

文章结构

- 点明核心
 - 算法改进
 - 前人方法的不足和改进优点
 - 给出算法流程/公式推导
 - 应用（是针对某一种数据还是普适性算法）
 - 算法应用/迁移
 - 应用场景
- 算法比较
 - Baseline（最少4种以上）： deepwalk、GCN等基本方法+本领域SOTA方法
 - 多维度比较： acc、F1、ROC等
- 应用场景
 - 要有2种以上应用场景
 - 有ground truth最好， 如果没有就要给出合理解释

基于GNN

- 直推式还是归纳式？
- 使用GraphSAGE的思想，更改信息聚合方式，根据任务不同使用LSTM、GRU等
- 融合多种信息聚合方式的图神经网络设计
- 比如deepwalk能否采用LSTM、GRU或attention机制

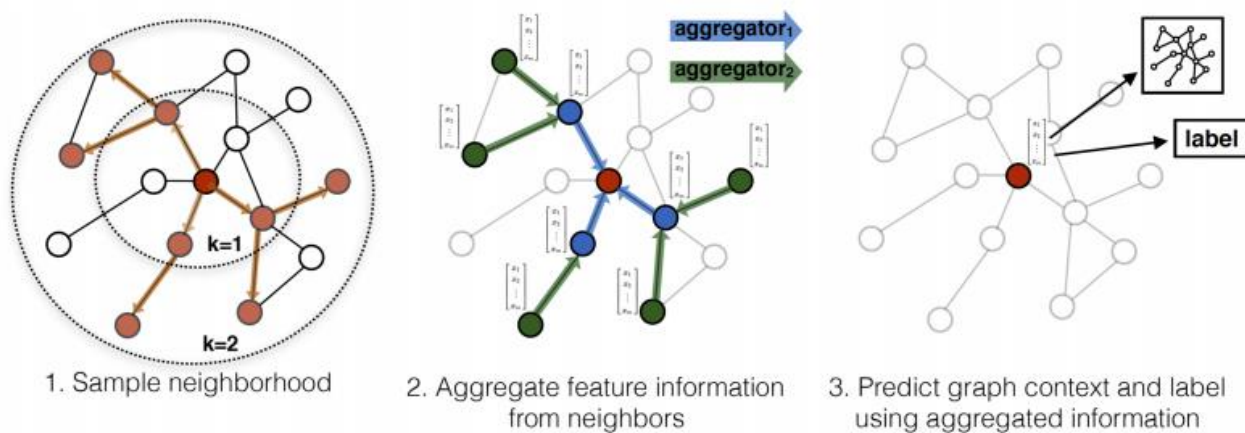


Figure 1: Visual illustration of the GraphSAGE sample and aggregate approach.

图构建算法

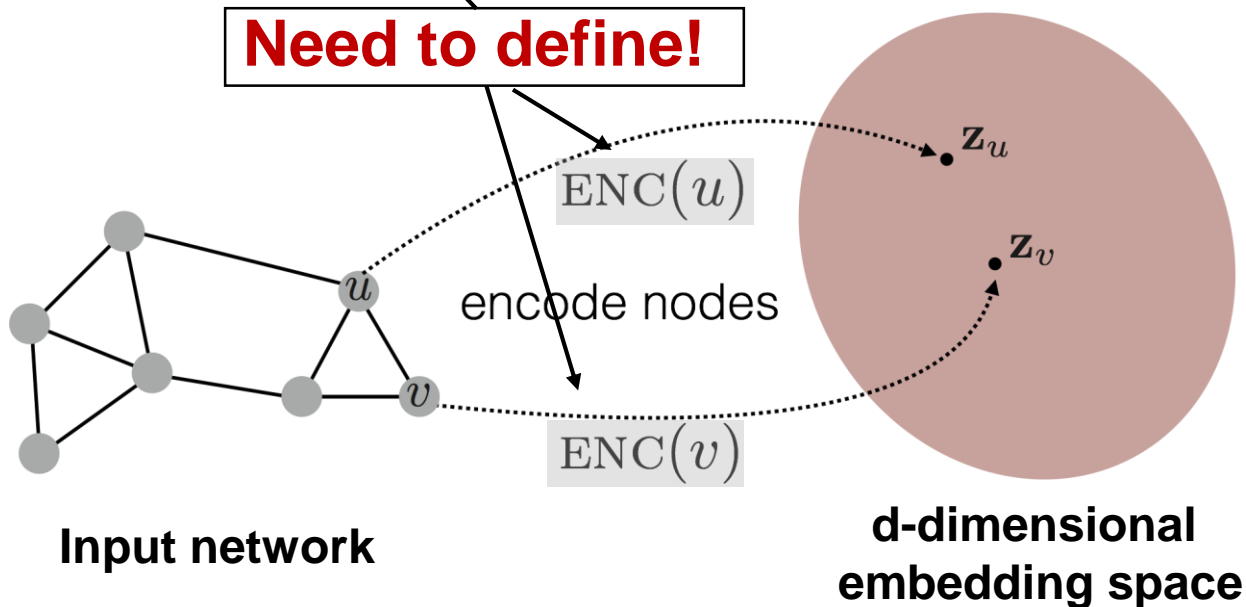
- 分类、聚类任务的样本往往不是graph的节点，而是相互绝缘的
- 如何构建一个图是最可信的（similarity和距离的度量，pearson、spearman、欧氏距离等）？
- 如何针对大规模节点构建图（Million、Billion级别）？

