

Machine learning with Graphs

Graph Structure

1. 图可以用来表示两种结构：

1. Networks(AKA Natural Graphs) : 底层本来就是图结构

- Social networks : 人与人之间关系的集合
- Communication and transactions : Electronic devices, phone calls, financial transactions
点与点之间建立联系，联系中有信息
- Biomedicine : Interaction between genes/proteins regulate life
- Brain connections : Our thoughts are hidden in the connections between billions of neurons

2. Graph(as a representations) : 存在 关系结构，可以用图来建模关系结构。不是天然的图，但天然可以用图来建模

- Information/knowledge : organized and linked : organized and linked
- Software (不太懂为啥)
- Similarity networks : connect similar data points
- Relational Structures : Molecules, Scene graphs, 3D shapes, Particle-based() physics simulations

2. Modern ML Toolbox is designed for simple sequences(NLP,Speech) and grids(CV)

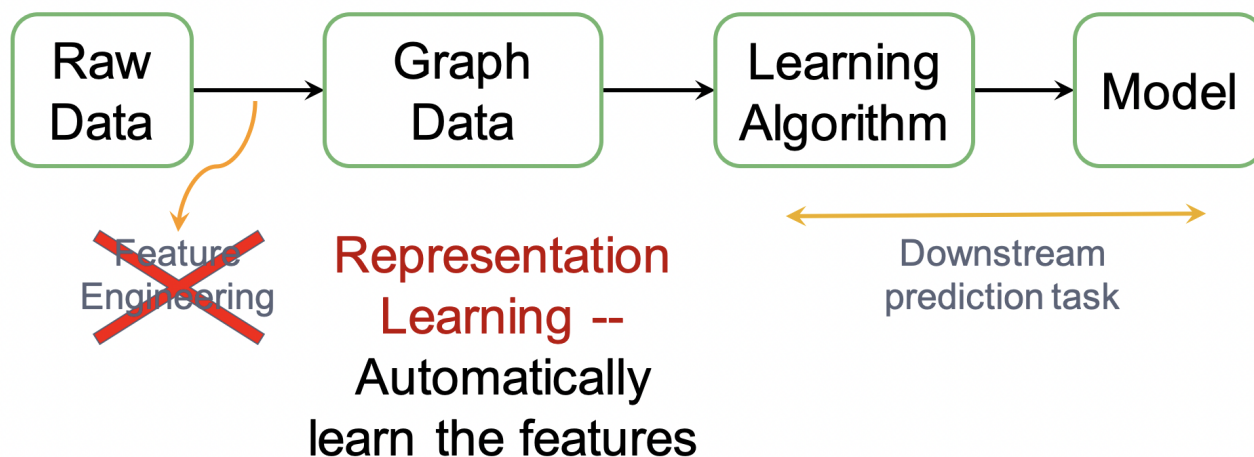
grid本质上还是二维或者说是高维的sequence

sequence,grid有 固定的拓扑结构，都可以实现固定大小。sequence直接截就行，grid可以拉伸压缩等

但是对于图：

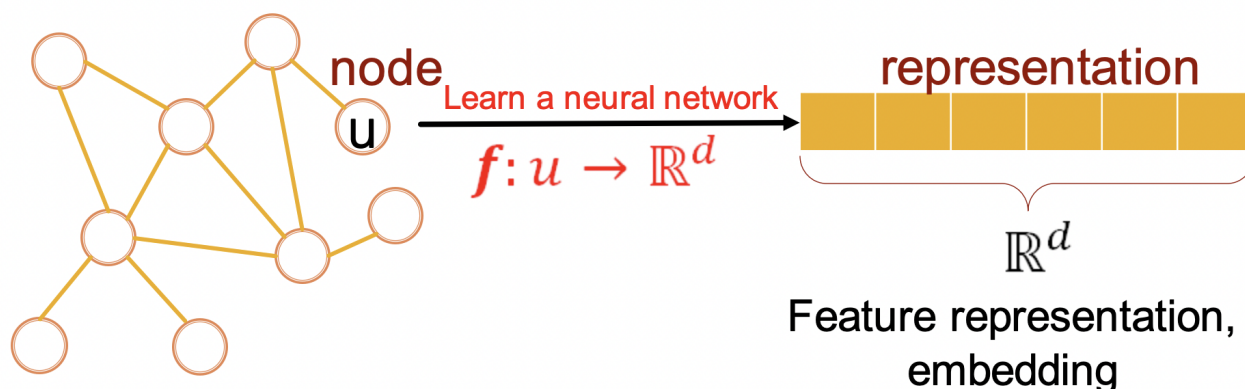
1. 任意大小和复杂拓扑： i.e. : 没有空间局部性(**spatial locality**):sequence有左右的概念，grid有上下左右
2. 没有参考点，没有固定的遍历顺序
3. 动态的，并且具有多个模态的特征

