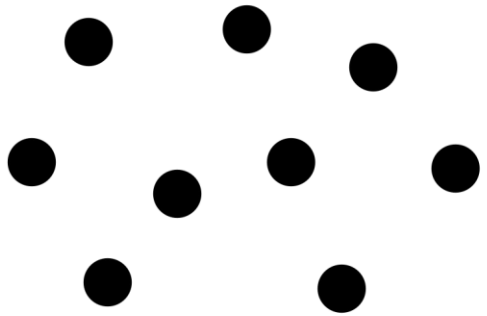
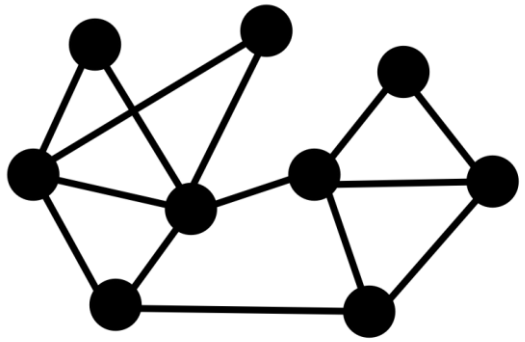
注意事项

- 一定要有充足的时间投入（建议时间投入比例1：10）
- 养成读代码的习惯，注重写代码或代码重现（论文、算法≠代码）
- 一定找到自己工作与GNN的切入点（即数据问题）
- 一定要有读论文的总结（别人用了哪些metrics、涉及到了哪些baseline）
- 学术问题在大群里提问

Graph Neural Network

What is a graph





What features?

Jurij Leskovec

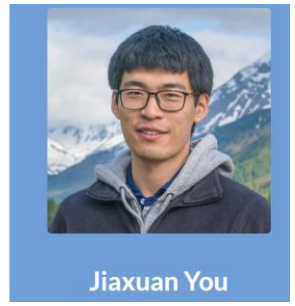
ASSOCIATE PROFESSOR OF COMPUTER SCIENCE

Web page: http://cs.stanford.edu/~jure

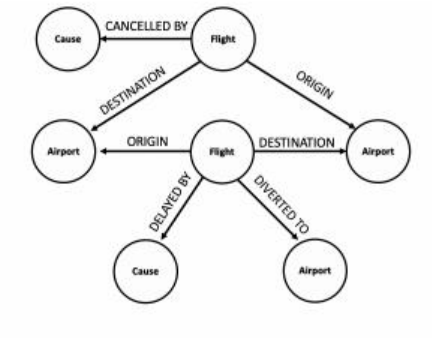
Jiaxuan You

Founding member
Kumo AI
Palo Alto, California
Email: jiaxuan@cs.stanford.edu
[Google Scholar] [Github]

Hi! I am a Founding member at Kumo AI. I received my Ph.D. and M.S. degrees from Department of Computer Science, Stanford University, advised by Prof. Jure Leskovec. I was supported by JPMC PhD Fellowship and Baidu Scholarship during my PhD. At Kumo AI, I aim to build a relational machine learning engine for cloud databases. **I'm on the 2022-2023 academic job market.**



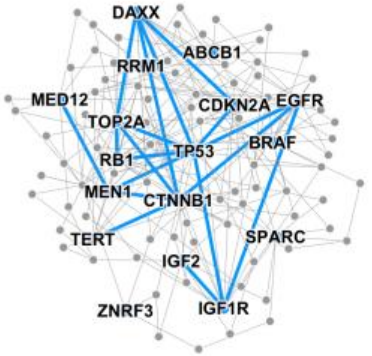
Why graph is so important?



