

## 中科院-华为• ICT 领域的数学专题研讨班 之一：图网络理论与应用

图网络模型、理论和算法在许多前沿领域显示了强大的基础性作用，并得到广泛应用。为了进一步探讨图分析、图计算、图神经网络等技术的理论与应用，中国科学院数学与系统科学研究院和华为技术有限公司于 2021 年 4 月 7 日联合举办图网络理论与应用专题研讨会（线上），欢迎各位专家参加。

我们邀请华为公司的专家，讲解图网络理论在华为公司产品和技术中的应用情况和需要解决的问题，希望各位专家共同讨论如何促进图网络理论的进一步发展，以及解决图网络理论在实际应用中的各类问题。

### 视频会议链接：

<https://welink-meeting.zoom.us/j/846655536> (ID: 846 655 536)

(注：1) 请把链接复制到浏览器中打开。第一次使用此会议系统可能需要安装相应的软件，具体见浏览器的提示；2) 进入界面时，如果要求输入姓名，请输入真实姓名，以方便记录会议相关的信息；3) 有进入会议相关的问题，请联系：华为技术有限公司，吴老师，13510030101)

**主持：** 闫桂英 中科院数学与系统科学研究院

报告题目	报告人	时间 (PM)
嘉宾讲话：中科院数学与系统科学研究院马志明院士		13:30-13:45
嘉宾讲话：华为全球技术合作副总裁艾超		13:45-14:00
1. Potential research topics on graph theory	李萍	14:00-14:30
2. 图算子的数学抽象	周敏奇	14:30-14:55
3. 语义二部图社群发现	杨宇言	14:55-15:20
4. 网络评估/关键节点评估	杨蕾蕾	15:20-15:45
5. 图神经网络相关问题	李洪敏	15:45-16:15
6.图神经网络在目标检测、神经网络架构搜索的一些应用和讨论	徐航	16:15-16:45
7.图神经网络在硬件设计中网表表征、推荐搜索等大规模实际场景中的应用难点	张莹雪 胡瑶晨	16:45-17:10
8.基于图根因推理算法	王强	17:10-17:35
讨论		
总结	马志明	

# 摘要

## 1、Potential research topics on graph theory

- 1) Research on the theory of graph limit (flag algebra, graphon and so on). For instance, the relation between information theory and graph limit.
- 2). Algorithmic and theoretic problems of partitioning graphs and hypergraphs.

## 2、图算子的数学抽象：

E.F Codd 通过关系代数来抽象关系模型上的操作（关系数据库的计算核心）；图算子数学抽象（需要新突破）：图模型能够更好表征真实世界，但是需要通过新的数学抽象，来抽象图上的操作（图路径匹配，图论算法等）；可能是图代数，或者其他数学抽象方式，图算子与关系操作的操作差别：关系代数无序，但是图路径遍历有序，关系代数是线性计算，但是图是多分支计算。

## 3、语义二部图社群发现：

在企业级应用中，社群识别类算法可以帮助社群营销、网络安全监控、企业推荐等场景中应用。研究发现多数企业场景都属于二部图的结构，一个节点往往伴随多个维度的特征。然而传统的社群图谱只涵盖人与其社会关系也就是同构图，划分方法更多依赖图的原始结构，缺少对语义特征的考虑。因此我们需要研究出一种基于语义二部图的社群划分算法，能够帮助在企业语义多特征二部图场景中的社群聚类划分应用，甚至支持重叠社群识别应用。问题：

- 1) 如何规避社群结果震荡？
- 2) 对于此类无监督任务，除了传统模块密度的指标，如何验证社群分布的合理性跟业务目标的契合度？

## 4、网络评估/关键节点评估：

在实际应用中，常需要在图中找一些关键的节点。例如：华为云的部署结构就是一张网络，网络上有一些节点是关键且脆弱的，一旦节点不可用，会影响到整张网络或者局部网络；公司的产品采购可以形成一张产品-采购件-供应商的网络，某些采购件/供应商是比较关键的节点，一旦供应出现问题，会大范围影响到产品和合同履行，影响到项目进度；需要提前识别出这些节点和风险，并制定相应的策略。问题：1) 网络结构多样化，特征多，节点评估算法也有很多，如何选择合适的算法？2) 网络的节点关系多，变化快，整张图进行计算，对资源消耗大，耗时长，有没有精度损失不大的采样算法？