Embedding loss function的改进

- 相似性度量方式的改变（这个改进方向较广泛，比如嵌入前后的KL距离等，cos距离、KL散度、MINE算法等）

Goal: $\text{similarity}(u, v) \approx \mathbf{z}_v^\top \mathbf{z}_u$

Need to define!



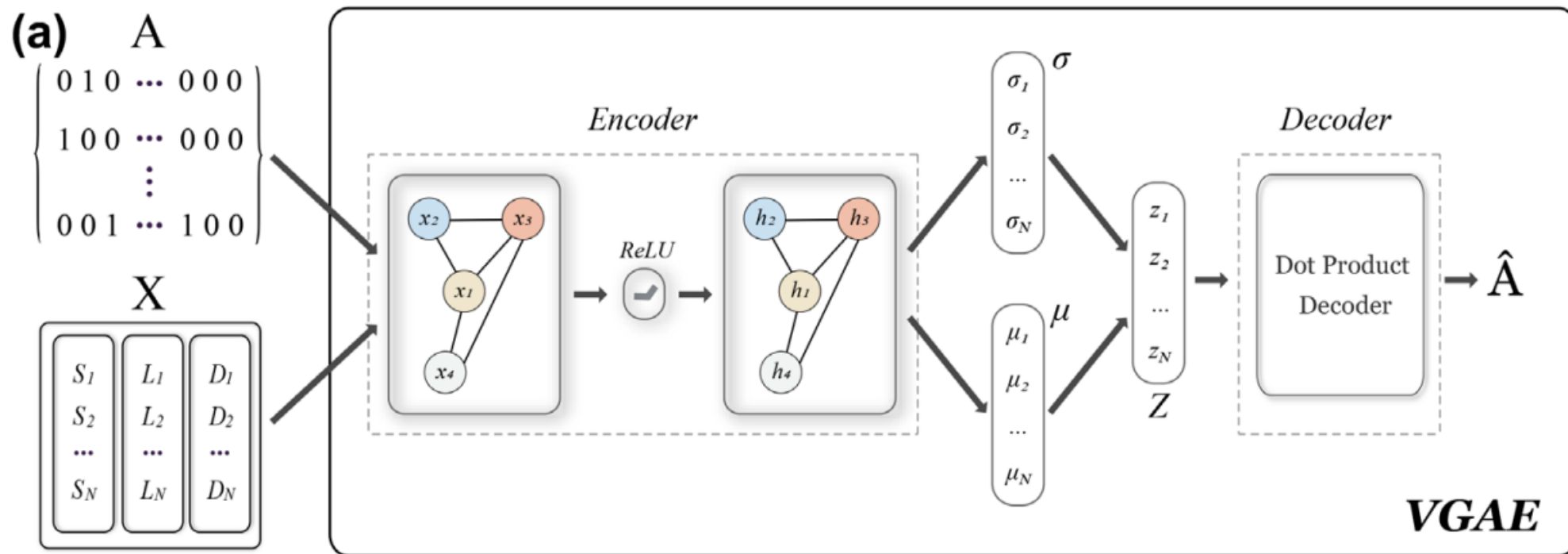
时空图网络

- 结合空间和时间信息进行时间序列上图的变化推断
- 空间：构建一幅图
- 时间：图的前进方式（改变方式）
- 分析：那些节点起到关键作用，预测下一时间点的空间图



