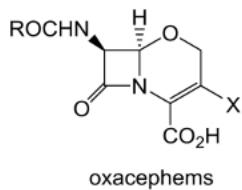
penicillins



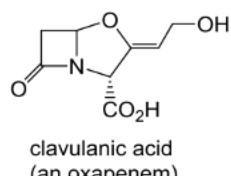
cephalosporins



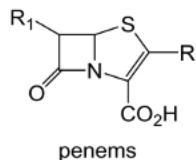
cephamycins



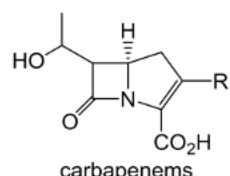
oxacephems



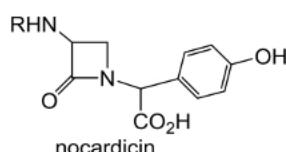
clavulanic acid
(an oxapenem)



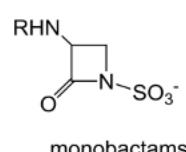
penems



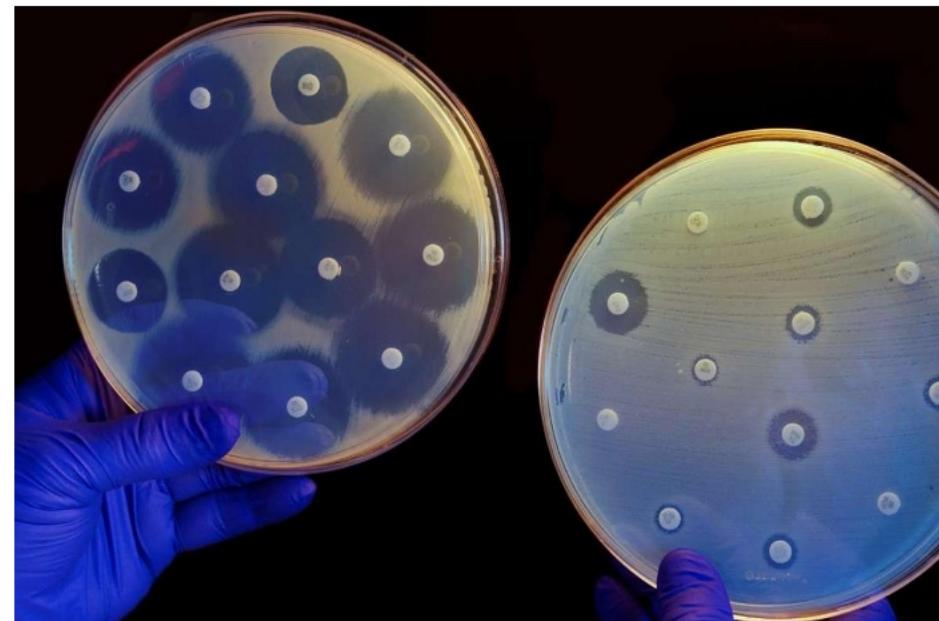
carbapenems



nocardicin



monobactams

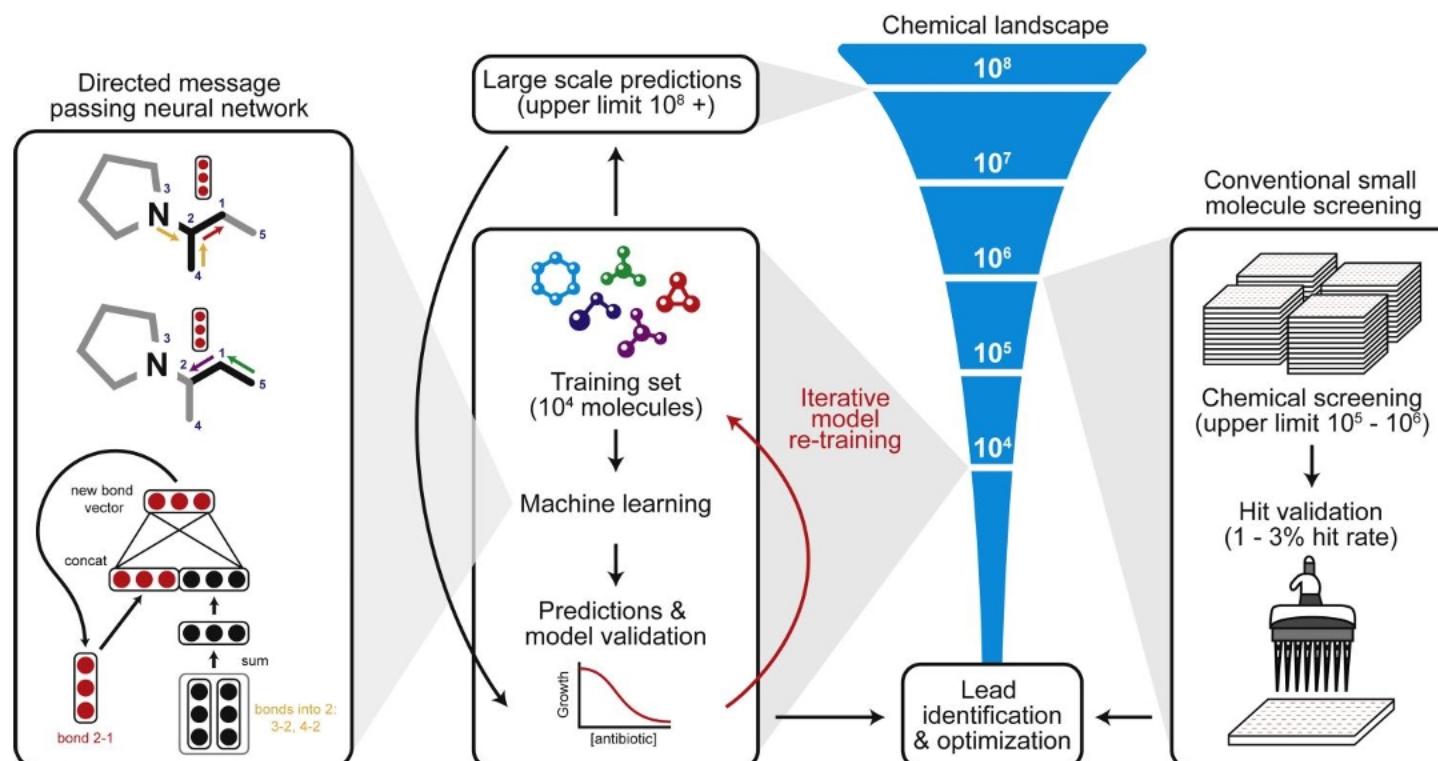


Konaklieva, Monika I. "Molecular targets of β -lactam-based antimicrobials: beyond the usual suspects." *Antibiotics* 3.2 (2014): 128-142.