## 5、图神经网络相关问题：

GNN over-smooth 问题：在图神经网络的训练过程中，随着网络层数的增加和迭代次数的增加，每个节点的隐层表征会趋向于收敛到同一个值。怎样解决这个问题，实现更深层的 GNN，提高 GNN 的特征表达能力？

可导的 Random walk embedding 设计：基于 Random walk 的 embedding 怎样实现和下游任务联合优化，或者设计可导的 random walk embedding？

**可解释性问题：** GNN 不可解释，怎么使得 GNN 提取的特征具有可解释性；把 GNN 用于图上的优化任务时，怎样解释 GNN 提取的特征和优化任务的关系？

## 6、图神经网络在目标检测，神经网络架构搜索的一些应用和讨论：

图神经网络由于其良好的易用性以及可解释性，在深度学习各种任务中都得到了应用以及验证。我们将介绍诺亚实验室的图神经网络在目标检测，神经网络架构搜索的两个工作：Reasoning-RCNN 以及 BONAS。在**目标检测任务**中，图神经网络用于捕捉和建模物体与物体之间的关系，并通过学到的关联关系的增强特征来增强检测效果。在**神经网络架构搜索**的任务中，图神经网络用于刻画不同神经网络的图结构，并将学到的神经网络的图特征用于下一步贝叶斯优化中，取得了良好的效果。还将讨论图神经网络在应用中的一些实际问题及难题。

## 7. 图神经网络在硬件设计中网表表征、推荐搜索等实际场景中的应用难点

图神经网络作为深度学习中较新的探索领域得到了很多的关注，也有不断涌出的新算法和新应用场景，然而其验证数据场景通常规模较小，对拓扑数据的假设较为简单（同构无向图）。在真实应用场景经常会遇到更复杂更大规模的拓扑数据类型，应用经典的图神经网络框架会遇到可拓展性差、由于邻居采样导致的模型收敛较慢、无法处理复杂交互关系和无法实现不同拓扑结构之间表征迁移等问题。我们将分享诺亚实验室在大规模实际场景中图神经网络应用所积累的经验 and 难点问题提炼，并利用图神经网络在硬件设计中网表表征以及推荐搜索两大重点应用场景作为具体实例，重点讨论当前遇到的难点问题和目前自研的解决策略。

## 8. 基于图根因推理算法

IT 的 AIOPS 场景中，涉及问题的自动发现，根因的智能定位：基于对象的静态关系（拓扑、依赖、部署、调用等关系）和动态指标/告警/异常的挂接关系等多维度数据：

1) 基于图时空的异常事件识别：**现状：**当前异常检测，多数通过时序的预测的方法识别异常，涉及灵敏度的问题，准确度不是太高。**期望：**A:基于图结构结合时空信息，能够自动的发现异常，提升异常检测的效果，B:通过图的时空信息，能够对当前检测的异常/告警降噪/汇聚。

2) 基于事件的根因推理：**现状：**在企业应用中，在多种因素相互影响的情况下，准确的识别出根因非常困难。并且在事件的传导链很长、中间状态缺失下准确的进行根因推理也是一个难点。**期望：**A:在信息完备情形下，异常事件/告警，推理异常/告警的根因，自动给出所有异常传播链，并输出不同传播图的置信度。B:模糊推理：在事件图（含长事件传导链、含中间状态缺失）中进行高准确率（90%+）的根因识别。

3) 图神经网络的小样本半监督学习：**现状：**企业应用中，标记样本非常稀缺。目前 AI 深度学习模型虽然精度高，但是需要大量的样本数据进行学习，但标注成本很高，而且如何获取大量标注样本一直是业界难题。**期望：**使用图神经网络进行小样本半监督学习，保证在少量样本基础上，达到和深度学习接近的精度（90%）。