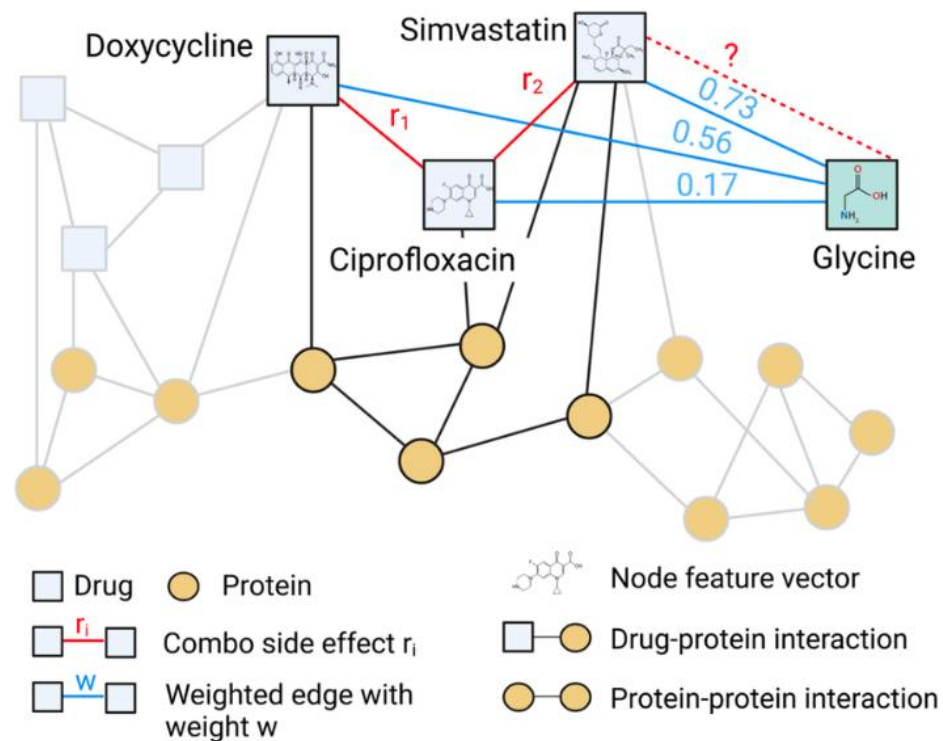
VGAE/graphGAN

- 仿真节点的生成（生成假样本）



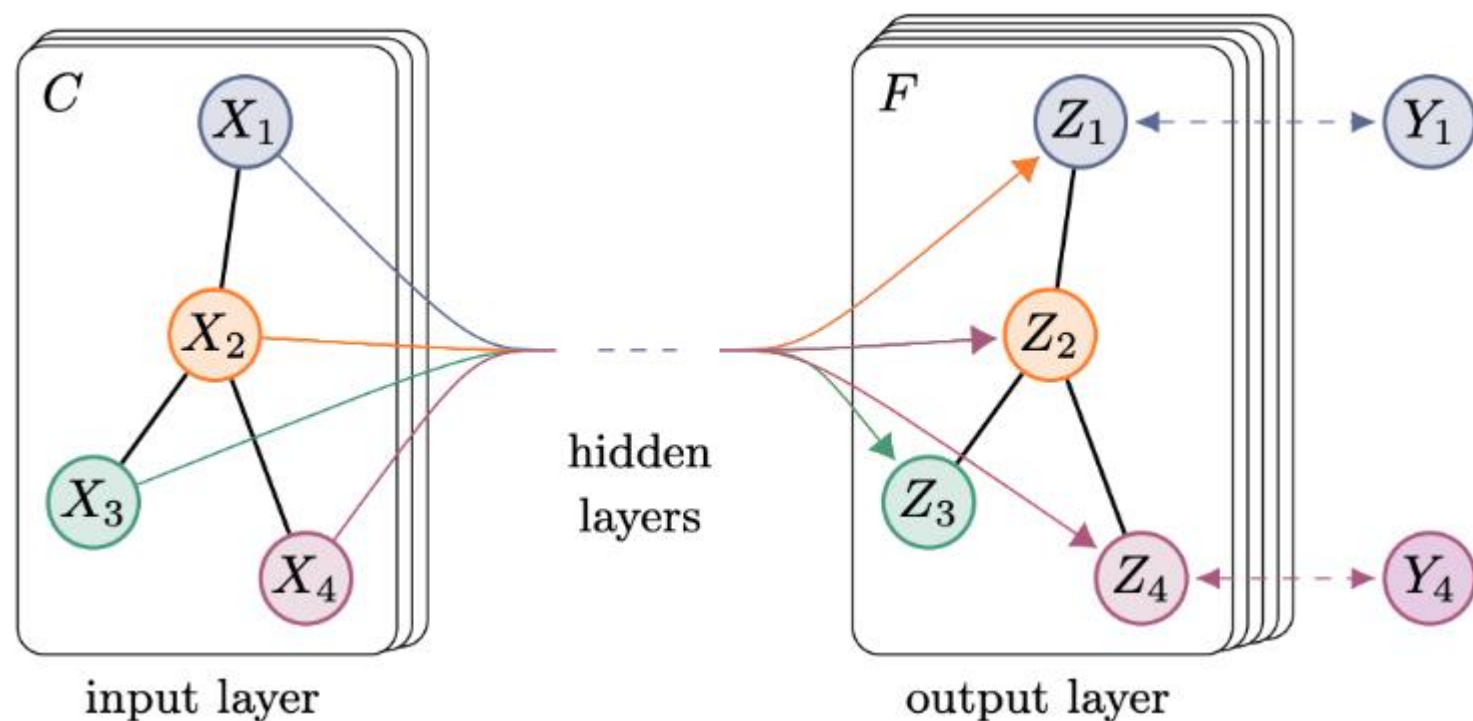
异构图模型在生物医学网络中的应用

- 使用MetaPath
- 使用attention机制



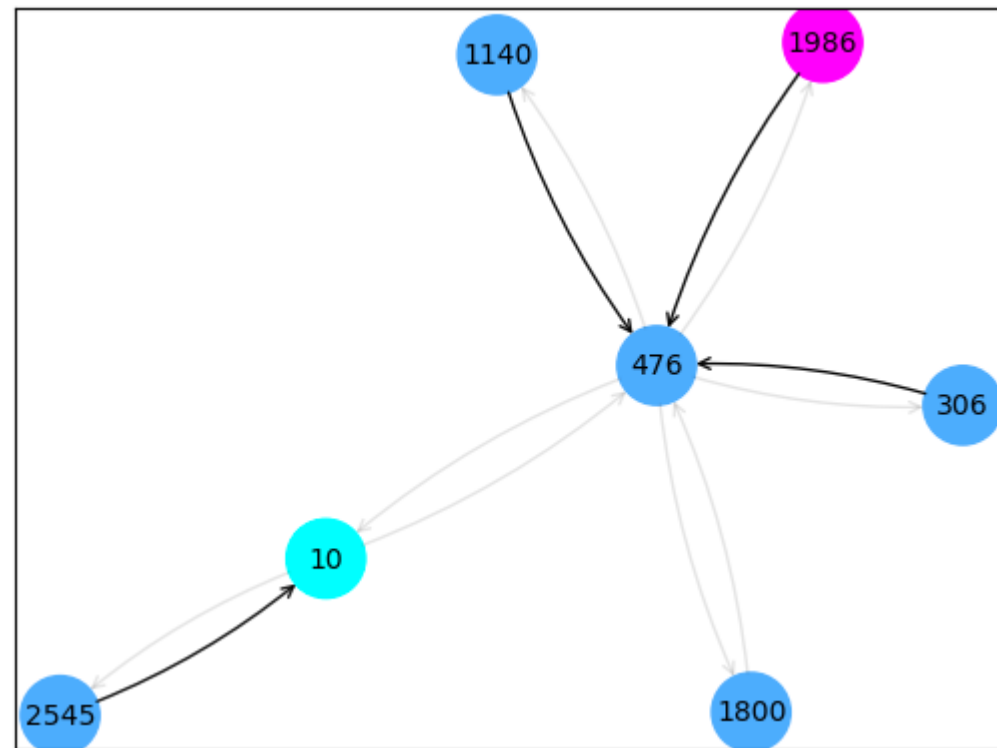
问题讨论：GCN嵌入层数的研究

- 浅层GCN和深层GCN问题的讨论
- 如何避免过平滑问题以及如何简化GCN操作