Image credit: [CNN](#)

Deep Learning for Antibiotic Discovery

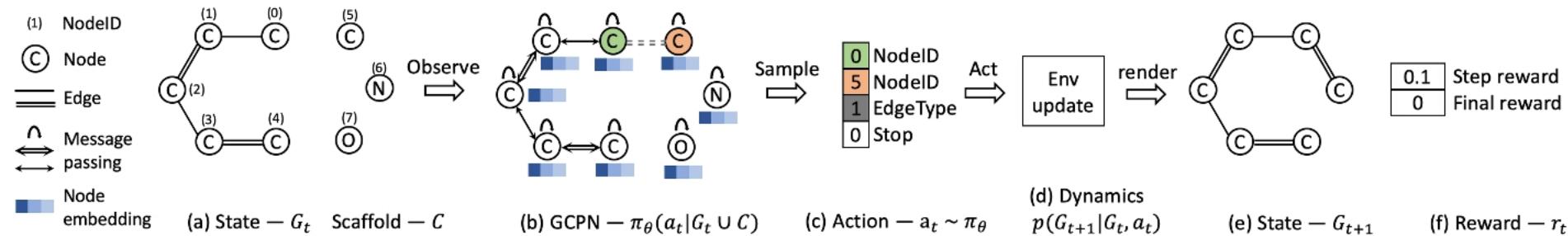
- A Graph Neural Network **graph classification model**
- Predict promising molecules from a pool of candidates



Stokes, Jonathan M., et al. "A deep learning approach to antibiotic discovery." Cell 180.4 (2020): 688-702.

Molecule Generation / Optimization

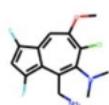
Graph generation: Generating novel molecules



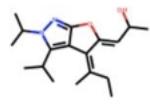
Use case 1: Generate novel molecules with high Drug likeness value



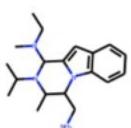
0.948



0.945



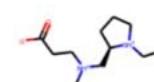
0.944



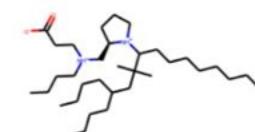
0.941

Drug likeness

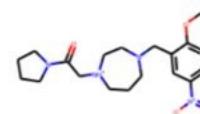
Use case 2: Optimize existing molecules to have desirable properties



-8.32



-0.71



-5.55

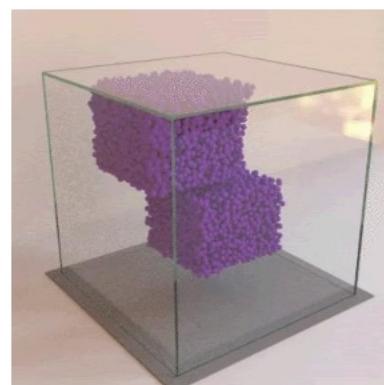
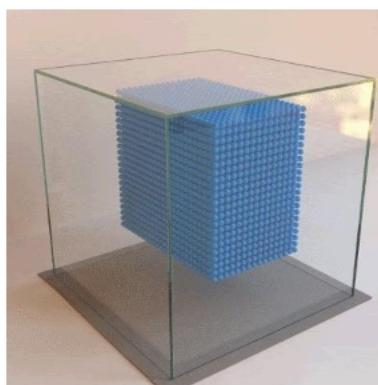


-1.78

Example (6): Physics Simulation

Physical simulation as a graph:

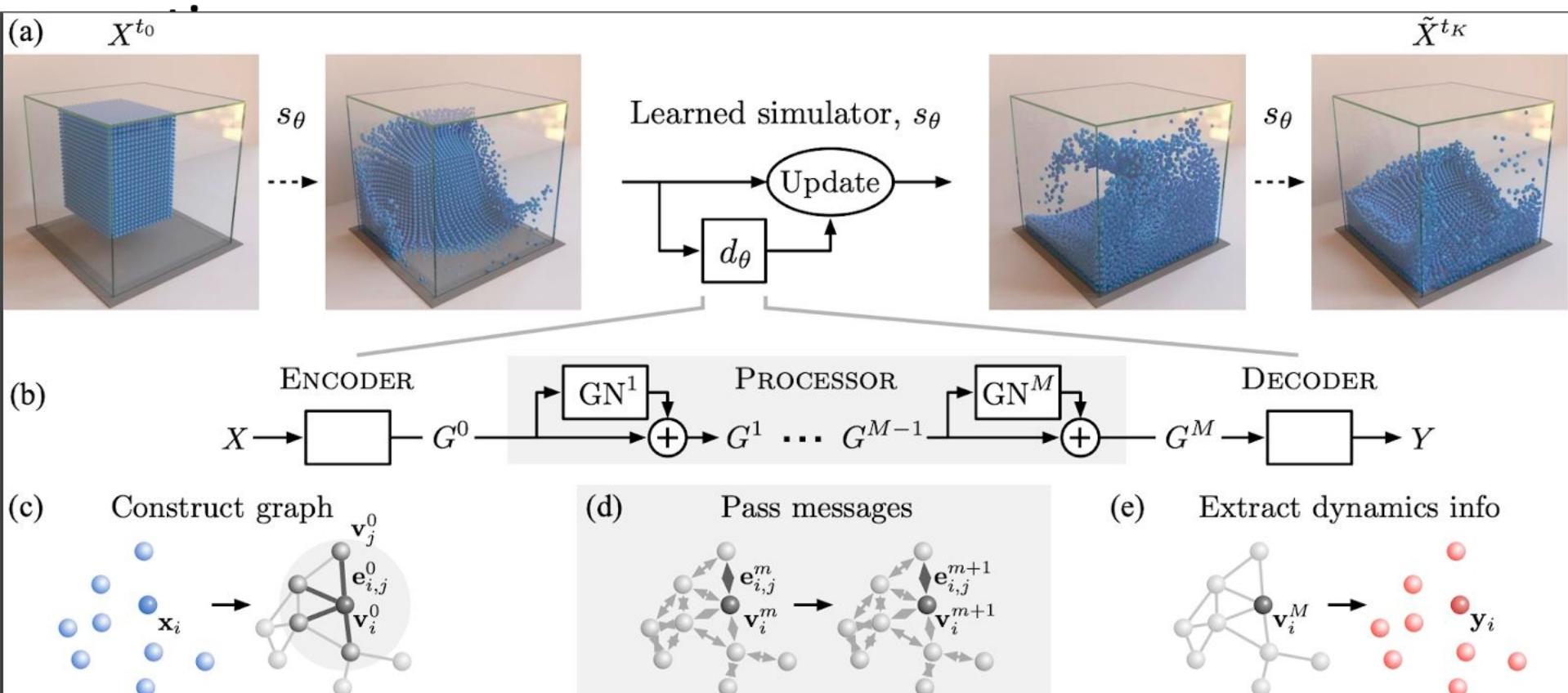
- **Nodes:** Particles
- **Edges:** Interaction between particles



