Graph Learning

Topic Preview

Anand Iyer

api@microsoft.com

USENIX OSDI/ATC, July 2021

Microsoft | Research

Graph Learning at **OSDI/ATC '21**

Two exciting sessions on graphs / learning on graphs

- **ATC: Wednesday July 14**
  - **Track 2: Searching for Tracks: Graphs**
  - From 10:30am to 12:00noon

- **OSDI: Friday July 16**
  - **Graph Embeddings and Neural Networks**
  - From 10:15am to 11:30am

All the times listed are in **Pacific Daylight Time (PDT)**

Graph Learning

=

Machine Learning

+

Graph-structured Inputs

Why Learn on **Graphs?**

Many real-world datasets are graphs!

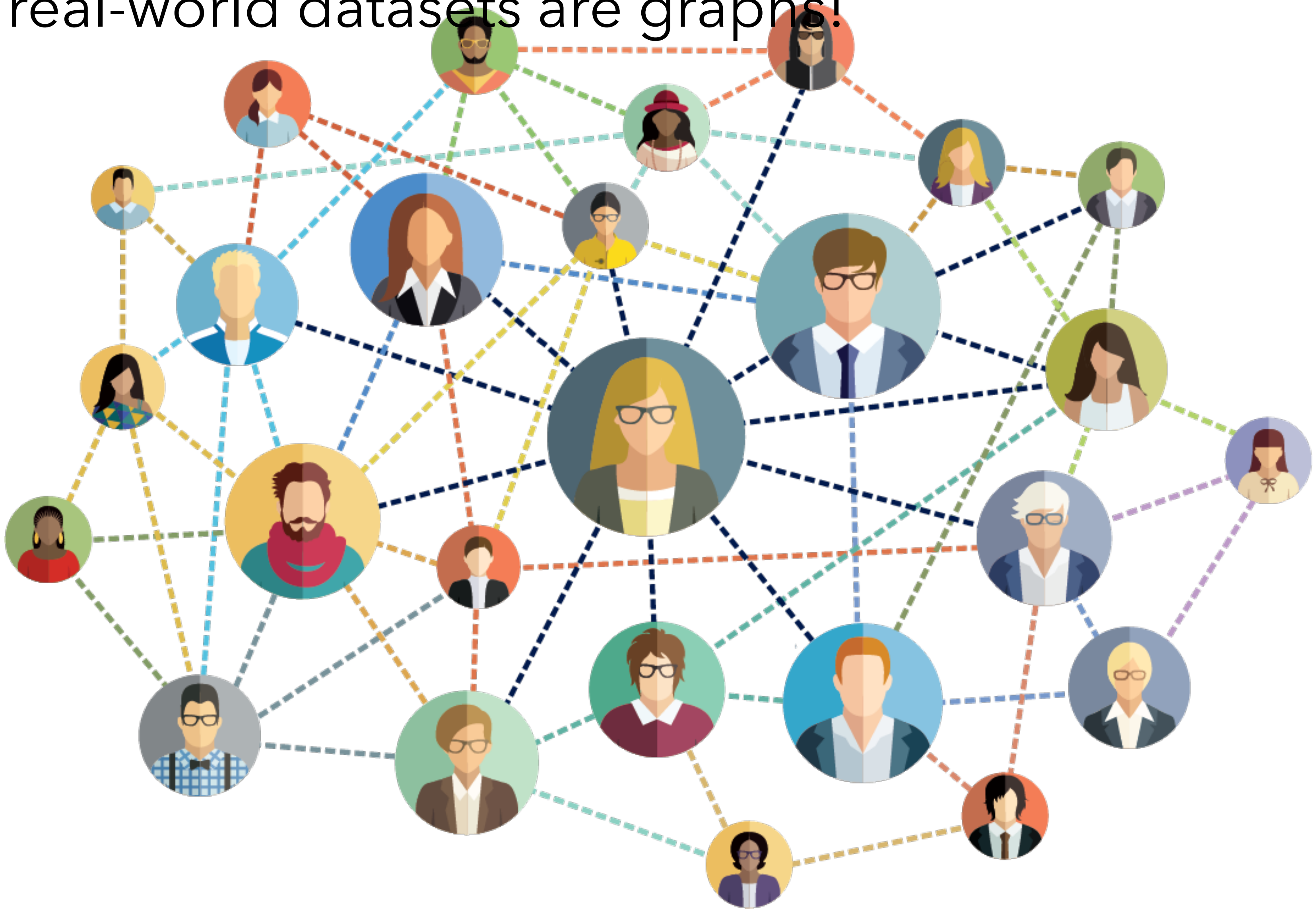


Image credits: PinClipart

Why Learn on **Graphs?**

Many real-world datasets are graphs!

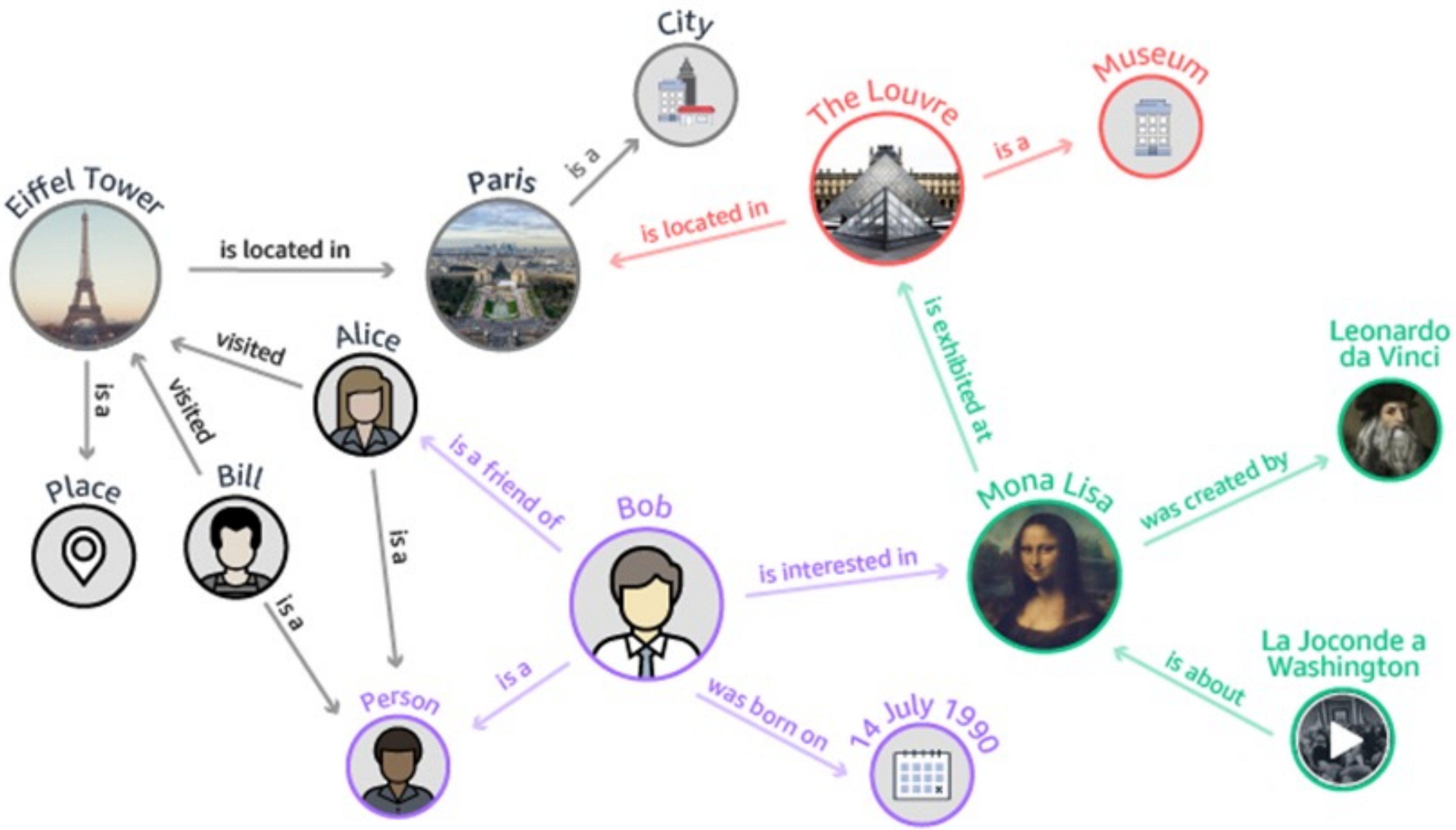


Image credits: Amazon

Why Learn on **Graphs?**

Many real-world datasets are graphs!

Image credits: PinClipart, NetworkWorld

Why Learn on **Graphs?**

Many real-world datasets are graphs!

