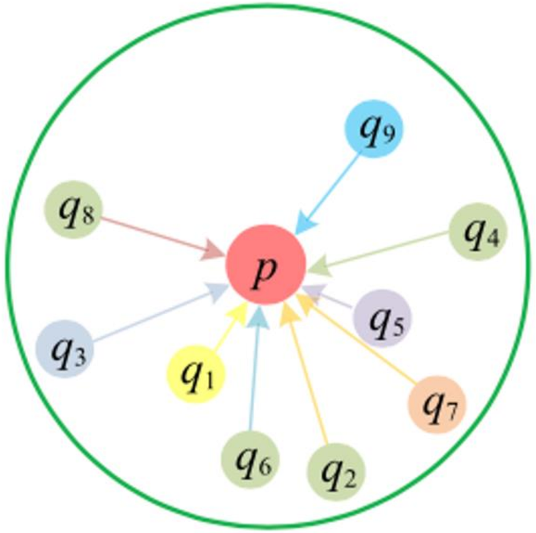
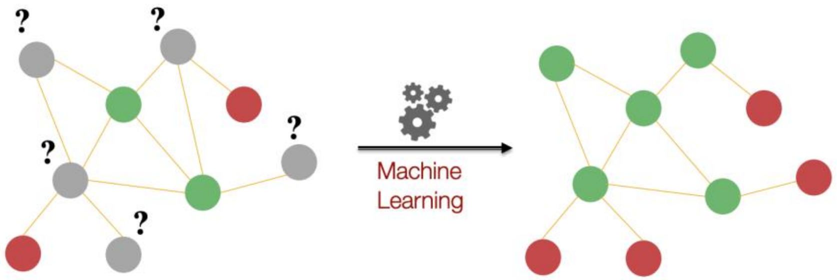
Introduction to Graph Machine Learning

Charu Sharma
Assistant Professor
Machine Learning Lab
IIIT Hyderabad



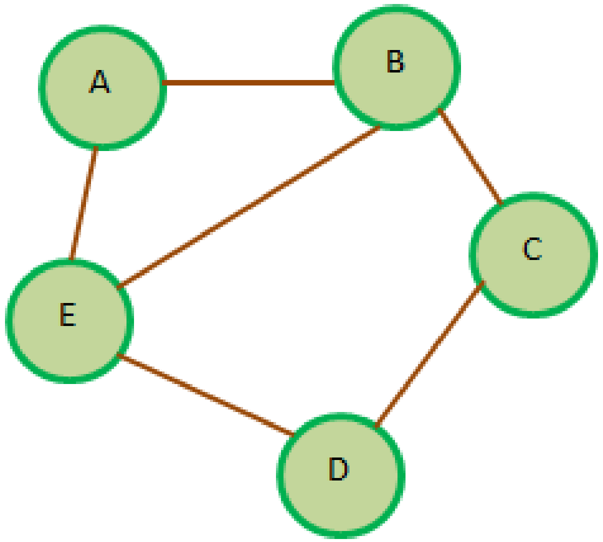
INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD



Outline

- What and Why?
- Examples
- Machine/Deep Learning for Graphs.
- Tasks
- Applications!
- Datasets
- Embeddings Methods

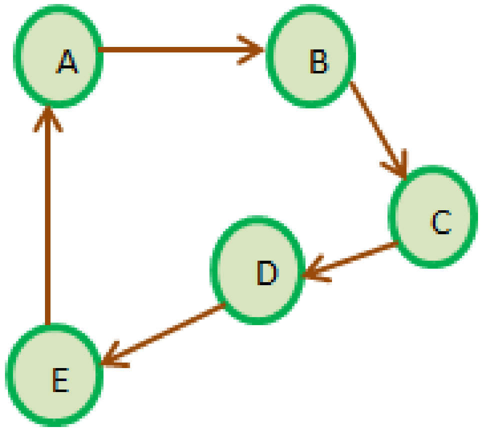
Graphs

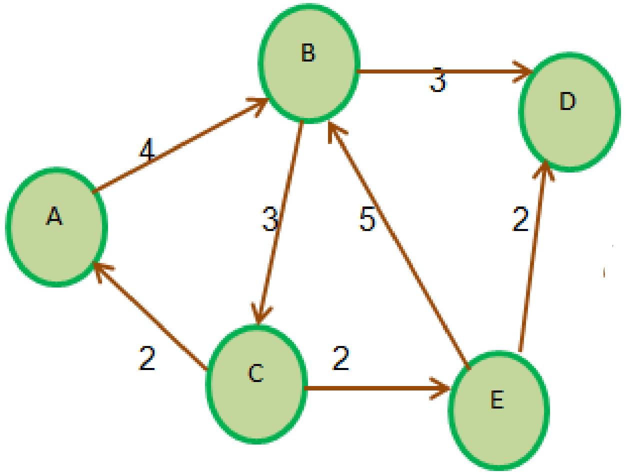


Undirected graph

	A	B	C	D	E
A	0	1	0	0	1
B	1	0	1	0	1
C	0	1	0	1	0
D	0	0	1	0	1
E	1	1	0	1	0

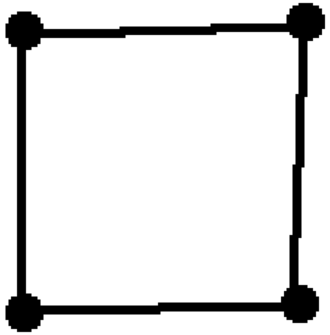
Adjacency Matrix





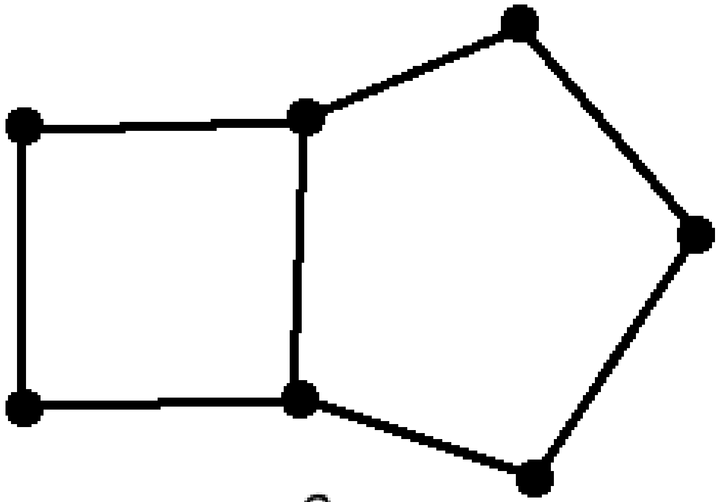
Subgraphs

- Subgraphs are the integral parts of a graph
- We can differentiate between graphs using subgraphs



S

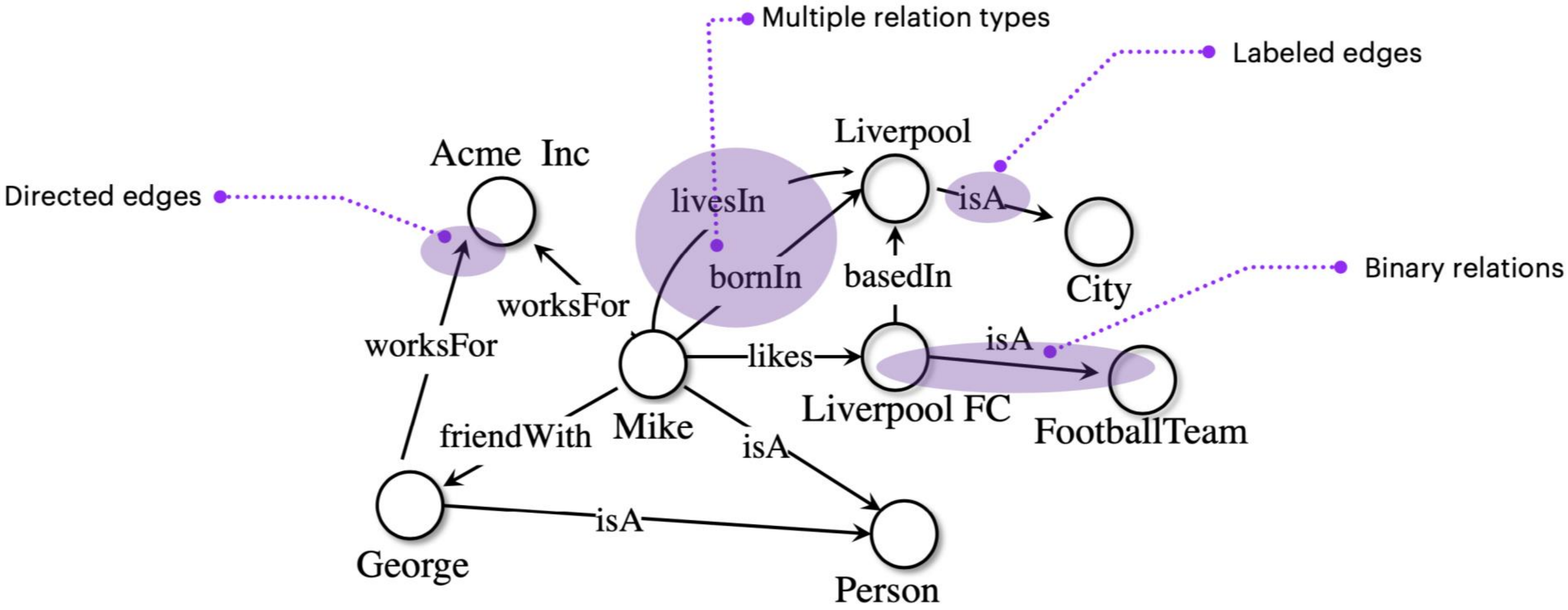
(Subgraph of G)



G

Knowledge Graph

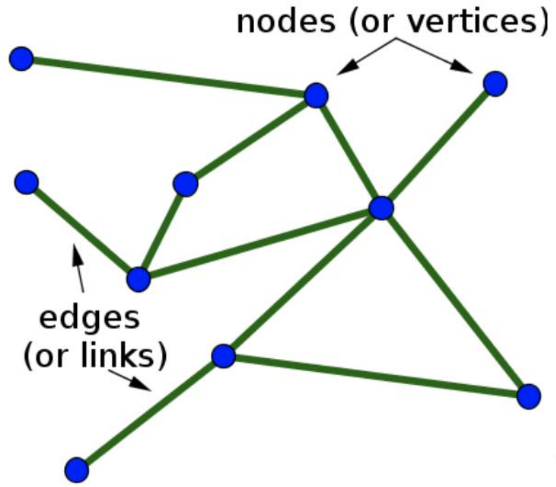
- Knowledge in graph form
  - Capture entities, types, and relationships
- Nodes are **entities.** Nodes are labeled with their **types.**
- Edges between two nodes capture **relationships** between entities.
- **KG is an example of a <u>heterogeneous</u> graph.**



Credits: ECAI20 tutorial

Why Graphs?

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

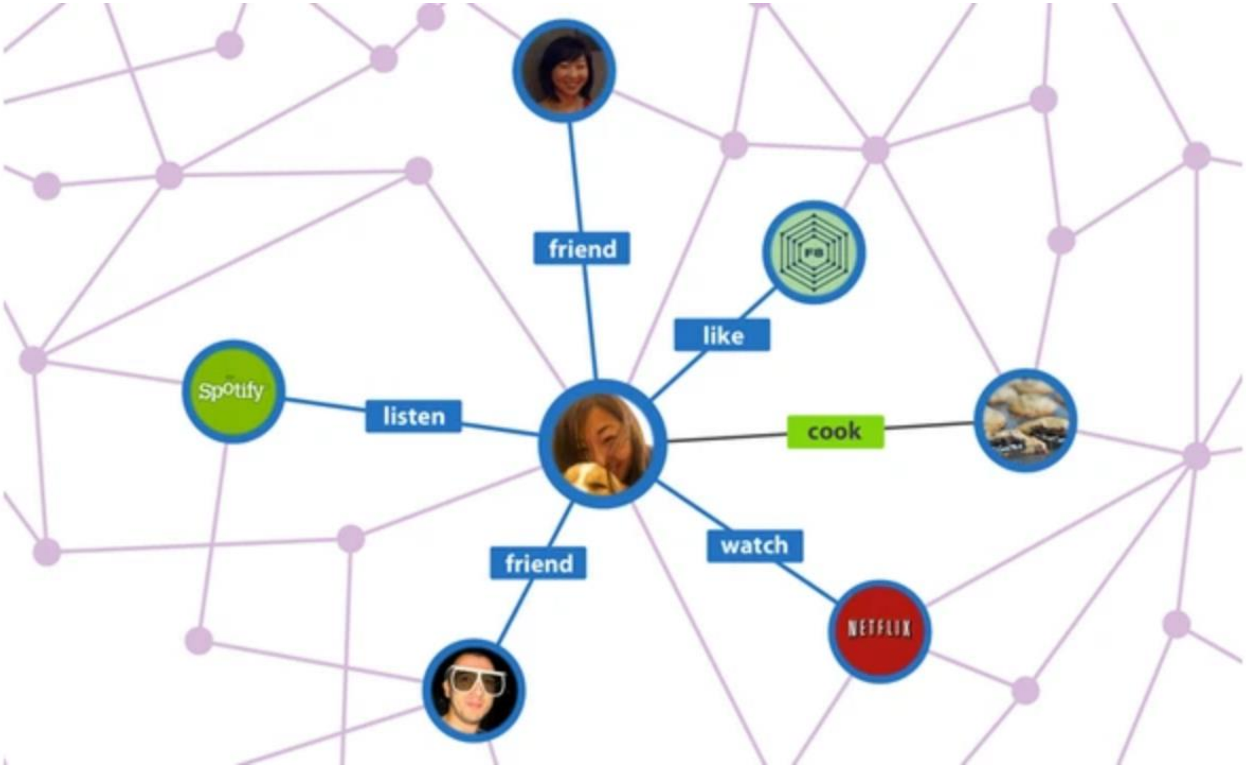


Outline

- Why and What?
- Examples
- Machine/Deep Learning for Graphs.
- Tasks
- Applications!
- Datasets
- Embeddings Methods