Simulation Learning Framework

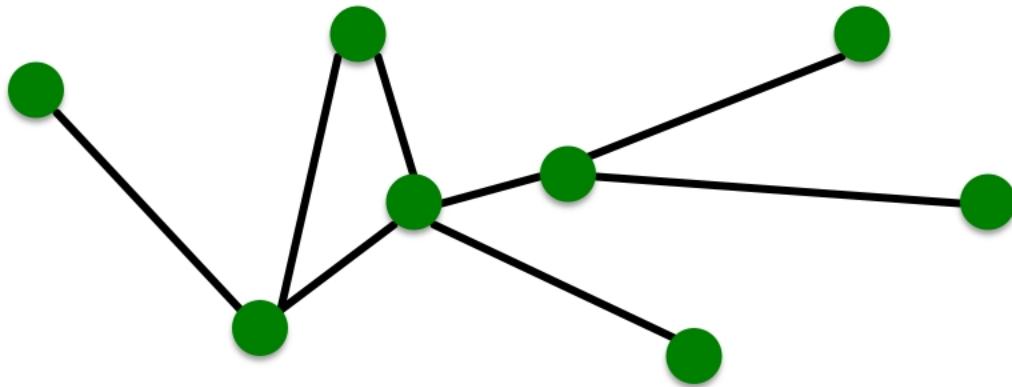
A graph evolution task:

- **Goal:** Predict how a graph will evolve over time



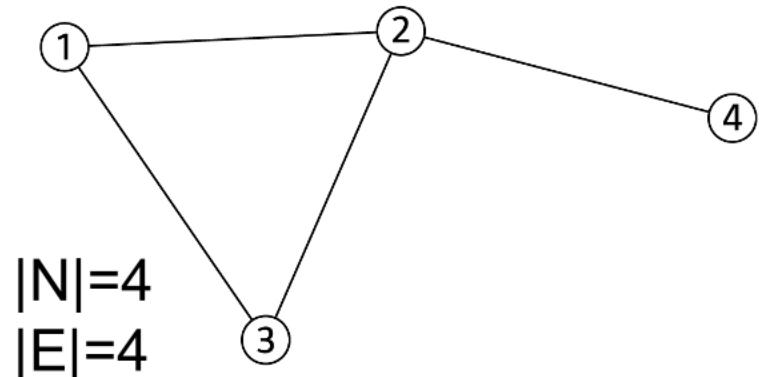
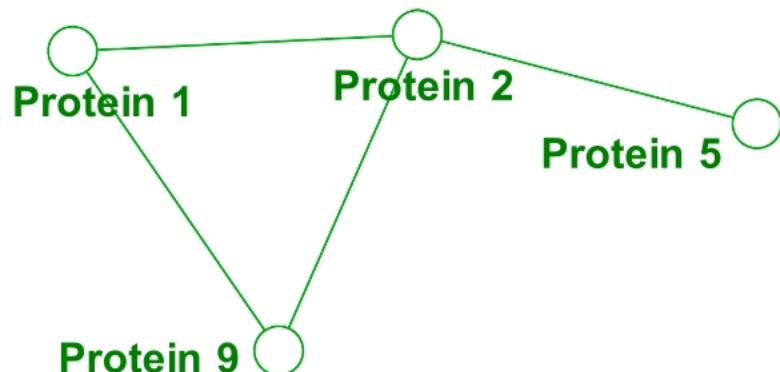
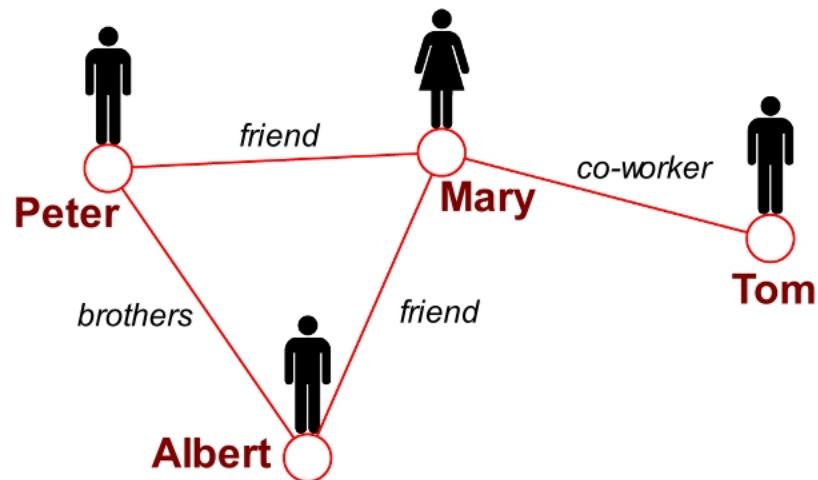
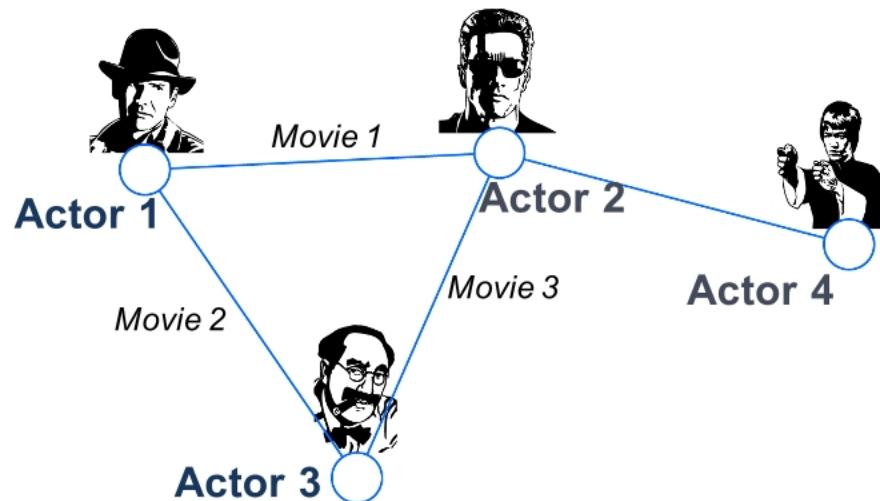
Choice of Graph Representation