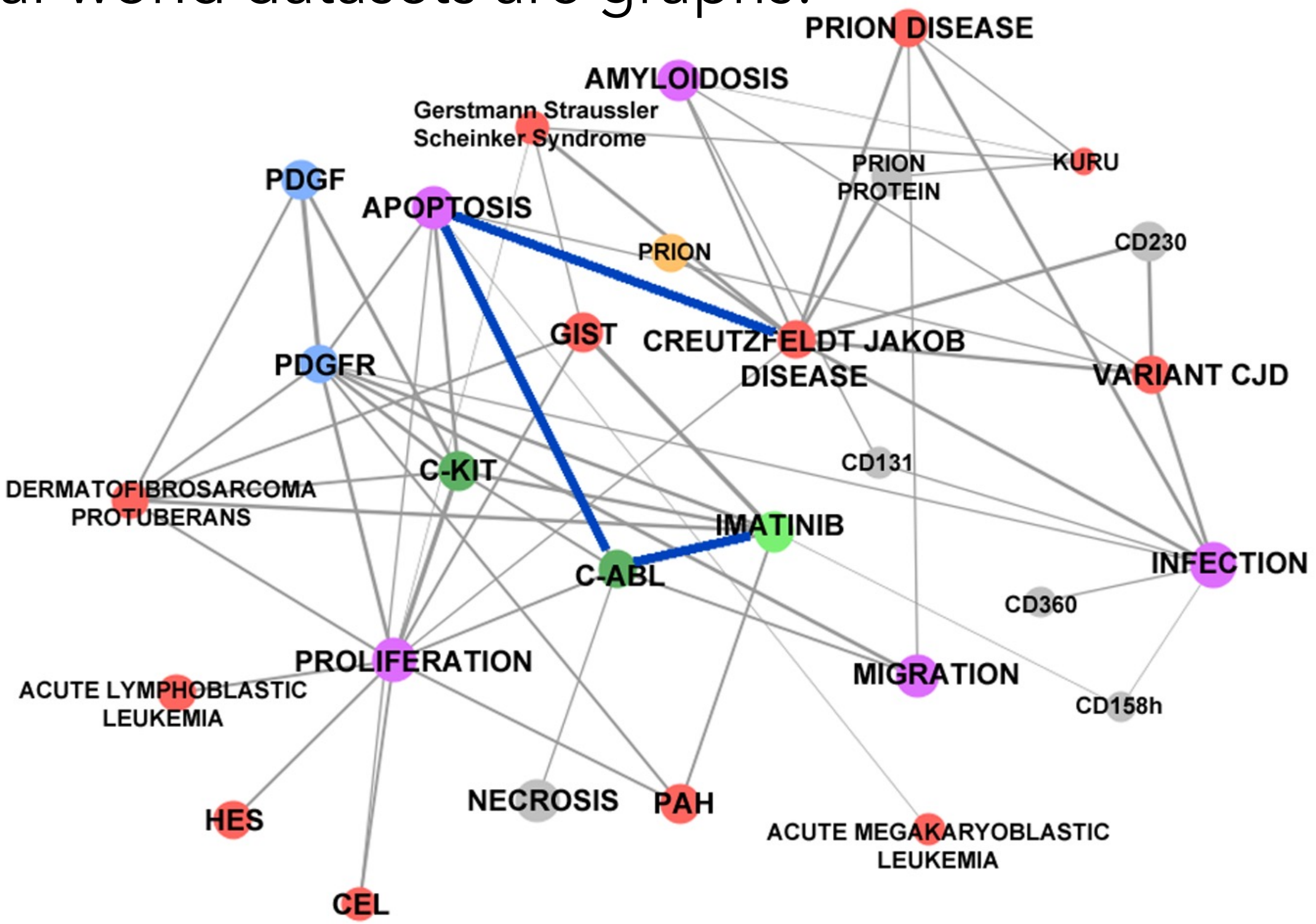


Image credits: PLOS

Graphs can capture the
rich & **complex relationships**
in the data in these domains

Graph learning can provide
significant performance boost!

Why Learn on **Graphs?**

Significant impact in a **wide variety** of domains

Chemistry
Discover chemical structures

Vision
Zero shot classification

NLP
Text classification, machine translation

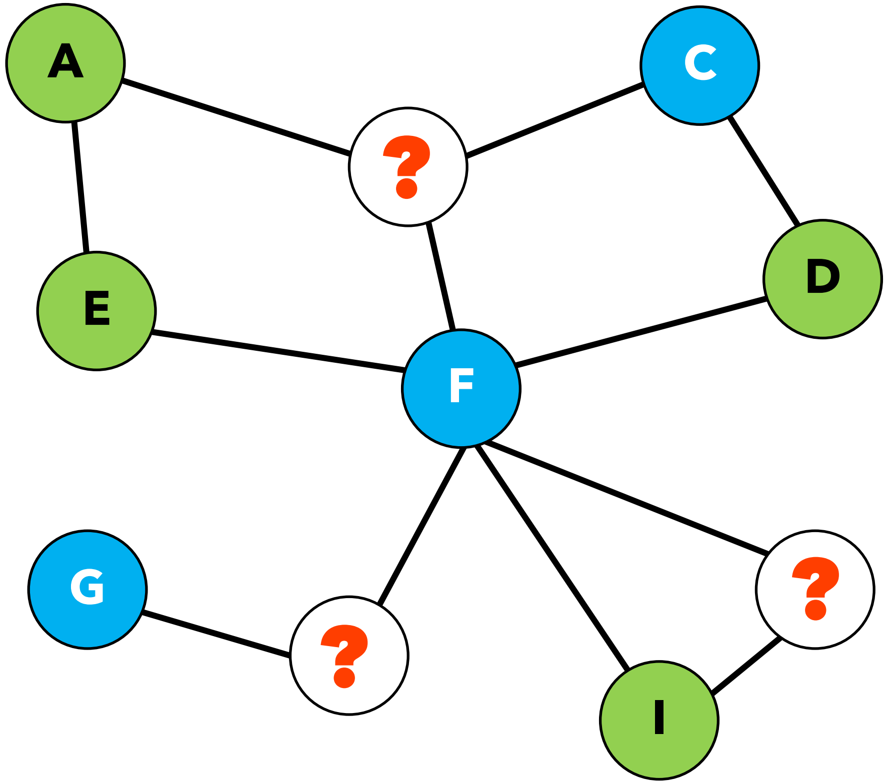
Biology
Disease classification, side effect prediction