Social Network

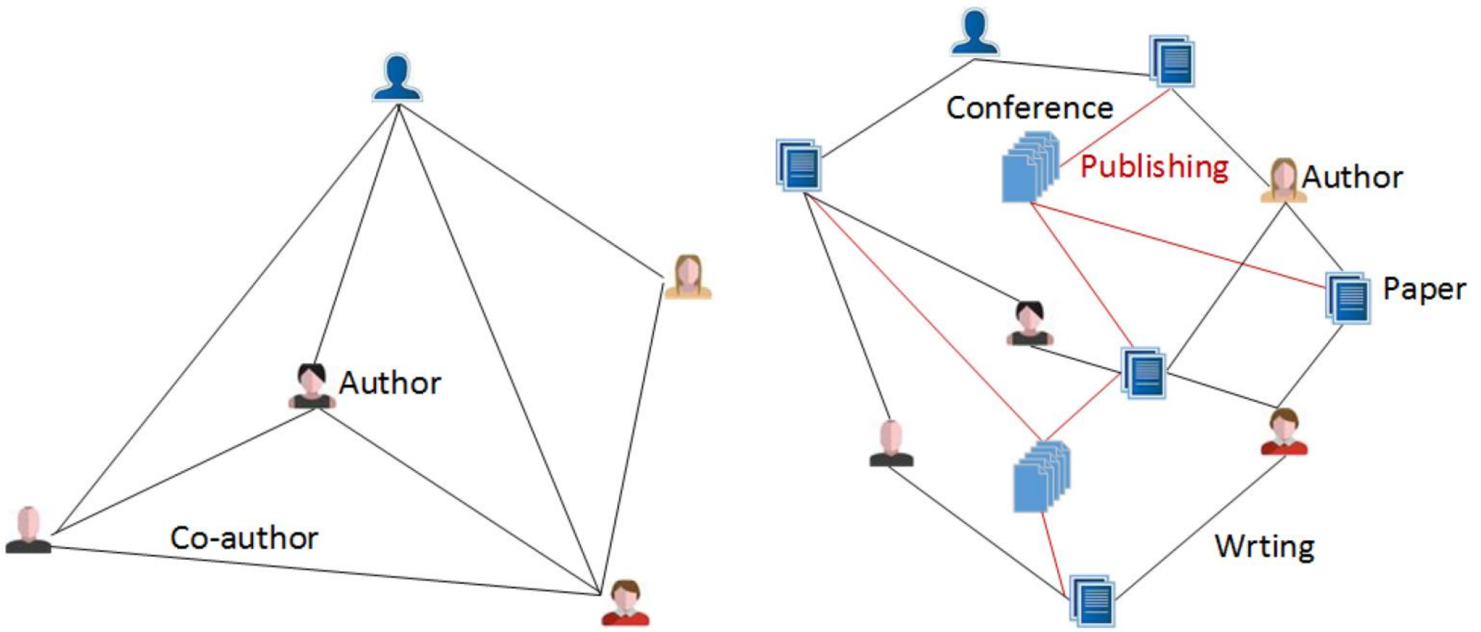
- **Node types**: account, song, post, food, channel
- **Relation types**: friend, like, cook, watch, listen



Credits: Stanford CS224W

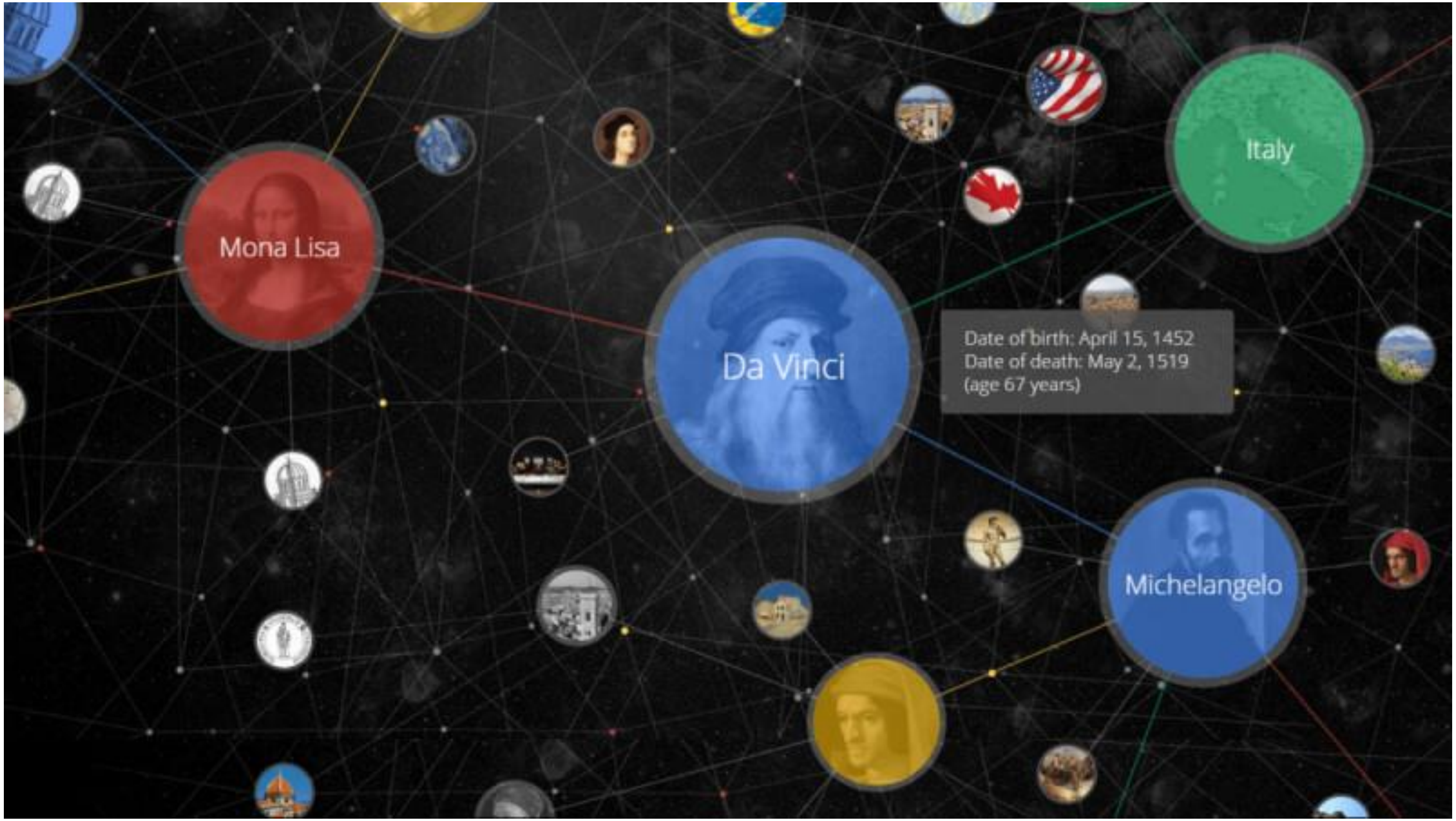
Bibliographic Network

- **Node types**: paper, title, author, conference, year
- **Relation types**: pubWhere, pubYear, hasTitle, hasAuthor, cite



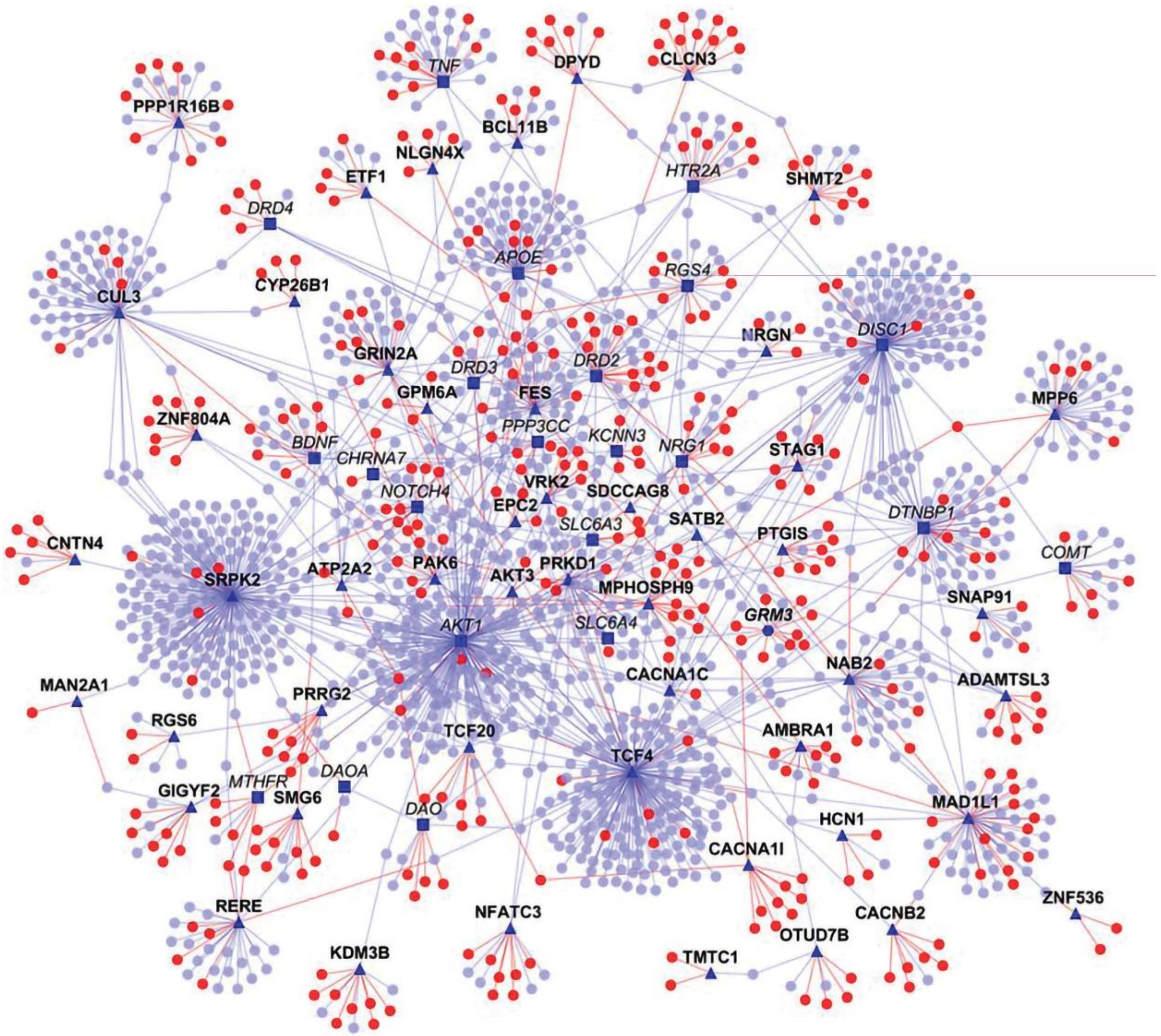
(a) Co-author Homogeneous Network (b) Bibliographic Heterogeneous Network