Components of a Network



- **Objects:** nodes, vertices N
- **Interactions:** links, edges E
- **System:** network, graph $G(N,E)$

Graphs: A Common Language



Choosing a Proper Representation

- If you connect individuals that work with each other, you will explore a **professional network**
- If you connect those that have a sexual relationship, you will be exploring **sexual networks**
- If you connect scientific papers that cite each other, you will be studying the **citation network**
- **If you connect all papers with the same word in the title, what will you be exploring?** It is a network, nevertheless



Image credit: [Euro Scientists](#)

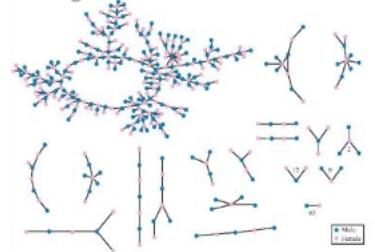
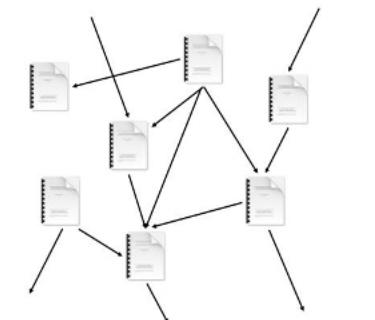


Image credit: [ResearchGate](#)



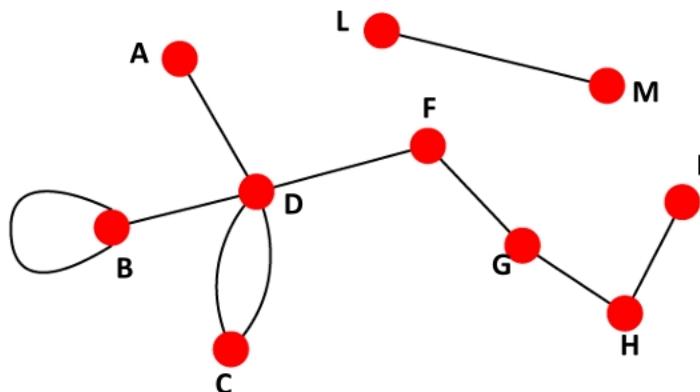
How do you define a graph?

- **How to build a graph:**
 - What are nodes?
 - What are edges?
- **Choice of the proper network representation of a given domain/problem determines our ability to use networks successfully:**
 - In some cases, there is a unique, unambiguous representation
 - In other cases, the representation is by no means unique
 - The way you assign links will determine the nature of the question you can study

Directed vs. Undirected Graphs

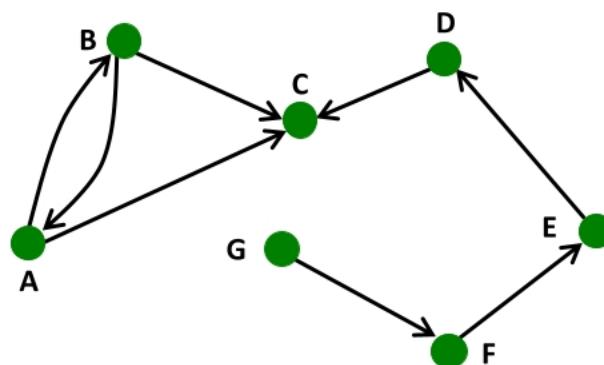
Undirected

- Links: undirected
(symmetrical, reciprocal)



Directed

- Links: directed
(arcs)



Examples:

- Collaborations
- Friendship on Facebook

Examples:

- Phone calls
- Following on Twitter

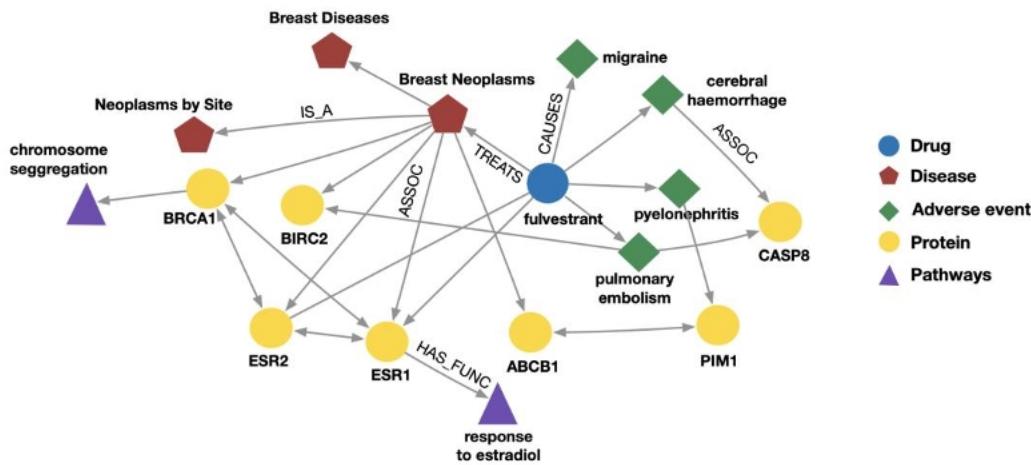
Heterogeneous Graphs

- A heterogeneous graph is defined as

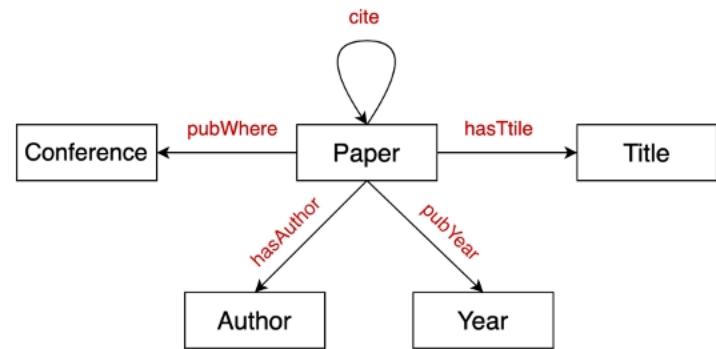
$$G = (V, E, R, T)$$

- Nodes with node types $v_i \in V$
- Edges with relation types $(v_i, r, v_j) \in E$
- Node type $T(v_i)$
- Relation type $r \in R$