Event Graphs

Image credit: SalientNetworks

Computer Networks



Disease Pathways

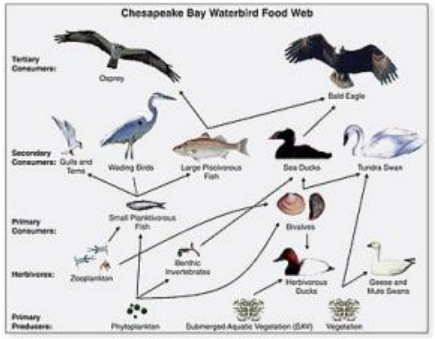


Image credit: Wikipedia

Food Webs

Image credit: Pinterest

Particle Networks

Image credit: visitlondon.com

Underground Networks

Why graph is so important?

Image credit: Medium

Social Networks

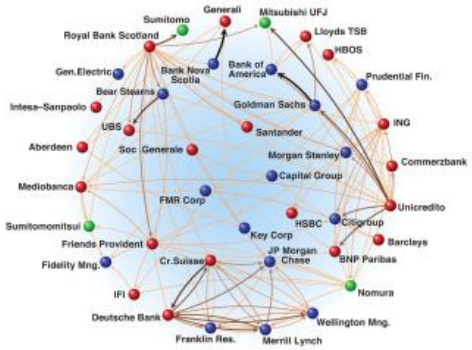
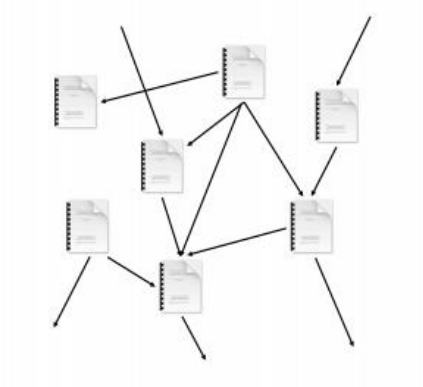


Image credit: Science

Economic Networks

Image credit: Lumen Learning

Communication Networks



Citation Networks

Image credit: Missoula Current News

Internet

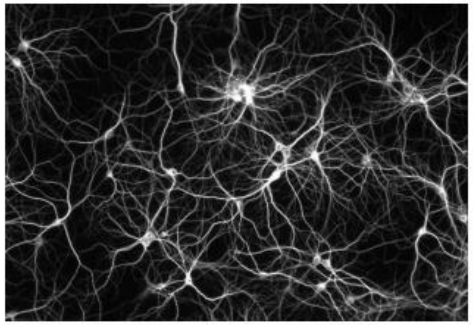


Image credit: The Conversation

Networks of Neurons

Why graph is so important?

