FGL 论文总结

组会想法

20211207

- 框架的思路：现有的一些组织或公司存在大量的图数据，可以用来训练强大的 GNN 模型，但是由于集中数据训练不现实（隐私约束，监管限制，商业竞争），所以需要借助联邦学习框架，可以在不直接获取原始数据的情况下，联合多个客户端进行全局模型的训练，但是由于各个客户端上的数据存在统计异质性（non-IID）和隐私泄露的风险。提出新的方法解决不同客户端上图数据的异质性问题
- 为什么隐私保护方法？传统联邦学习设置的隐私安全性不足。增强 FL 设置中的安全：算法私有化，安全聚合，同态加密，差分隐私
- 缓解异质性问题的方法：元学习，关系注意力和聚合，全局自监督，聚类，单独的网络架构。 知识蒸馏，原型，超网络…见word文档，原型一文总结模型/数据异质性的方法

讨论

传统的FL架构能否适应Graph数据？需要做什么改进？

- 可以在某个具体场景中挖掘共有的问题，现在那些隐私安全，数据异质性，通信问题都是联邦学习本身中存在的问题，联邦和图结合又会有什么新问题呢？思考阅读论文。
- 寻找客观存在的问题但是没有人解决，可以从联邦学习或者图学习中的一个根本的假设入手，随着不同场景的需要，可能这个假设根本就不满足不现实。例如，联邦学习中部分客户端上的数据不足，分布异质性，可能有的节点上有数据有的没有数据，通信问题或者中心服务器的设置不合理。 Setup
- 可以分别从联邦学习和图学习最新的论文中的问题去入手，分别列举出两个方向的问题，去找问题的共同点或者互补点，这样如果能找到很巧妙的结合点

目前针对FL挖掘的问题较多，Graph方面欠缺

20211214

- 暂时不考虑推荐和知识图相关的方向。
- 框架：总结当前联邦学习和图学习结合的框架中解决现有问题的方法，思考该方法的缺陷，如何更好地弥补缺陷。
- 数据方面：从解决一般数据存在的问题，移植到图数据会出现的问题，将图数据应用到联邦学习又会出现什么问题。
- 这段时间一直看的是联邦学习和图结合的论文，下来应该再多看一些图神经网络相关的论文，对这部分不是很熟悉。

FL Challenges

1. 现有的联邦图神经网络基于一个集中的服务器来协调训练过程，这在许多现实应用中是不可接受的，比如跨竞争银行构建金融风险控制模型，没有银行愿意将自己的信息提交到公共的中心服务器。大多数之前的工作都假定有一个中心服务器来聚合从客户机节点收集的模型信息，这样的强有力的假设限制了现实应用中图数据的联邦学习模型的进一步发展。 第三方可信组织 推翻中央服务器的假设
2. 图数据的分散分布，在每个客户端上的数据分布不均衡，非独立同分布，联邦平均之后的全局模型精度会受到严重影响。是否可以针对不同的客户端的模型进行个性化训练，对联邦学习过程的参数聚合时候做算法优化。 已有很多PFL的工作，主要针对架构的异质性
3. 很少考虑中心服务器的通信成本。中央服务器必须处理比客户端多几十倍的通信负载。
4. 解决的问题和联邦学习类似，但是加上隐私安全保护或者移植已有的方法降低 FGL 中的通信开销，也算创新点（组合策略）。 创新性太小
5. GNN 模型高性能的重要原因是拥有高质量的图数据用来训练，包括丰富的节点特征和完整的相邻信息，现实应用中很少满足这种需求。 图链接预测，节点自监督
6. 在 VFGL 中，对于数据垂直划分时普遍存在的 GNN 的问题，目前的研究较少。 金融诈骗检测，知识图嵌入
7. 隐私保护

GNN Challenges

1. class imbalance 问题。现有 GNN 默认节点 samples 的类别满足 balance 的条件，但是在现实世界中可能存在 class imbalance 的情况，如果直接训练 GNN model 可能存在依赖偏好导致模型性能欠佳。
2. graph-level 的图分类任务。当建立一个用于图分类的 GNN model 时，训练集中的图数据假定满足同分布。然而在现实世界中，同一数据集中的图可能具有差异性很大的不同结构，即图数据彼此之间可能是非独立同分布的。
3. 解决图结构和特征异质性问题。
4. 图数据的半监督图学习