Many Graphs are Heterogeneous Graphs



- Drug
- Disease
- Adverse event
- Protein
- Pathways



Biomedical Knowledge Graphs

Example node: Migraine

Example edge: (fulvestrant, Treats, Breast Neoplasms)

Example node type: Protein

Example edge type (relation): Causes

Academic Graphs

Example node: ICML

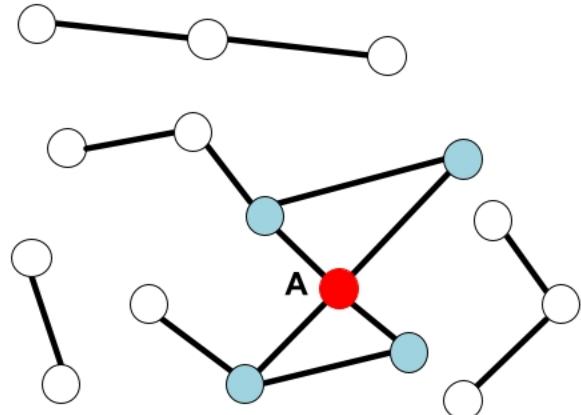
Example edge: (GraphSAGE, NeurIPS)

Example node type: Author

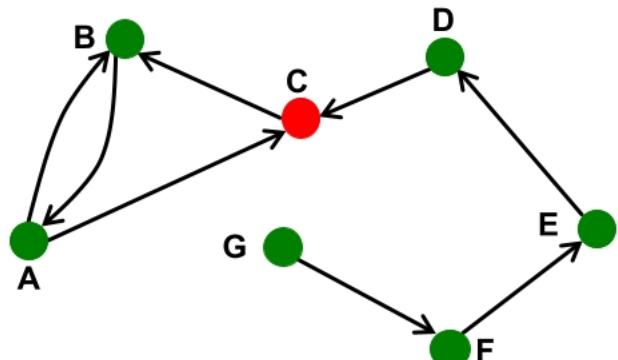
Example edge type (relation): pubYear

Node Degrees

Undirected



Directed



Source: Node with $k^{in} = 0$

Sink: Node with $k^{out} = 0$

Node degree, k_i : the number of edges adjacent to node i

$$k_A = 4$$

Avg. degree: $\bar{k} = \langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i = \frac{2E}{N}$

In directed networks we define an **in-degree** and **out-degree**. The (total) degree of a node is the sum of in- and out-degrees.

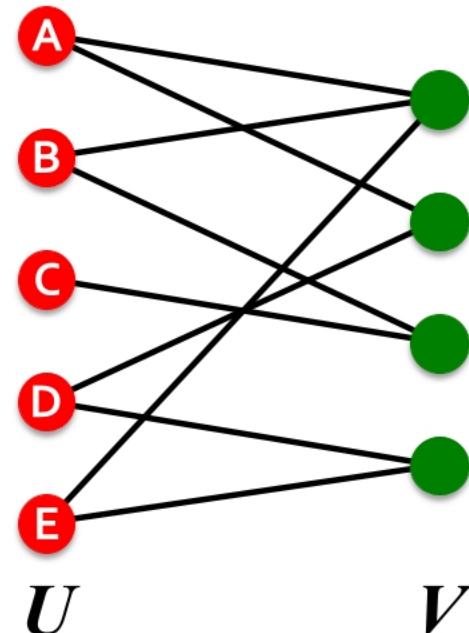
$$k_C^{in} = 2 \quad k_C^{out} = 1 \quad k_C = 3$$

$$\bar{k} = \frac{E}{N}$$

$$\bar{k}^{in} = \bar{k}^{out}$$

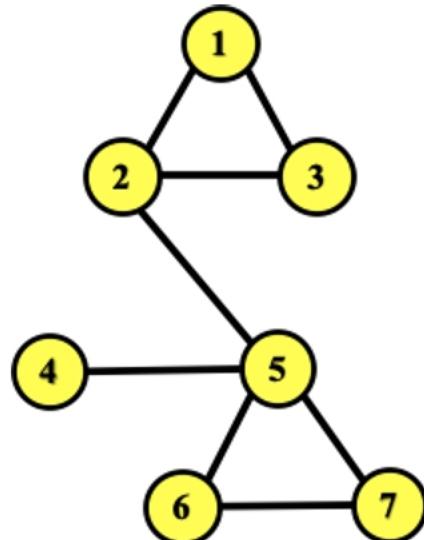
Bipartite Graph

- **Bipartite graph** is a graph whose nodes can be divided into two disjoint sets U and V such that every link connects a node in U to one in V ; that is, U and V are **independent sets**
- **Examples:**
 - Authors-to-Papers (they authored)
 - Actors-to-Movies (they appeared in)
 - Users-to-Movies (they rated)
 - Recipes-to-Ingredients (they contain)
- **“Folded” networks:**
 - Author collaboration networks
 - Movie co-rating networks

