



中国科学技术大学
University of Science and Technology of China

《人工智能数学原理与算法》

第4章：图神经网络

4.1 图学习基础

王翔

xiangwang@ustc.edu.cn



01 图数据

02 图数据的表示

03 图学习的任务

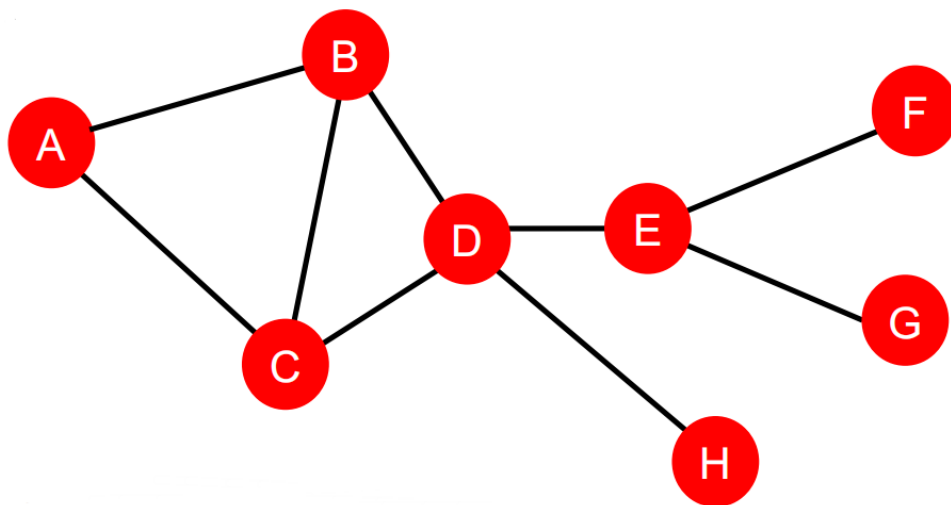
目录



图数据

图(Graph)是一种描述实体节点及其关系/交互/链接的通用语言：

- 现实中很多种类型的数据都可以表示为图结构。



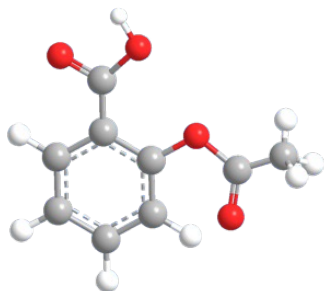
图数据举例：微观世界

DeepMind AlphaFold团队
2021年首次在Nature发文：
将蛋白质结构预测问题建模为
图推理问题

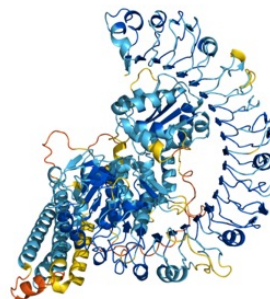
同年，AI预测蛋白质结构入选
Science年度科学突破



AlphaFold领导团队因为其在蛋白质结构预测的贡献，
获**2024年诺贝尔化学奖**

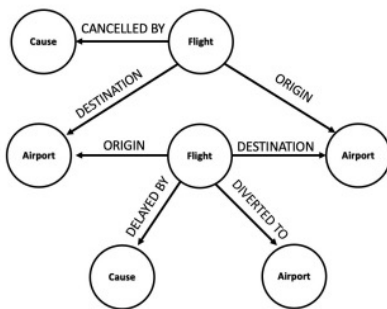


化学分子



蛋白质结构

图数据举例：宏观世界1

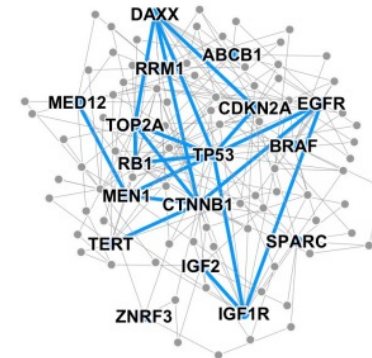


事件图



Image credit: [SalientNetworks](#)

计算机网络



疾病路径

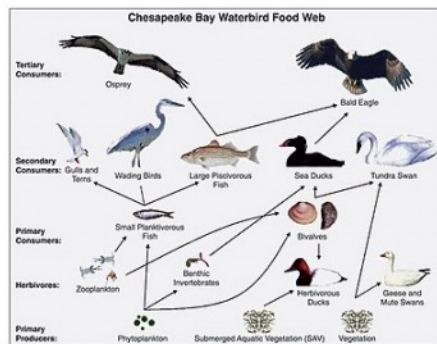


Image credit: [Wikipedia](#)

食物网



Image credit: [Pinterest](#)

粒子网络

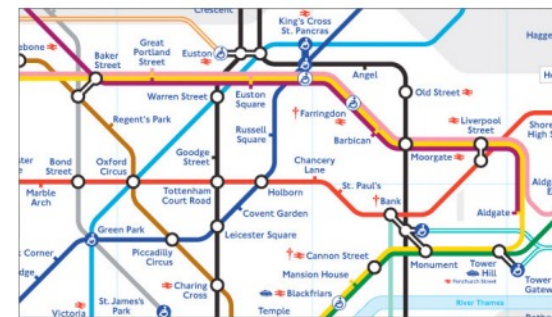


Image credit: [visitlondon.com](#)

地下网络

图数据举例：宏观世界2



Image credit: [Medium](#)

社交网络

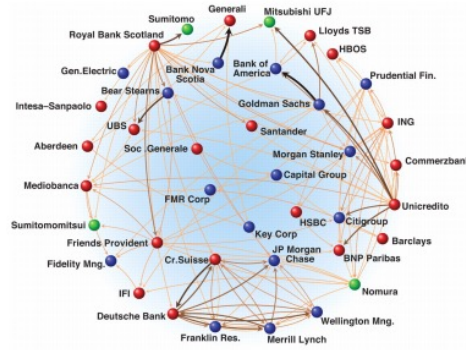


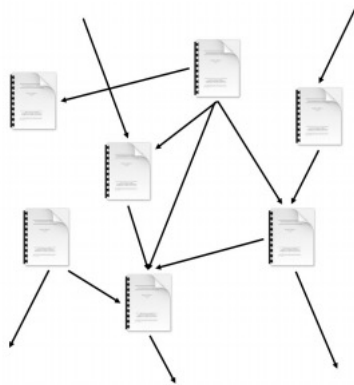
Image credit: [Science](#)

经济网络



Image credit: [Lumen Learning](#)

通信网络



引用网络



Image credit: [Missoula Current News](#)

互联网

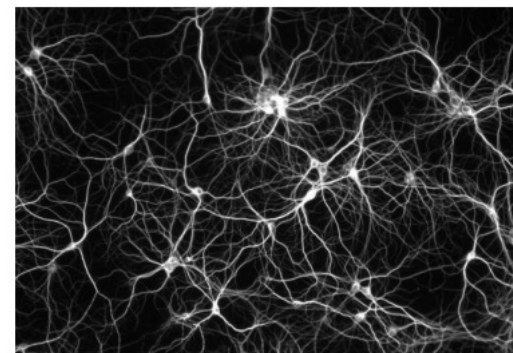


Image credit: [The Conversation](#)

神经元网络

图数据举例：宏观世界3

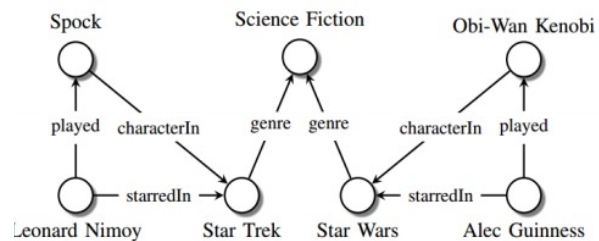


Image credit: [Maximilian Nickel et al](#)

知识图谱

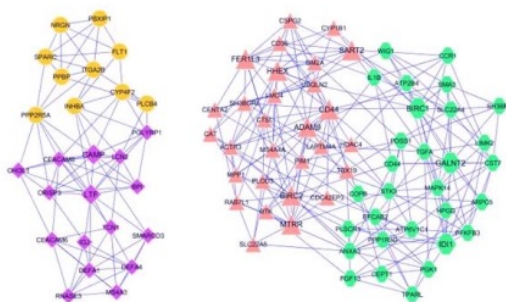


Image credit: [ese.wustl.edu](#)

调控网络

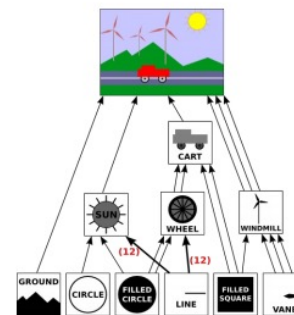


Image credit: [math.hws.edu](#)

场景图

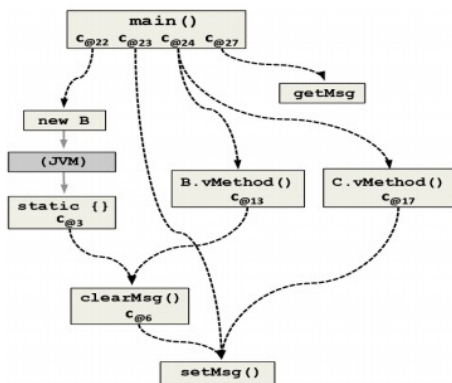


Image credit: [ResearchGate](#)

代码图

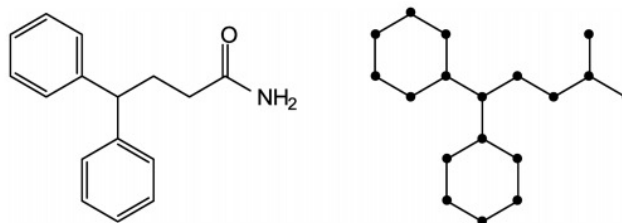


Image credit: [MDPI](#)

分子图

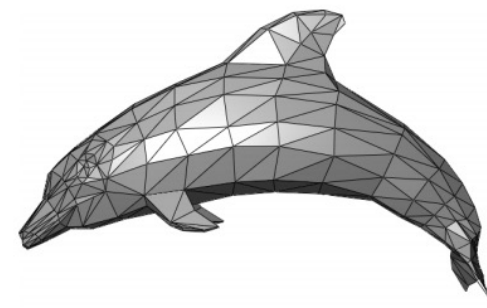


Image credit: [Wikipedia](#)