1. 按照任务分类

1.1. Graph Classification

- （聚类 解决图结构的结构和特征异质性）Federated Graph Classification over Non-IID Graphs_NeurIPS_2021 [Paper] [Code]
- （无服务器 多任务 证明不需要集中的拓扑）SpreadGNN Serverless Multi-task Federated Learning for graph neural networks_ICML_2021 [Paper] [Code]

1.2. Node Classification

- (处理 non-IID 图数据，贝叶斯优化调整超参数) ASFGNN Automated separated-federated graph neural network_PPNA_2021 [Paper] [No Code]
- (半监督节点分类) GraphFL: A Federated Learning Framework for Semi-Supervised Node Classification on Graphs_Arxiv 2020 [Paper] [No Code]
- (节点分类 子图 邻居缺失) Subgraph Federated Learning with Missing Neighbor Generation_NeurIPS_2021 [Paper] [Code]
- (自动学习 神经架构搜索)FL-AGCNS Federated Learning Framework for Automatic Graph Convolutional Network Search_arxiv_2021 [Paper] [No Code]
- (去中心 节点分类 减少通信负担 联邦+图)Decentralized Federated Graph Learning with Traffic Throttling and Flow Scheduling_IWQOS 2021 [Paper] [No Code]
- (全局自监督学习)FedGL_Federated Graph Learning Framework with Global Self-Supervision_arxiv_2021 [Paper] [No Code]
- (去中心服务器)Decentralized Federated Graph Neural Networks_IJCAI 2021 [Paper] [No Code]
- (同态加密)A Vertical Federated Learning Framework for Graph Convolutional Network_arxiv_2021 [Paper] [No Code]
- ~~*FedGLF: Federated Graph Learning Framework for Node Classification*~~

1.3. Position Predicting

- Federated Dynamic GNN with Secure Aggregation_Arxiv_2020 [Paper] [No Code]

1.4. Image Classification

- (聚类 知识蒸馏 联邦+图)Cluster_Driven Graph Federated Learning Over Multiple Domains_CVPRW_2021 [Paper] [No Code]

1.5. Traffic flow prediction

- (联邦+GNN 交通流预测)Cross-Node Federated Graph Neural Network for Spatio-Temporal Data Modeling_arxiv_2021 [Paper] [Code]
- (多用户数据序列 安全聚合) Federated Dynamic GNN with Secure Aggregation arxiv_2020 [Paper] [Code]

1.6. Social Recommendation

- (社交推荐)Federated Social Recommendation with Graph Neural network_arxiv 2021

[Paper] [No Code]

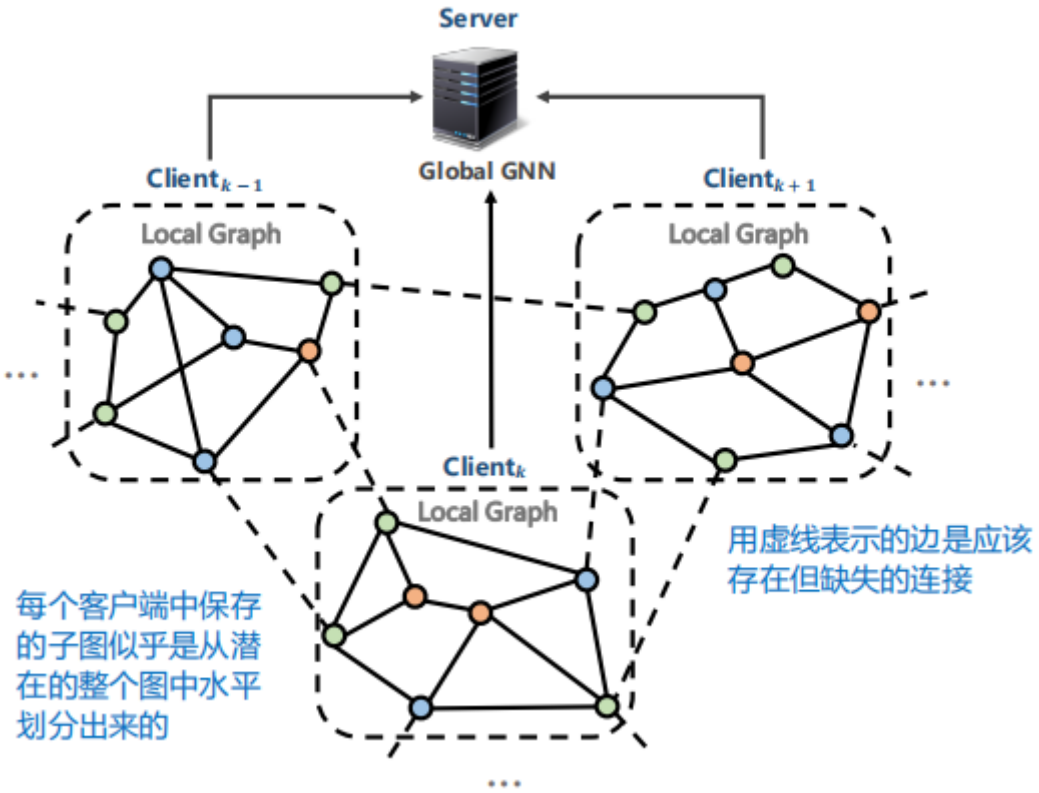
- (隐私推荐 联邦+GNN 解决异质性，个性化，隐私问题)FedGNN Federated Graph Neural Network for privacy-preserving recommendation_ICML 2021 [Paper] [No Code]