Physics
Particle simulation, interaction networks

Many others
Program synthesis, knowledge graphs

Graph Learning: **Node Classification**

Predict labels for **new** or previously **unknown** nodes



Input Graph

Graph Learning: **Node Classification**

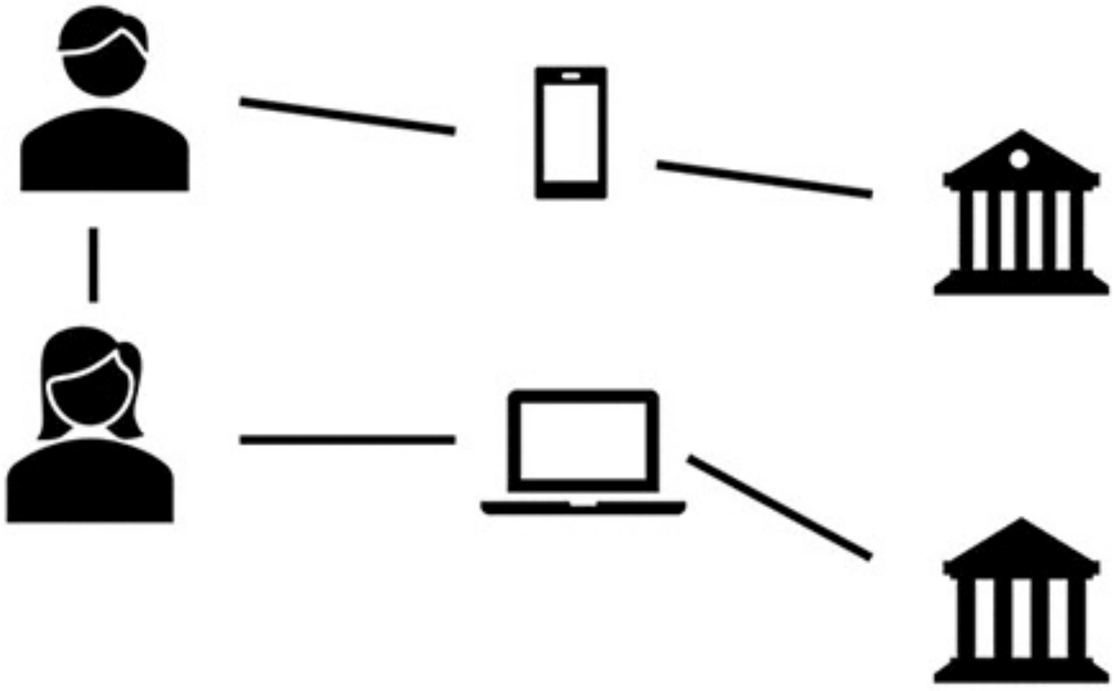


Image courtesy: Amazon

Graph Learning: **Node Classification**

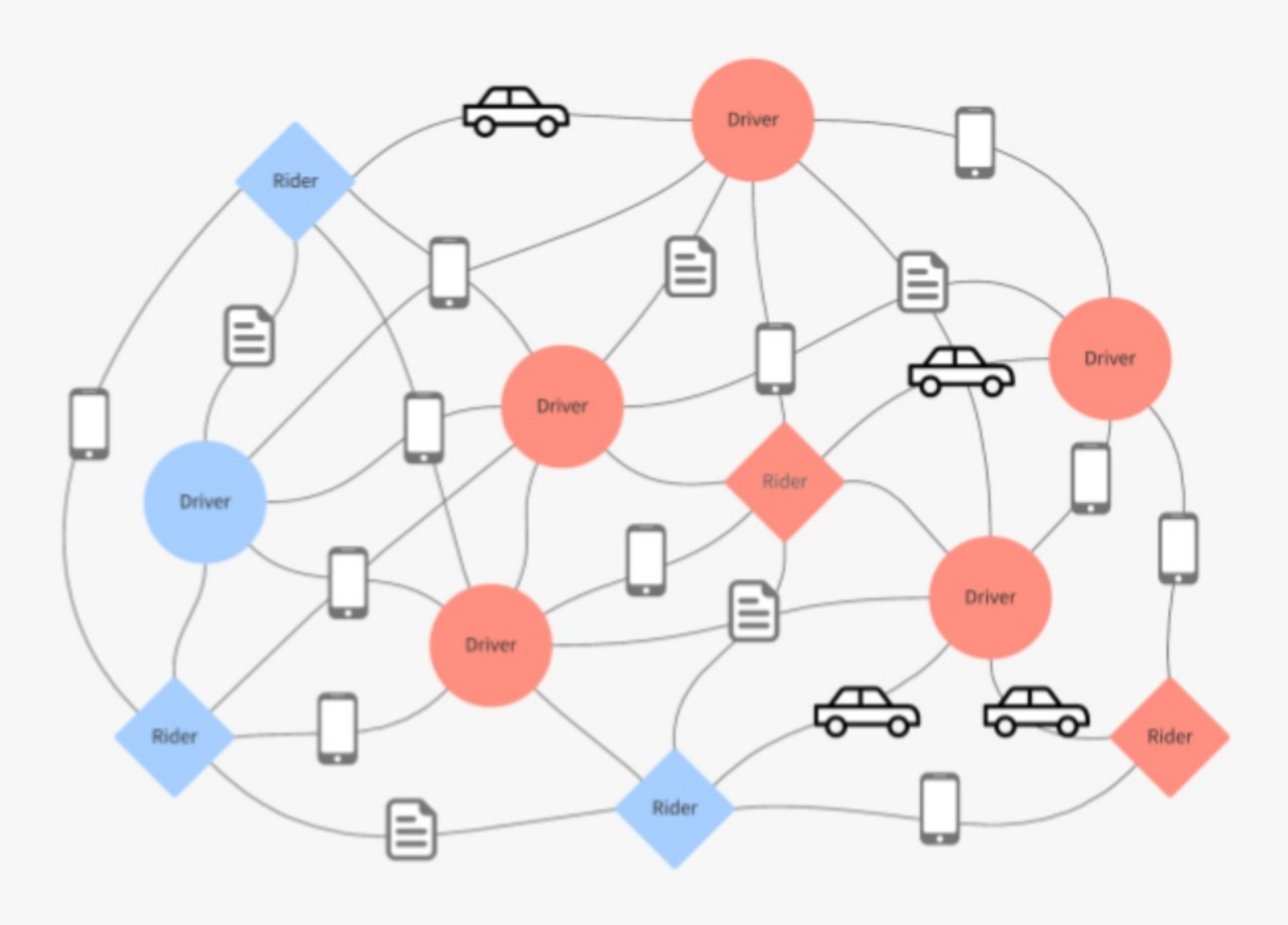
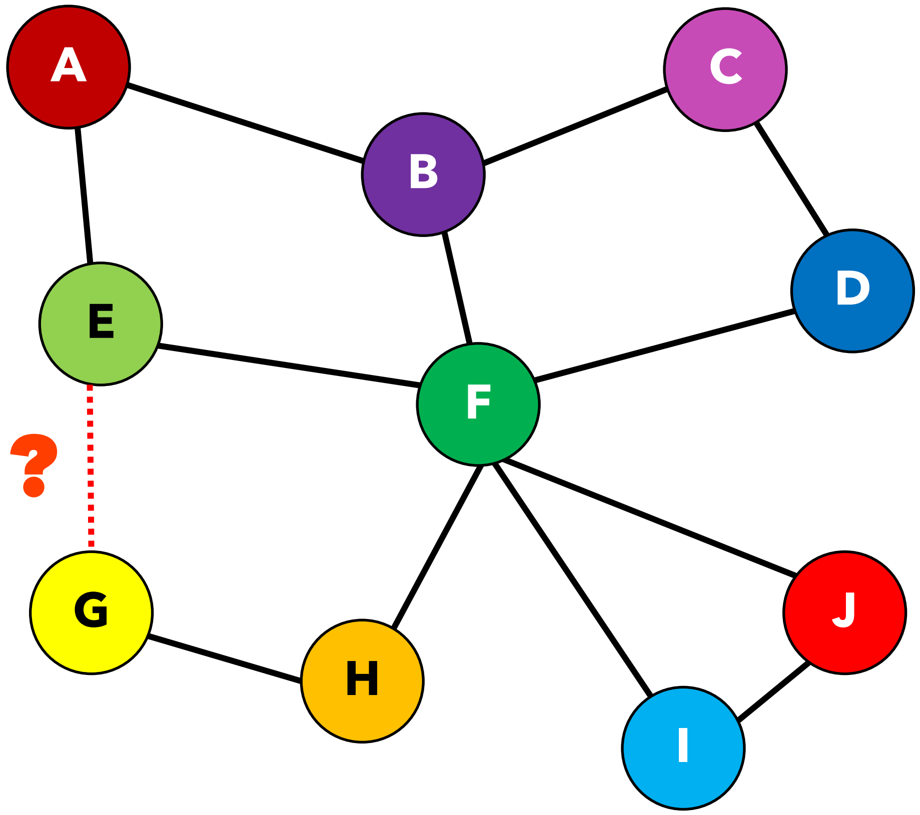