3. 传统ML中，**feature engineering**占用很多精力。图神经网络就是一种表示学习，自动学习图的良好表示，用于下游算法，跳过feature engineering的过程。



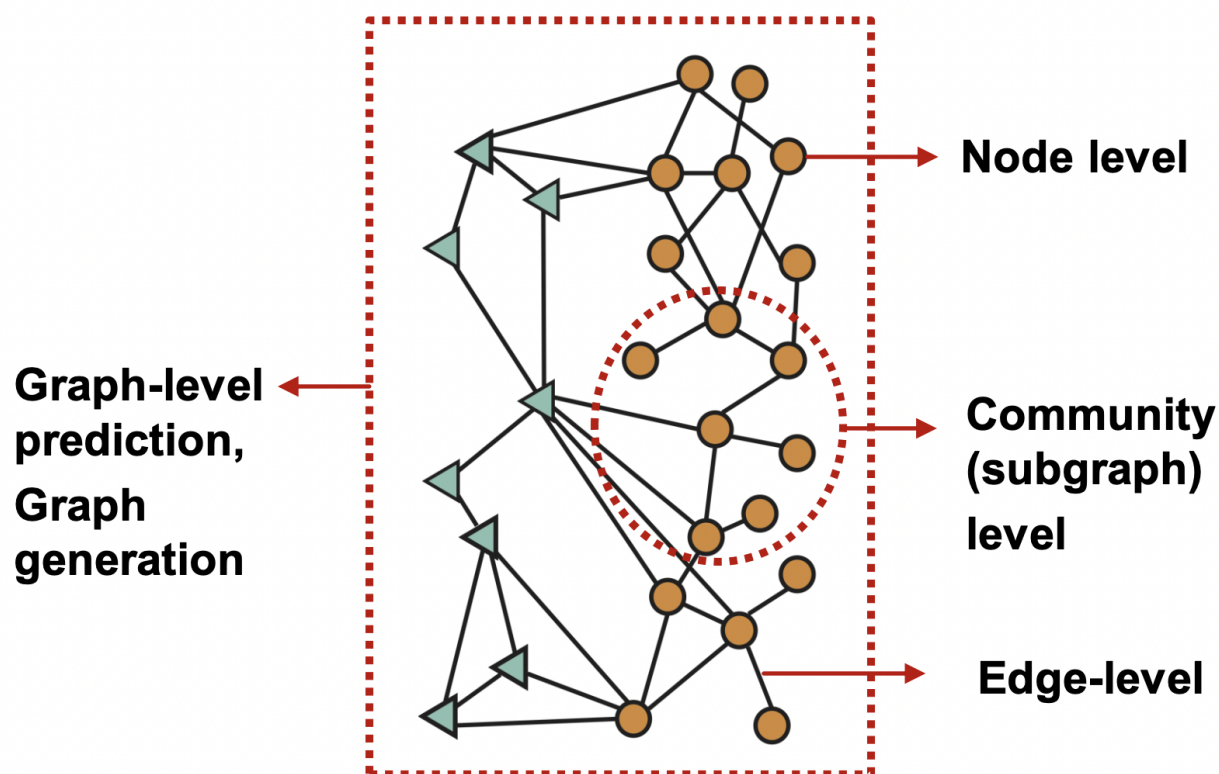
4. 图表示学习就是将图上的点映射为d维嵌入。相似的点的嵌入也应该紧密。

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