1.7. GraphFL Benchmark System

- ICLR 2021 FedGraphNN: A Federated Learning Benchmark System for Graph Neural Networks [Paper] [Code]

2. 按照结构分类

2.1. Horizontal Intra-graph FL



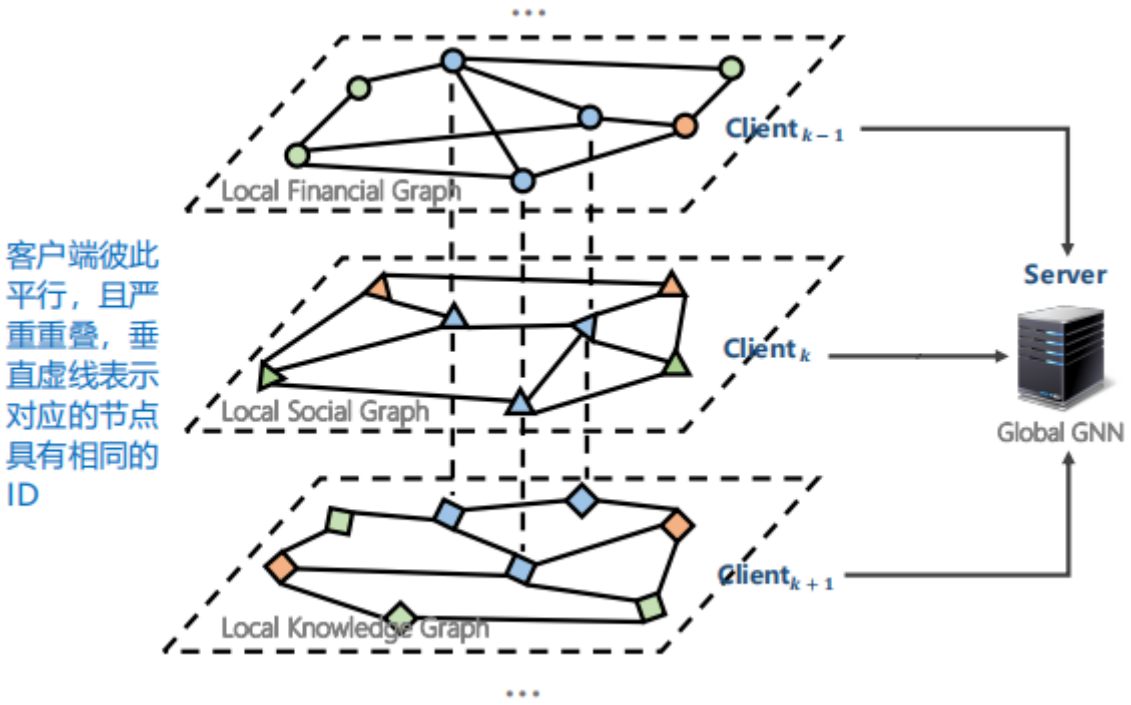
子图横向联邦学习在现实世界中非常常见。例如社交 app 中，每个用户都有一个本地社交网络。开发人员能够设计出基于横向联邦的好友推荐算法，避免侵犯用户的隐私数据。