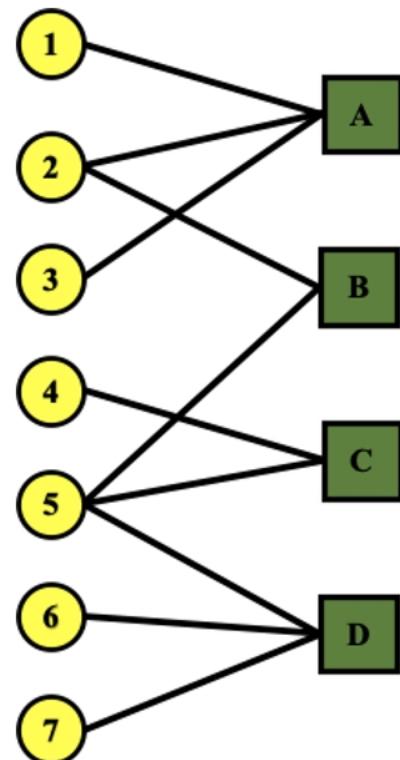


Folded/Projected Bipartite Graphs

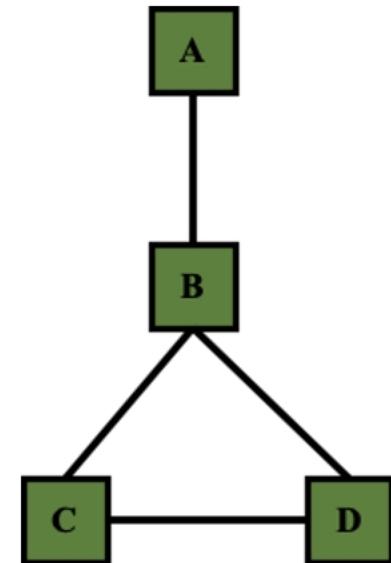
Projection U



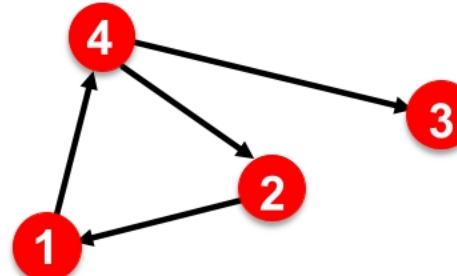
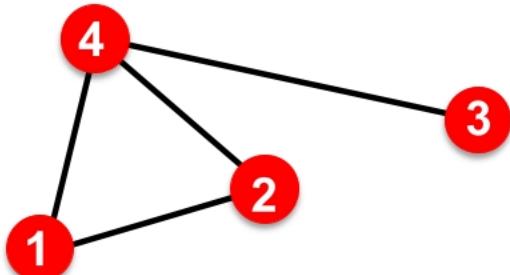
U V



Projection V



Representing Graphs: Adjacency Matrix



$A_{ij} = 1$ if there is a link from node i to node j

$A_{ij} = 0$ otherwise

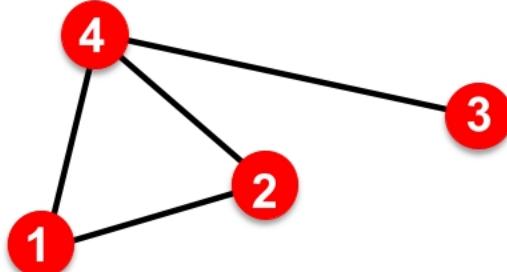
$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Note that for a directed graph (right) the matrix is not symmetric.

Adjacency Matrix

Undirected



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

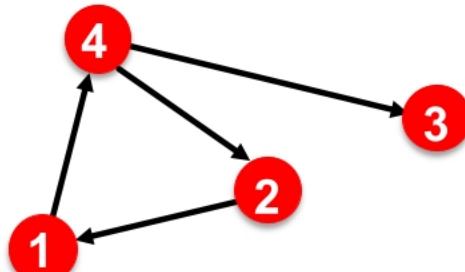
$$\begin{aligned} A_{ij} &= A_{ji} \\ A_{ii} &= 0 \end{aligned}$$

$$k_i = \sum_{j=1}^N A_{ij}$$

$$k_j = \sum_{i=1}^N A_{ij}$$

$$L = \frac{1}{2} \sum_{i=1}^N k_i = \frac{1}{2} \sum_{ij}^N A_{ij}$$

Directed



$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

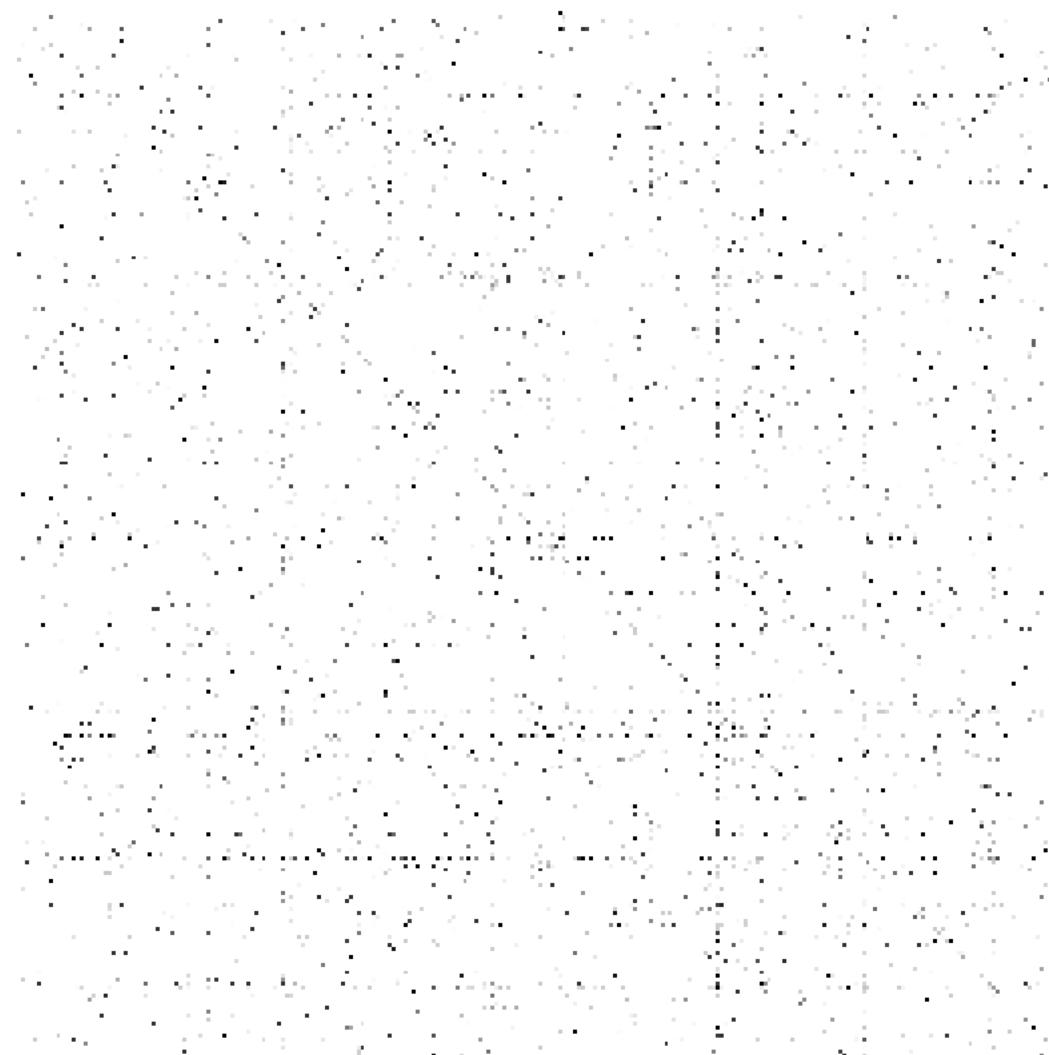
$$\begin{aligned} A_{ij} &= A_{ji} \\ A_{ii} &= 0 \end{aligned}$$

$$k_i^{out} = \sum_{j=1}^N A_{ij}$$

$$k_j^{in} = \sum_{i=1}^N A_{ij}$$

$$L = \sum_{i=1}^N k_i^{in} = \sum_{j=1}^N k_j^{out} = \sum_{i,j}^N A_{ij}$$

Adjacency Matrices are Sparse



Networks are Sparse Graphs

Most real-world networks are **sparse**

$$E \ll E_{\max} \text{ (or } k \ll N-1)$$

NETWORK	NODES	LINKS	DIRECTED/ UNDIRECTED	N	L	$\langle k \rangle$
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email Addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