Different types of tasks

图上的任务大概有4种：点 → 边 → 子图 → 全图



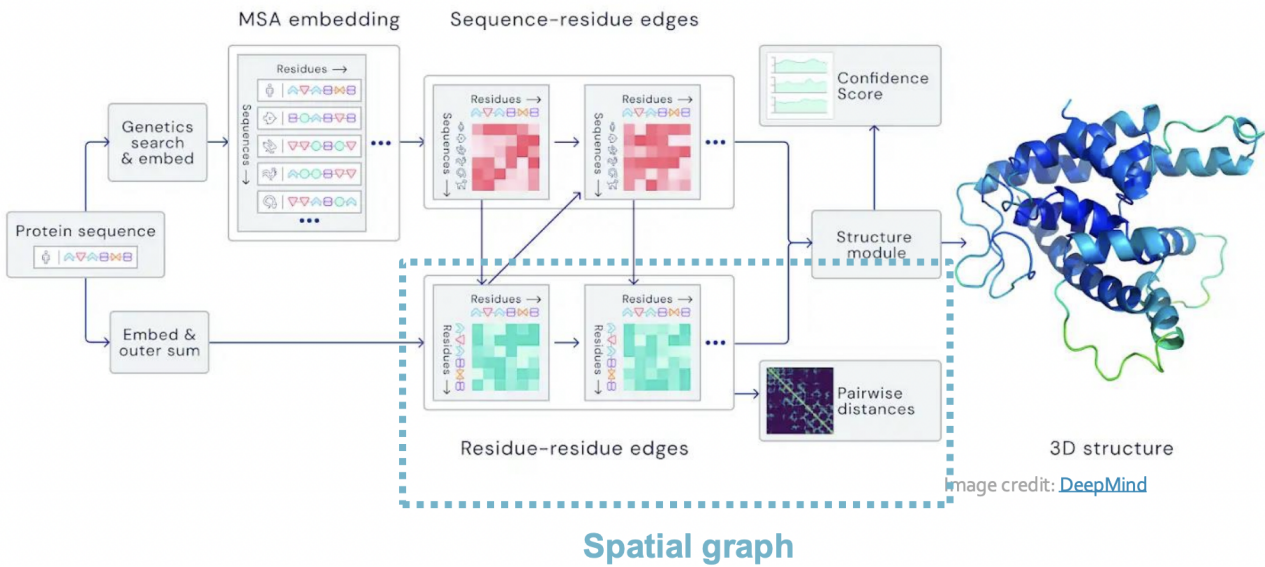
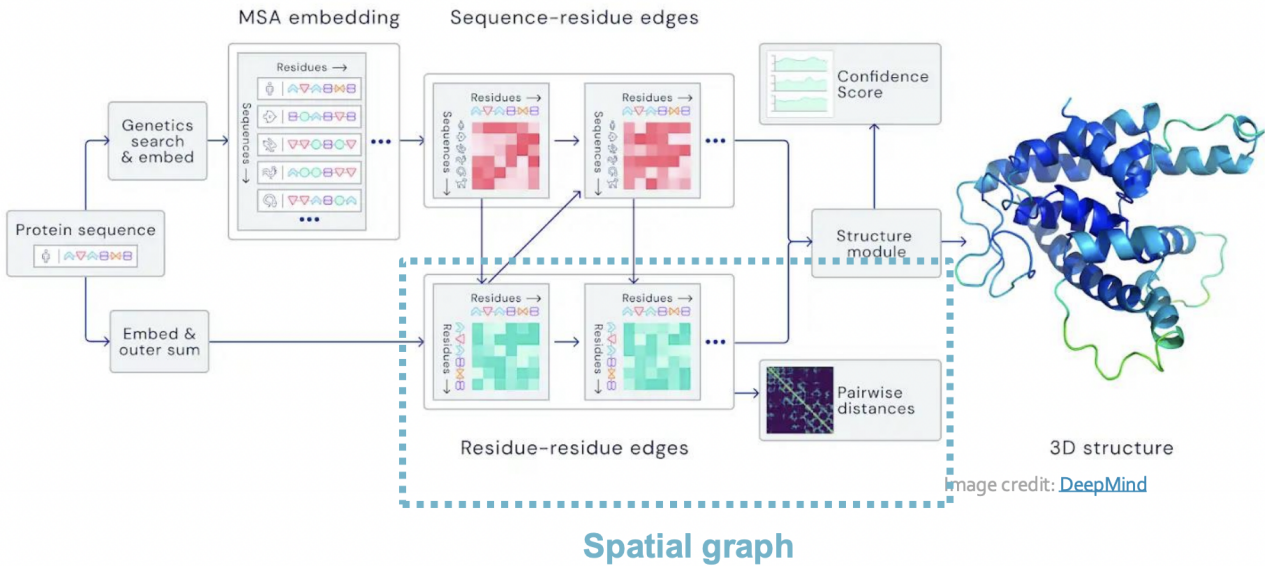
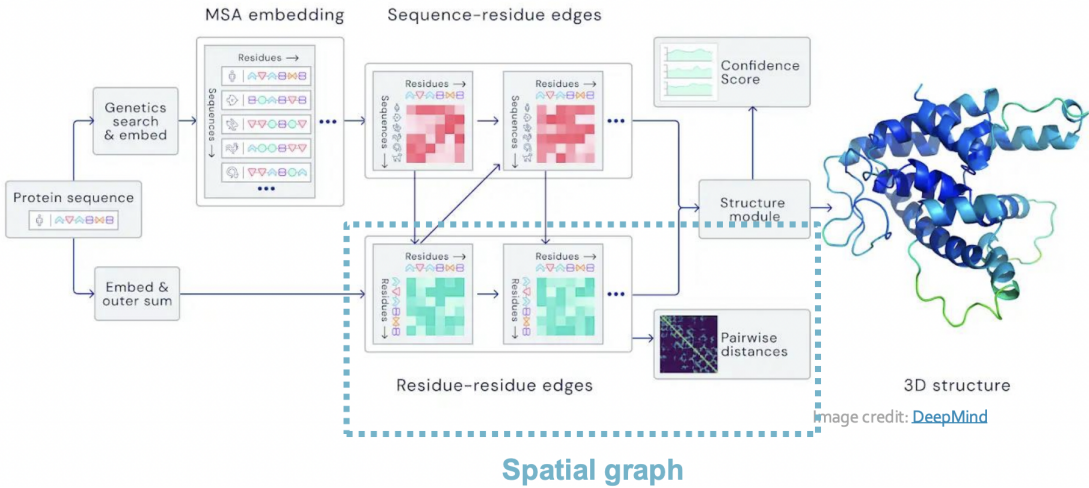
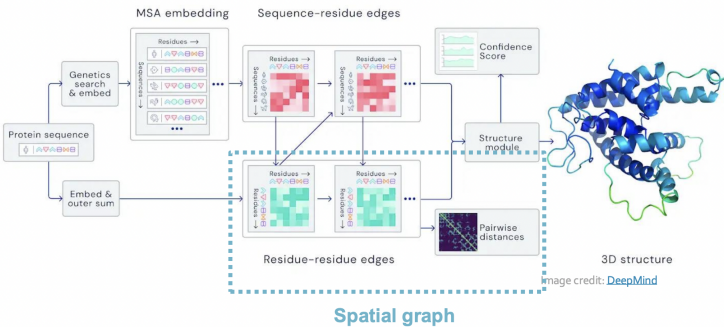
一些经典的图上的任务