Image courtesy: Uber

Graph Learning: **Edge Prediction**

Predict existence of **link** between existing nodes



Input Graph

Graph Learning: **Edge Prediction**

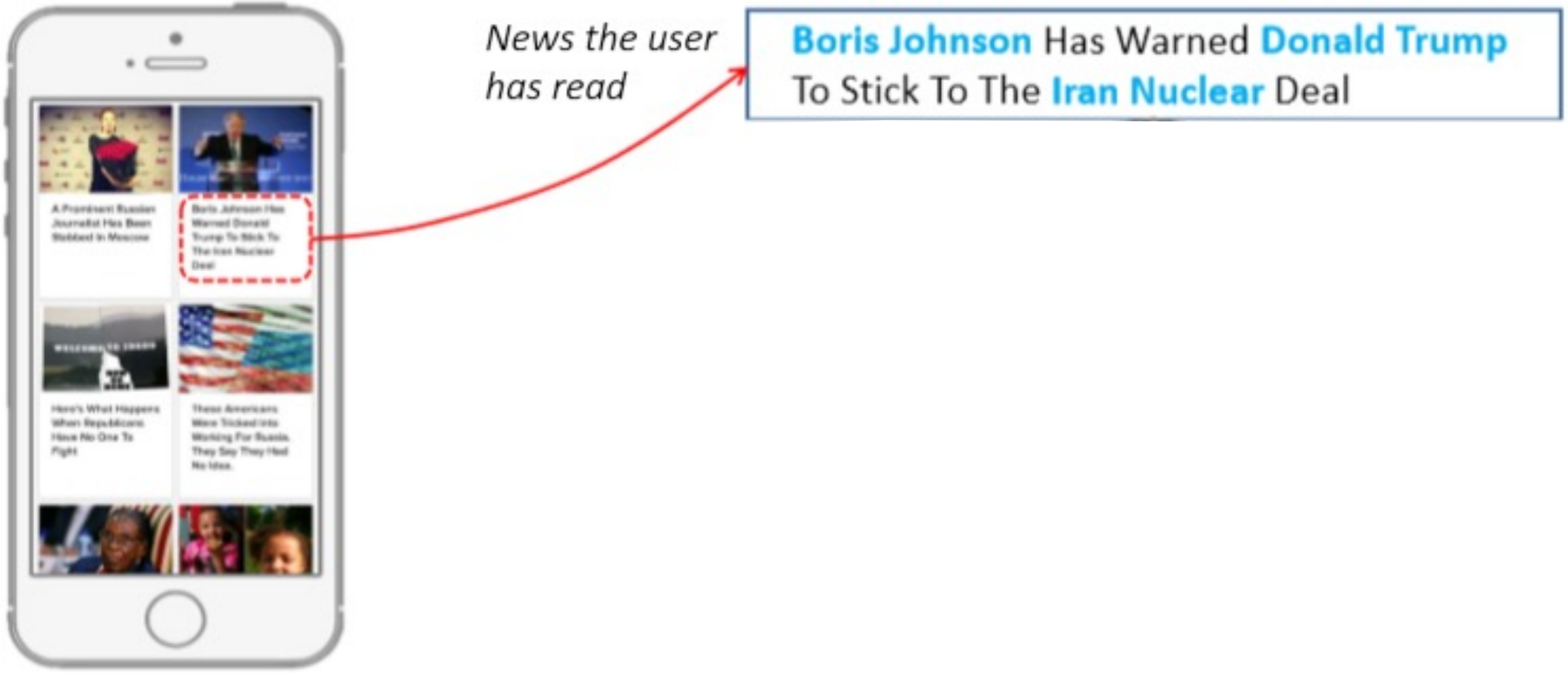
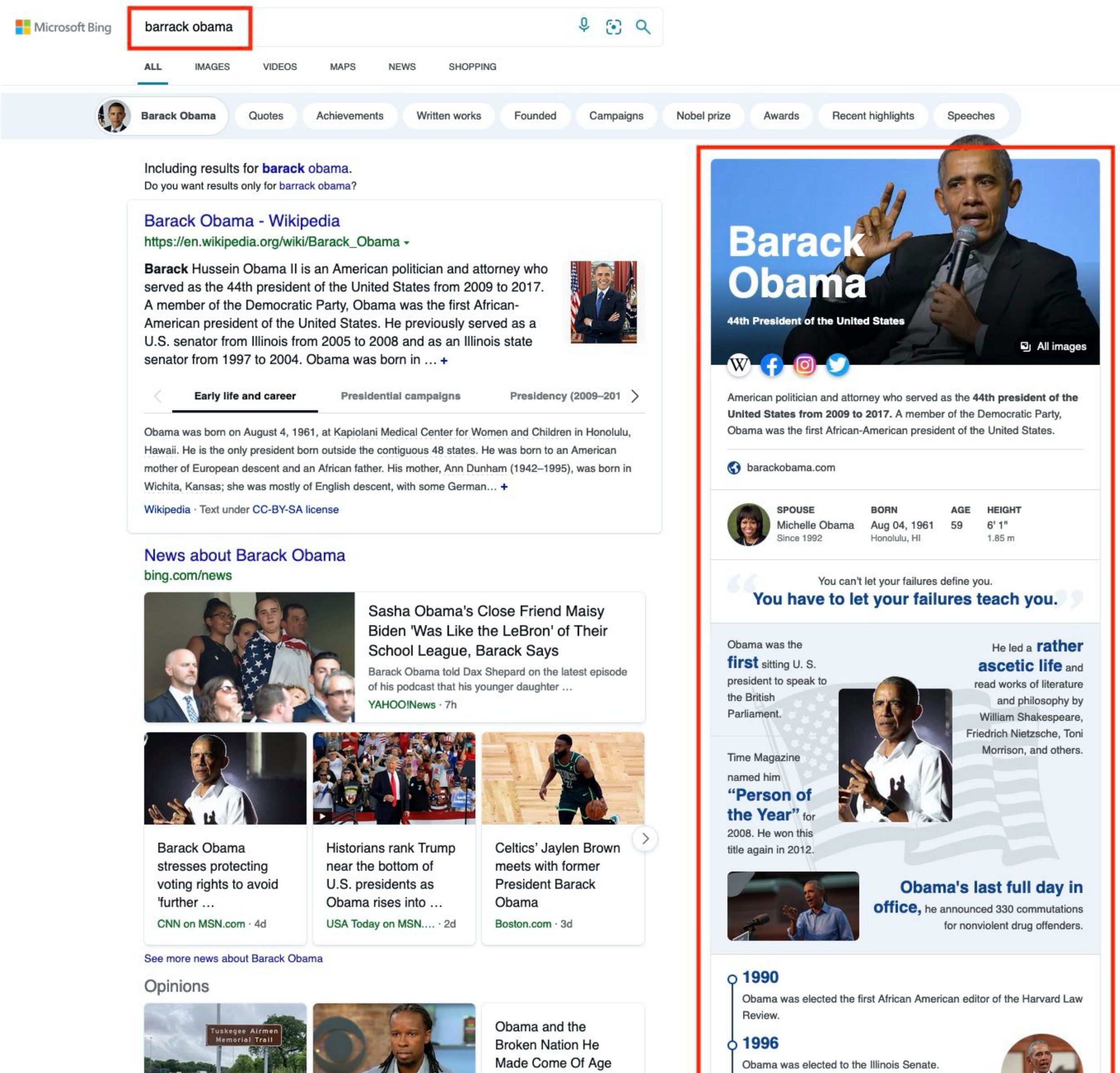
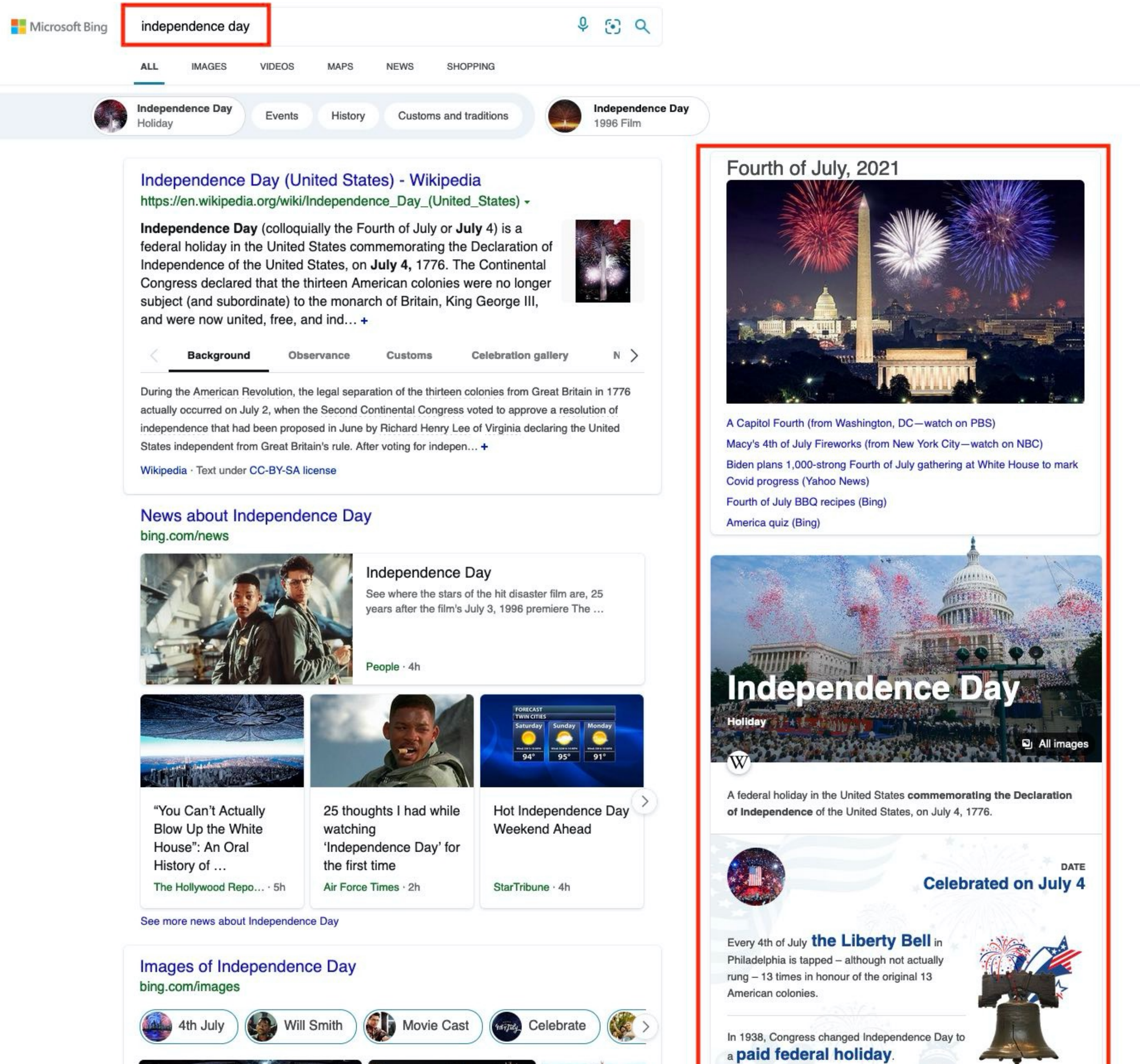