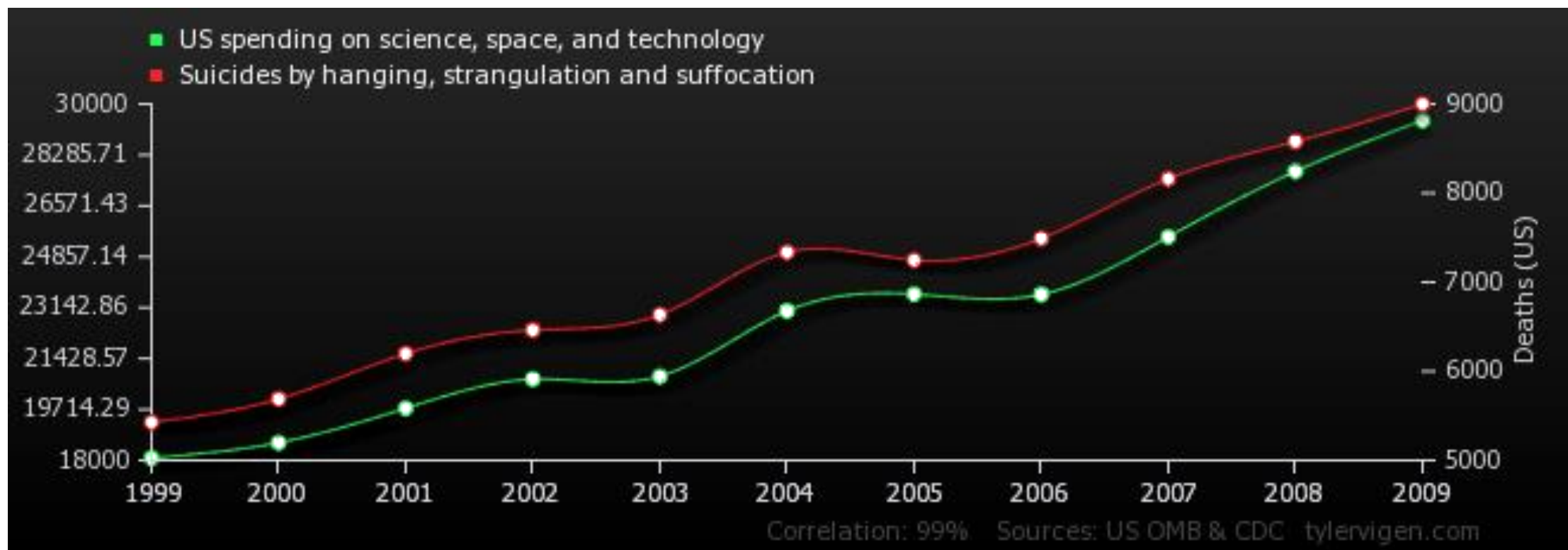
网络解释性问题

- 那些subgraph对于你的任务最重要
- 基于扰动 (GNNExplainer)
- 基于时序
- 对于具体的生物医学任务极其重要



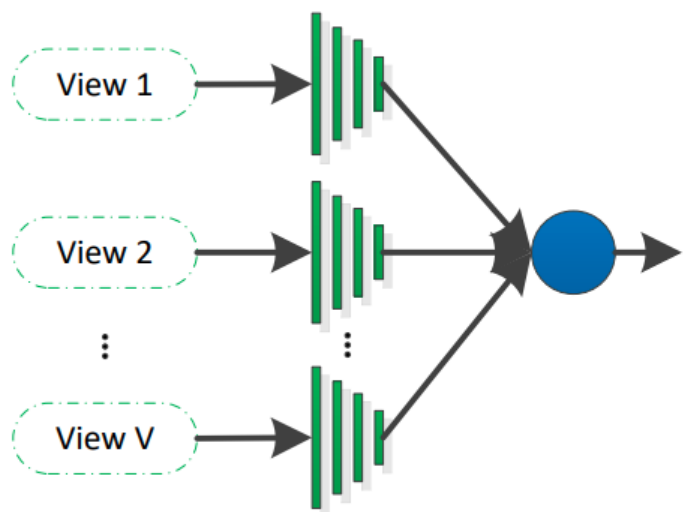
因果推断

- 什么问题需要做因果推断？
- 如何应对Spurious Correlations问题？

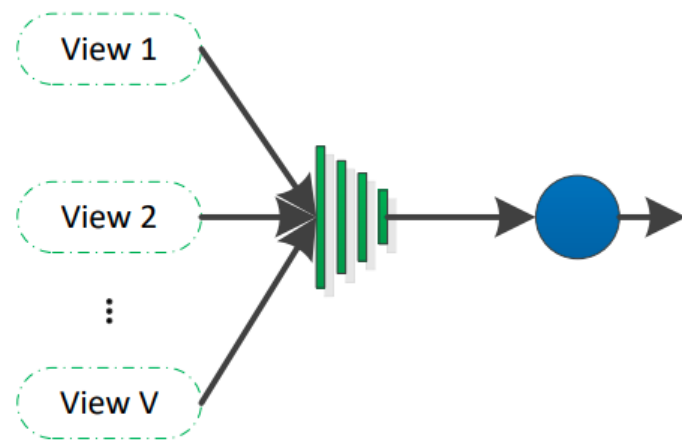