三维形状

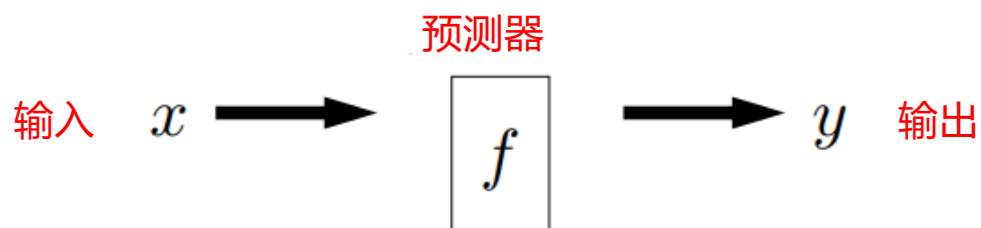
图学习 = 面向图数据的机器学习&深度学习

复杂的领域具有丰富的结构，可以表示为图数据。

通过明确地建模图结构，我们可以获得更好的性能！

主要问题：

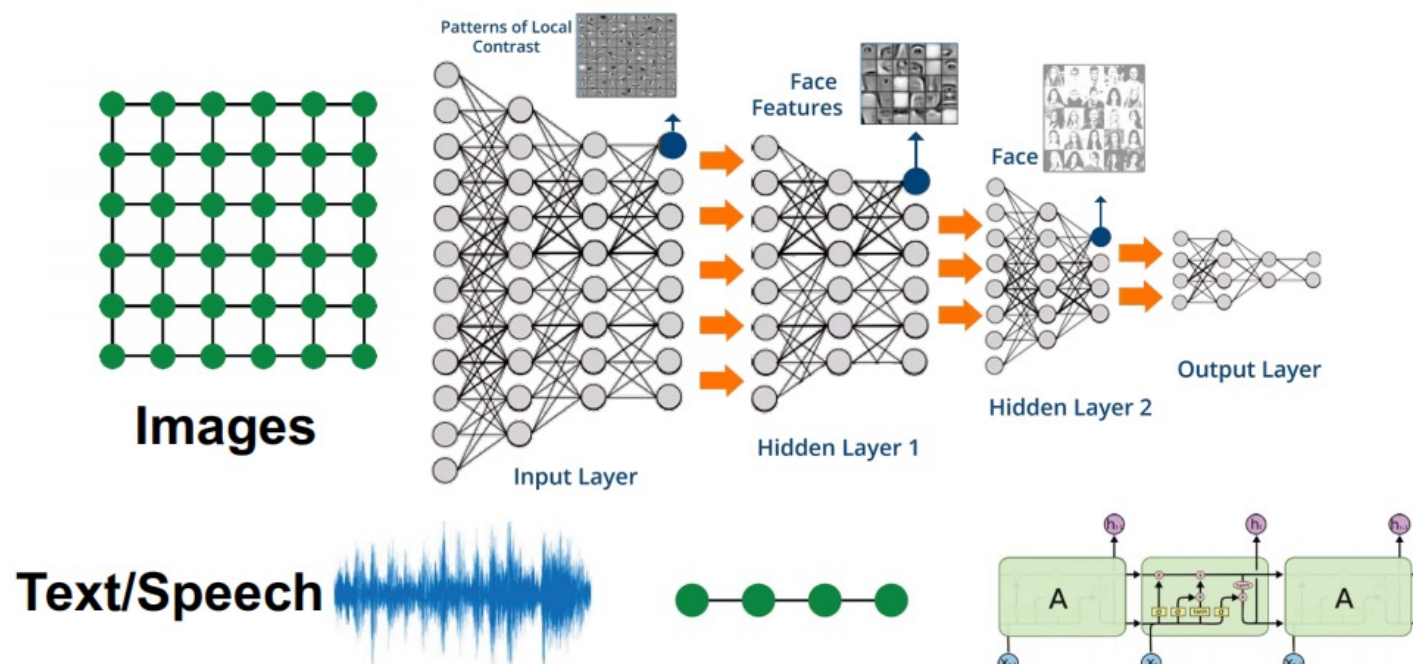
如何利用图数据进行更好的预测？



传统神经网络的局限：

传统神经网络大多是为简单的**非结构化数据**设计的：

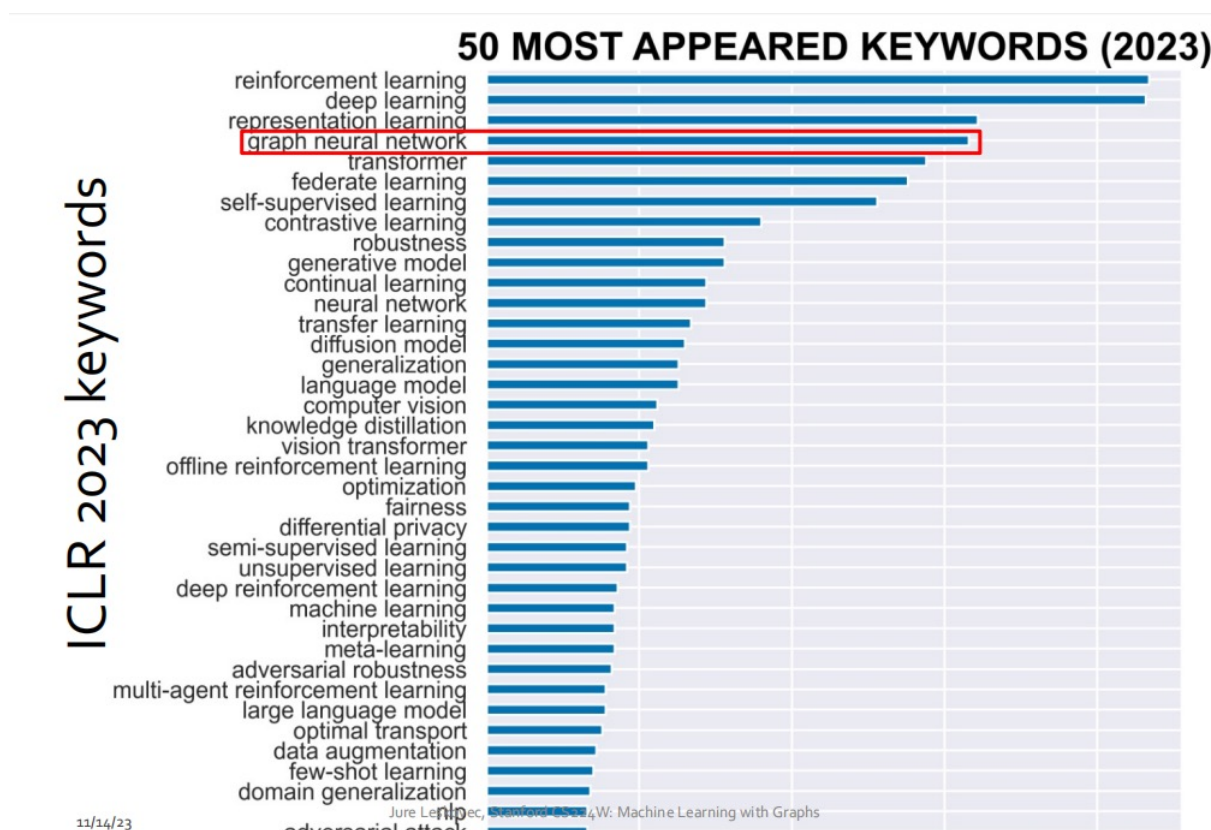
- 表格数据（如贷款数据）
- 向量数据（如自然语言、图像）



图神经网络 = 面向图数据的神经网络

我们如何开发更广泛适用的神经网络？

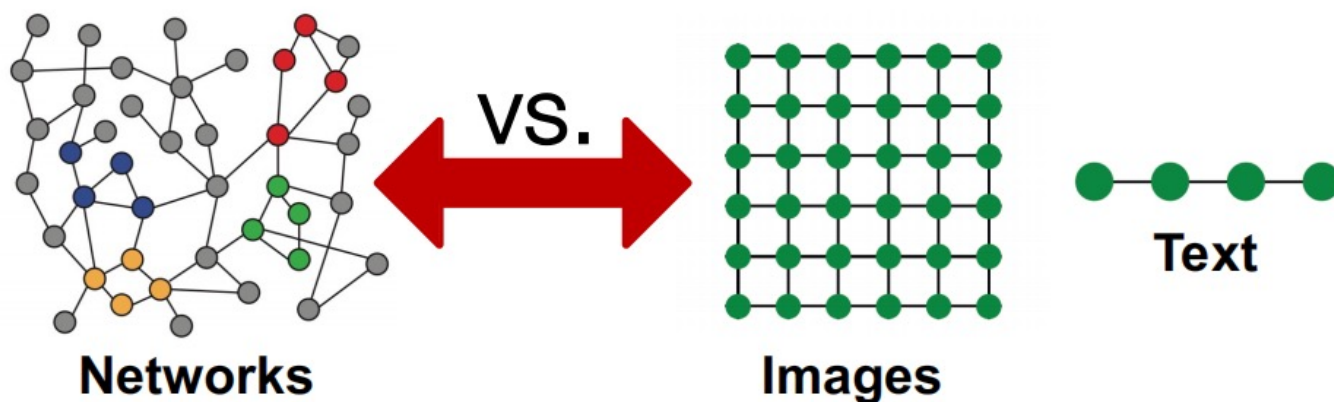
面向图数据的神经网络是深度学习的前沿方向之一。



图神经网络的难点 = 图数据的特点

不同于表格和向量数据，图结构非常复杂：

- 任意大小和复杂的拓扑结构



- 没有固定的节点顺序或参考点。
- 常常是动态的，并且具有多模态特征。



01 图数据

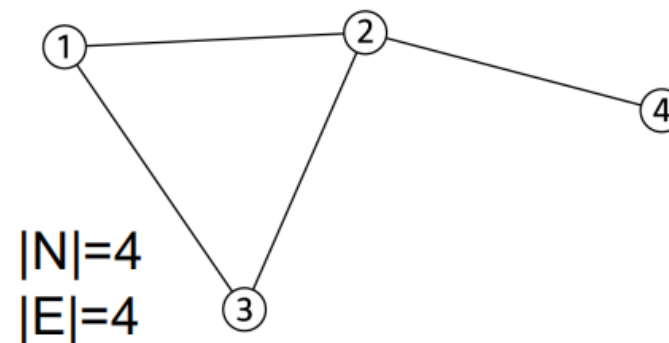
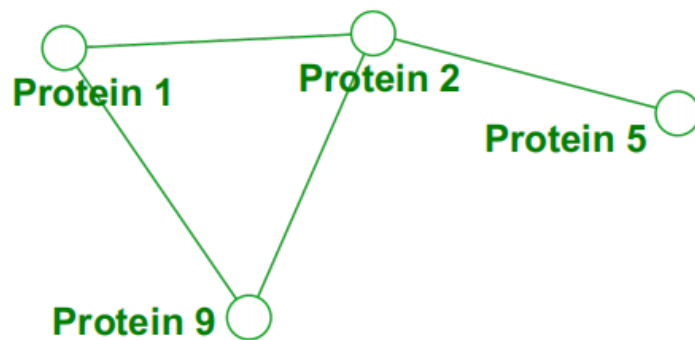
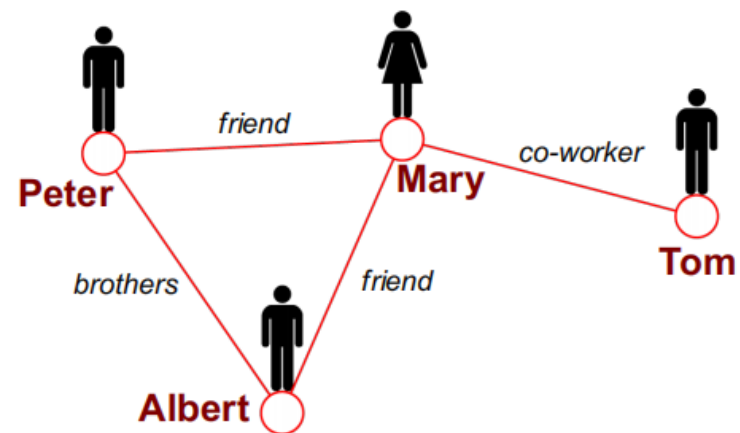
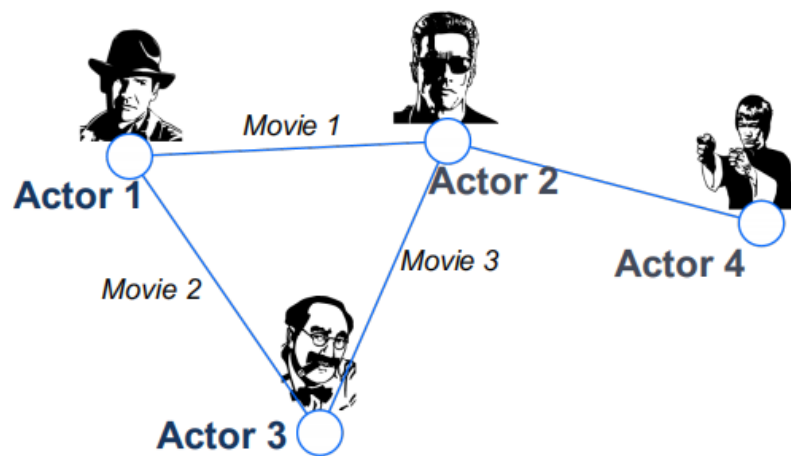
02 图数据的表示