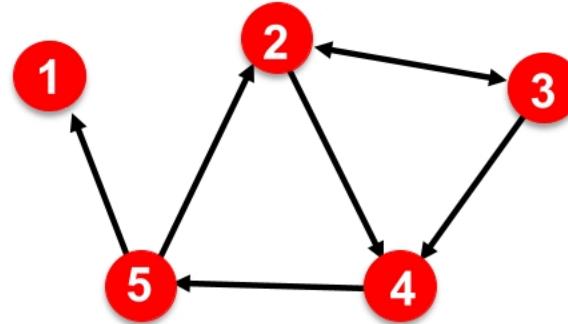
Consequence: Adjacency matrix is filled with zeros!

(Density of the matrix (E/N^2): WWW=1.51x10⁻⁵, MSN IM = 2.27x10⁻⁸)

Representing Graphs: Edge list

- Represent graph as a **list of edges**:

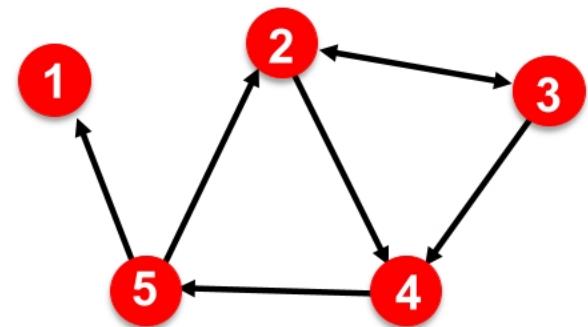
- (2, 3)
- (2, 4)
- (3, 2)
- (3, 4)
- (4, 5)
- (5, 2)
- (5, 1)



Representing Graphs: Adjacency list

■ **Adjacency list:**

- Easier to work with if network is
 - Large
 - Sparse
- Allows us to quickly retrieve all neighbors of a given node
 - 1:
 - 2: 3, 4
 - 3: 2, 4
 - 4: 5
 - 5: 1, 2



Node and Edge Attributes

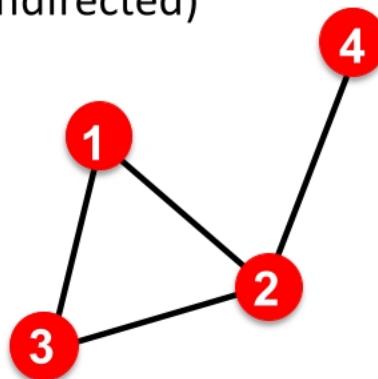
Possible options:

- Weight (*e.g.*, frequency of communication)
- Ranking (best friend, second best friend...)
- Type (friend, relative, co-worker)
- Sign: Friend vs. Foe, Trust vs. Distrust
- Properties depending on the structure of the rest of the graph: Number of common friends

More Types of Graphs

■ Unweighted

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0$$

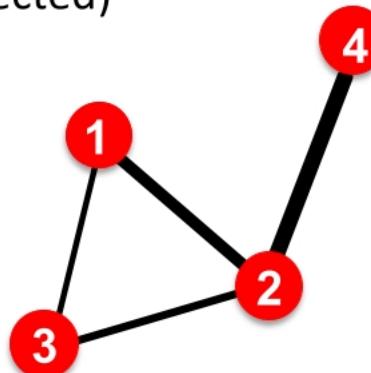
$$A_{ij} = A_{ji}$$

$$E = \frac{1}{2} \sum_{i,j=1}^N A_{ij} \quad \bar{k} = \frac{2E}{N}$$

Examples: Friendship, Hyperlink

■ Weighted

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0$$

$$A_{ij} = A_{ji}$$

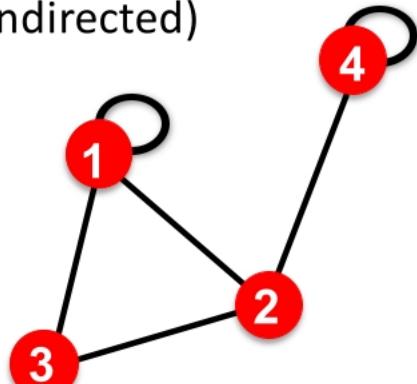
$$E = \frac{1}{2} \sum_{i,j=1}^N \text{nonzero}(A_{ij}) \quad \bar{k} = \frac{2E}{N}$$

Examples: Collaboration, Internet, Roads

More Types of Graphs

■ Self-edges (self-loops)

(undirected)



$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$A_{ii} \neq 0$$

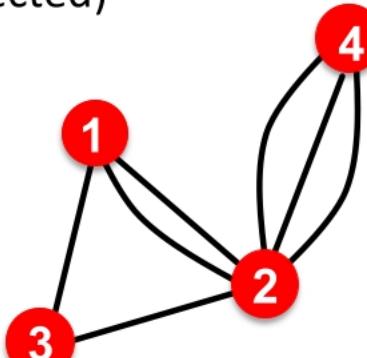
$$A_{ij} = A_{ji}$$

$$E = \frac{1}{2} \sum_{i,j=1, i \neq j}^N A_{ij} + \sum_{i=1}^N A_{ii}$$

Examples: Proteins, Hyperlinks

■ Multigraph

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0$$

$$E = \frac{1}{2} \sum_{i,j=1}^N \text{nonzero}(A_{ij})$$

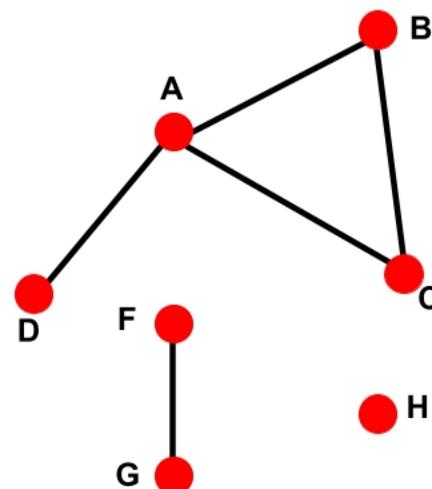
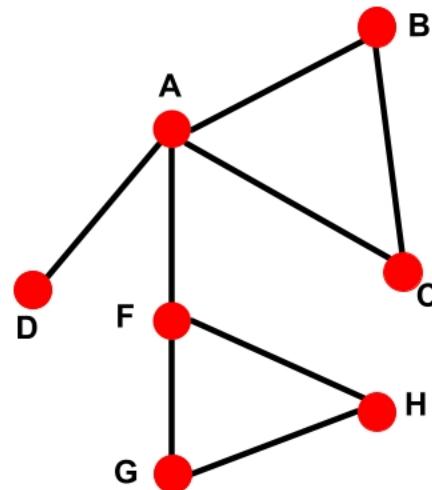
$$A_{ij} = A_{ji}$$

$$\bar{k} = \frac{2E}{N}$$

Examples: Communication, Collaboration

Connectivity of Undirected Graphs

- **Connected (undirected) graph:**
 - Any two vertices can be joined by a path
- A disconnected graph is made up by two or more connected components