有关于多模态嵌入问题

- 一个图够吗？
- 多模态图数据如何进行融合？
- Multiview-learning



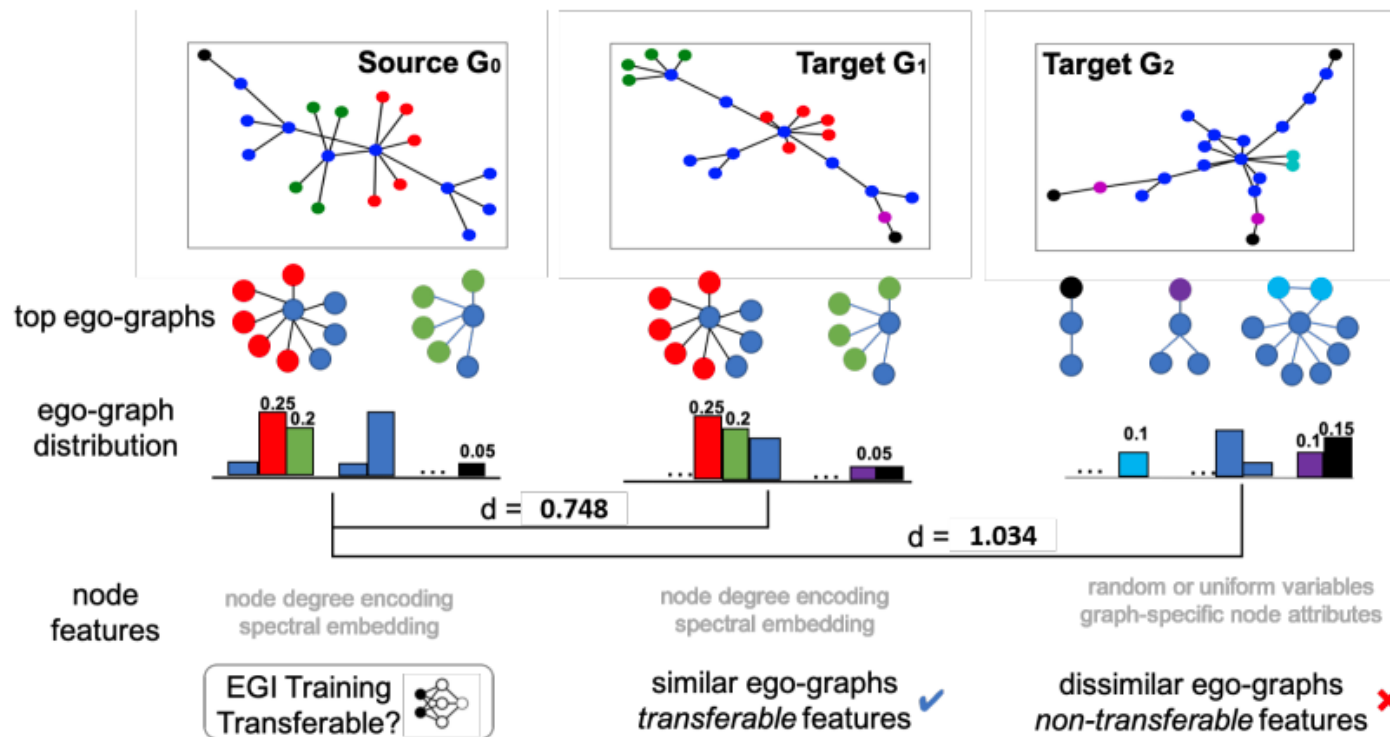
(a) One-view-one-network strategy



(b) Multi-view-one-network strategy

网络的迁移学习

- 可能更偏应用一些
- 需要对可迁移目标进行规划和判断



Baseline和数据集

- Biosnap
 - 20余种生物医学graph数据
- BioNEV
 - 11种图嵌入方法baseline
 - 7种生物医学graph数据
- 社交网络
 - Cora、CiteSeer、PubMed、Reddit、BlogCatalog、YouTube
- 推荐系统
 - [Recommender Systems Datasets \(ucsd.edu\)](https://www.cs.ucsd.edu/~davis/recommender-systems-datasets/)
 - MovieLens
 - Last.fm