Image courtesy: Microsoft

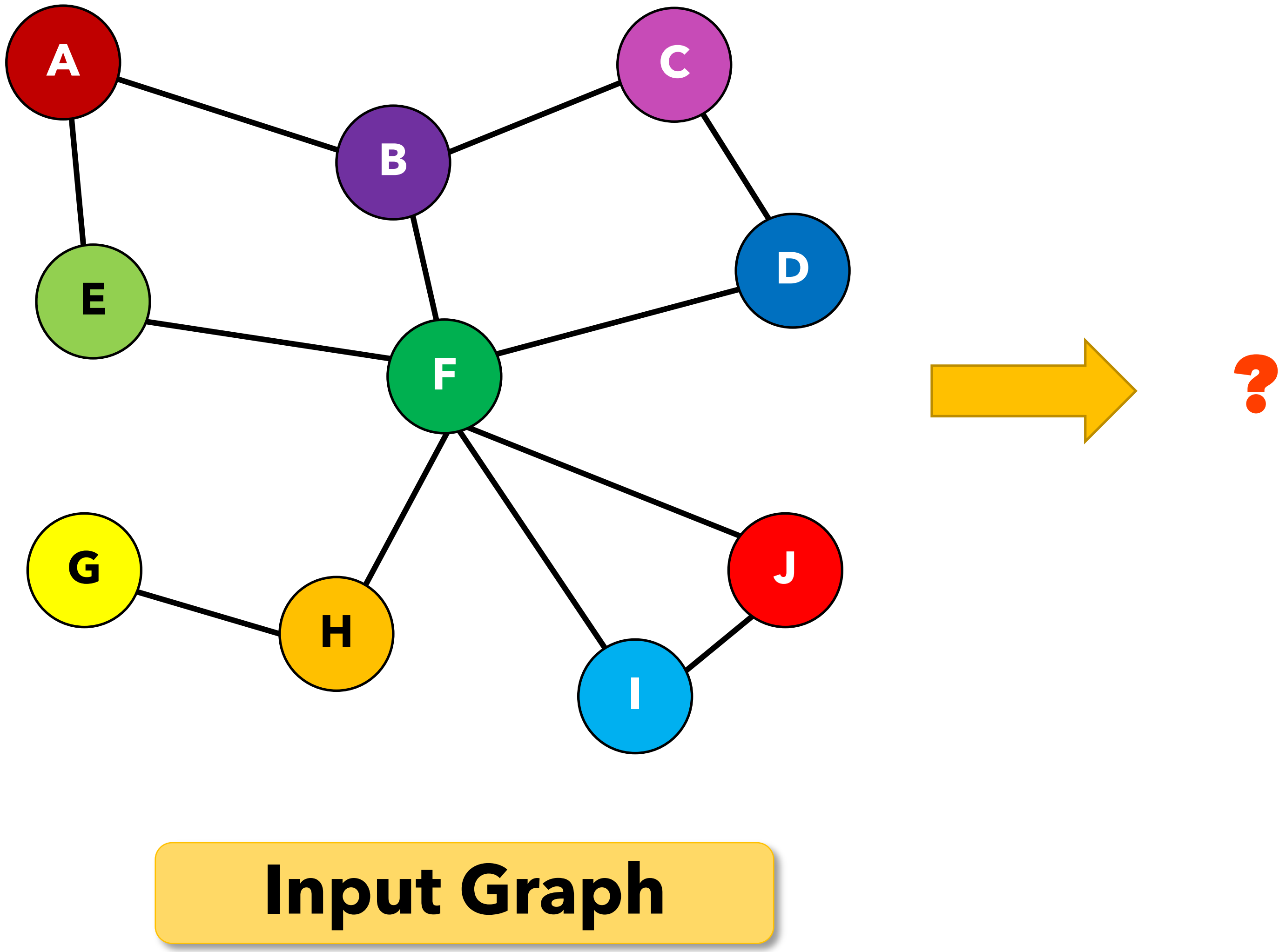
Graph Learning: **Edge Prediction**





Graph Learning: **Graph Classification**

Predict label for a **graph** or **subgraph**



Graph **Classification**

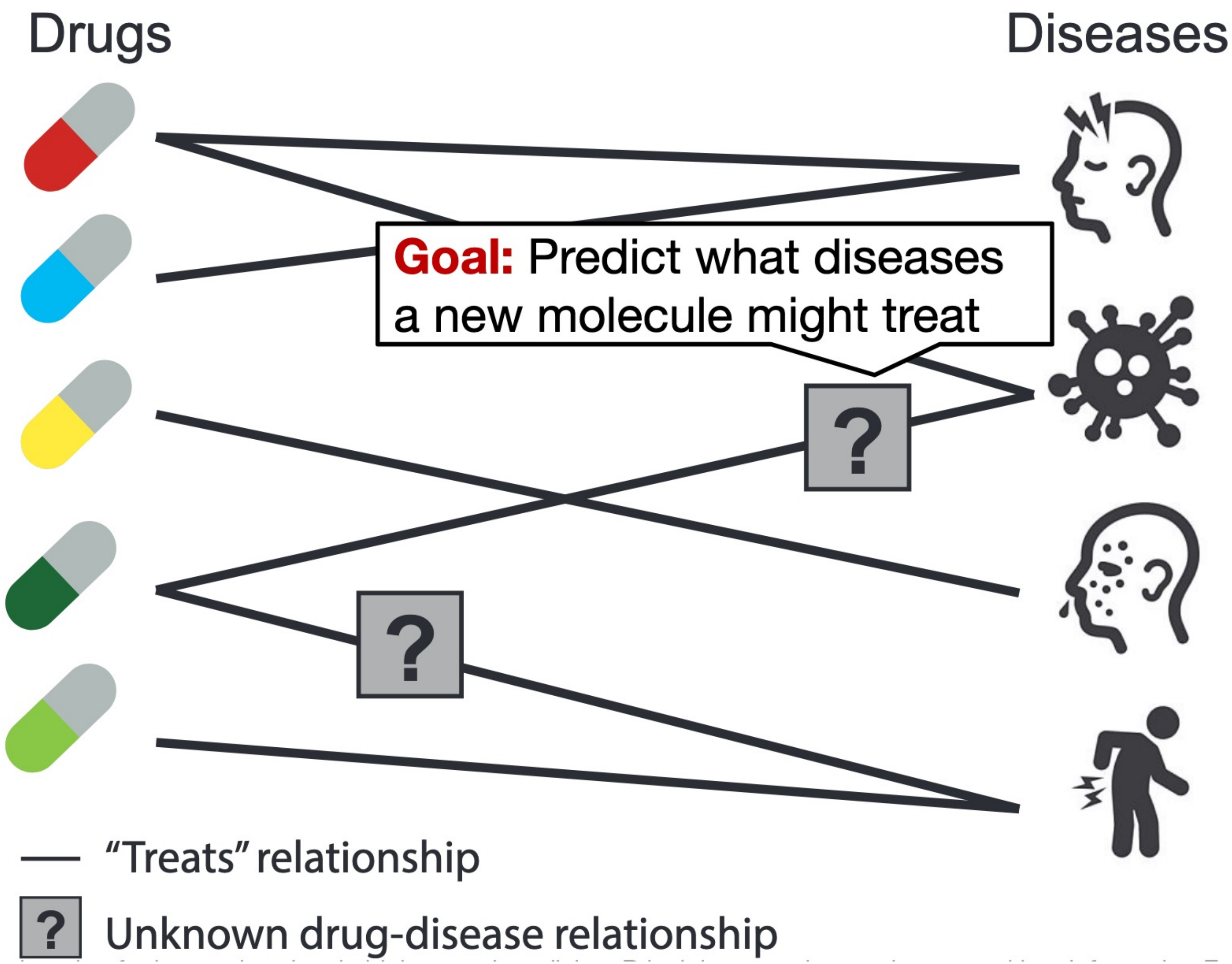
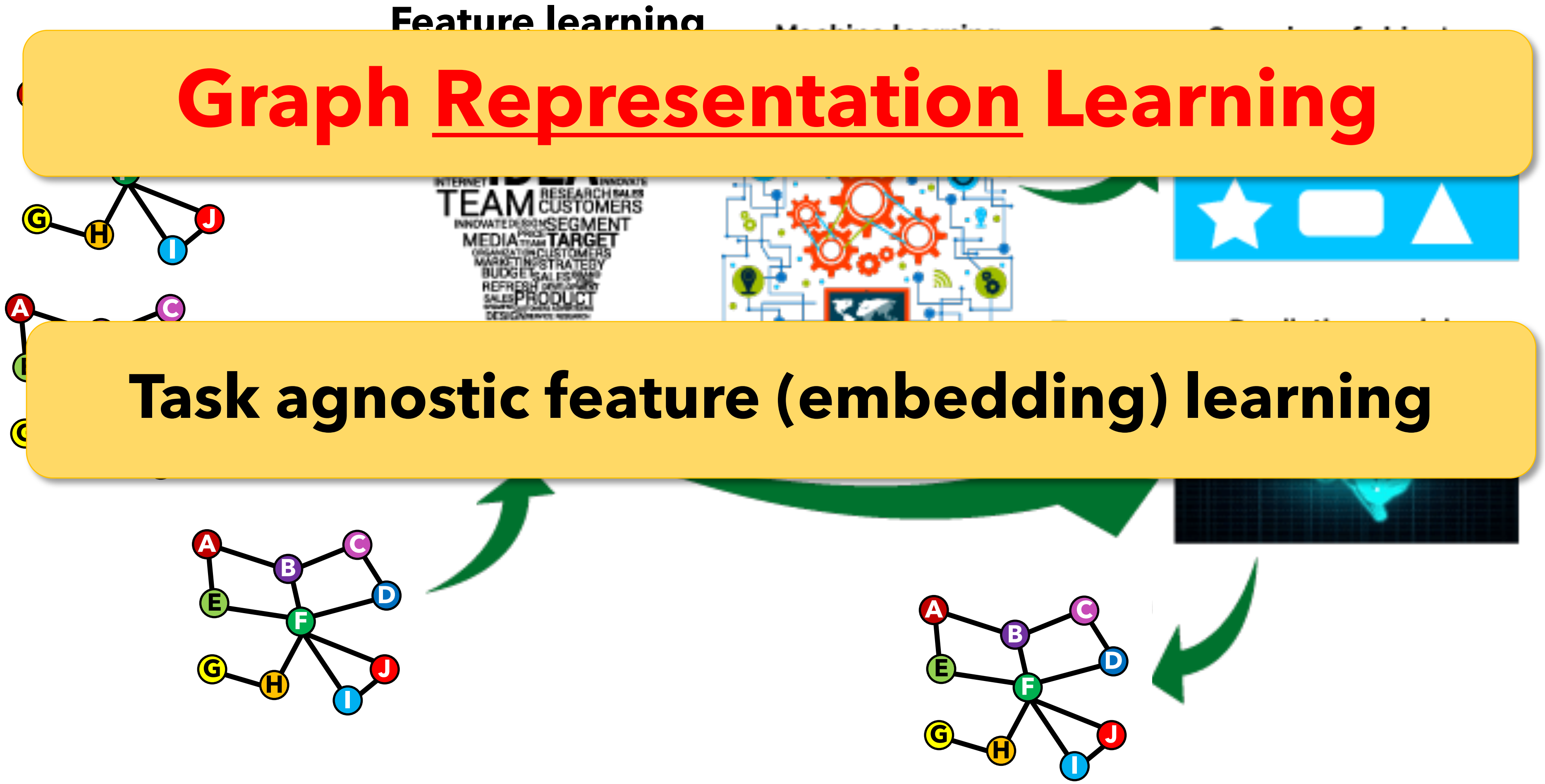