Google Knowledge graph



Credits: Stanford CS224W

Protein-Protein Interaction Network



Investigation of the complex biological activities in the cell

Molecular Graphs

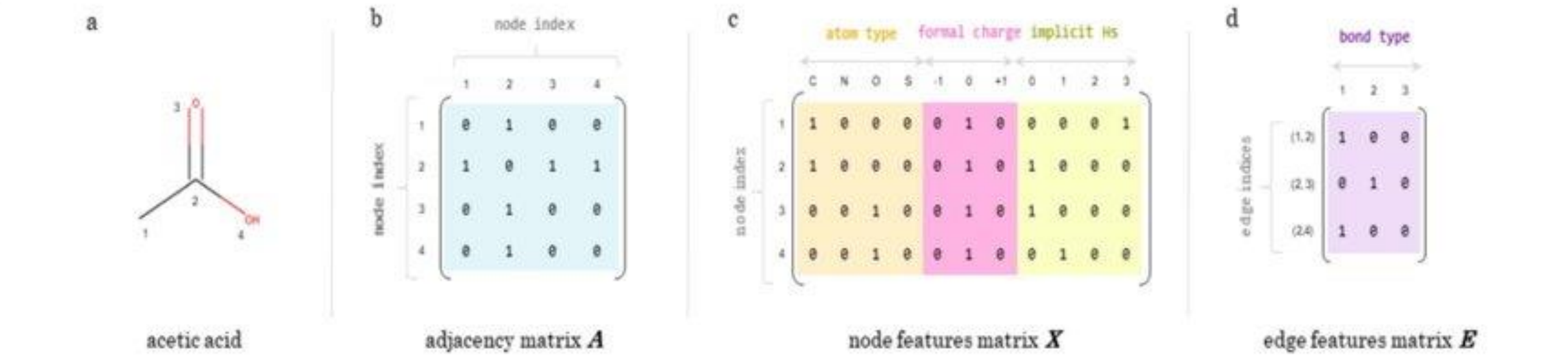
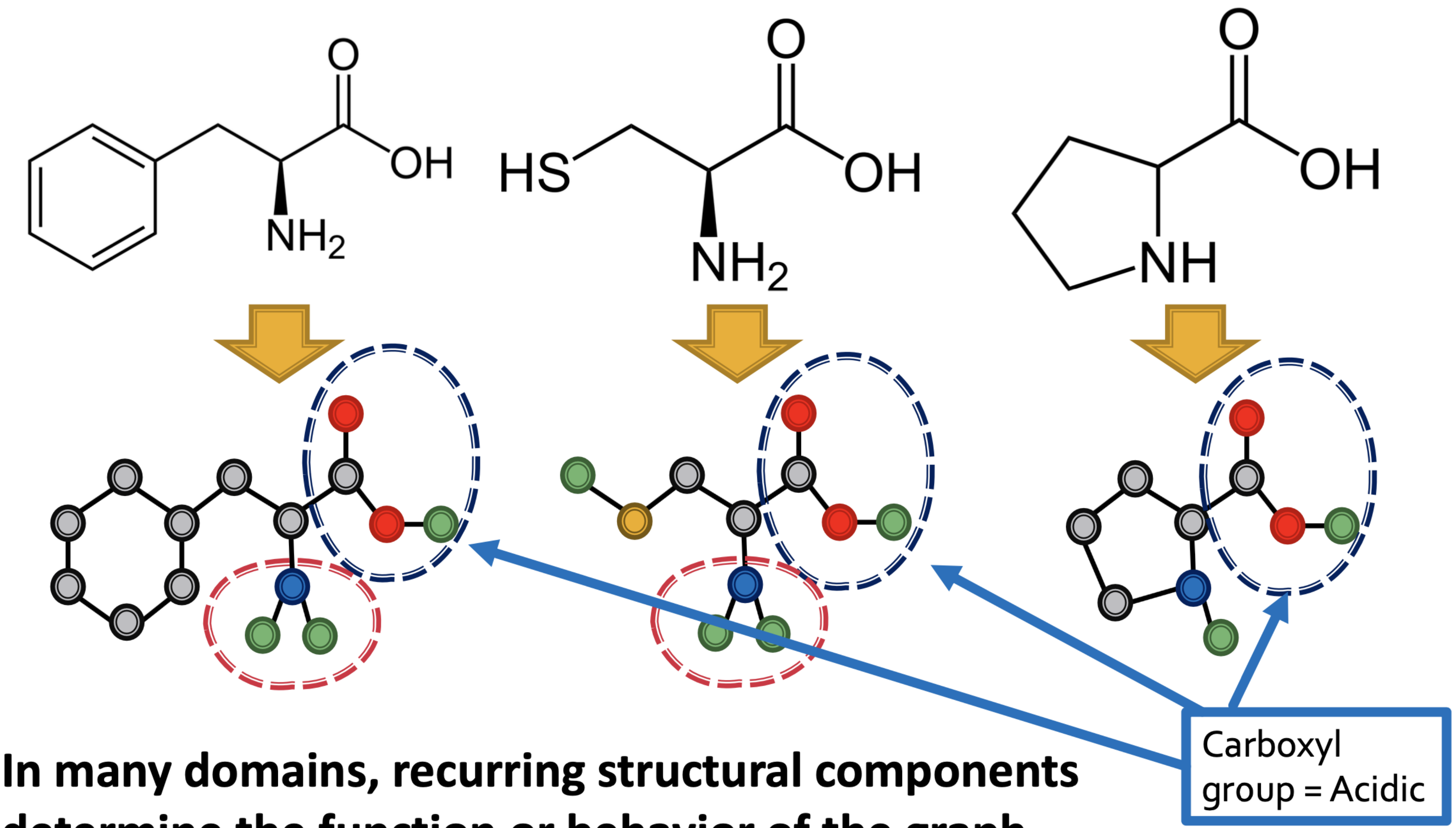


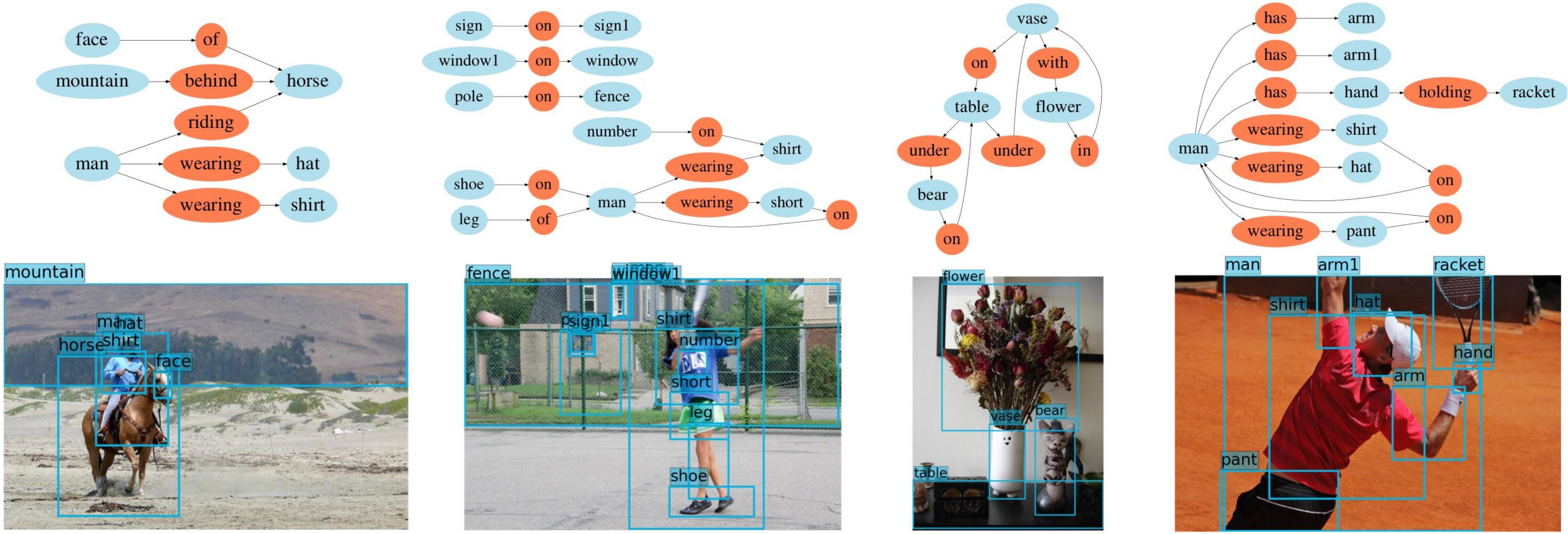
Fig. 1 Example graph representation for acetic acid. **a** Graph representation of acetic acid with nodes numbered from one to four. **b** Example adjacency matrix, ***A***, for an acetic acid graph with the corresponding node ordering on the left. **c** Example node features matrix, ***X***, which one-hot encodes a few selected properties. **d** Example edge features matrix, ***E***, where each edge feature vector is a one-hot encoding of single, double, or triple bonds. "Implicit Hs" stands for the number of implicit hydrogens on a given node

Molecular (sub)graphs



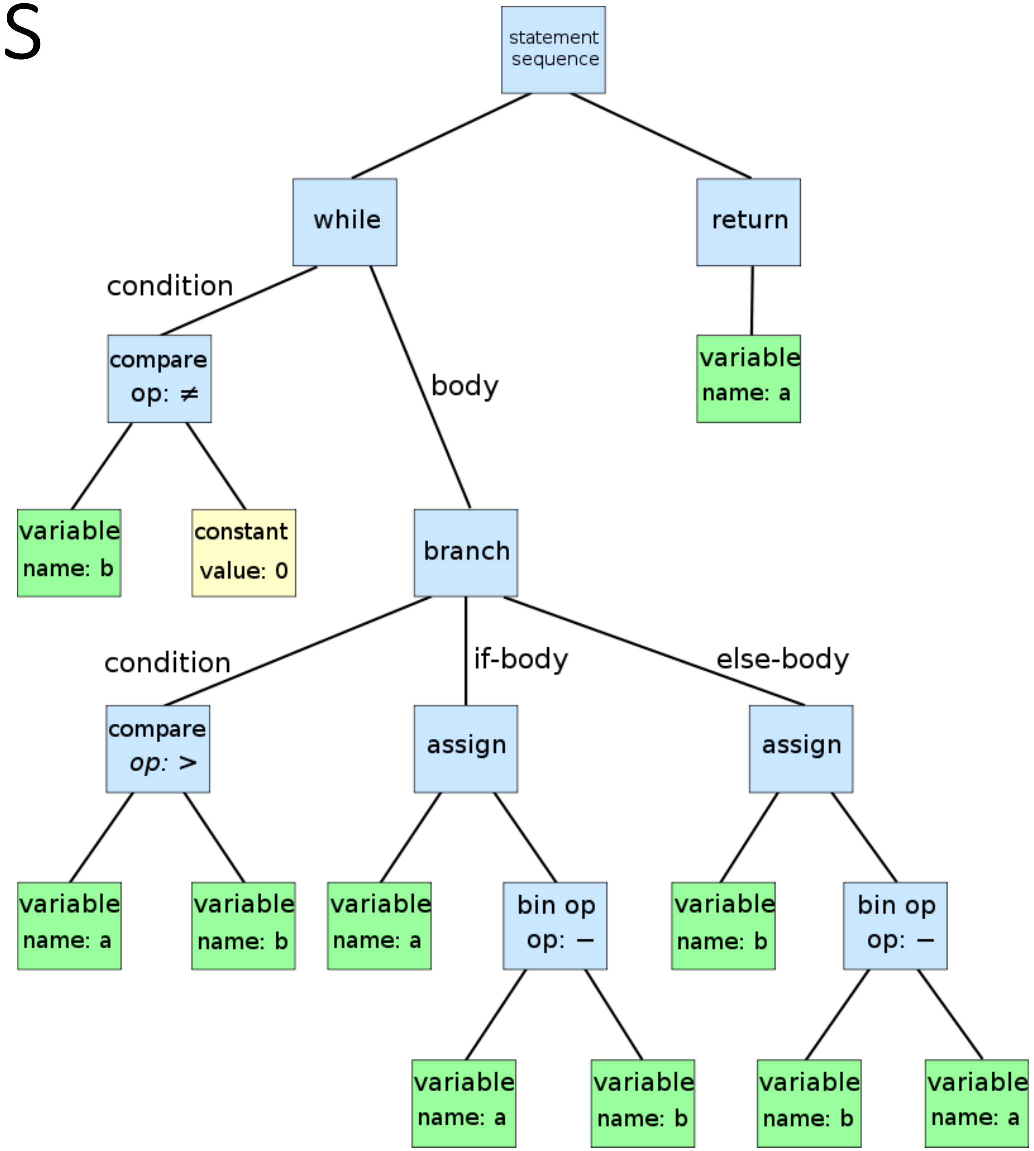
In many domains, recurring structural components determine the function or behavior of the graph

Scene Graphs



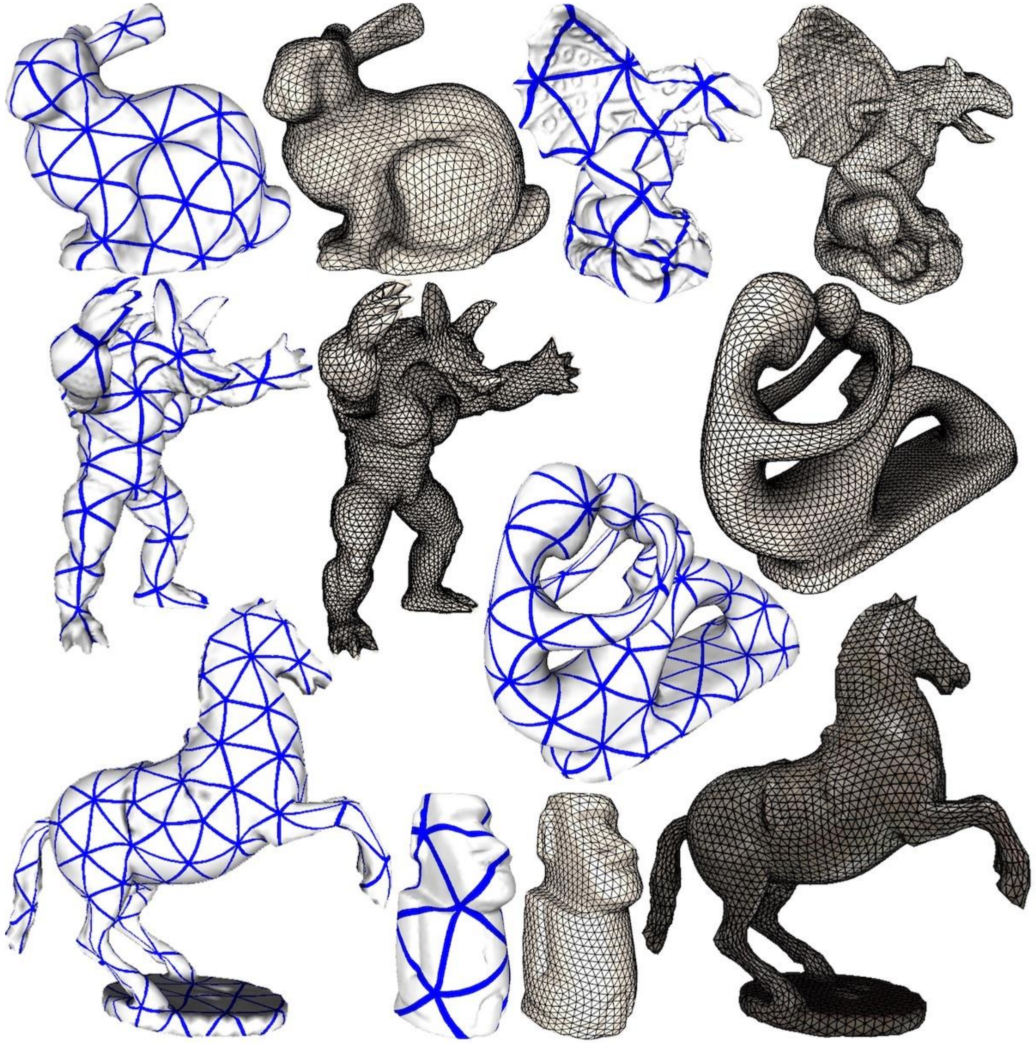
Credits: Scene Graph Generation by Iterative Message Passing, CVPR 2017.