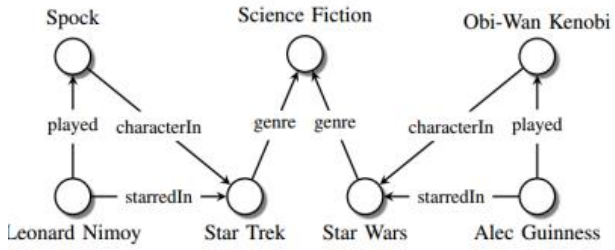


Image credit: Maximilian Nickel et al

Knowledge Graphs

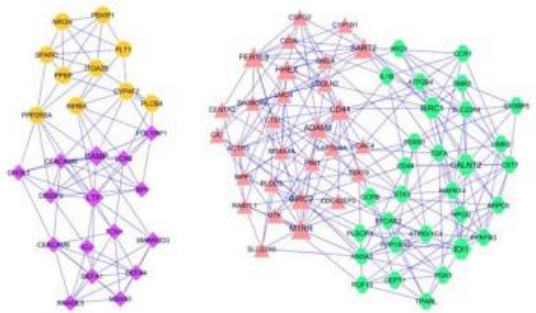


Image credit: ese.wustl.edu

Regulatory Networks

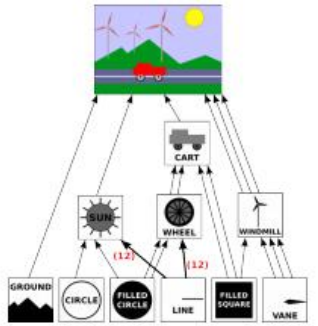


Image credit: math.hws.edu

Scene Graphs

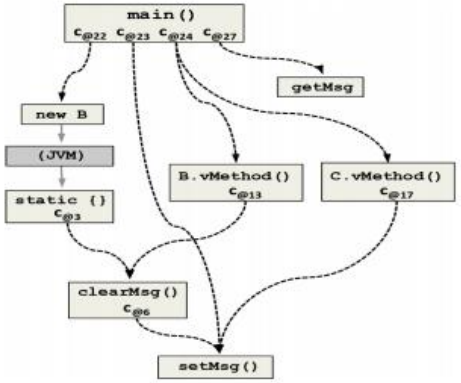


Image credit: ResearchGate

Code Graphs

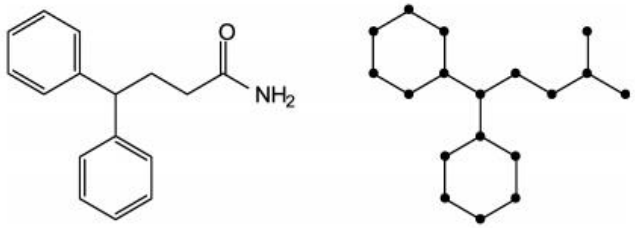


Image credit: MDPI

Molecules

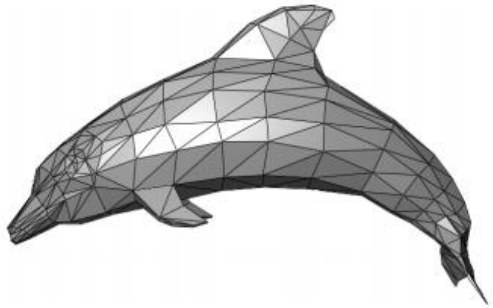
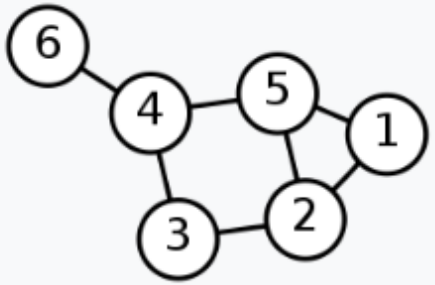