03 图学习的任务

目录



如何表示图数据呢？



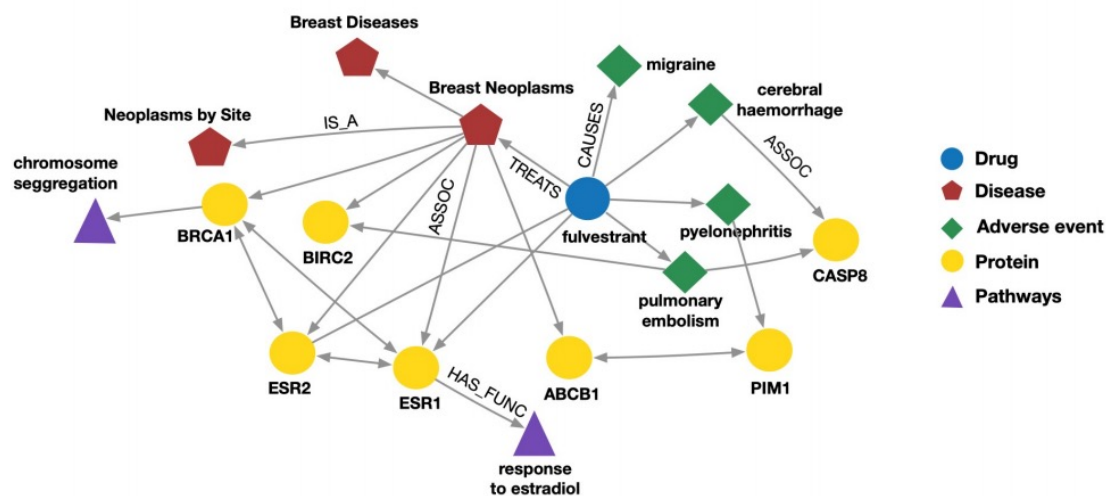
异构图

- 异构图(Heterogeneous Graph)的定义为:

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

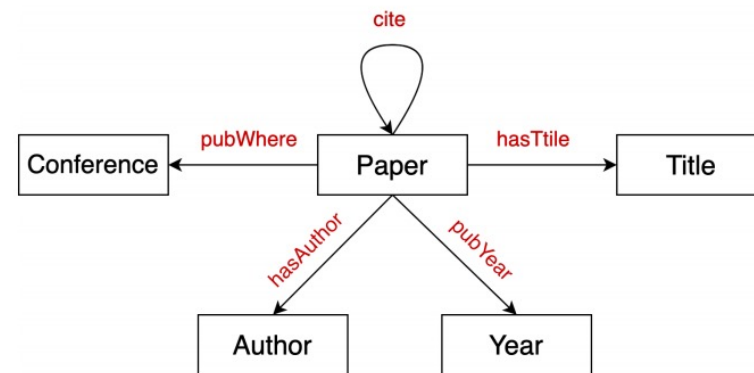
- 不同节点类型的节点 $v_i \in V$
- 不同关系类型的边或链接 $(v_i, r, v_j) \in E$
- 节点类型 $T(v_i)$
- 关系类型 $r \in R$
- 节点和边都具有属性/特征

异构图示例



生物医学知识图谱

示例节点：Migraine偏头痛
示例边：(fulvestrant氟维司群, Treats治疗, Breast Neoplasms乳腺肿瘤)
示例节点类型：Protein蛋白质
示例边类型(关系)：Causes导致



学术图谱

示例节点：ICML
示例边：(GraphSAGE, NeurIPS)
示例节点类型：Author
示例边类型(关系)：pubYear

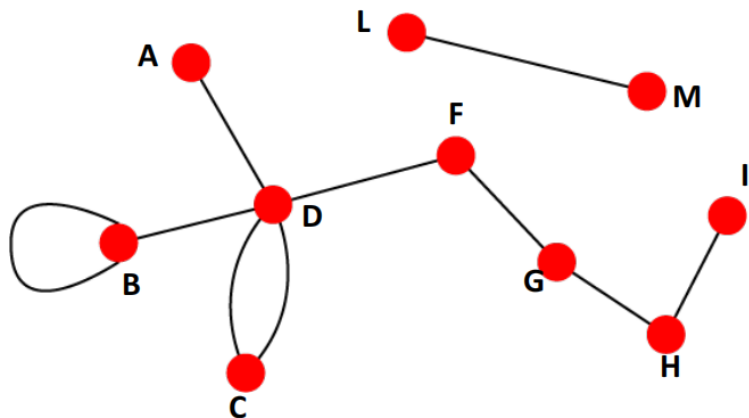
图数据的合适表示

- 如何构建一个图：
 - 节点是什么？
 - 边是什么？
- 对给定领域/问题选择合适的图表示决定了能否成功地构建神经网络：
 - 在某些情况下，存在一种唯一且明确的表示方法
 - 在其他情况下，表示方法并非唯一的。
 - 边的表示方式将决定研究问题的性质。

有向图vs无向图

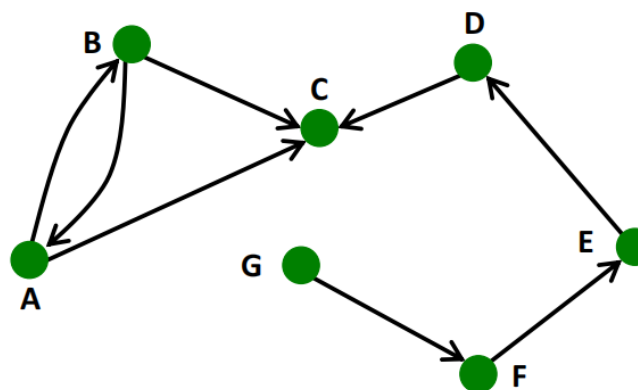
无向图

- 边：无向的
(对称的, 相互的)



有向图

- 边：有向的



- 其他考虑因素：权重、属性、类型、特征

二分图

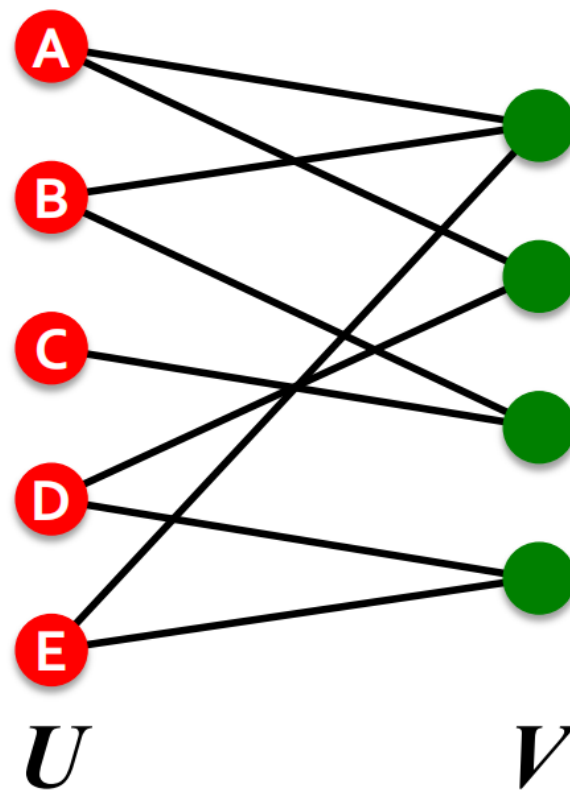
二分图(Bipartite graph)是一个节点可以分为两个不相交集合 U 和 V 的图，使得每条边连接 U 中的一个节点到 V 中的一个节点；也就是说， U 和 V 是独立集。

- 示例：

- 作者-论文（他们撰写的）
- 演员-电影（他们出演的）
- 食谱-成分（它们包含的）

- “折叠”网络：

- 作者合作网络
- 电影共同评分网络





01 图数据

02 图数据的表示

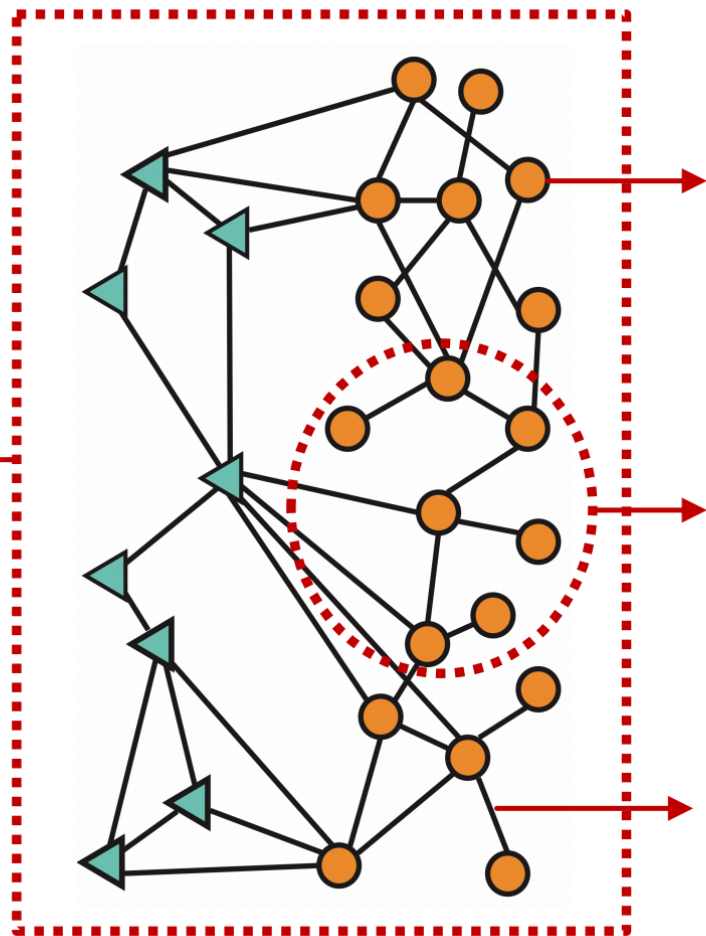
03 图学习的任务

目录



图学习的任务

图级预测、图生成
(Graph-level prediction,
Graph generation)