Code Graphs



Credits: Medium.com

3D Meshes



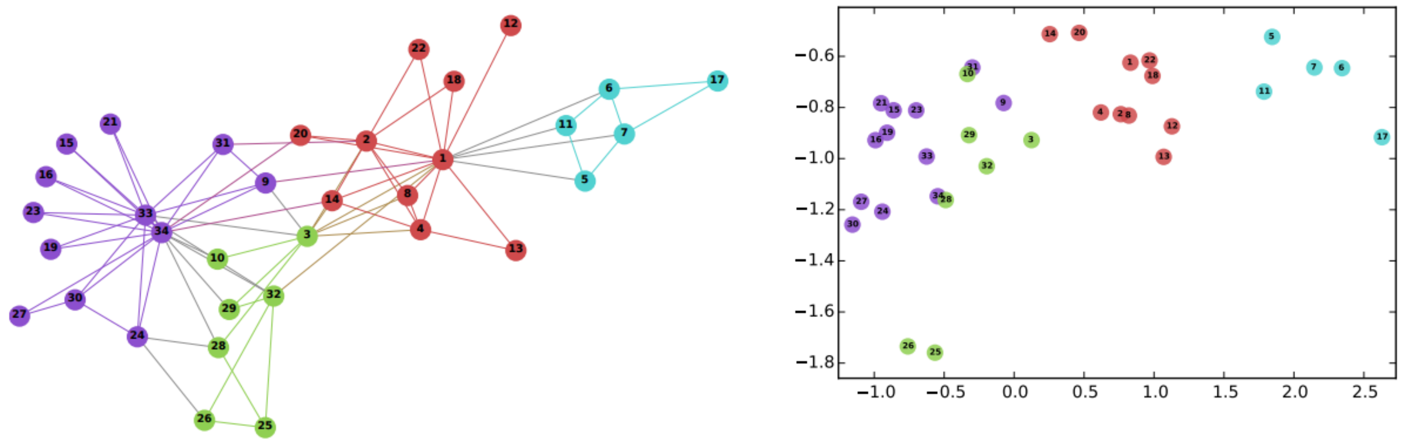
Credits: MeshLab

Outline

- Why and What?
- Examples
- Machine/Deep Learning for Graphs
- Tasks
- Applications!
- Datasets
- Embeddings Methods

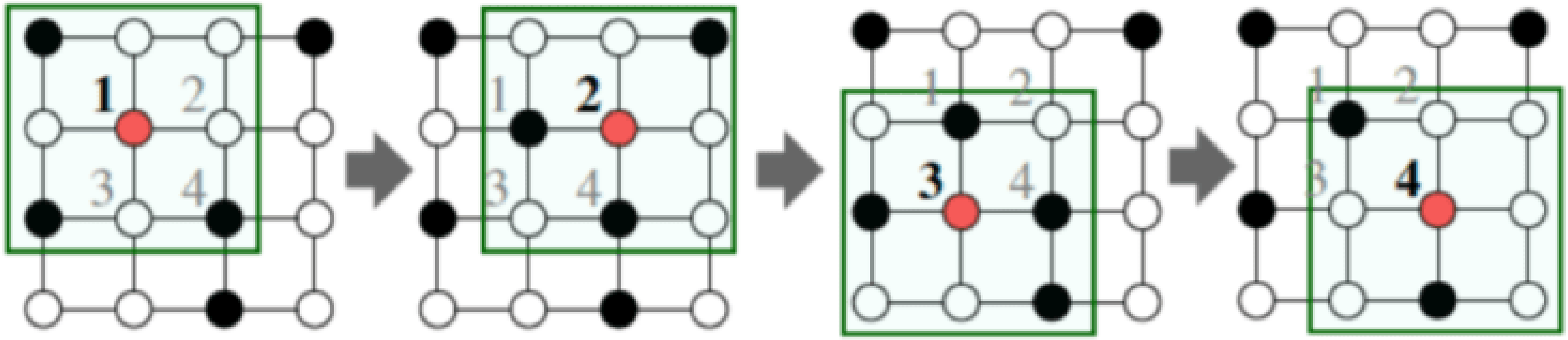
Graphs for Machine Learning

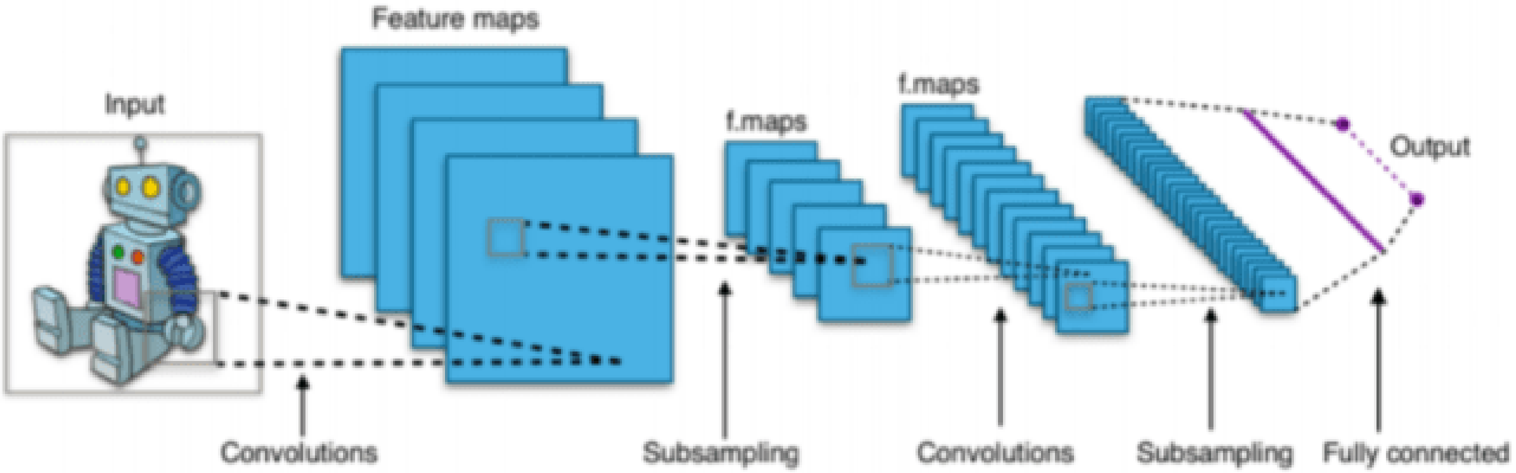
- How do we take advantage of relational structure for better prediction?
- Complex domains have a rich relational structure, which can be represented as a **relational graph**
- **By explicitly modeling relationships we achieve better performance!**



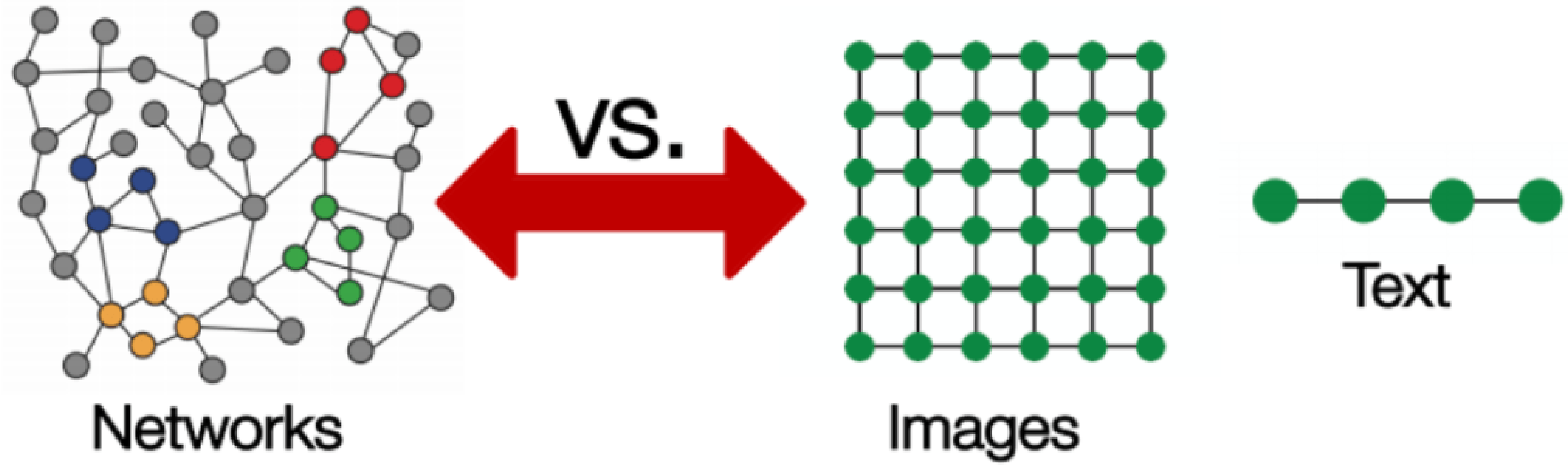
(a) Input: Karate Graph (b) Output: Representation

CNNs for Graphs



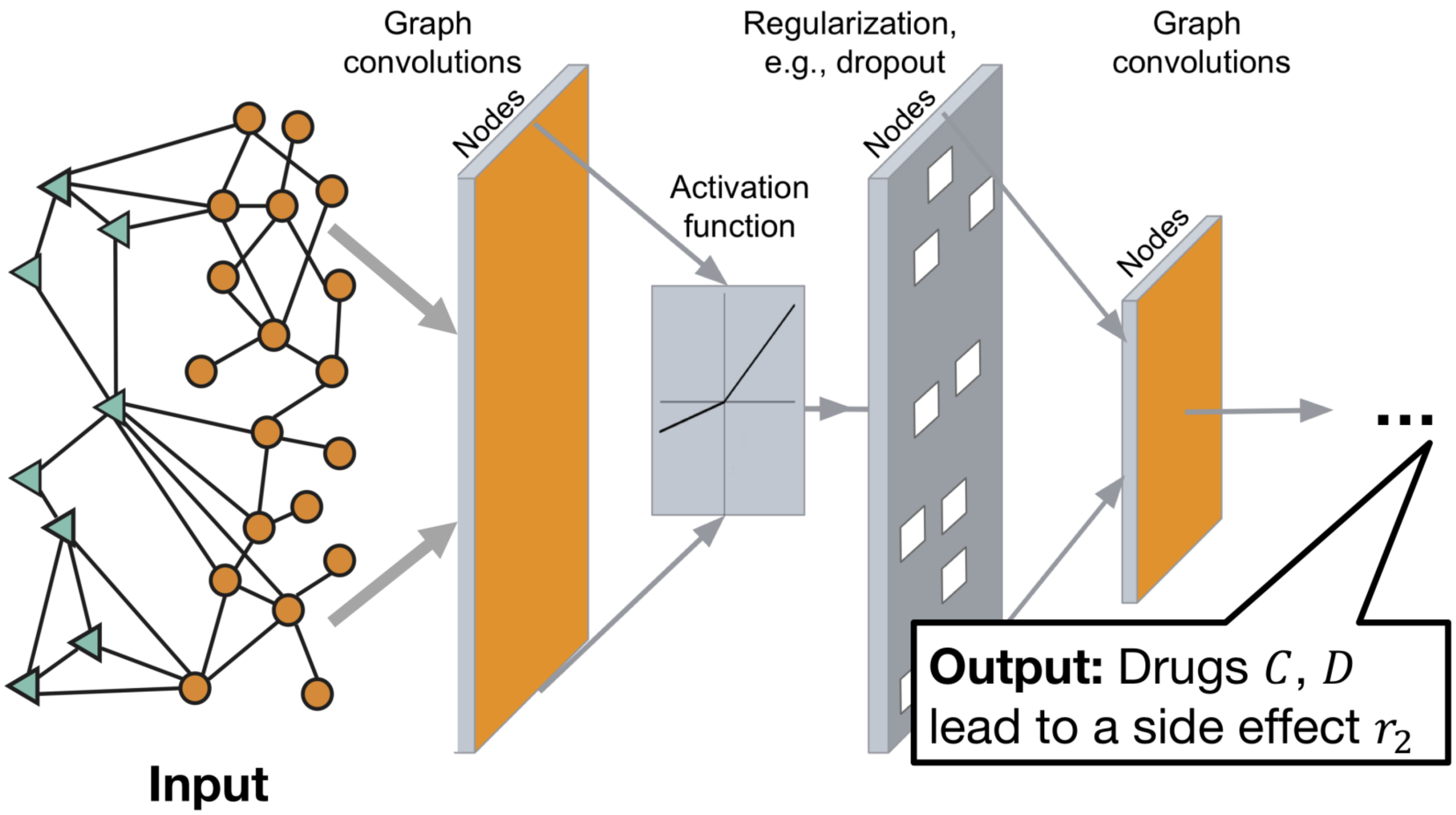


Challenges with Graphs

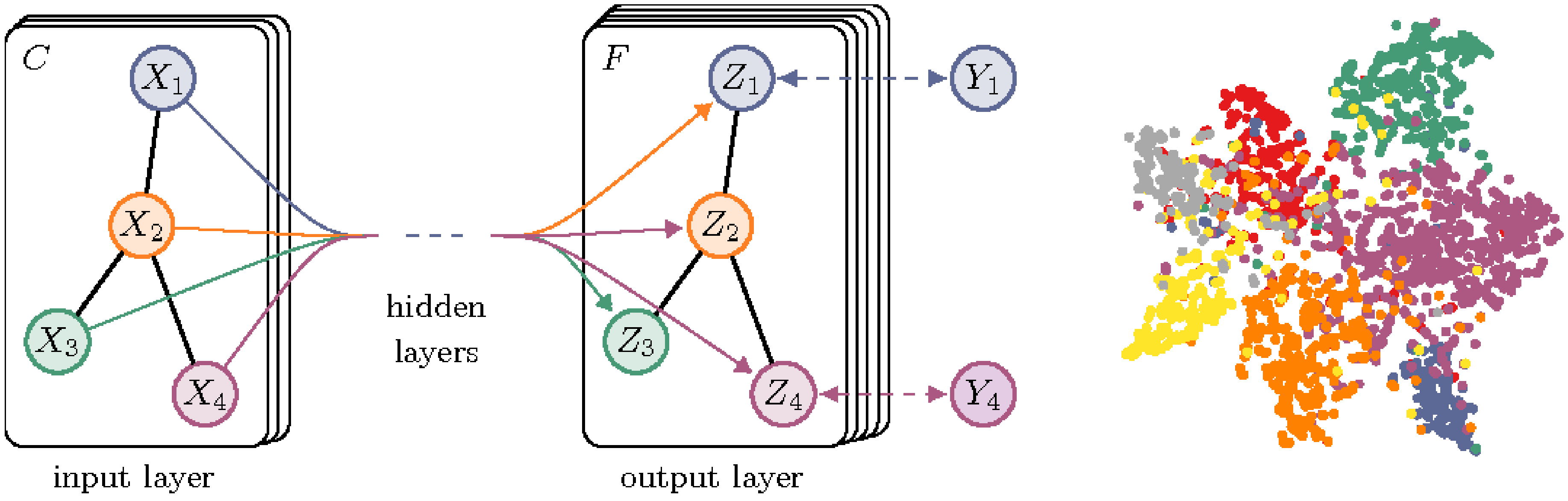


- No spatial locality like grids
- Arbitrary size and complex topological structure
- No fixed node ordering