Image credit: Wikipedia

3D Shapes

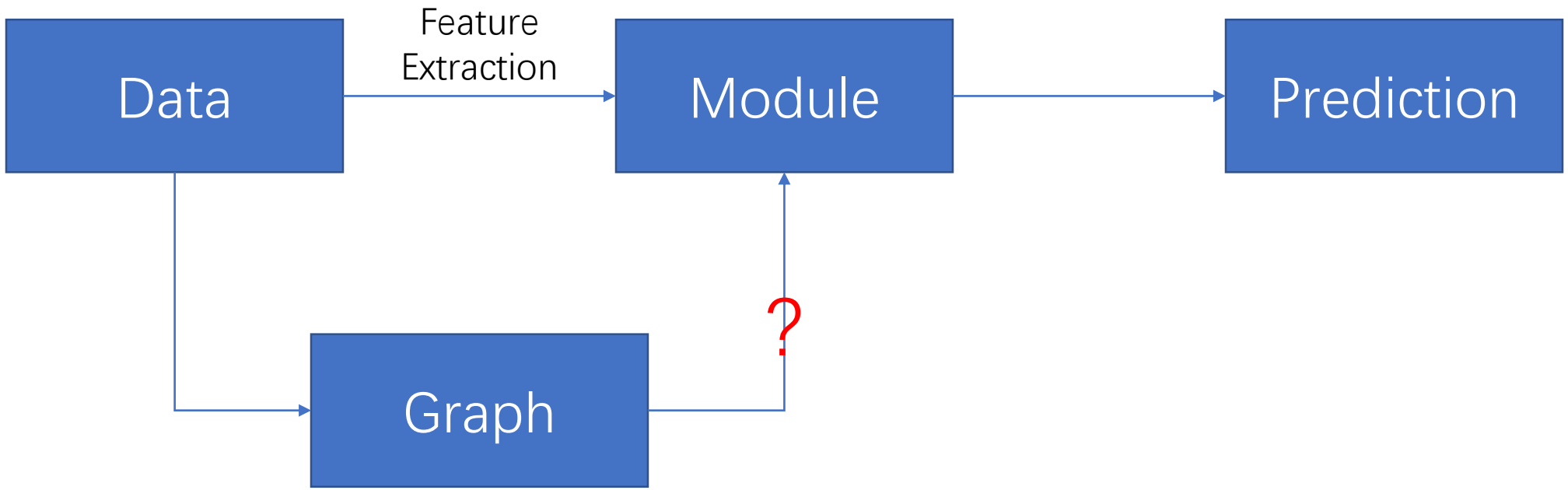
Features of a graph

图	度数矩阵	邻接矩阵	调和矩阵
	$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$
	D	A	L=D-A

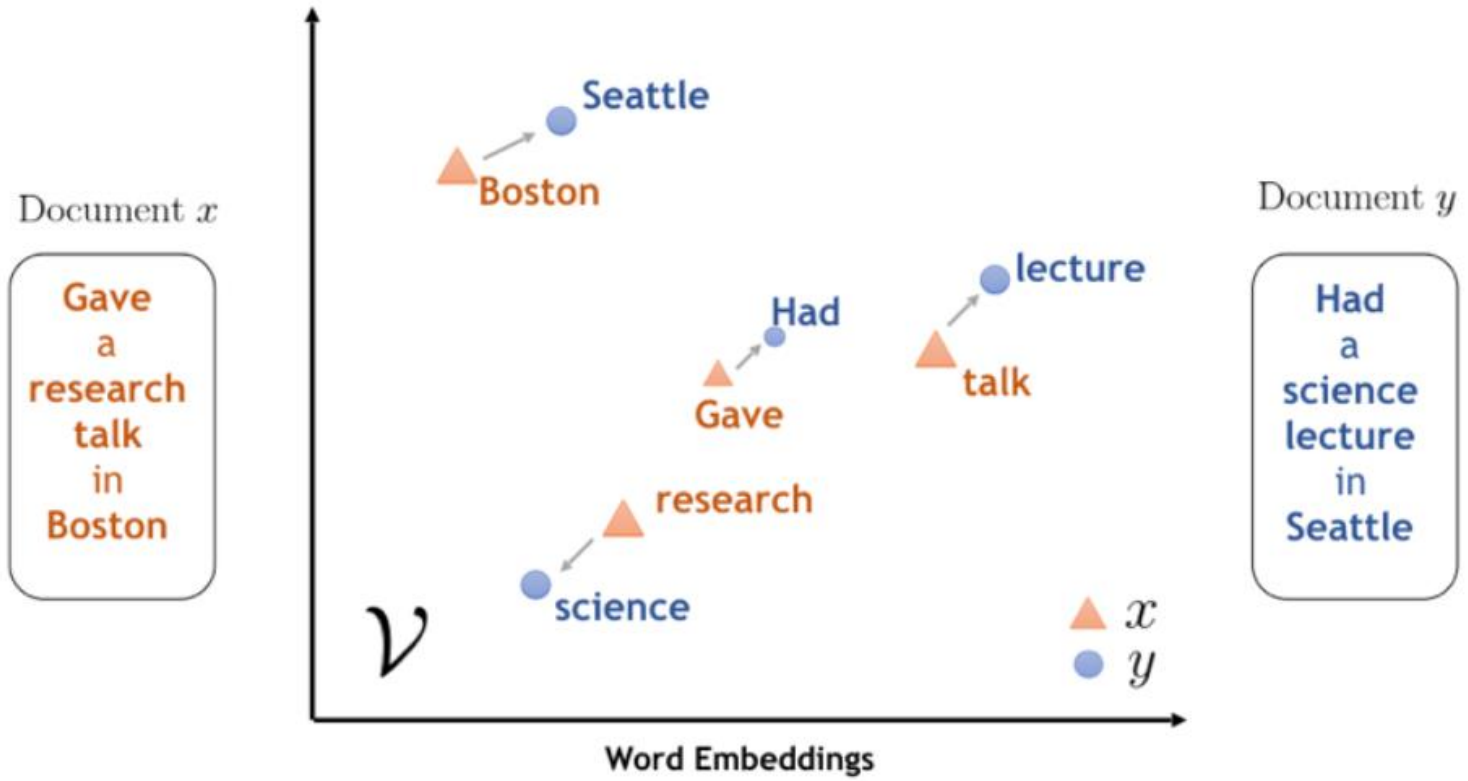
拉普拉斯矩阵的性质:

- 特征值都是非负实数
- 特征向量都是实向量，且正交
- 正则化表示$L_{sym} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$

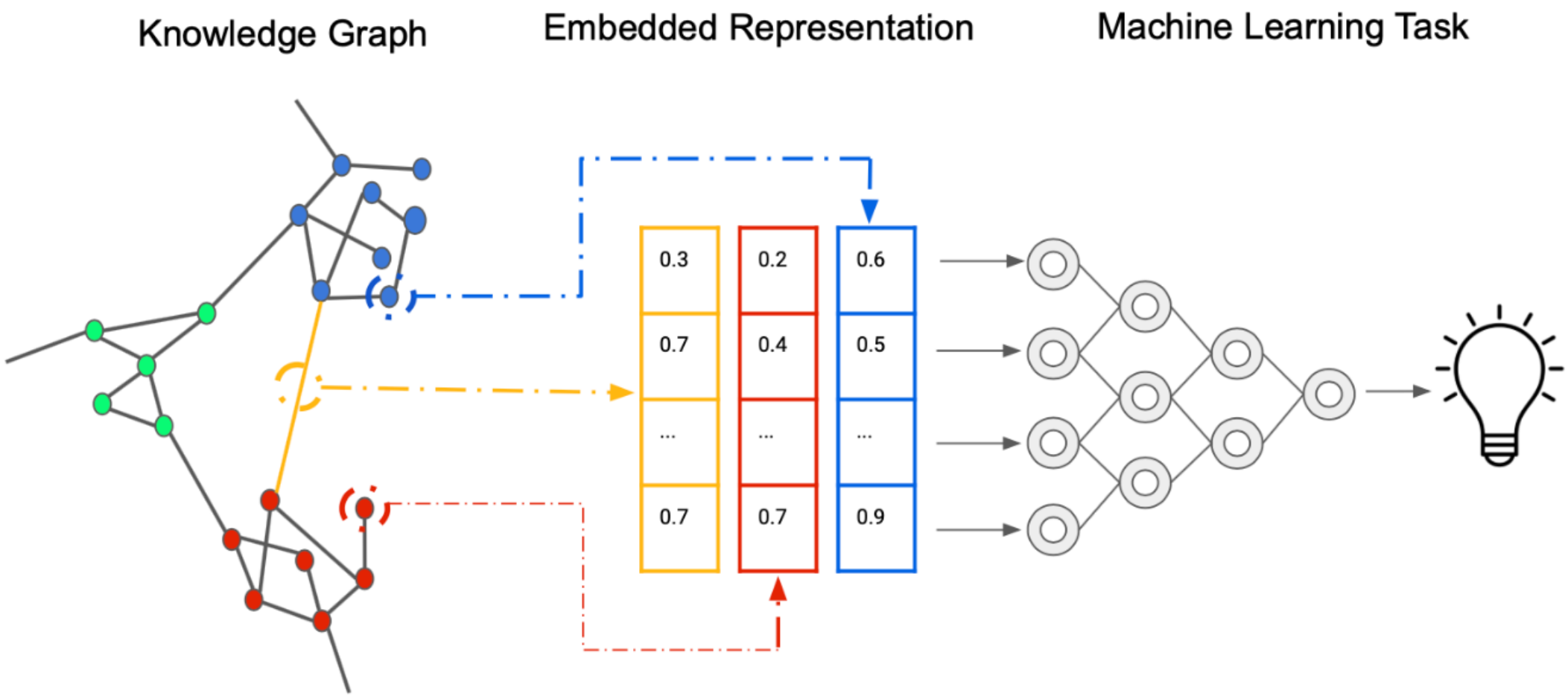
What we will do?