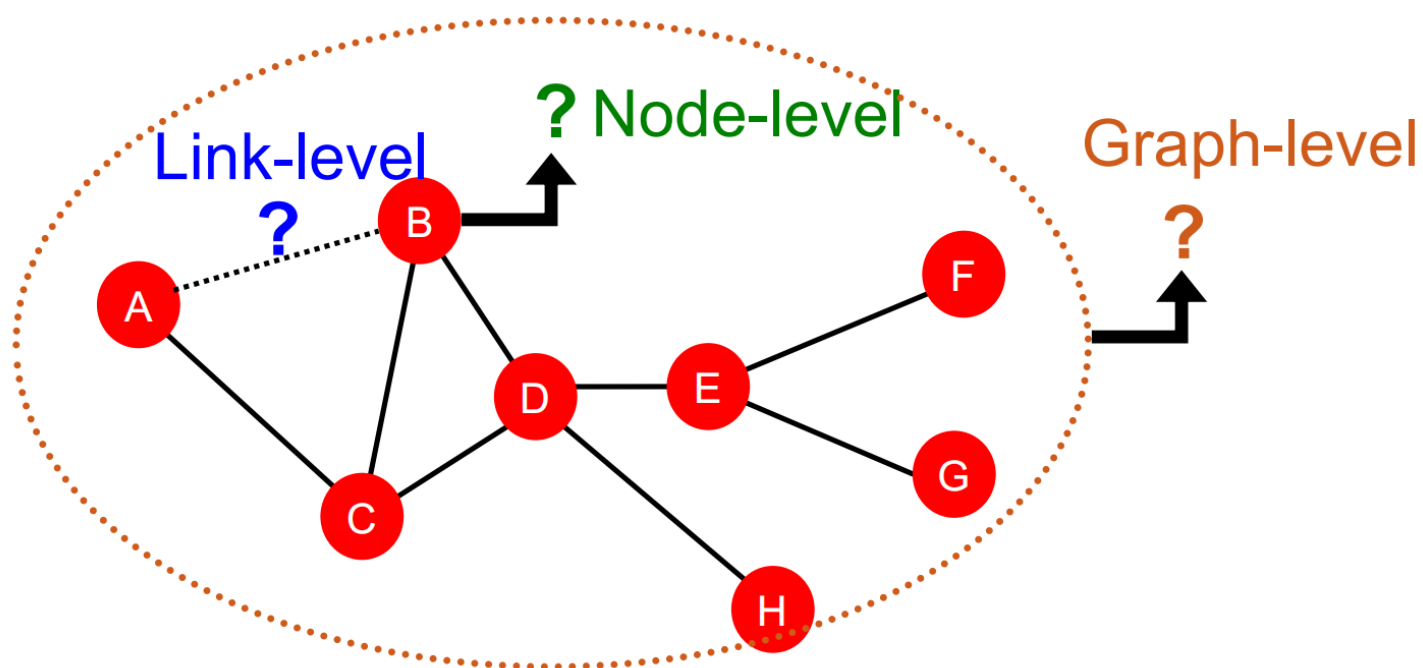
节点级(Node-level)

社区(子图)级
(Community (subgraph) level)

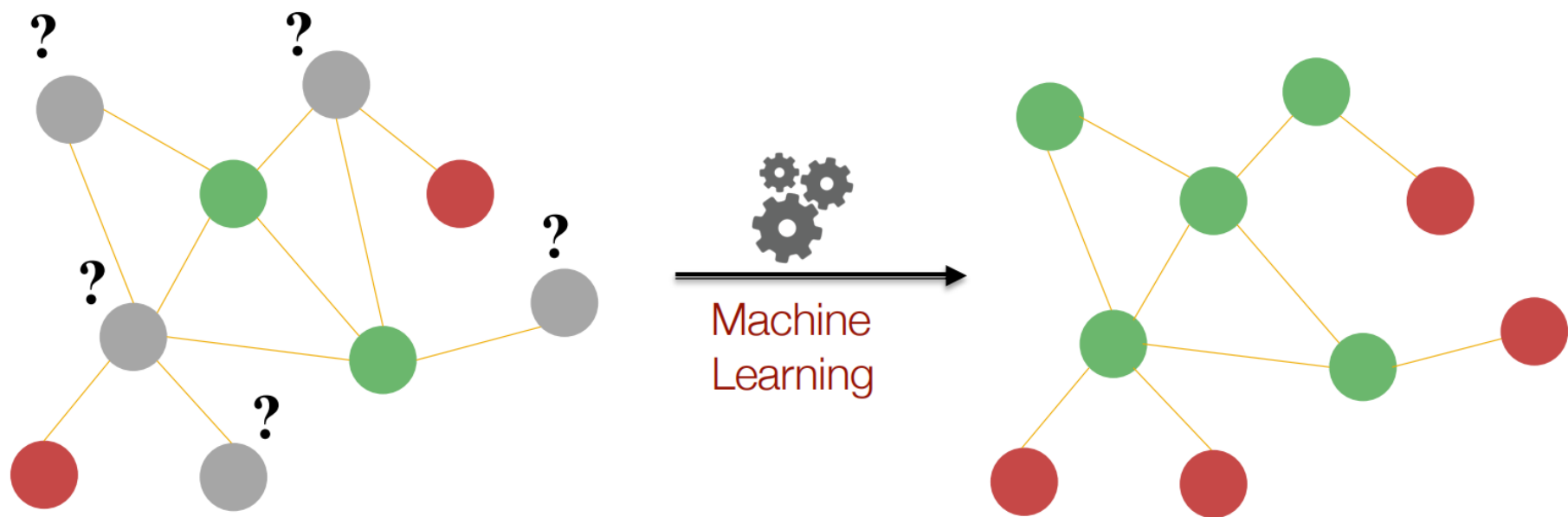
边级(Edge-level)

图学习的任务

- 节点级(Node-level)预测
- 链接级(Link-level)预测
- 图级(Graph-level)预测



节点级任务

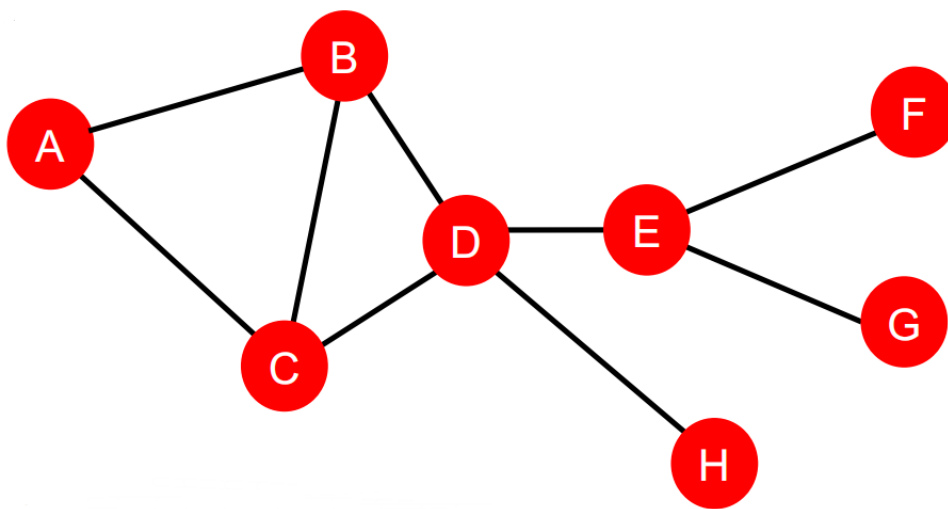


节点分类

如何表示节点=抽取节点特征

目标：描述节点在图中的结构和位置

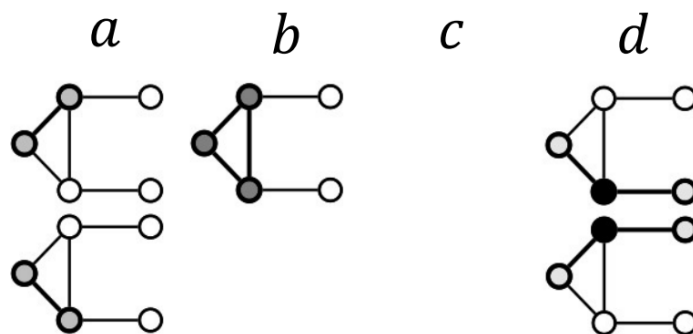
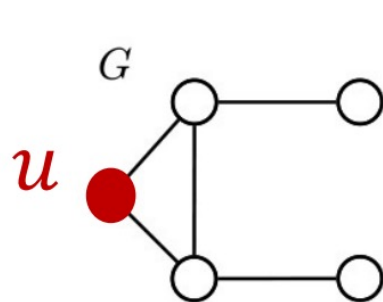
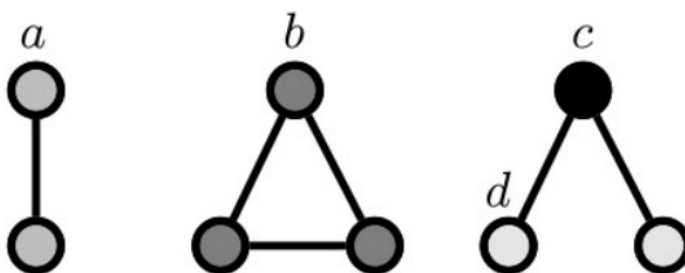
- 节点度数
- 节点重要性和位置
 - 例如，穿过节点的最短路径数量
 - 例如，到达其他节点的平均最短路径长度
- 节点周围的子结构



节点子图：Graphlets

- Graphlets: 给定节点的子图计数向量

最多包含3个节点的所有可能的Graphlets



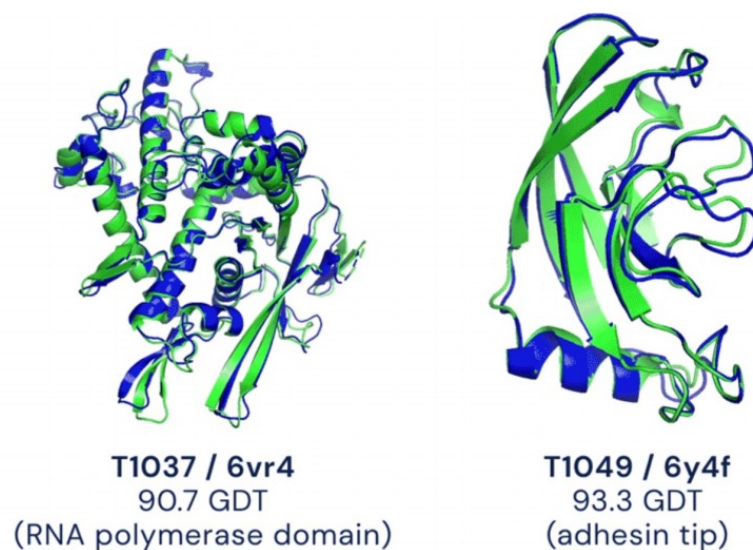
节点 u 的 Graphlets 特征向量:
a b c d
[2,1,0,2]