Deep Learning for Graphs



Graph Representation Learning



(a) Graph Convolutional Network

(b) Hidden layer activations

- Map nodes to d-dimensional **embeddings** such that similar nodes in the network are embedded close together (no feature engineering required)
- Learning a network,

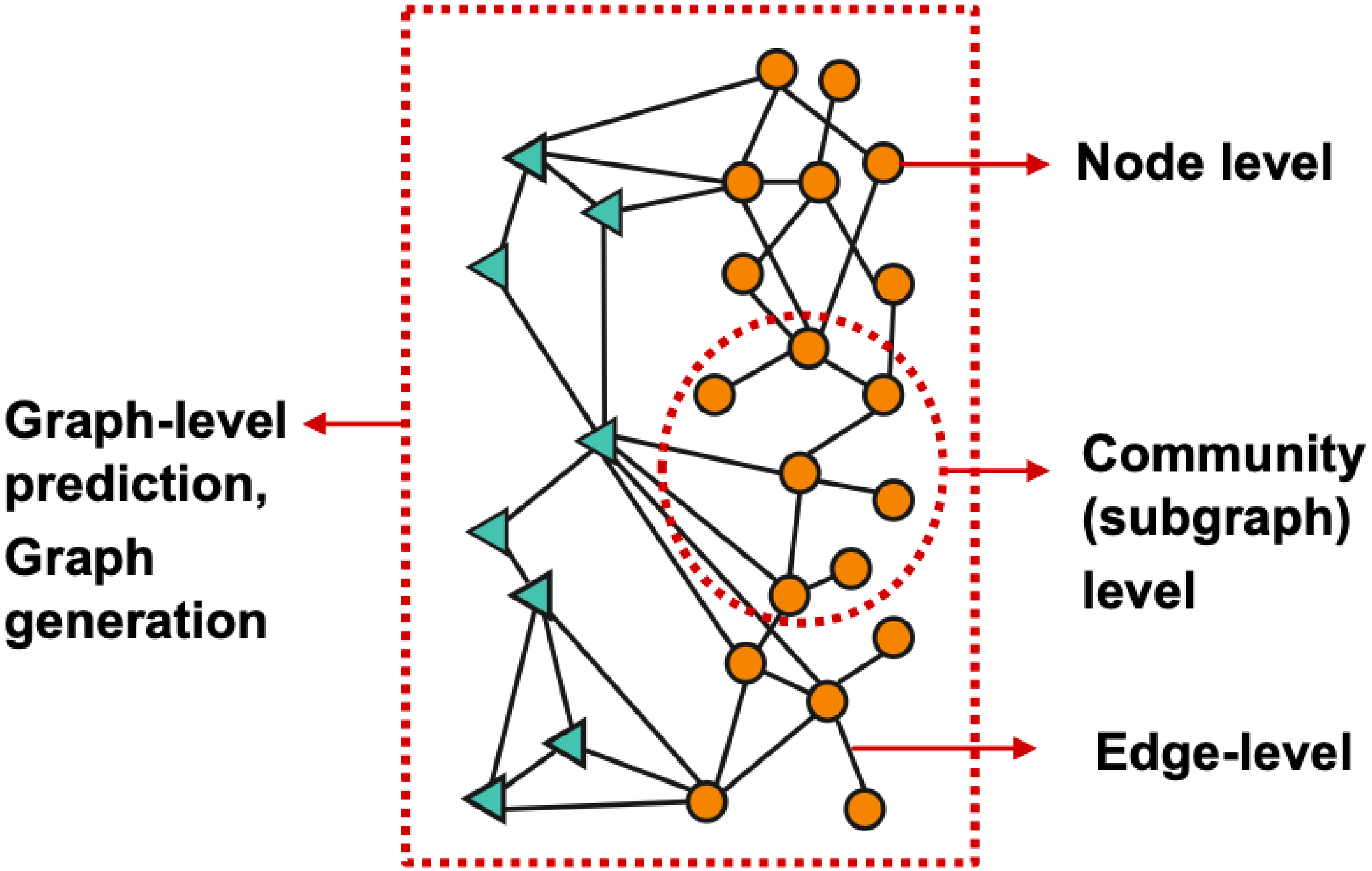
$$f: u \rightarrow R^d$$

Credits: GCN, ICLR 2017

Outline

- Why and What?
- Examples
- Machine/Deep Learning for Graphs.
- Tasks
- Applications!
- Datasets
- Embeddings Methods

Categories of Tasks



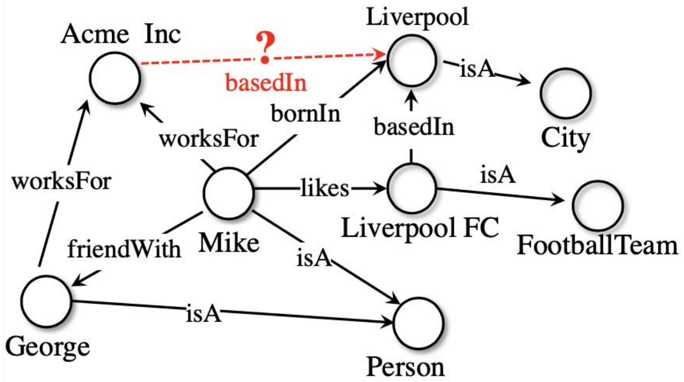
Tasks for ML in Graphs

- **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
- **Graph generation**: Drug discovery
- **Graph evolution**: Physical simulation, time series

Tasks for KGs

LINK PREDICTION / TRIPLE CLASSIFICATION

- Knowledge graph completion
- Content recommendation
- Question answering

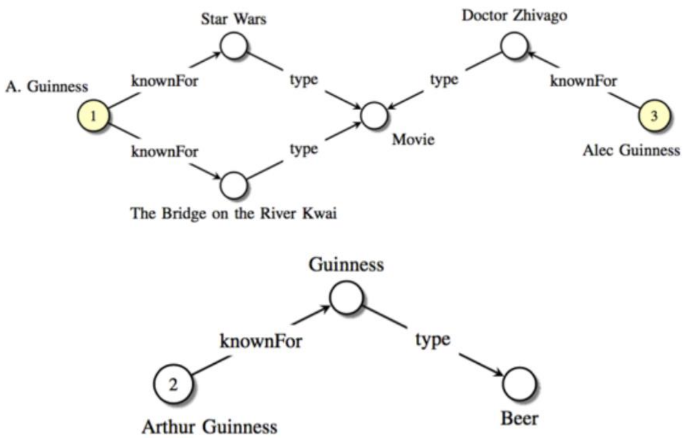


COLLECTIVE NODE CLASSIFICATION / LINK-BASED CLUSTERING

- Customer segmentation

ENTITY MATCHING

- Duplicate detection
- Inventory items deduplication



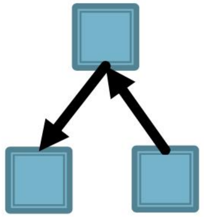
Pic from [Nickel et al. 2016a]

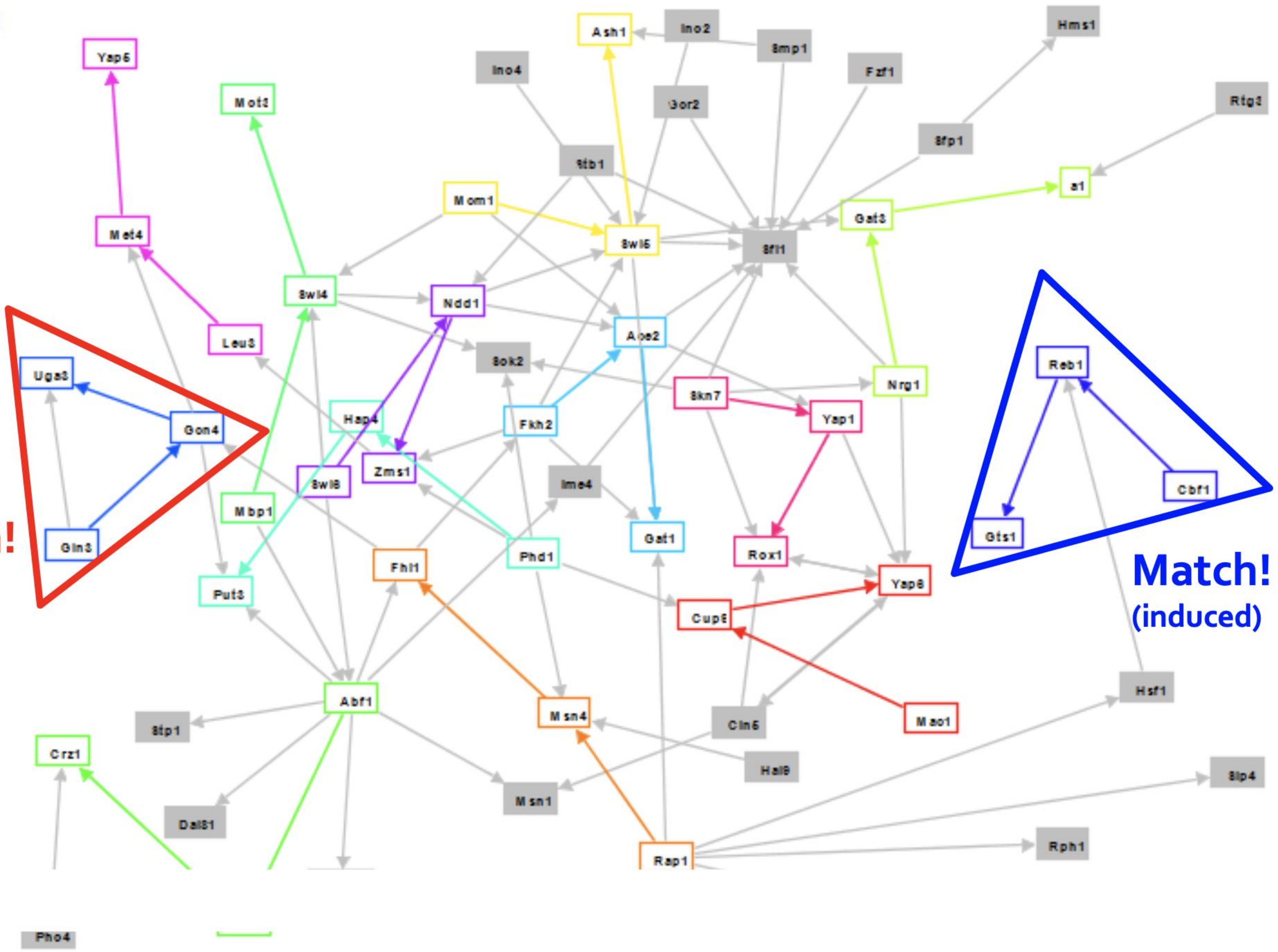
Outline

- Why and What?
- Examples
- Machine/Deep Learning for Graphs.
- Tasks
- Applications!
- Datasets
- Embeddings Methods

Motifs

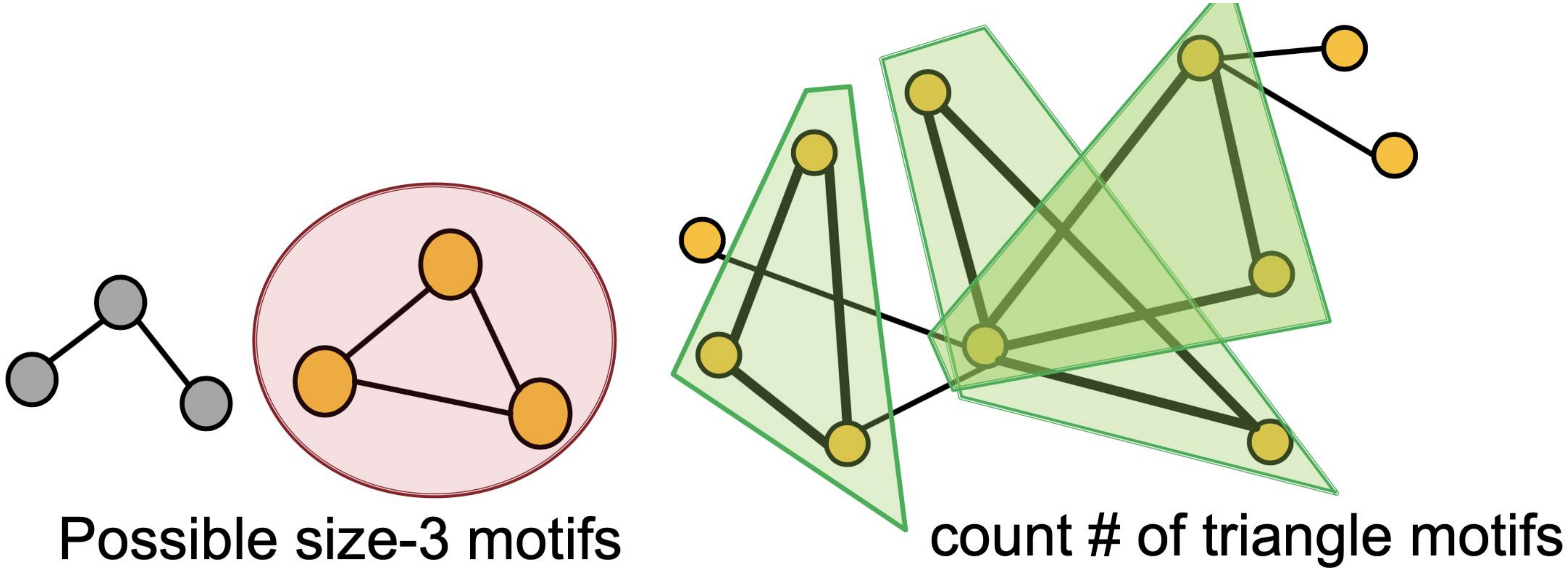
Induced subgraph of interest (aka Motif):