- 点分类: 预测点的属性 分类在线用户/商品
- 边预测: 预测两个点之间是否有缺失的边 知识图谱补全
- 图分类: Molecule property prediction
- 聚类: Social circle detection
- 图生成: Drug discovery
- 图进化: Physical simulation(物理模拟).

Node-Level ML Tasks

根据氨基酸序列预测3D结构: Alphafold

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

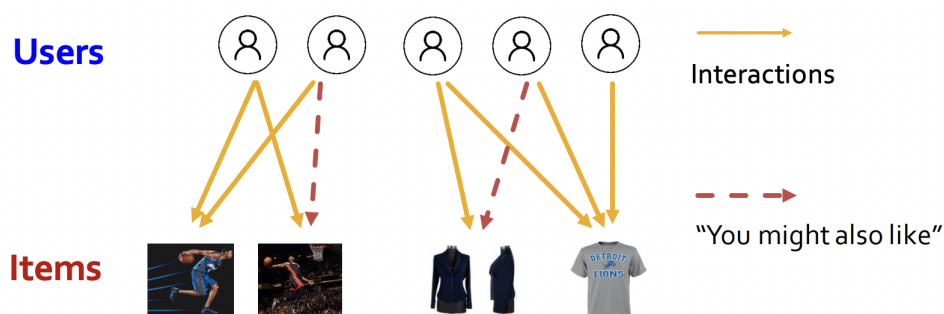


预测3D结构不是在预测边，不要搞混了，是在预测 **点的位置**。另外，Alphafold并非是建立了一个GNN，而是借鉴了GNN的MPNN的思想，也就是“Spatial graph”

Edge-level ML Tasks

1. 推荐系统

- Nodes: Users and items
- Edges: User-item interactions



注意这是一个 **二分图**，边只存在user和item之间，user与user之间，item与item之间，是没有边的。
(如果存在边，这三种边的性质肯定是不一样的，就成了异构图)

PinSage:Graph-based Recommender

- Task:Learn node embedding z_i such that $z_u \cdot z_i$

Choice of a graph representation