节点 u 的 Graphlets

节点级任务示例：蛋白质折叠

基于蛋白质的氨基酸序列计算预测其三维结构：

对于每个节点，预测其三维坐标



● Experimental result
● Computational prediction

Image credit: [DeepMind](#)

节点级任务示例：AlphaFold

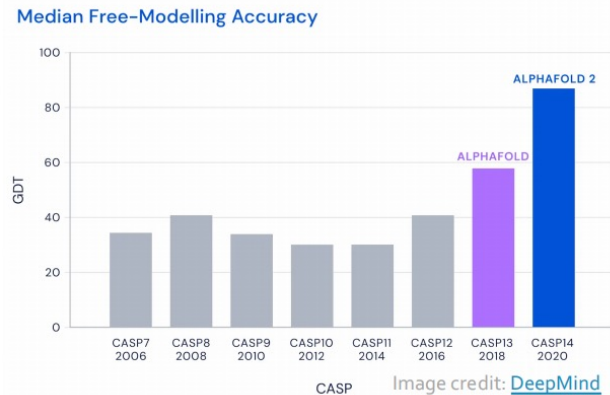


Image credit: [SingularityHub](#)

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

节点级任务示例：AlphaFold解决蛋白质折叠

- **关键概念：**“空间图(Spatial graph)”
 - **节点：** 蛋白质序列中的氨基酸
 - **边：** 氨基酸（残基）之间的接近性

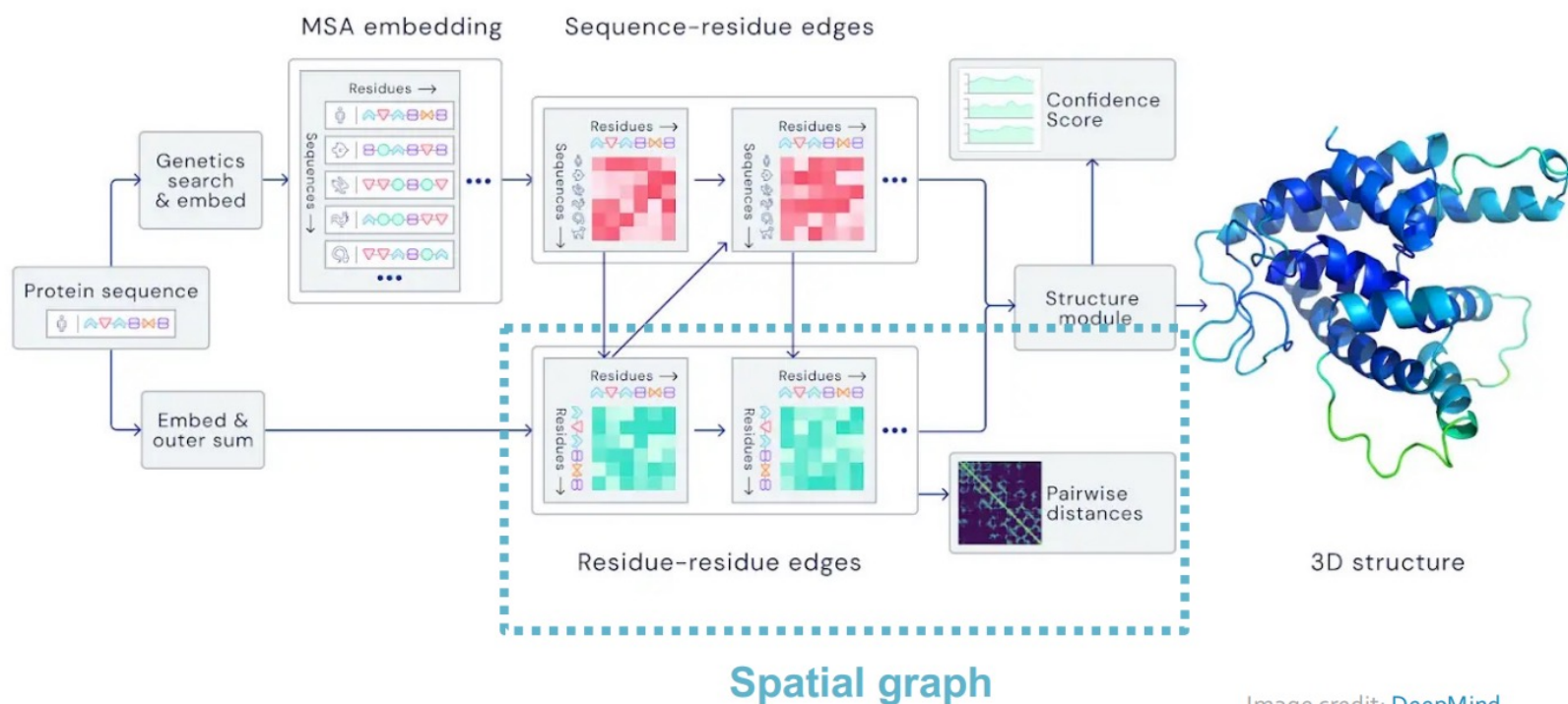
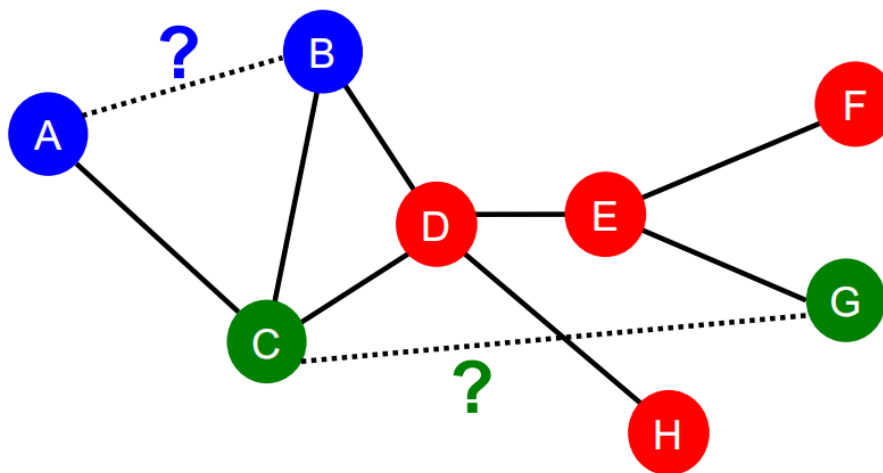


Image credit: [DeepMind](#)

链接级任务

- 链接级预测任务旨在基于现有链接预测新的/缺失/未知链接。
- 在测试时，会对节点对（没有现有链接的）进行排名，并预测前K个节点对
- 任务：对链接（“节点对”）做出预测



链接级任务

两种链接预测任务的表述：

- 1) 随机缺失的链接：

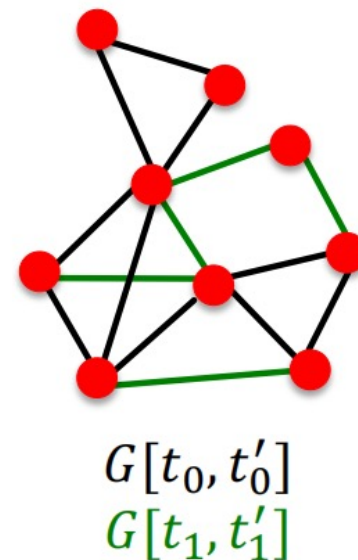
- 移除一组随机链接，然后预测它们。

- 2) 时间上的链接：

- 给定由时间 t_0' 之前的所有边定义的图 $G[t_0, t_0']$ ，输出一个排名列表 L ，其中包含不在 $G[t_0, t_0']$ 中的边，这些边预计会在时间 $G[t_1, t_1']$ 出现。

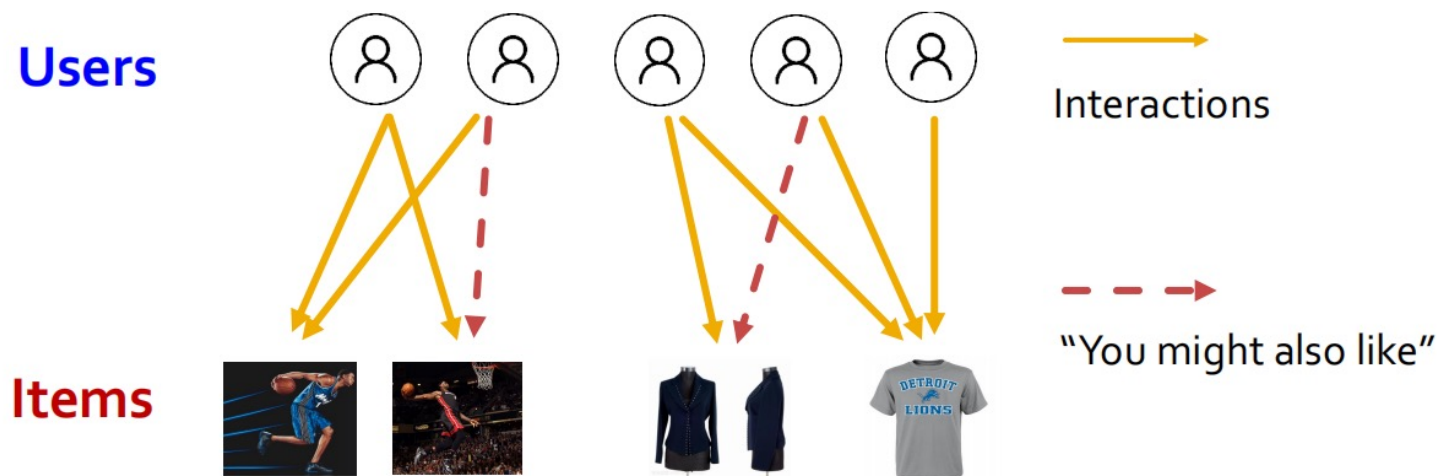
- 评估：

- $n = |E_{new}|$ ：#在测试期间 $[t_1, t_1']$ 出现的新边的数量
- 取 L 的前 n 个元素并计算正确边的数量。



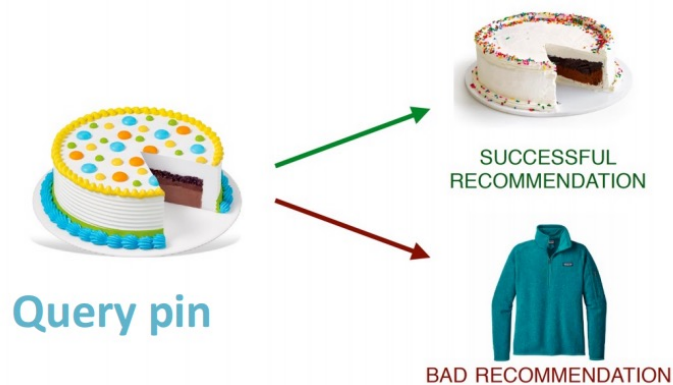
链接级任务示例1: 推荐系统

- **用户与物品的交互：**
 - 观看电影、购买商品、听音乐。
 - **节点：** 用户和物品
 - **边：** 用户-物品的交互
- **目标：** 推荐用户可能喜欢的物品