Finding subgraphs

- Different sizes of subgraphs
- Count subgraphs of each size



Correspondence problem in images

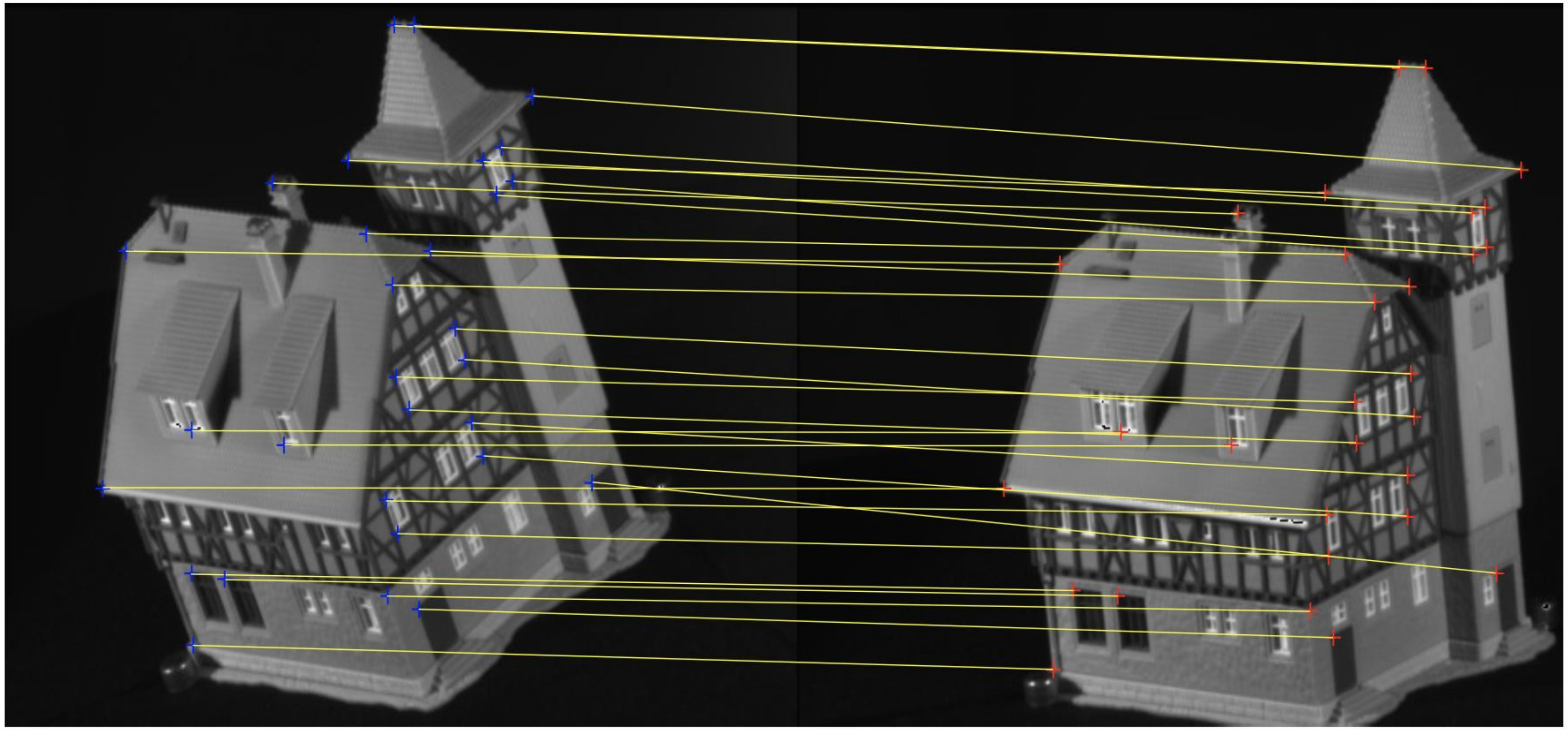
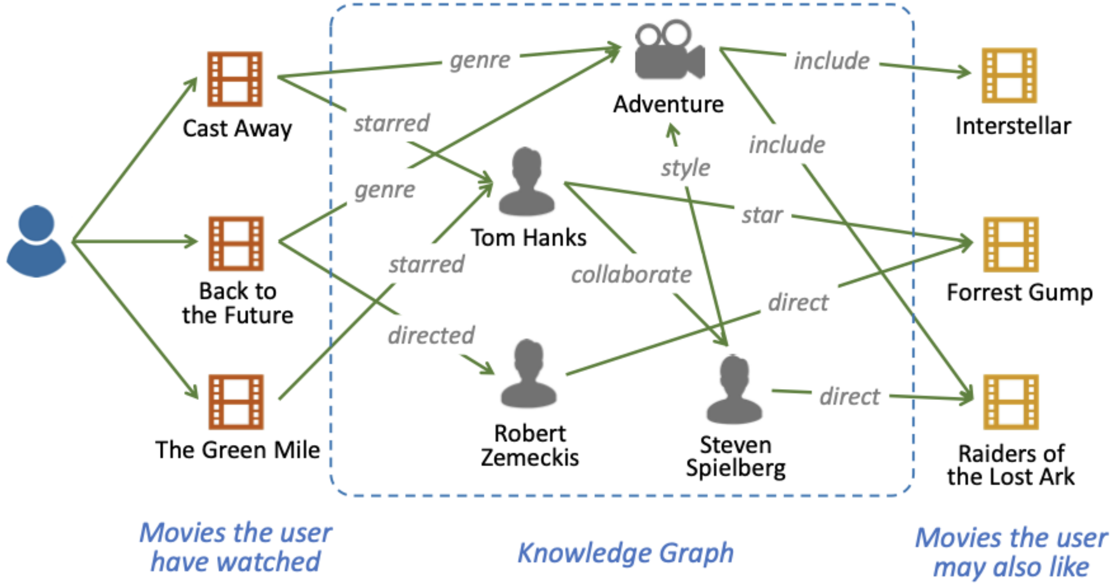


Figure: Instance of a Bijective Mapping[4].

Recommendation in E-commerce

- Suggest relevant items to users

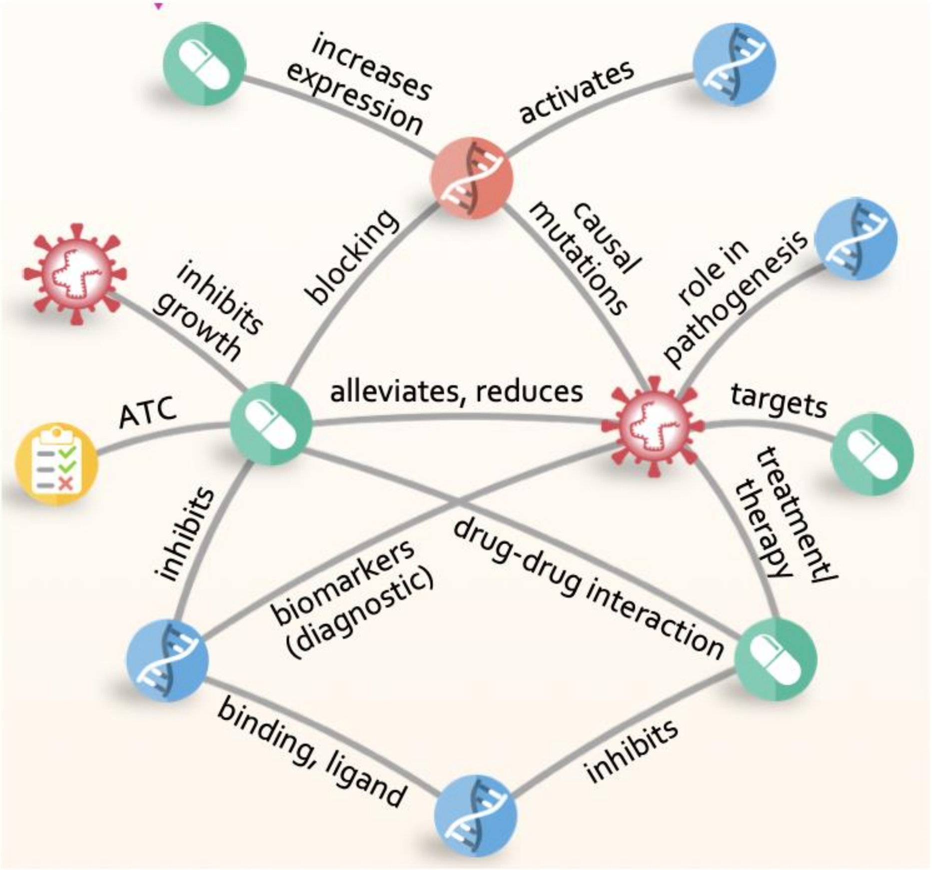


Hongwei Wang et al. RippleNet: Propagating user preferences on the knowledge graph for recommender systems. *CIKM* 2018.

Credits: AAAI22 tutorial

Drug Repurposing

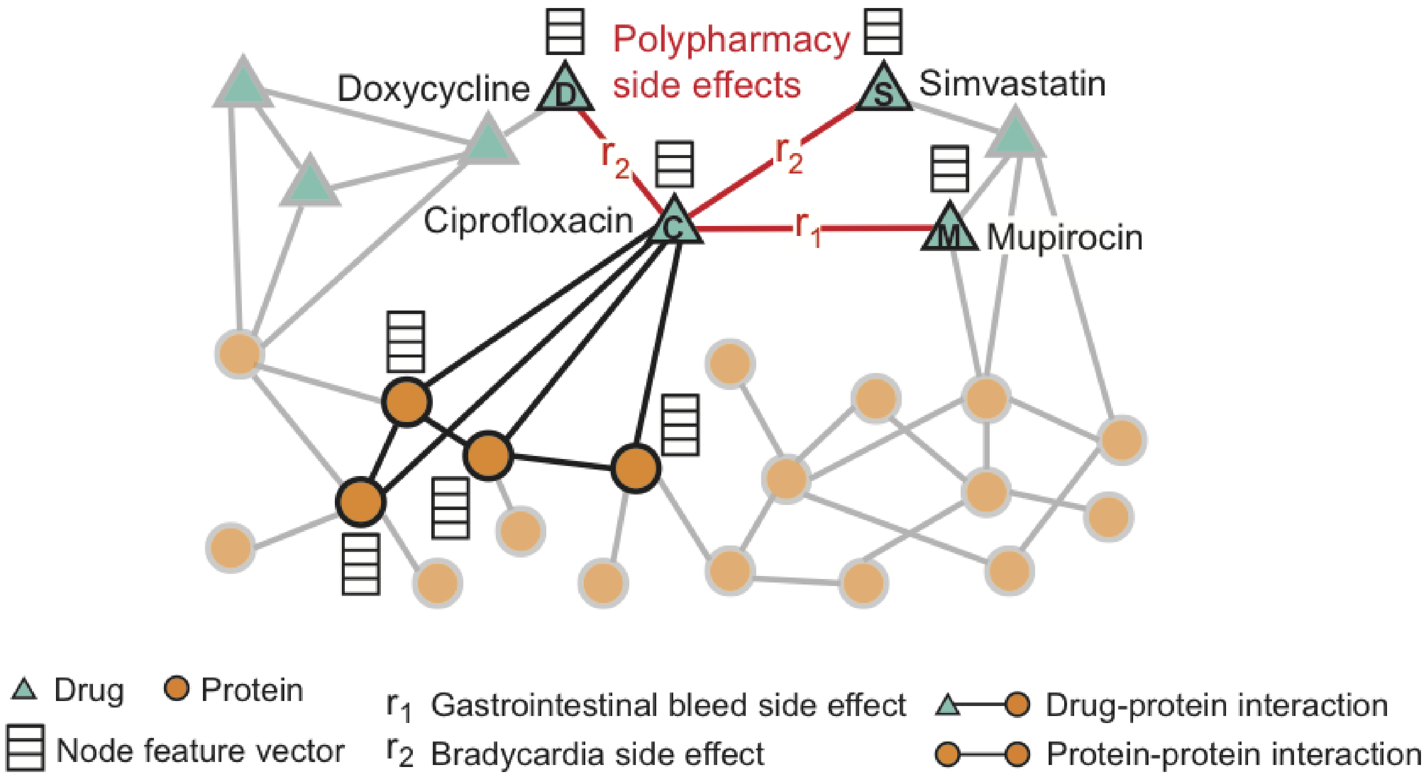
- Predicting effective (approved) drugs given a disease



Xiangxiang Zeng et al. Repurpose open data to discover therapeutics for COVID-19 using deep learning. *Journal of proteome research* 2020.