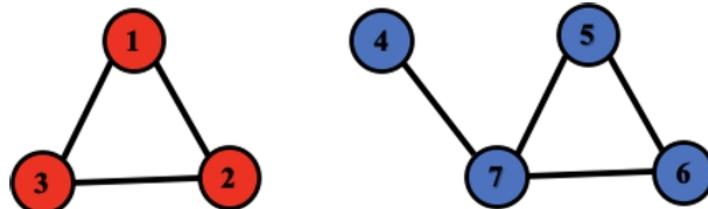
Largest Component:
Giant Component

Isolated node (node H)

Connectivity: Example

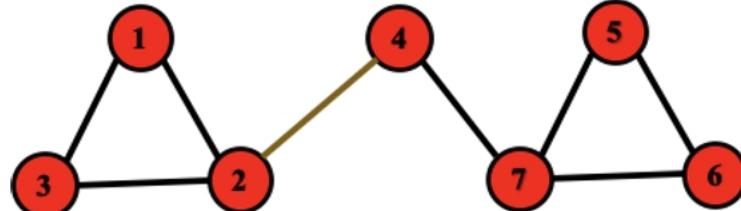
The adjacency matrix of a network with several components can be written in a block-diagonal form, so that nonzero elements are confined to squares, with all other elements being zero:

Disconnected



$$\begin{bmatrix} \textcolor{red}{\begin{matrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{matrix}} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \textcolor{blue}{\begin{matrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{matrix}} \end{bmatrix}$$

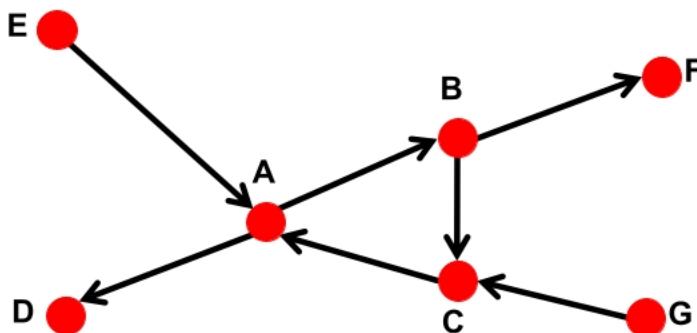
Connected



$$\begin{bmatrix} \textcolor{red}{\begin{matrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{matrix}} & \begin{matrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{matrix} \\ \begin{matrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \textcolor{blue}{\begin{matrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{matrix}} \end{bmatrix}$$

Connectivity of Directed Graphs

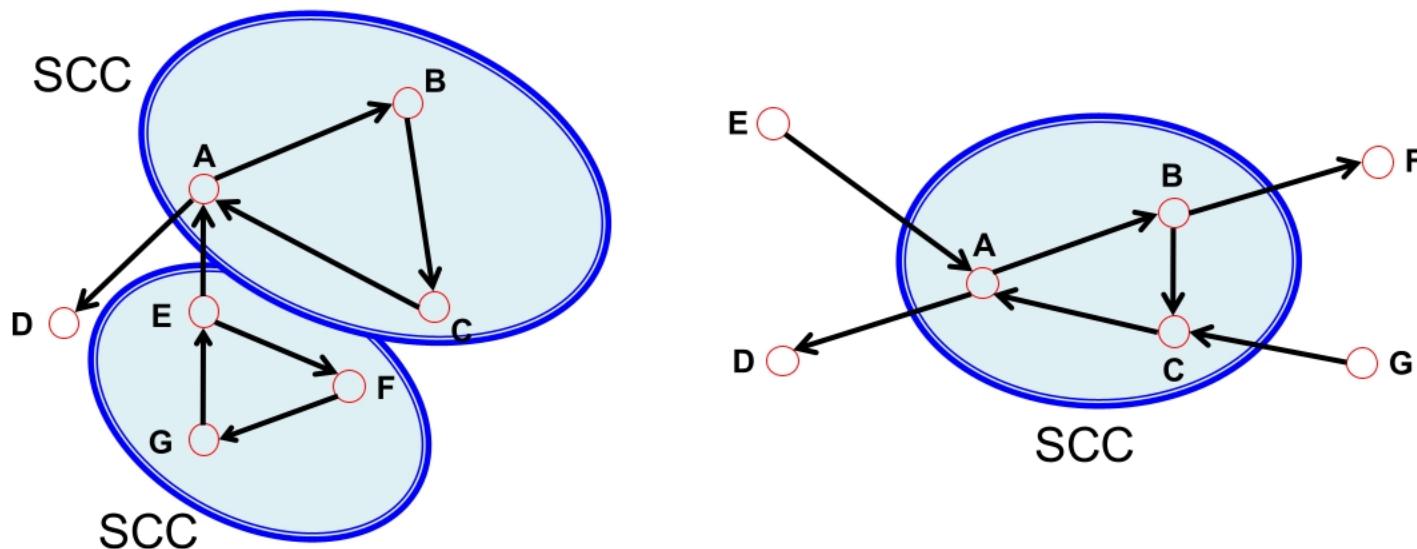
- **Strongly connected directed graph**
 - has a path from each node to every other node and vice versa (e.g., A-B path and B-A path)
- **Weakly connected directed graph**
 - is connected if we disregard the edge directions



Graph on the left is connected but not strongly connected (e.g., there is no way to get from F to G by following the edge directions).

Connectivity of Directed Graphs

- Strongly connected components (SCCs) can be identified, but not every node is part of a nontrivial strongly connected component.



In-component: nodes that can reach the SCC,

Out-component: nodes that can be reached from the SCC.