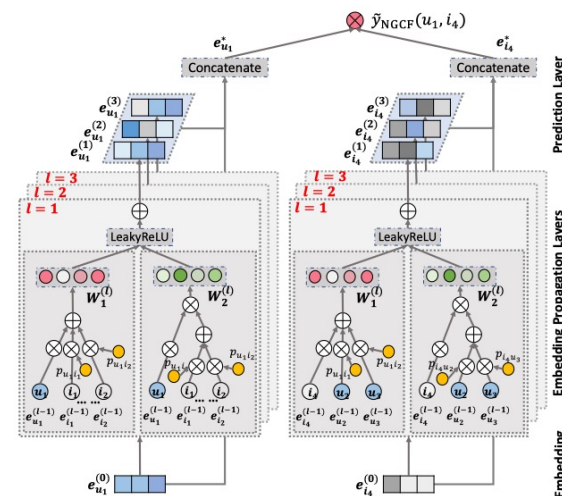
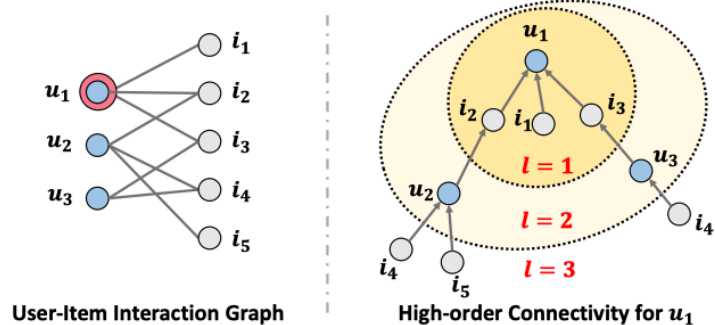
链接级任务1：基于图的推荐

任务： 向用户推荐相关的内容



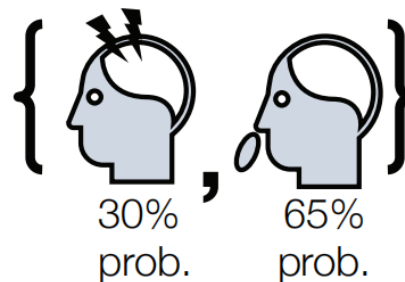
任务： 学习节点嵌入 z_i , 使得：
 $d(z_{\text{cake1}}, z_{\text{cake2}}) < d(z_{\text{cake1}}, z_{\text{sweater}})$

预测图中两个节点是否相关



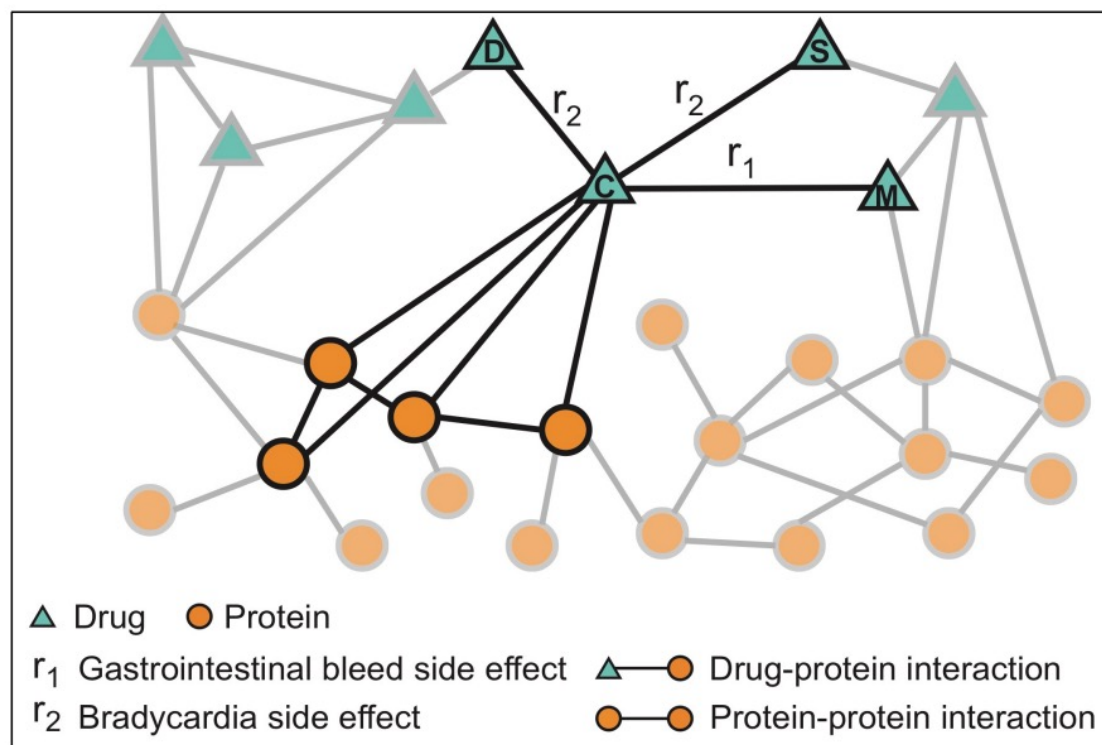
链接级任务2: 药物副作用

- 许多患者服用**多种药物**来治疗**复杂或共存的疾病**:
 - 46% 的70-79岁人群服用超过5种药物
 - 许多患者服用超过20种药物来治疗心脏病、抑郁症、失眠等
- **目标**: 给定一对药物, 预测不良副作用

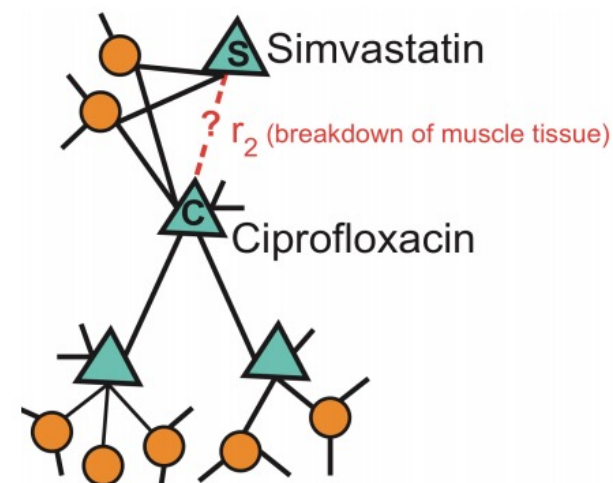


链接级任务2: 生物医学图链路预测

- **节点:** 药物和蛋白质
- **边:** 交互



- **查询:** 辛伐他汀和环丙沙星同时服用时, 肌肉组织分解的可能性有多大?



链接级任务2: 生物医学图链路预测

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	

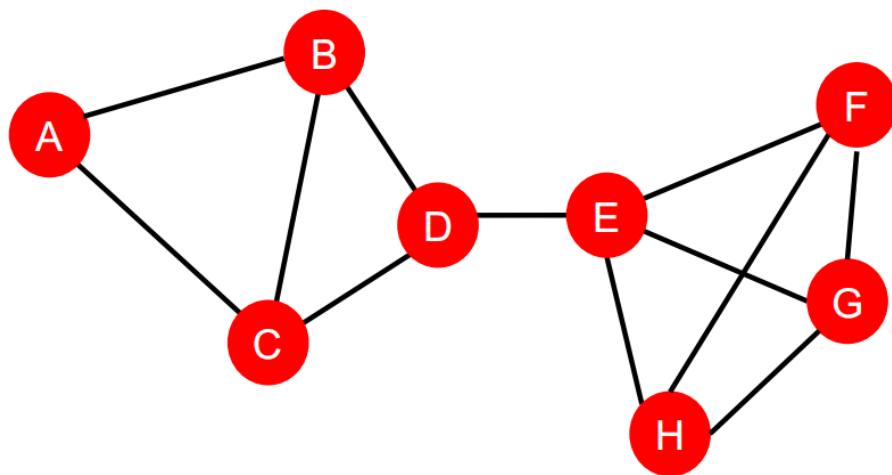
病例报告:

严重的横纹肌溶解, 可能是由于阿托伐他汀与氨氯地平和替格瑞洛之间的药物相互作用所致。

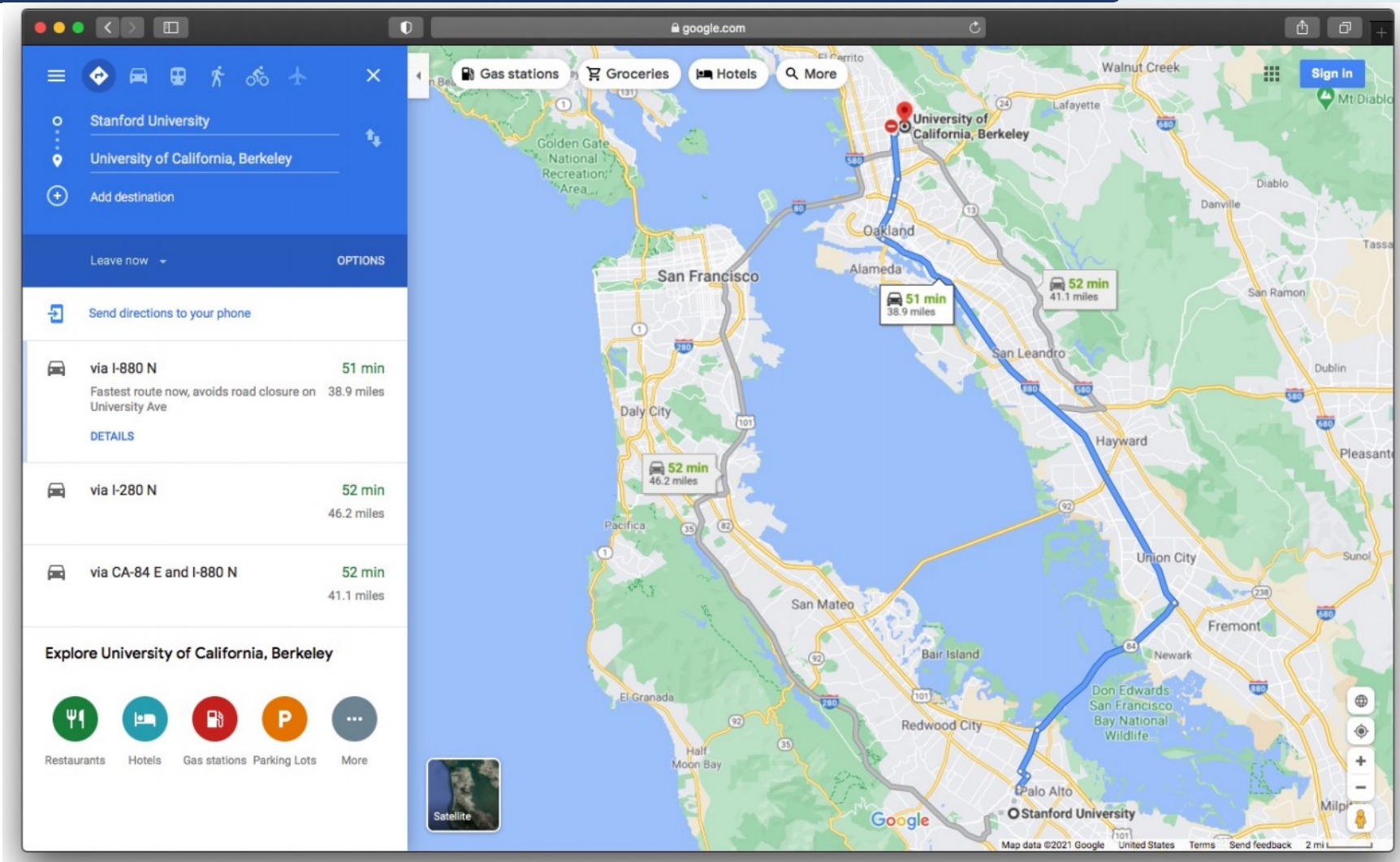
图级任务

- **目标：** 我们想要对整个图或图的子图进行预测。

- **例如：**



图级任务1：交通预测



图级任务1：道路网络图

- **节点：**道路段
- **边：**道路段之间的连接性
- **预测：**到达时间（ETA）

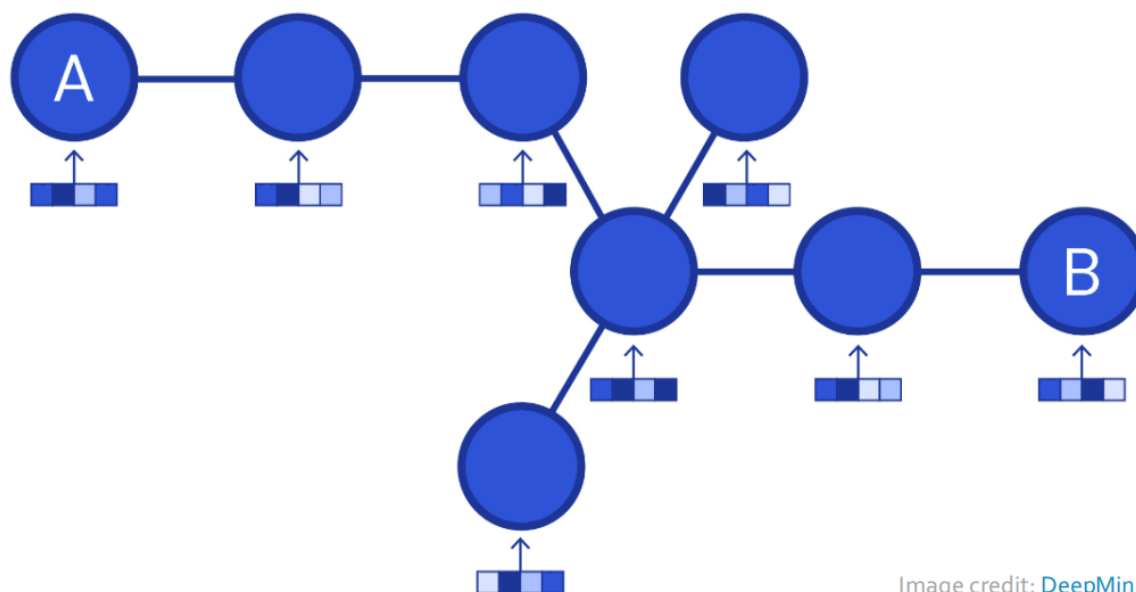
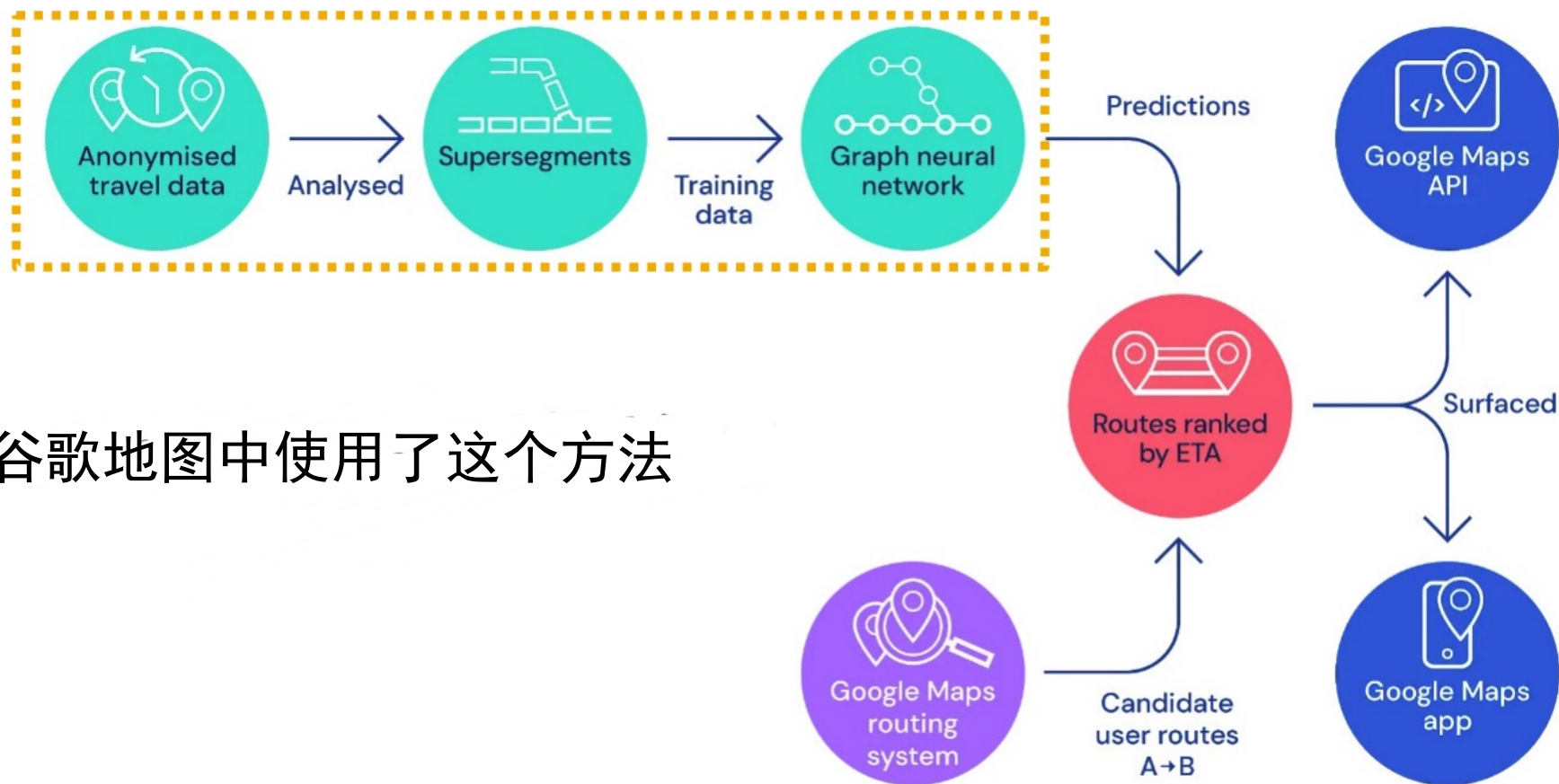


Image credit: [DeepMind](#)

图级任务1：通过图神经网络进行交通预测

- 使用图神经网络来预测到达时间



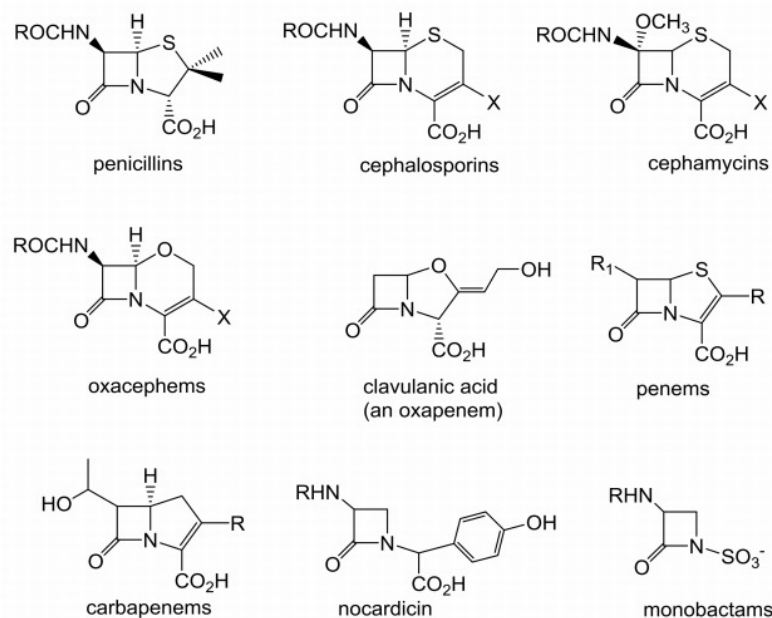
- 谷歌地图中使用了这个方法

THE MODEL ARCHITECTURE FOR DETERMINING OPTIMAL ROUTES AND THEIR TRAVEL TIME.