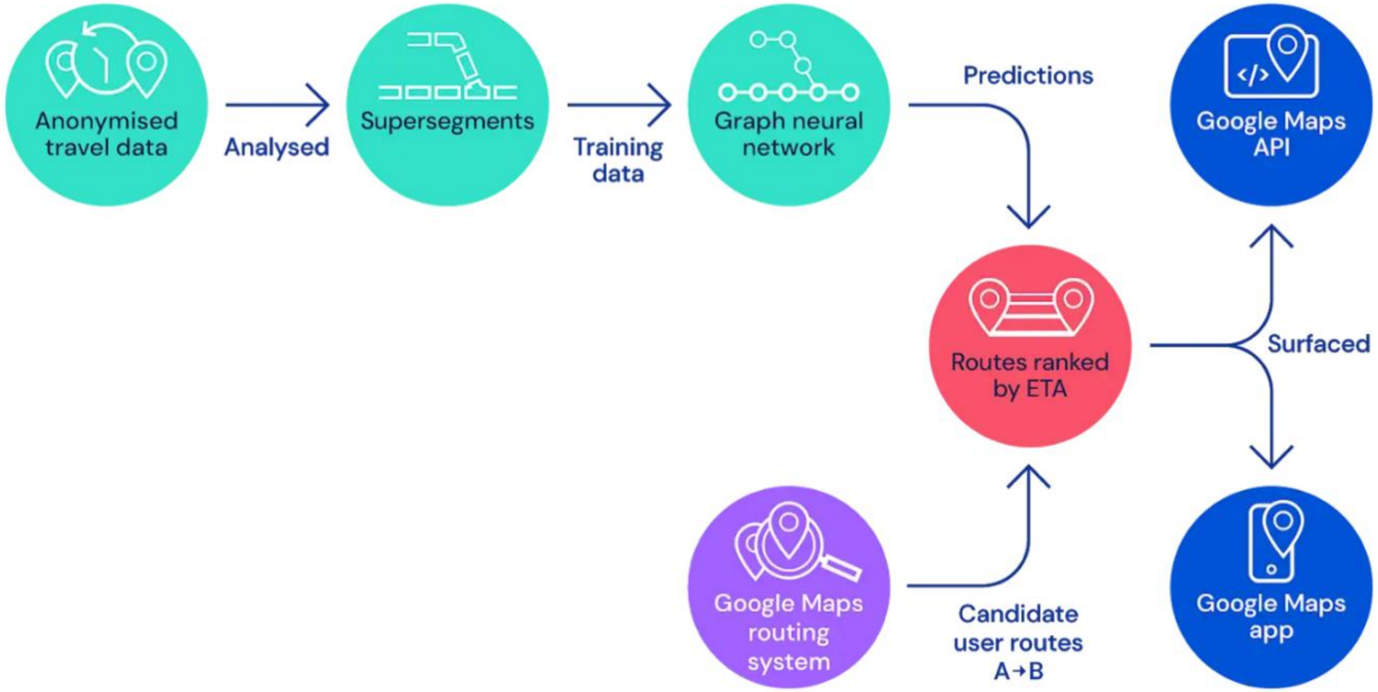
Polypharmacy Side effects

- **Given a pair of drugs predict adverse side effects.**
- How likely will Simvastatin and Ciprofloxacin, when taken together, break down muscle tissue?



Traffic Prediction using GNNs

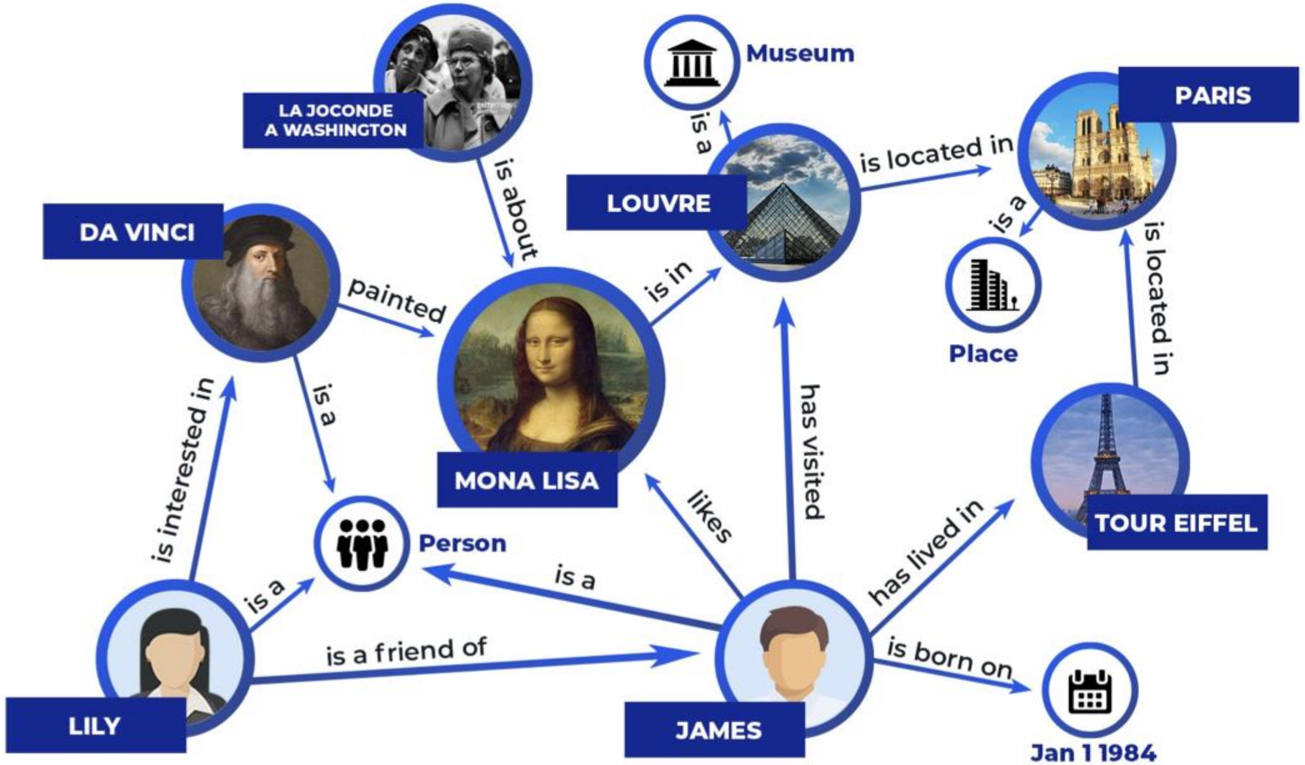
- **Nodes:** Road segments
  Edges: Connectivity between road segments
- **Predicting Time of Arrival with Graph Neural Networks**



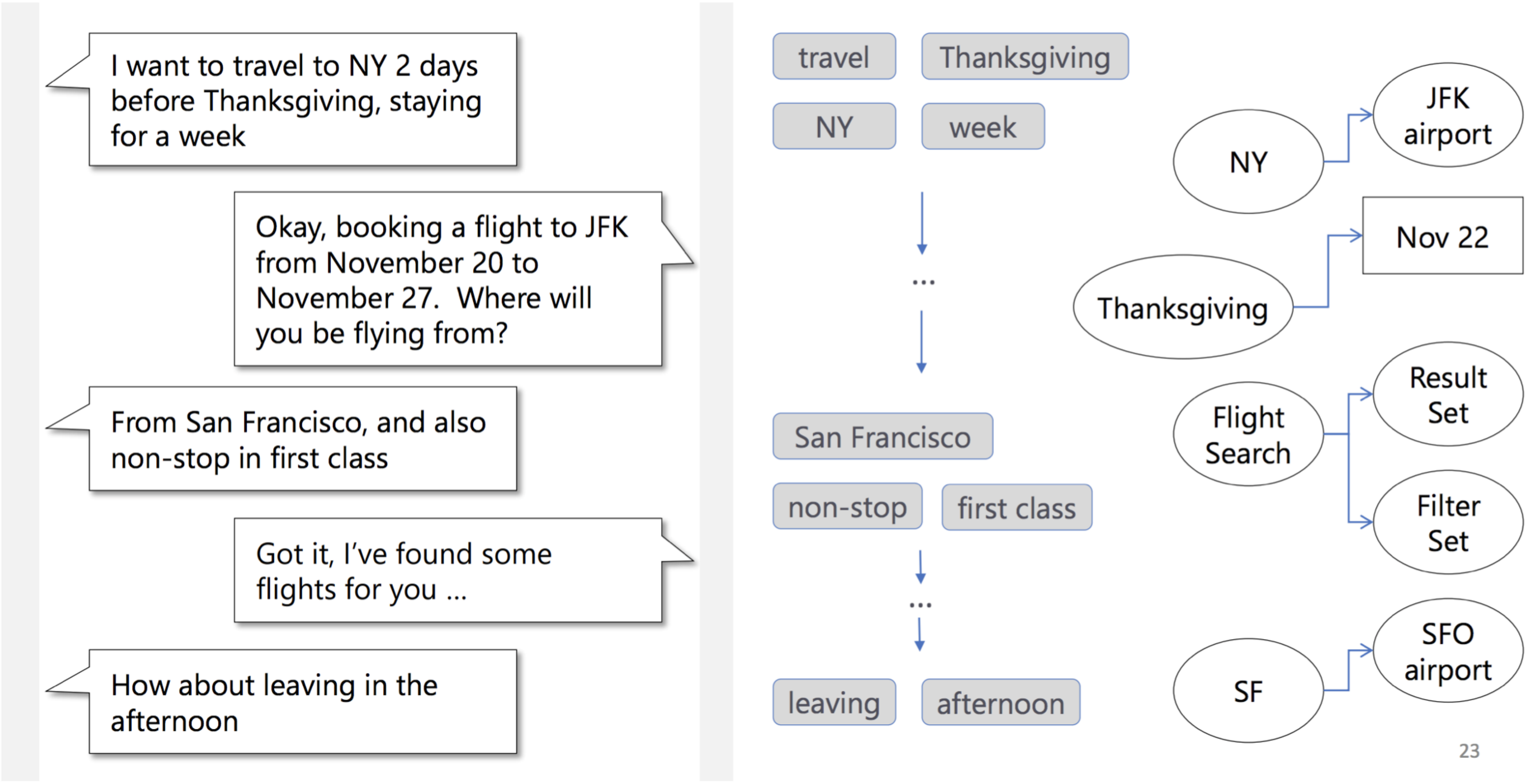
By DeepMind

Question Answering

Question: **“What are all the country capitals in Africa?”**



Question Answering



Information Retrieval

- Knowledge graphs are used to understand the meanings of query terms and identify documents that match the meanings