发现客户端本地子图之间的潜在结构关系是横向图内联邦的一个重要挑战。

- （缺失子图的边）Subgraph Federated Learning with Missing Neighbor Generation_NIPS 2021 [Paper] [Code]
- （隐私推荐 解决异质性，个性化，隐私问题）FedGNN: Federated Graph Neural Network for Privacy-Preserving Recommendation _ICML 2021[Paper] [No Code]

- （全局自监督信息）FedGL: Federated Graph Learning Framework with Global Self-Supervision_Arxiv 2021 [Paper] [No Code]
- （处理 non-IID 超参数优化）ASFGNN: Automated Separated-Federated Graph Neural Network_PPNA 2021 [Paper] [No Code]
- （分布式训练 GCN）Distributed Training of Graph Convolutional Networks_TSIPN 2021 [Paper] [No Code]
- （分裂学习 交替优化 交通流预测）Cross-Node Federated Graph Neural Network for Spatio-Temporal Data Modeling_KDD 2021 [Paper] [Code]

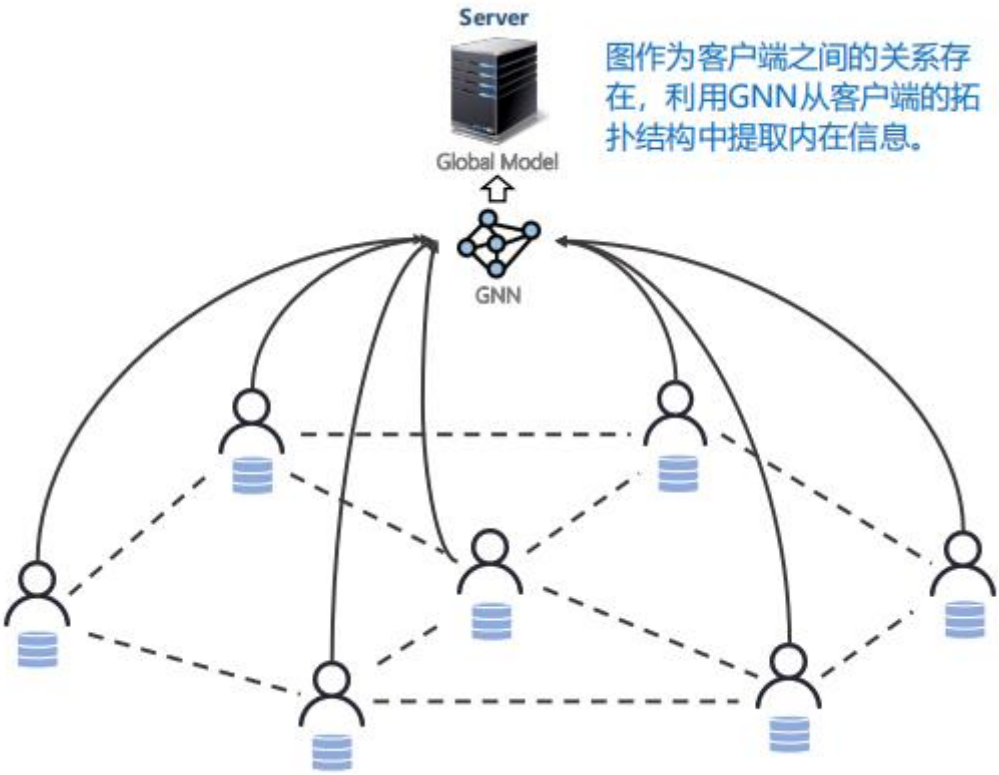
2.2. Vertical Intra-graph FL



子图纵向联邦下客户端数据彼此并行，彼此节点 ID 重叠，但是不共享同一节点特征域和标签域。全局模型不是唯一的（取决于有多少客户端有标签），这表明子图纵向联邦支持多任务学习。可以应用于不同组织之间的合作。