Image courtesy: Marinka Zitnik

Graph **Learning**



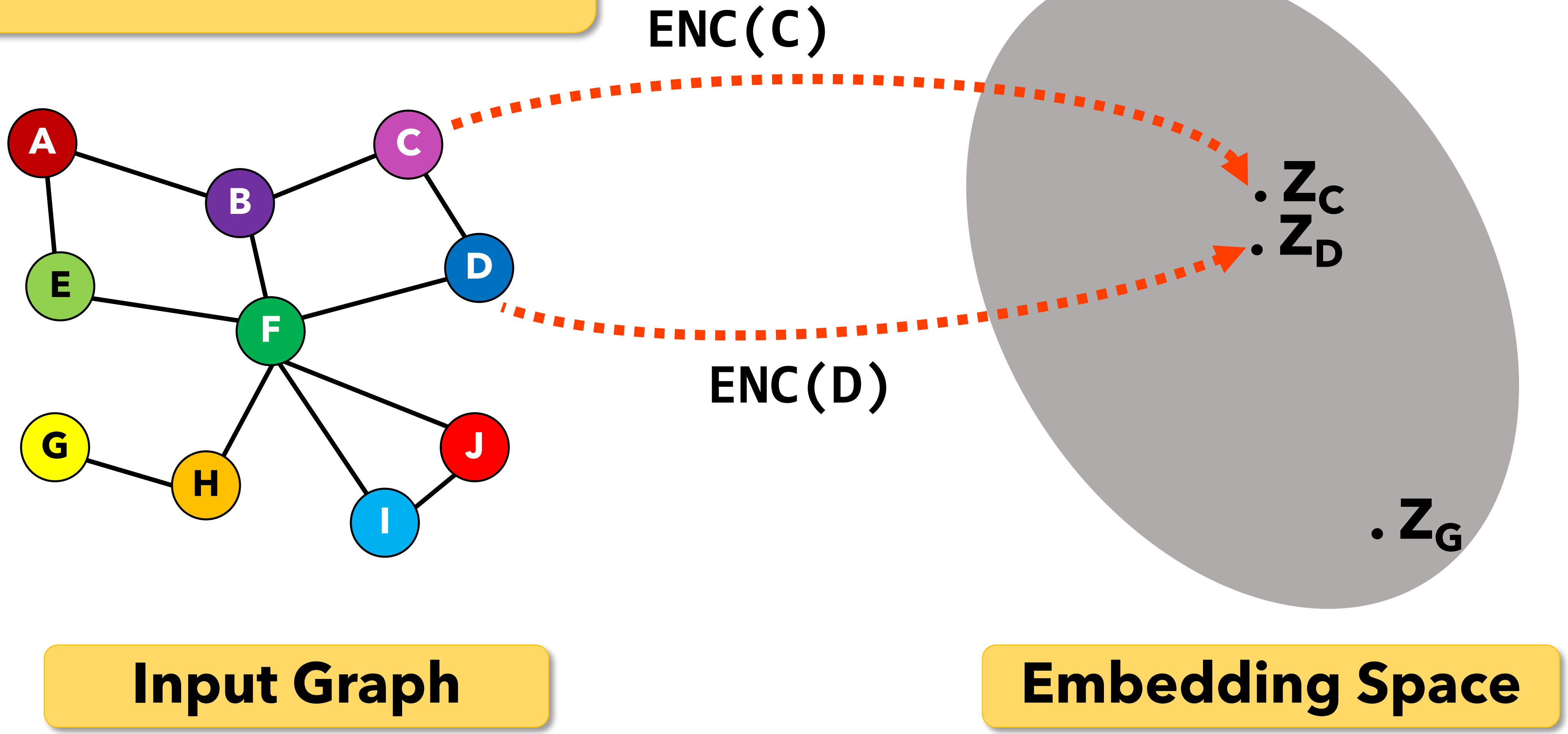
Graph <u>Representation</u> Learning

Task agnostic feature (embedding) learning

Image courtesy: towardsdatascience.com

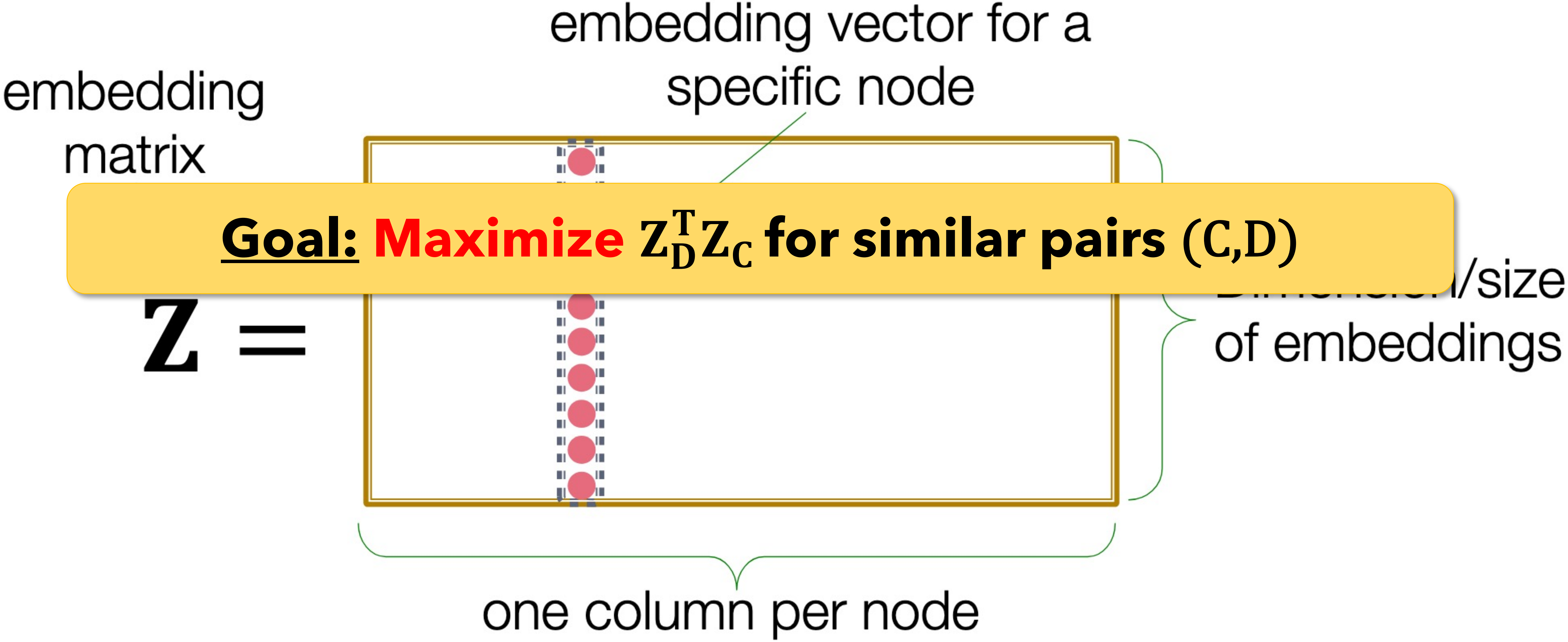
Graph **Embeddings**

$$\text{similarity}(C, D) \cong z_D^T z_C$$



Shallow **Encoding**

Encoder: simple **embedding lookup**



Goal: Maximize $z_D^T z_C$ for similar pairs (C,D)