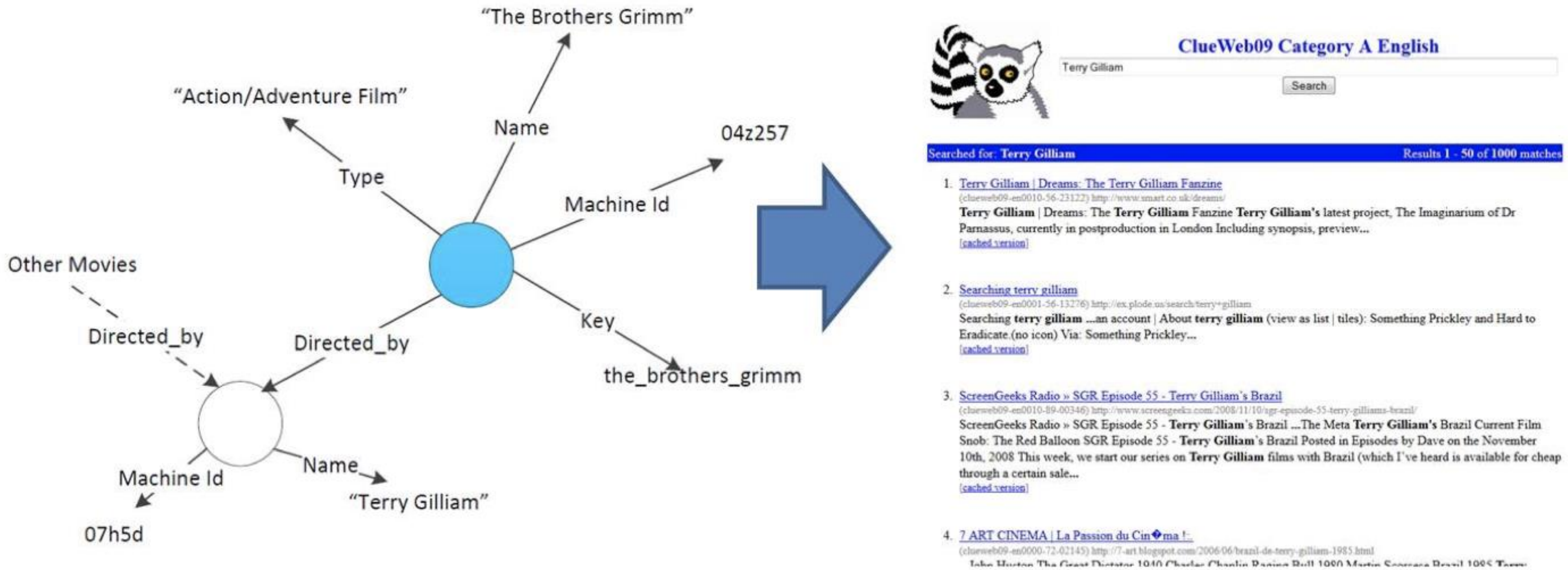
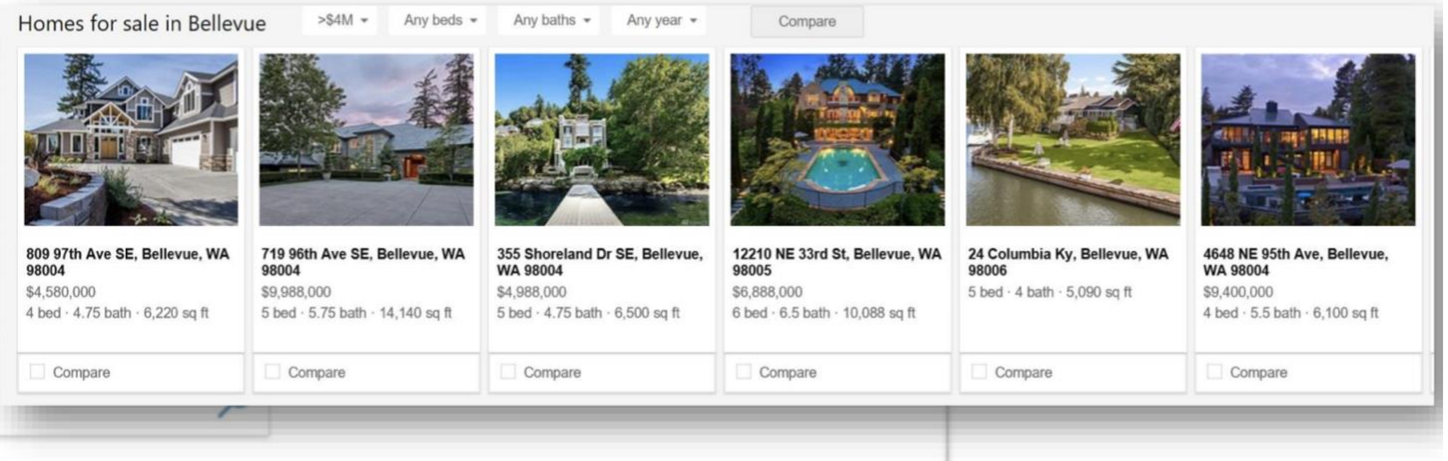
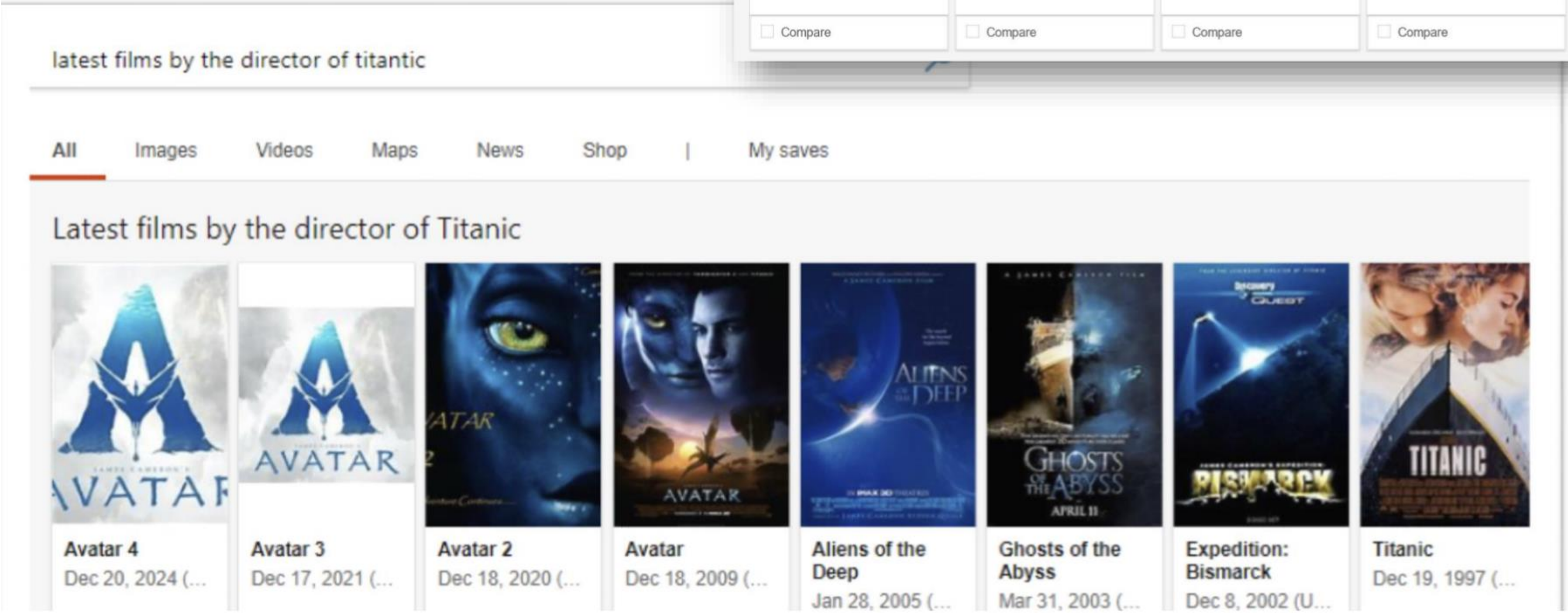


Figure from http://www.cs.cmu.edu/~callan/Projects/IIS-1422676/

Serving Information





Reasoning

- Knowledge graphs are usually incomplete. Many facts are missing
- A fundamental task: **predicting missing links (or facts) by reasoning on existing facts**
- The Key Idea: leverage **logic rules** for reasoning on knowledge graphs implicitly or explicity
- Example:

Barack_Obama **BornIn** United_States

$\Downarrow$

Barack_Obama **Nationality** American

Parents of Parents are Grandparents

Outline

- Why and What?
- Examples
- Machine/Deep Learning for Graphs.
- Tasks
- Applications!
- Datasets
- Embeddings Methods

Popular Graph Datasets

Table 5: Node embedding datasets

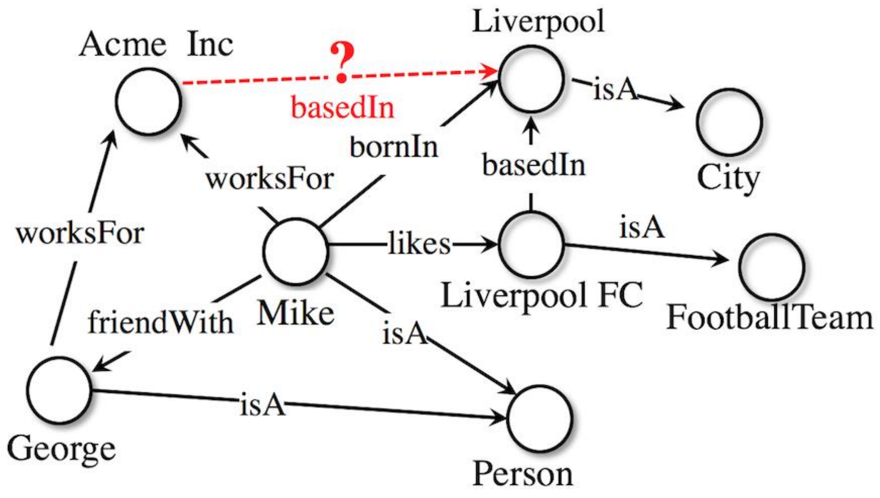
Dataset	$\|V\|$	$\|E\|$	k	**#Classes**
BlogCatalog	10'312	333'983	N/A	39
Wiki	4'777	184'812	N/A	40
PPI	3'890	76'584	N/A	50
Cora	2'708	5'429	1'433	7
Citeseer	3'327	4'732	3'703	6
Pubmed	19'717	44'338	500	3
Reddit	231'443	11'606'919	602	41
NELL	65'755	266'144	5'414	210

Table 6: Graph embedding datasets

Dataset	**Avg.** $\|V\|$	**Avg.** $\|E\|$	**#Graphs**	k	**#Classes**
MUTAG	17.93	19.79	188	1	2
ENZYMES	32.63	62.14	600	18	6
DD	284.32	715.66	1'178	1	2
COLLAB	74.49	2'457.78	5'000	N/A	3
IMDB	13	65.94	1'000	N/A	3

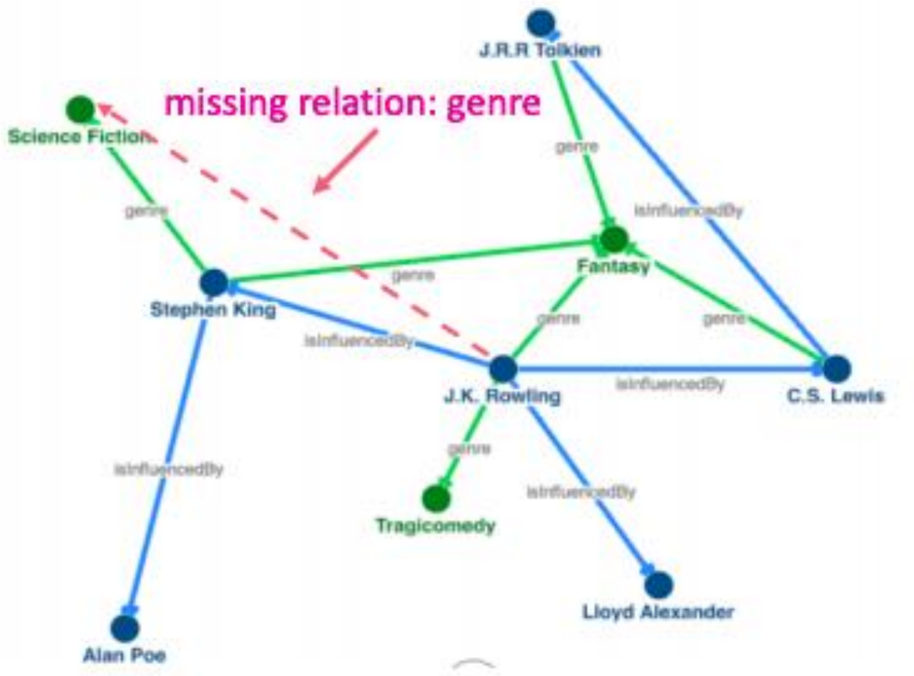
KG Datasets

- Publicly available KGs:
  - FreeBase, Wikidata, Dbpedia, YAGO, NELL, etc.
- Absence of a fact does not imply fact is false. We simply do not know.
- Common characteristics:
  - **Massive**: millions of nodes and edges
  - **Incomplete**: many true edges are missing



KG Completion

- Given an enormous KG, can we complete the KG / predict missing relations?
  - For a given (**head**, **relation**), we predict missing **tails**.



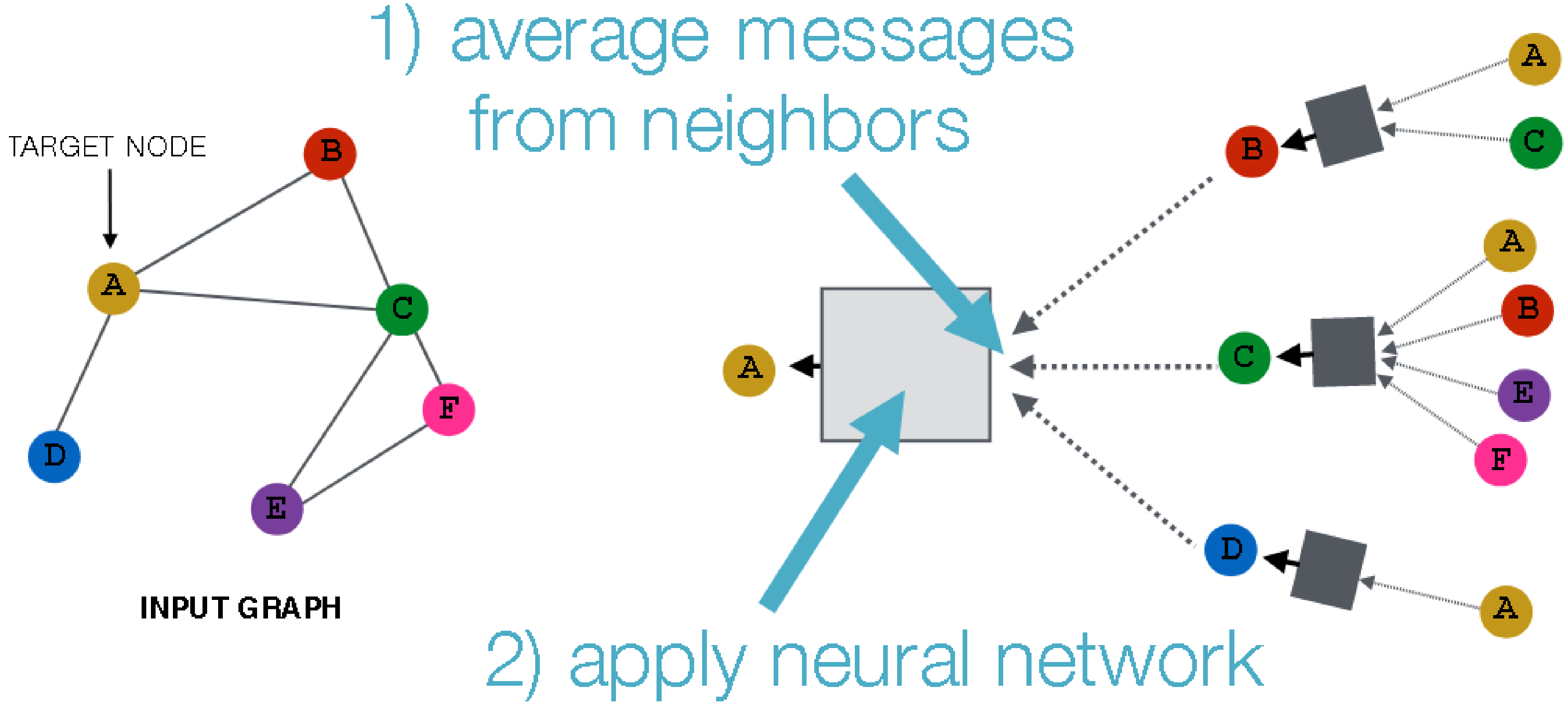
Outline

- Why and What?
- Examples
- Machine/Deep Learning for Graphs.
- Tasks
- Applications!
- Datasets
- Embeddings Methods

Graph Representation Learning Methods

- Node Representation/Graph Feature based Methods
  - DeepWalk, node2vec
- Graph Neural Networks (GNNs)
  - GCNs, Graph Attention Networks
- Knowledge Graph Embeddings (KGE)
  - Non-graph based models: TransE, TransR
  - Graph-based KG Embeddings: RGCN

General Perspective on GNNs



Thank you!