Image credit: [DeepMind](#)

图级任务2：药物发现

- 抗生素是小分子图
 - **节点：** 原子
 - **边：** 化学键



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

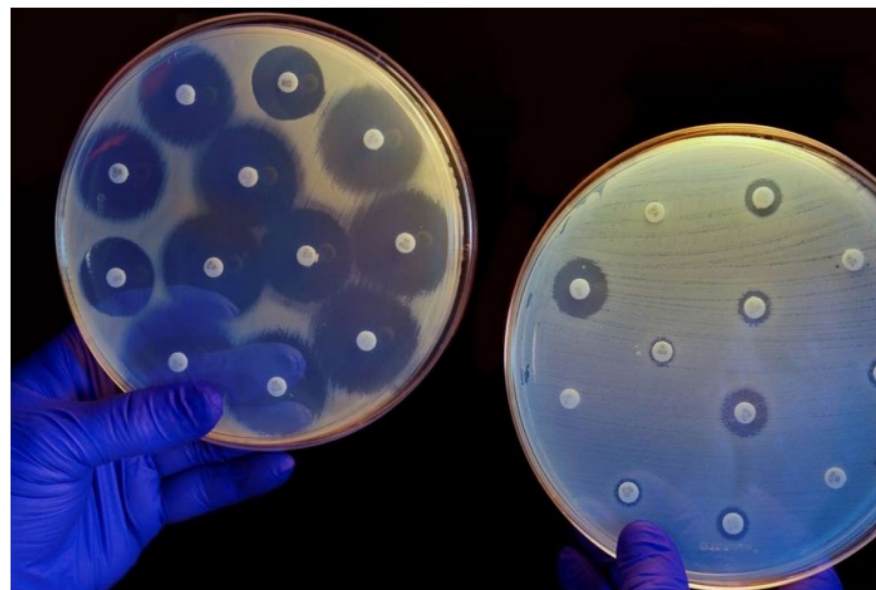
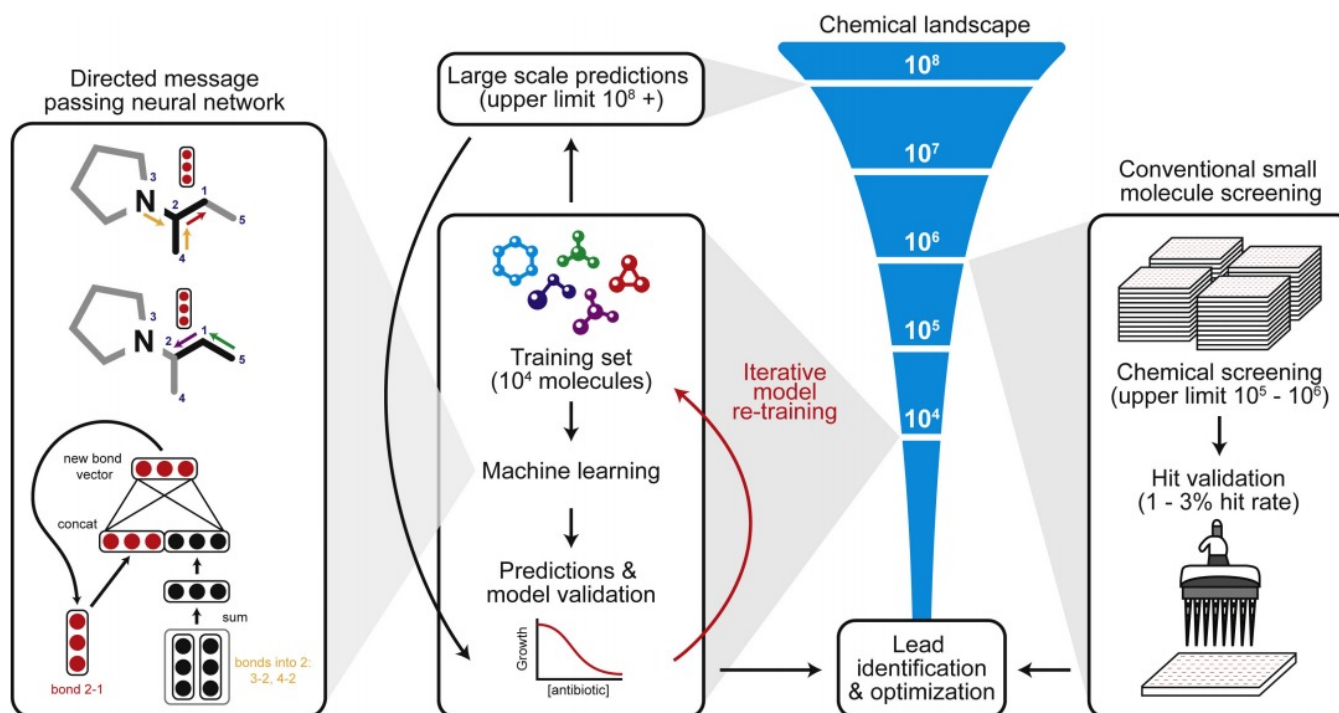


Image credit: [CNN](#)

图级任务2：深度学习在抗生素发现中的应用

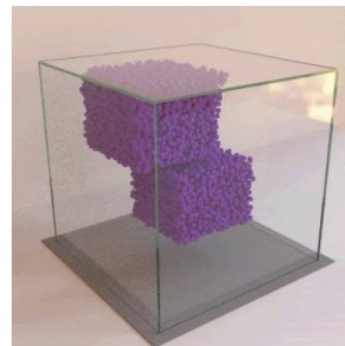
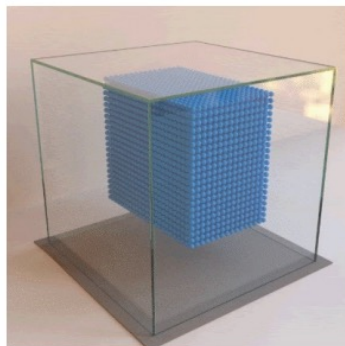
- 选用图神经网络分类模型
- 使用模型从候选池中预测有潜力的分子。



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

图级任务3：物理模拟

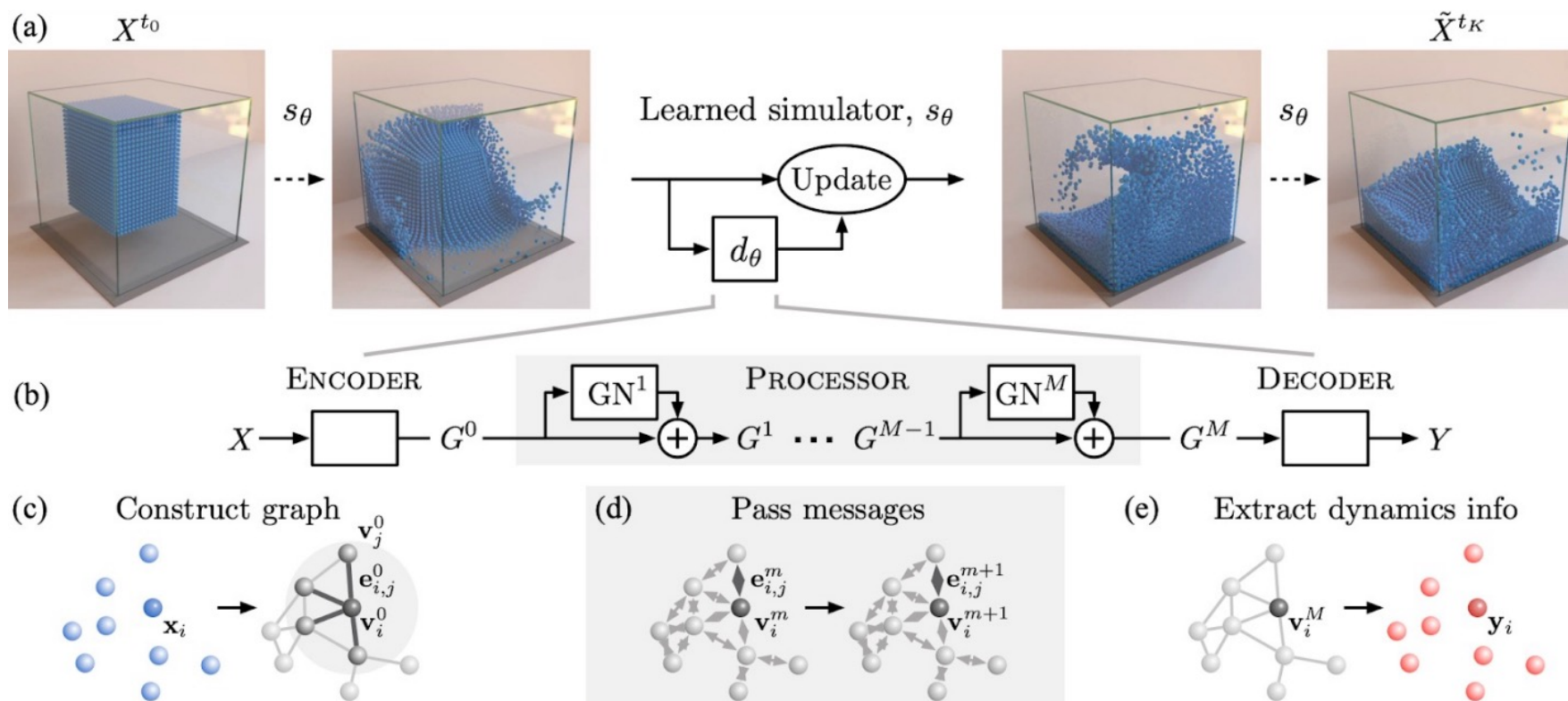
- 将物理模拟视作图：
- **节点**：粒子
- **边**：粒子之间的相互作用



图级任务3：模拟学习框架

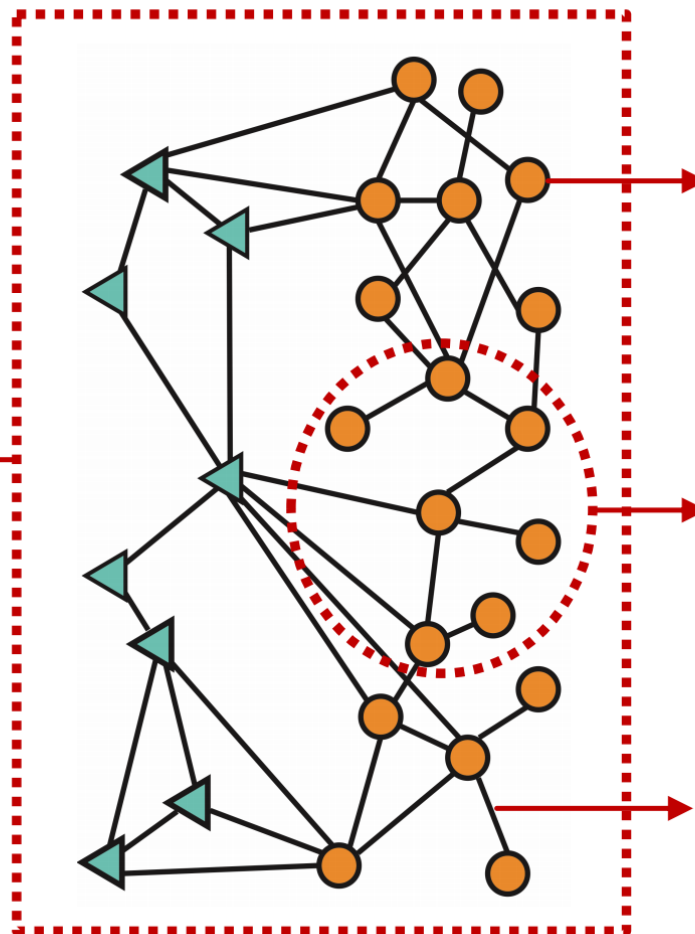
一个图进化任务

- **目标：** 预测图将如何随时间演变。



总结

图级预测、图生成
(Graph-level prediction,
Graph generation): 预
测整个图的属性或生
成新的图



节点级(Node-level):
预测单个节点的属性
或类别

社区(子图)级
(Community (subgraph) level):
预测图中特定社区或子图的
属性

边级(Edge-level): 预
测两个节点之间是否
存在链接