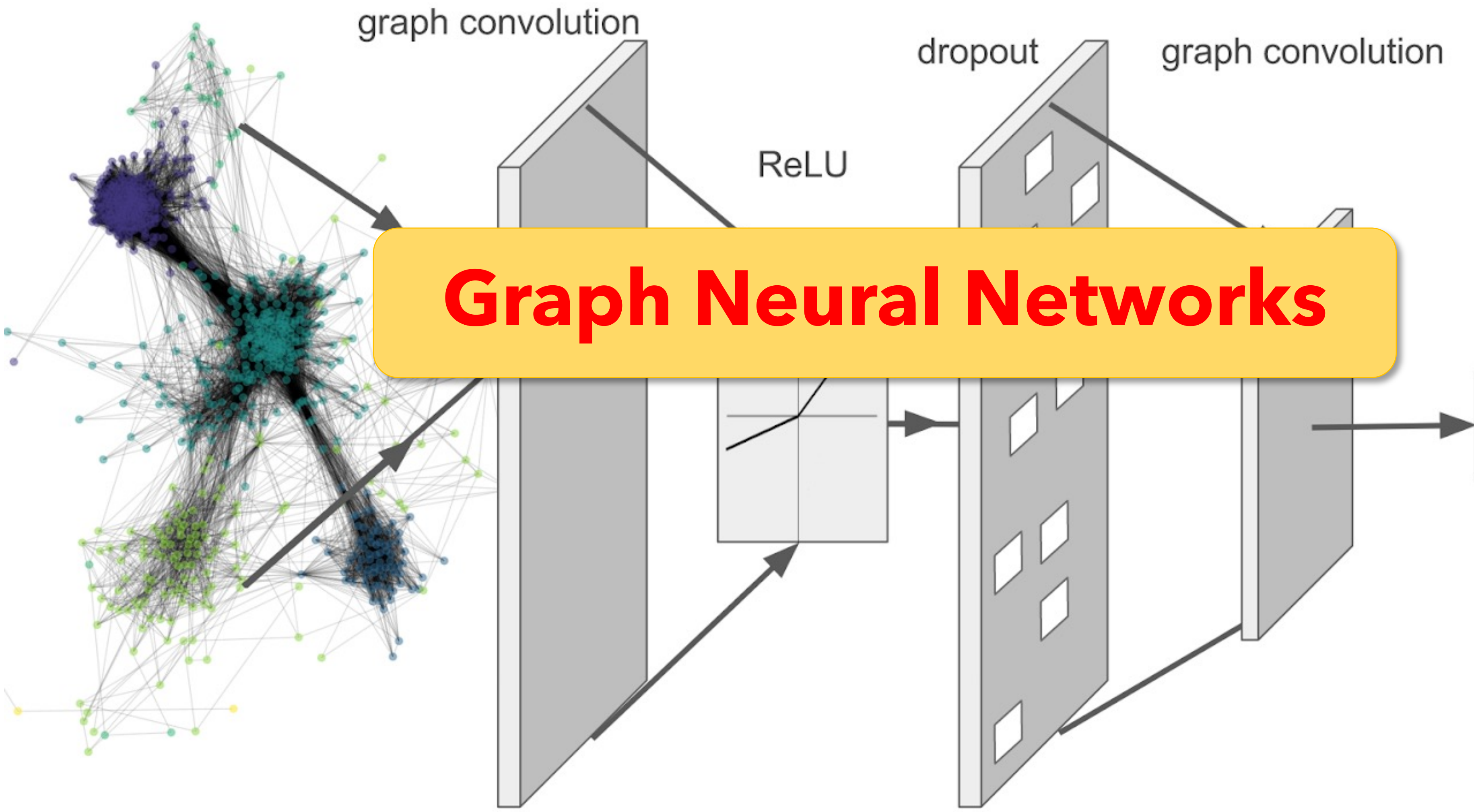
Image courtesy: Prof. Jure Leskovec

Shallow **Encoding**

- Simple approach
  - **Many proposals:** DeepWalk, node2vec, DistMult, ComplEx, …
  - Techniques differ in how **similarity** is defined
- Every node gets a **unique** embedding vector
  - Nothing shared across nodes
- **Transductive**: cannot generate embeddings for nodes not present at training
- Does not incorporate node features

Deep **Encoding**

Encoder: multiple layers of non-linear **transformations** on the graph

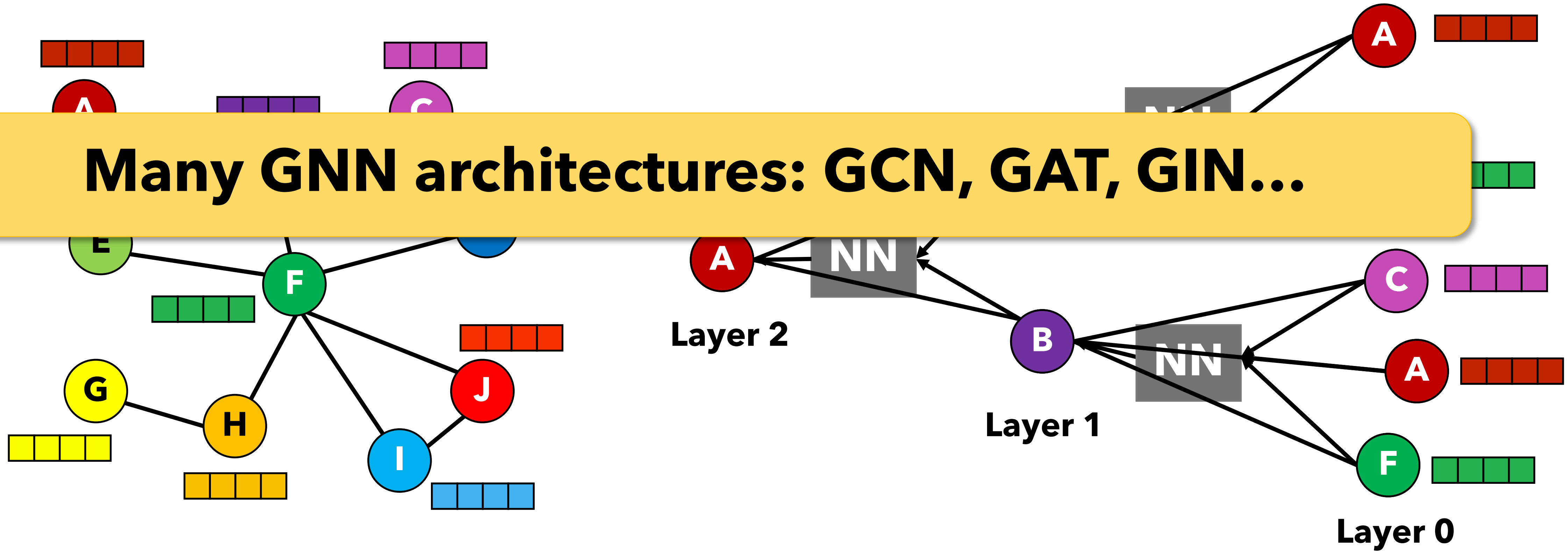


Graph Neural Networks

Embeddings

Deep **Encoding**

Graph Structure: **What** (to propagate?)
Neural Network: **How** (information is transformed)



Many GNN architectures: GCN, GAT, GIN...

Input Graph

2-hop neighborhood of A

Deep **Encoding**

$$h_v^0 = x_v$$

Initial 0-th layer embeddings = node features

Trainable weight matrices

$$h_v^{(l+1)} = \sigma\left(\boldsymbol{W_l} \sum_{u \in N(v)} \frac{h_u^{(l)}}{|N(v)|} + \boldsymbol{B_l}\, h_v^{(l)}\right), \forall l \in \{0, \dots, L-1\}$$

Non-linearity (e.g., ReLU)

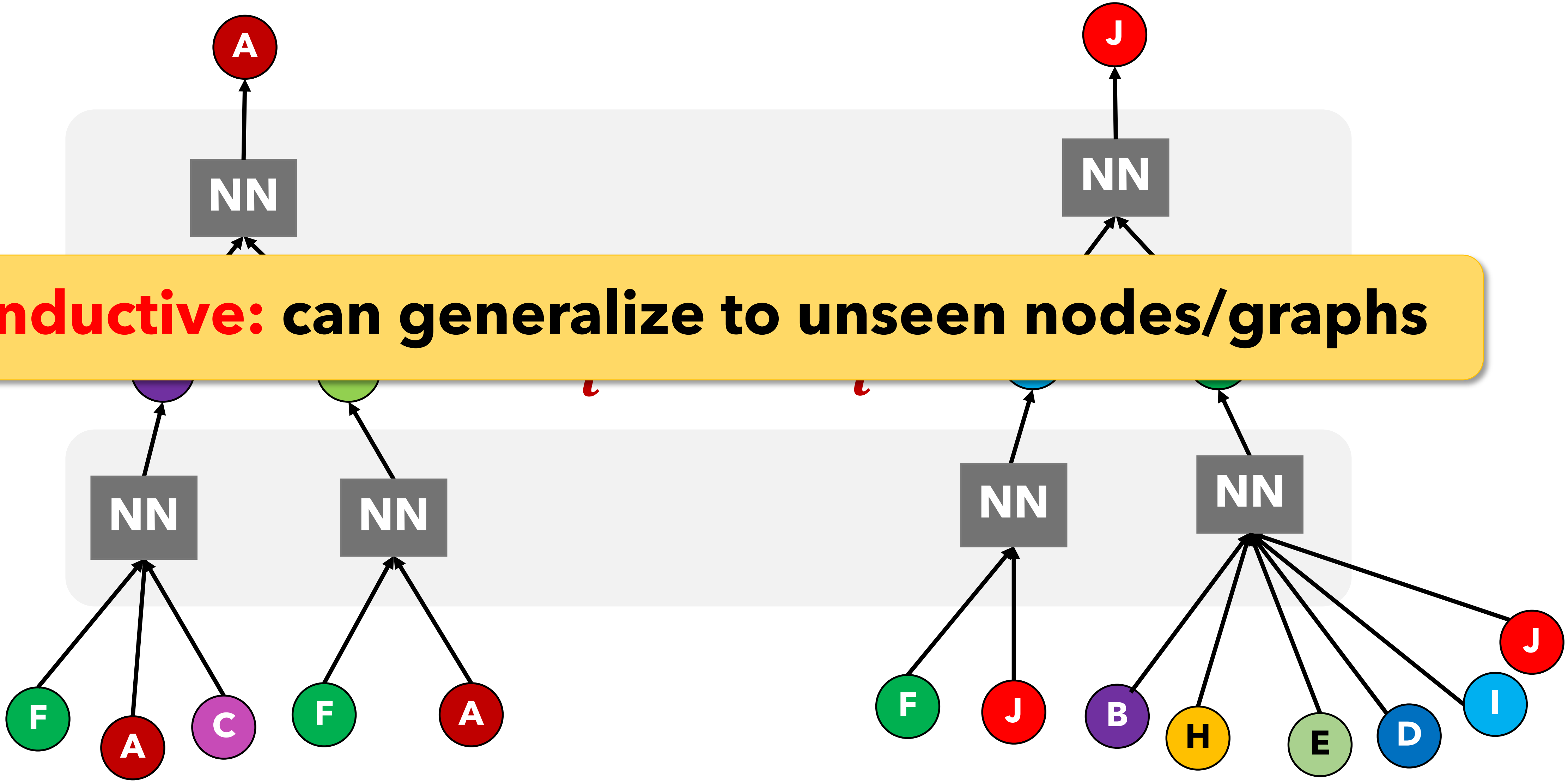
Average of neighbor's previous layer embeddings