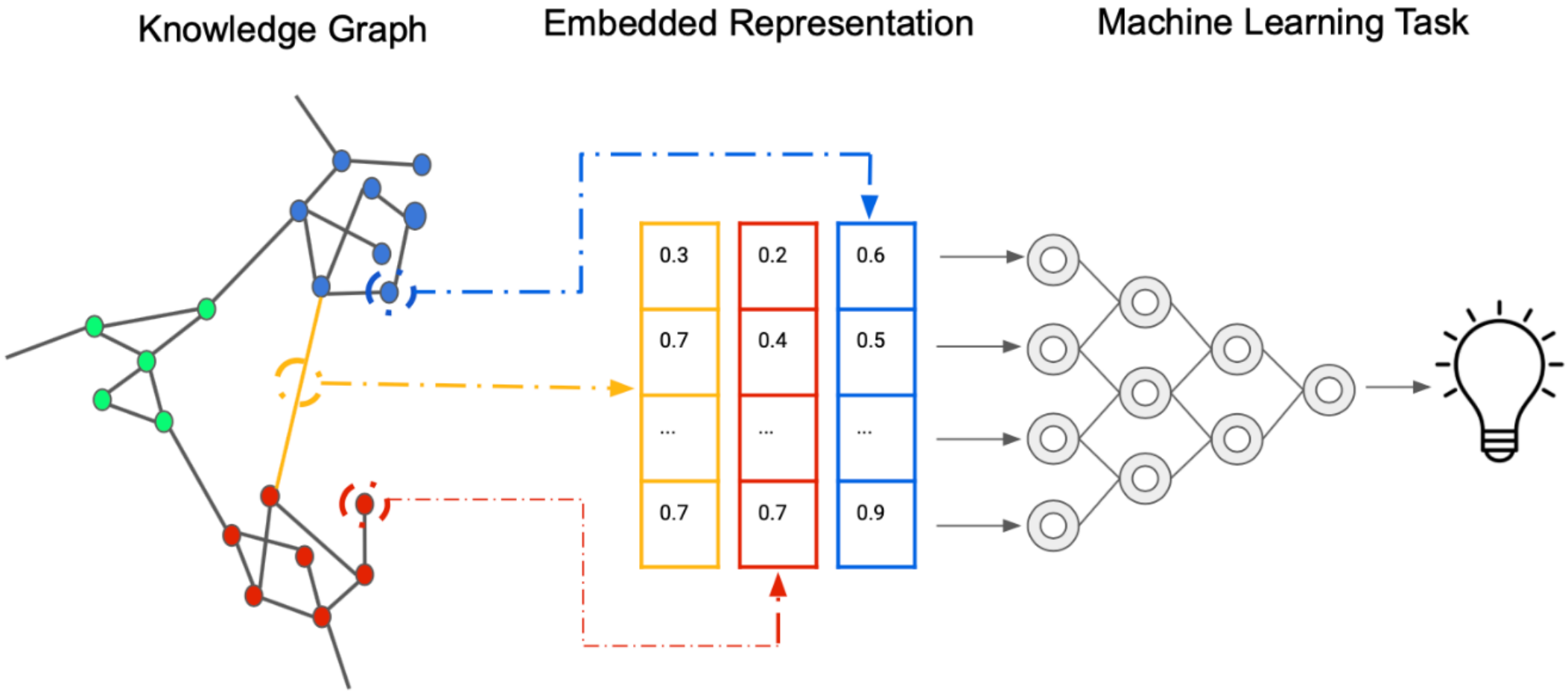
Word embedding



What we will do?