- (纵向联邦+GCN 同态加密) A Vertical Federated Learning Framework for Graph Convolutional Network_arxiv_2021 [Paper] [No Code]
- (数据隔离问题 差分隐私) Vertically Federated Graph Neural Network for Privacy-Preserving Node Classification_Arxiv 2021 [Paper] [No Code]
- (私有数据收集 易受到攻击) Graph-Fraudster: Adversarial Attacks on Graph Neural Network Based Vertical Federated Learning_arxiv_2021 [Paper] [No Code]

2.3. Graph-structured FL



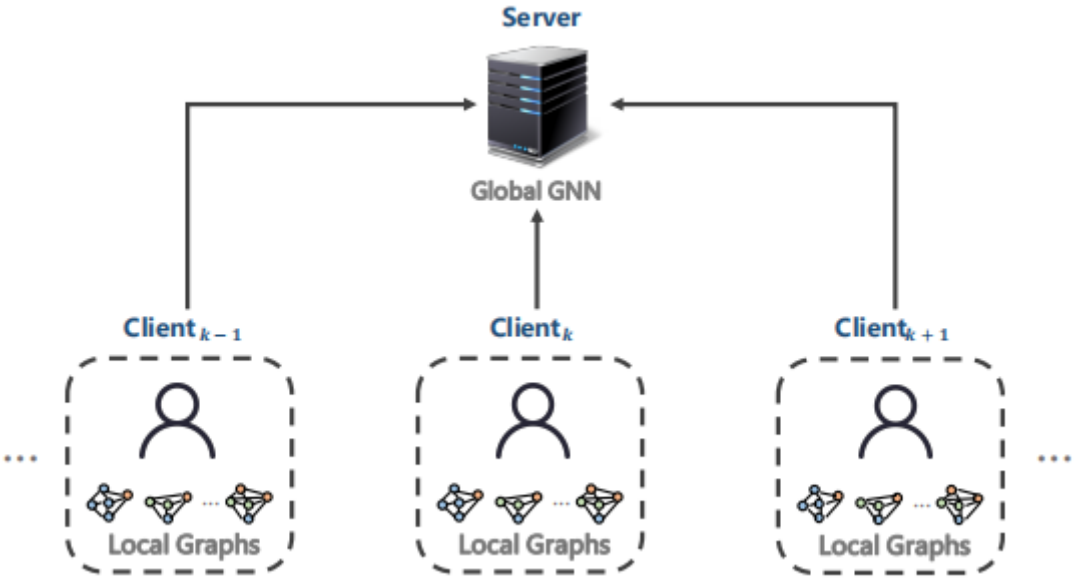
除了数据之外，图还可以表示本地客户端之间的关系，特殊的联邦优化方法

将图神经网络和连联邦学习结合来解决 non-IID 的问题，将多个客户端建模为图结构，每个 client 是一个 node，通过客户端的拓扑结构聚合本地模型，共享传递不同节点的数据信息，训练全局模型。典型的应用是联邦交通流量预测，监测设备分布在不同的地理位置，GNN 用于捕获设备之间的空间依赖关系。

- （无中心服务器 分散并行 SGD）Decentralized Federated Graph Neural Networks_IJCAI 2021 [Paper] [No Code]
- SpreadGNN: Serverless Multi task Federated Learning for Graph Neural Networks _ICML 2021 [Paper] [Code]
- （分布式优化 GCN）Distributed Training of Graph Convolutional Networks_TSIPN 2021 [Paper] [No Code]
- （多 server 算法私有化 差分隐私）A Graph Federated Architecture with Privacy Preserving Learning_Arxiv 2021 [Paper] [No Code]
- Cross-Node Federated Graph Neural Network for Spatio-Temporal Data Modeling_KDD 2021 [Paper] [Code]
- （聚类 知识蒸馏）Cluster-driven Graph Federated Learning over Multiple Domains_CVPR 2021 [Paper] [No Code]