Embedding of *v* at layer *l*

$$z_v = h_v^{(L)}$$

Embedding after L layers

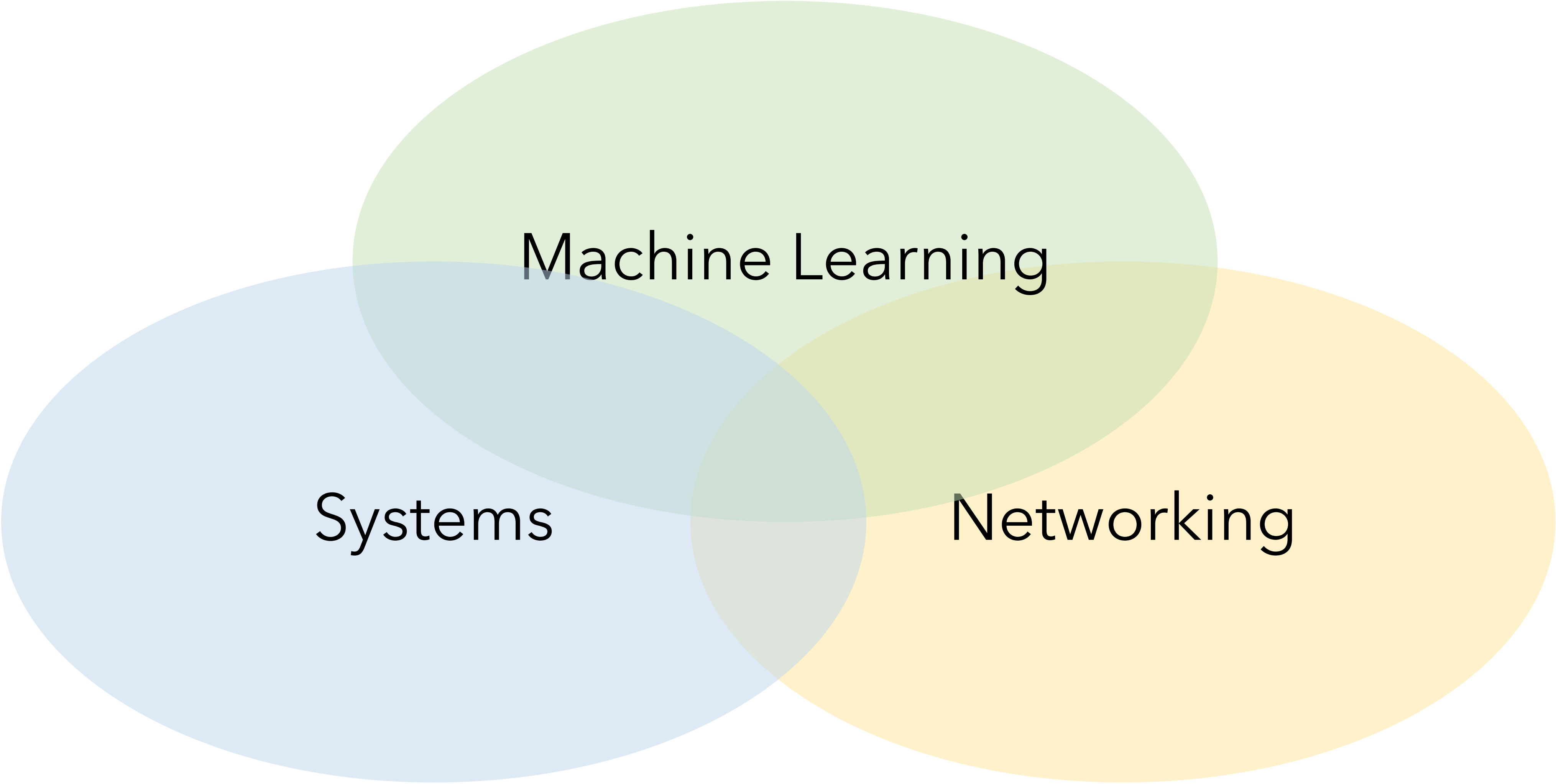
Deep **Encoding**



Inductive: can generalize to unseen nodes/graphs

Graph Learning **Research**

Active area of research



Graph Learning: **Trends**

Large Graphs

Billions of nodes,
trillions of edges

Hundreds or thousands of features

New Models

DistMult, ComplEx, GCN,
GAT, GIN, …

More **sophisticated**,
complex architectures

Graph Learning: **Systems Challenges**

- **Scalability**
  - How do we learn on **massive datasets**?
  - How do we handle dynamicity?
- **Efficiency**
  - Deep learning is **resource hungry**.
  - How do improve resource utilization and reduce cost?
- **Security**
  - Many domains have **sensitive data** (e.g., financial networks, healthcare)
  - How do we ensure robustness?

Graph Learning @ ATC '21

Wednesday, July 14
Track 2: 10:30 am – 12 noon

Deep Encoders

Scalability

Data movement

Graph learning techniques **move** large amounts of **data from storage to host memory**, incurring significant latency and power usage

A **graph learning accelerator inside storage** can provide **an order of magnitude** speedup!

GLIST: Towards In-Storage Graph Learning

Cangyuan Li [1,2], Ying Wang [1,2], Cheng Liu [1,2], Shengwen Liang [1,2], Huawei Li [1,2,3], Xiaowei Li [1,2]
SKLCA, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China [1]
University of Chinese Academy of Sciences, Beijing, China [2]
Peng Cheng Laboratory, Shenzhen, China [3]

Graph learning training needs **expensive GPUs** incurring high cost

Decomposing the training pipeline into fine-grained tasks allows training on a **combination of serverless threads and cheap CPUs** **faster than** **using GPUs**!

Dorylus: Affordable, Scalable, and Accurate GNN Training with Distributed CPU Servers and Serverless Threads

John Thorpe[†♣] Yifan Qiao[†♣] Jonathan Eyolfson[†] Shen Teng[†] Guanzhou Hu[†‡] Zhihao Jia[§]
Jinliang Wei[*] Keval Vora[♭] Ravi Netravali[♯] Miryung Kim[†] Guoqing Harry Xu[†]
UCLA[†] *University of Wisconsin*[‡] *CMU*[§] *Google Brain*[*] *Simon Fraser*[♭] *Princeton University*[♯]

Deep Encoders

Efficiency

Cost

Friday, July 16
10:15 – 11:30 am PDT

Graph Learning @ OSDI '21

One-size-fit-all optimizations do not work for emerging graph learning architectures, resulting in **poor performance**

Run-time that **adapts to the given workload and graph learning architecture** can improve performance **up to 4x**!

GNNAdvisor: An Adaptive and Efficient Runtime System for GNN Acceleration on GPUs

Yuke Wang, Boyuan Feng, Gushu Li, Shuangchen Li, Lei Deng, Yuan Xie, and Yufei Ding
University of California, Santa Barbara

Deep Encoders
Efficiency
Resource Utilization

Friday, July 16
10:15 - 11:30 am PDT

Graph Learning @ OSDI '21

Existing systems are bottlenecked by **data movement**, resulting in **inefficient** training and **poor performance**

Minimizing disk access by caching and data ordering and interleaving data movement with computation enables training of **billion edge graphs** on a **single** machine!

Marius: Learning Massive Graph Embeddings on a Single Machine

Jason Mohoney, Roger Waleffe, Henry Xu,* Theodoros Rekatsinas, Shivaram Venkataraman
University of Wisconsin-Madison

Shallow Encoders

Efficiency

Data movement

Friday, July 16
10:15 - 11:30 am PDT

Graph Learning @ OSDI '21

Existing distributed graph learning systems are bottlenecked by **communication**, resulting in **poor scalability**

A **hybrid training** approach can significantly **reduce** the need for communication, gracefully scale to **multiple billion edge graphs** and speed up training by up to **7x**!

P^3: Distributed Deep Graph Learning at Scale

Swapnil Gandhi*
Microsoft Research

Anand Padmanabha Iyer
Microsoft Research

Deep Encoders
Scalability
Communication

Friday, July 16
10:15 – 11:30 am PDT

Graph Learning @ OSDI '21

Graph Learning **Summary**

- Emerging field with **active** research
  - Inter-disciplinary: machine learning, systems and networking
  - Tremendous impact in many domains
- Significant opportunities for systems researchers
  - Increasing interest in the industry
  - Many open systems challenges
- Do checkout the papers in this year's ATC and OSDI!

Acknowledgements: Content based on Prof. Jure Leskovec's CS224W course at Stanford and Prof. William Hamilton's "*Graph Representation Learning*" book.