What we will do?



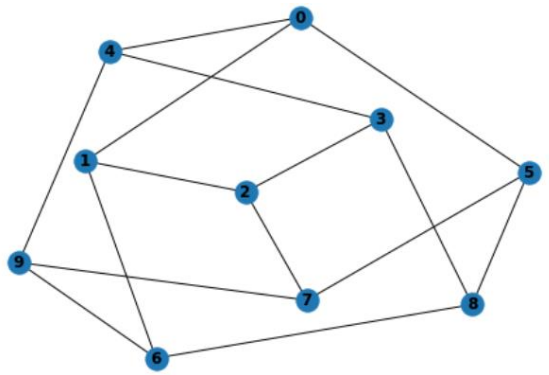
Task dependent or task independent?

核心工具介绍

NetworkX（与pytorch、pyg协同）

- 如何从dataframe创建一个graph
- 如何获取图的node、edge信息（比如手动构建一条路径）
- 如何获取图的邻接矩阵（计算GCN）
- 如何将NetworkX的graph对象转化为pyg对象
- …

Node1	Node2	Edge



sklearn和Pytorch

- Sklearn中常用的分类、回归方法
- 各种metrics（acc、F1等）和常用方法train_test_split等

- 如何构建Dataset和DataLoader
- 如何写模型的类，包括forward方法
- nn、Functional模块的常用层：Linear、各种activation
- 常见的损失函数：crossentropy、L1、MSE等
- Routine的训练过程

PyTorch Geometric（PyG）

- 如何直接创建pyg的network对象
- Pyg常用的模块（GCN、GAT等）
- PyG Documentation — pytorch_geometric documentation (pytorch-geometric.readthedocs.io)
- Loading Graphs from CSV — pytorch_geometric documentation (pytorch-geometric.readthedocs.io)

TORCH_GEOMETRIC.NN

Contents

- Convolutional Layers
- Aggregation Operators
- Normalization Layers
- Pooling Layers
- Unpooling Layers
- Models
- Encodings
- Functional
- Model Transformations
- Dense Convolutional Layers
- Dense Pooling Layers
- DataParallel Layers

一些关键的论文和数据

Network Embedding

- DeepWalk: Online Learning of Social Representations
- node2vec: Scalable Feature Learning for Networks
- Semi-Supervised Classification with Graph Convolutional Networks
- Inductive Representation Learning on Large Graphs
- Graph Attention Networks
- metapath2vec: Scalable Representation Learning for Heterogeneous Networks