2.4. Inter-graph

客户端的每个样本都是整图数据，全局模型执行图级任务。最典型的应用是分子图分类，生物化学领域。



大部分是在传统 FL 中遗留的问题并且在图中变得更加复杂。如 no-IID 数据分布问题、通信效率和鲁棒性等。

3. 问题挑战

3.1. 隐私保护

- (多服务器 图结构联邦+隐私保护学习) A Graph Federated Architecture with Privacy preserving learning_arxiv_2021[Paper] [No Code]
- (安全聚合免受推理攻击) Federated Dynamic GNN with Secure Aggregation_arxiv_2020 [Paper] [Code]
- (社交推荐)Federated Social Recommendation with Graph Neural network_arxiv 2021 [Paper] [No Code]
- (隐私推荐 联邦+GNN 解决异质性，个性化，隐私问题)FedGNN Federated Graph Neural Network for privacy-preserving recommendation_ICML 2021 [Paper] [No Code]
- (纵向联邦+GCN 同态加密)A Vertical Federated Learning Framework for Graph Convolutional Network_arxiv_2021 [Paper] [No Code]

3.2. 数据异质性

几乎在 FL 设置中是不可避免的问题，会影响模型的收敛速度和精度。

图结构的属性包括度分布、平均路径长度、平均聚类系数等。研究这些属性的 no-IID 可能是解决图域中 no-IID 问题的一个重要方面。

- (聚类 图分类 解决图结构的结构和特征异质性) Federated Graph Classification over Non-IID Graphs_NeurIPS_2021 [Paper] [Code]
- (聚类 知识蒸馏 联邦+图)Cluster_Driven Graph Federated Learning Over Multiple Domains_CVPRW_2021 [Paper] [No Code]
- (半监督节点分类) GraphFL: A Federated Learning Framework for Semi-Supervised Node Classification on Graphs_Arxiv 2020 [Paper] [No Code]
- (全局自监督学习)FedGL_Federated Graph Learning Framework with Global Self-Supervision_arxiv_2021 [Paper] [No Code]
- (社交推荐)Federated Social Recommendation with Graph Neural network_arxiv 2021 [Paper] [No Code]
- (隐私推荐 联邦+GNN 解决异质性，个性化，隐私问题)FedGNN Federated Graph Neural Network for privacy-preserving recommendation_ICML 2021 [Paper] [No Code]

3.3. 通信开销

在现实中应用联邦学习算法时，通信和内存消耗是一个关键的瓶颈。对基于联邦学习的推荐系统来说，服务器和客户之间传输的模型可能很大，其中用户-项表示层占据了大部分的模型参数，并且表示参数的大小随着用户-项目规模的不断增加而线性增长。这对通信和内存消耗都是不利的。当前模型量化、修剪、提炼是模型压缩的有效方法。因此，GNN 的压缩技术也是 FGL 的一个潜在研究方向。

- (去中心 节点分类 减少通信负担)Decentralized Federated Graph Learning with Traffic Throttling and Flow Scheduling_IWQOS 2021[Paper] [No Code]

4. 传统 FL 架构的变体

4.1. Multi-server

- (多服务器 图结构联邦+隐私保护学习) A Graph Federated Architecture with Privacy preserving learning_arxiv_2021 [Paper]
- (去中心化 节点分类)Decentralized Federated Graph Learning with Traffic Throttling and Flow Scheduling_IWQOS 2021 [Paper] [No Code]

4.2. Serverless

- (无服务器 多任务 联邦+GNN) SpreadGNN: Serverless Multi task Federated Learning for Graph Neural Networks_ICML 2021 [Paper] [Code]
- (去中心服务器 点对点客户端)Decentralized Federated Graph Neural Networks_IJCAI 2021 [Paper] [No Code]

5. 其他

5.1. Personalized Federated Learning

- Personalized Federated Learning using Hypernetworks_ICML 2021 [Paper] [Code]

5.2. Federated Learning on Non-IID

- Federated Learning on Non-IID Data Silos: An Experimental Study_Arxiv 2021 [Paper] [Code]

5.3. Knowledge Graphs

- (联邦+虚拟知识图 日志分析) Virtual Knowledge Graphs for Federated Log Analysis_ARES_2021 [Paper] [No Code]
- (知识图嵌入学习) FedE Embedding Knowledge Graphs in Federated Setting_arxiv_2020 [Paper] [Code]
- (知识图嵌入+差分隐私) Differentially Private Federated Knowledge Graphs Embedding_arxiv_2021 [Paper] [Code]