Datasets

Stanford Biomedical Network Dataset Collection

Jure Leskovec @ Stanford

https://cseweb.ucsd.edu/~jmcauley/datasets.html

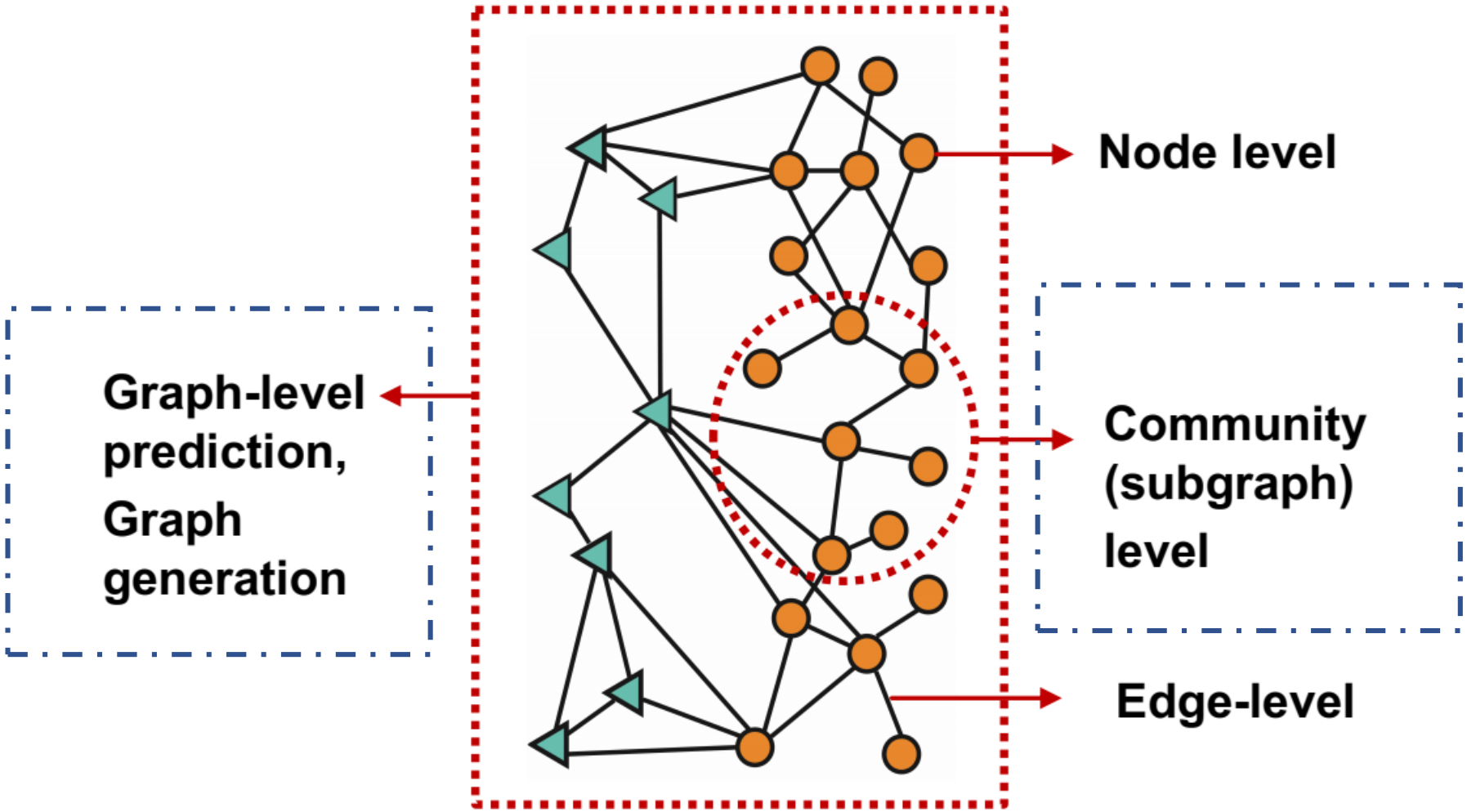
Table 1: Summary of the datasets used in our experiments.

	Cora	**Citeseer**	**Pubmed**	**PPI**
Task	Transductive	Transductive	Transductive	Inductive
# Nodes	2708 (1 graph)	3327 (1 graph)	19717 (1 graph)	56944 (24 graphs)
# Edges	5429	4732	44338	818716
# Features/Node	1433	3703	500	50
# Classes	7	6	3	121 (multilabel)
# Training Nodes	140	120	60	44906 (20 graphs)
# Validation Nodes	500	500	500	6514 (2 graphs)
# Test Nodes	1000	1000	1000	5524 (2 graphs)

GAT Tasks

图方法的任务

Tasks


Node level
Graph-level prediction, Graph generation
Community (subgraph) level
Edge-level

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

项目具体流程

- 先掌握2-3种baseline方法（deepwalk、GCN等，传统方法不要忘记）
- 确定每位学员的具体任务（大方向，非具体）
  - 确定数据
  - 确定方向（模型改进、模型迁移）
- 讲解进阶方法（会对具体的代码进行讲解，学员需要去确定一下采用哪种方法去做）
  - GAT、metapath…
- 具体案例讲解
- 每位学员的具体论文情况说明