



初心系国脉 奋进i时代

Holding on Original Aspiration, Advancing in Global i-Era

基于局部代理模型的图神经网络可解释性研究

Enhancing Explainability of Graph Neural Networks Using Local Proxy Model

汇报人: 王以可¹ 导师: 刘立 副教授²

¹2022214508@stu.cqupt.edu.cn

²liliu@cqupt.edu.cn

重庆邮电大学 计算机学院

2024 年 12 月 15 日

目 录

1 选题背景与意义

- 图神经网络简介
- 图神经网络解释的意义

2 代理模型的构建

- 简化原图模型
- 对原图进行图采样

3 解释模型的应用

- 解释模型的选取
- 解释子图的生成

4 解释子图的评价

- 解释子图的评价方法
- 评价指标

5 实验

- 数据集介绍
- 实验结果

6 未来计划与安排



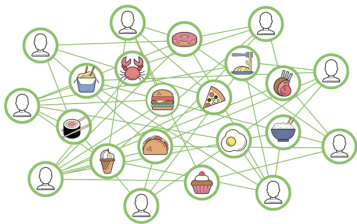
选题背景与意义

- 图神经网络简介
- 图神经网络解释的意义



图神经网络简介

图神经网络（GNNs）是一种专门用于处理图结构数据的深度学习模型。图结构数据是由节点（vertices）和边（edges）组成的，常见于社交网络、分子结构、交通网络等领域。GNNs的核心思想是通过节点的特征信息和图结构来学习节点的代表（representation）。如图 1.1。



(a) 推荐系统的图模型



(b) 社交网络的图模型

图 1.1: 图网络模型结构示例图



选题背景与意义

- 图神经网络简介
- 图神经网络解释的意义



图神经网络解释的意义

图神经网络（Graph Neural Networks, GNNs）的解释性在于它们能够帮助我们理解为什么模型会做出某些决策，以及模型如何从数据中提取和利用信息。

- **理解模型决策：**解释性可以帮助我们理解模型是如何做出决策的，这对于建立对模型的信任，以及在实际应用中确保模型的公平性和透明度至关重要。
- **重要决策问题：**在某些领域，如医疗和金融，模型必须能够解释其决策过程，以满足法规要求。在医疗领域，模型的决策可能会直接影响到患者的生命。不同的预测会导致不同的治疗方案。



图神经网络中的代理模型的意义

- 简化复杂的拓扑结构

大图通常具有复杂的拓扑结构，复杂的拓扑结构带来不可避免的噪声，给图神经网络的训练和解释带来挑战。通过使用小图分解和解释大图，可以将复杂问题拆分为易处理的小问题，从而提高模型的可解释性和效率。

- 提高模型解释的效率

在处理具有数十万个节点的大规模图时，训练可解释模型（例如 **PGE**，基于图的解释模型）是极具挑战性的，其计算资源和时间开销极为庞大。通过引入代理模型，可以将原始复杂图进行简化，从而显著降低模型训练的复杂度和时间成本。这不仅使得训练过程更加高效，也使得解释生成的速度和质量得到提升。



代理模型的构建

- 简化原图模型
- 对原图进行图采样



图的分割方法

简言之，无向图的分割按照切割方式分为边分割 (点分区) 和点分割 (边分区)，即切割图时，来确定是切割的点或边。

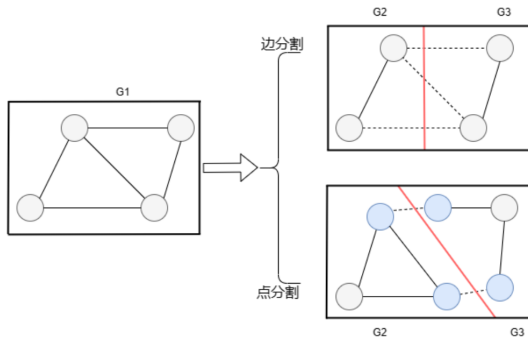


图 2.1: 两种分割方法



基于社区检测算法的构建

我们尝试改进社区发现算法，但发现其时间复杂度较高，对千级节点规模的 Cora 数据集图 2.2 进行切分需要约 45 分钟。此外，社区发现算法更倾向于聚类，而不能很好地保留大图的结构信息。

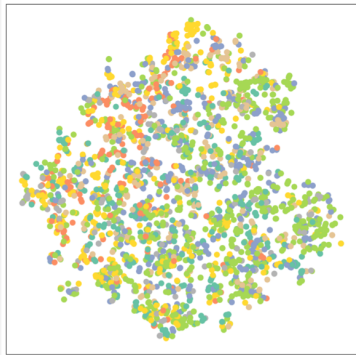


图 2.2: Cora 数据集可视化



基于 METIS 图划分的构建

METIS 是一种层次化的分割算法 (multi-level partitioning), 核心思想: 对于给定原图结构, 持续的降低原图的大小, 然后达到一定程度对于缩减后的图结构进行分割, 最后将分割后的小图还原成原始的图结构保证每份子图的均衡性。

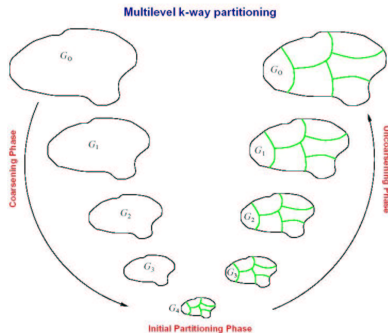


图 2.3: METIS 图划分



METIS 示意图

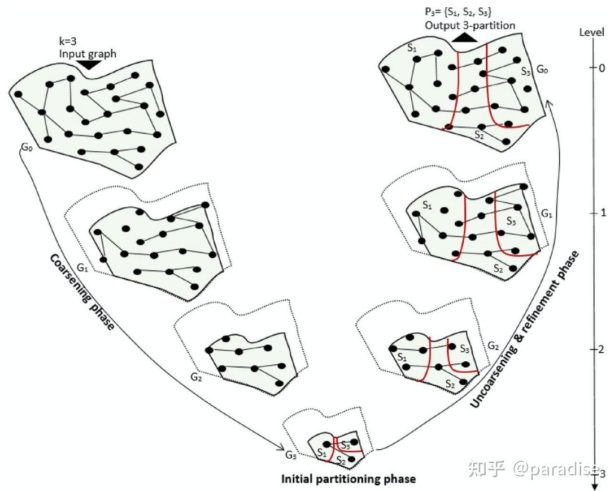


图 2.4: METIS 图划分



METIS 伪代码

Algorithm 1: METIS_GraphPartitioning(Graph G , int num_partitions)

Input: Graph G (包含顶点和边的图), num_partitions (划分的组数)**Output:** Partitioning P (划分结果, 每个顶点属于哪个划分组)

// 图粗化

CoarseGraph $\leftarrow G$;**while** CoarseGraph 的顶点数 $>$ 目标规模 **do**└ CoarseGraph \leftarrow CoarsenGraph(CoarseGraph);

// 图划分

InitialPartition \leftarrow InitialPartitioning(CoarseGraph, num_partitions);CurrentPartition \leftarrow InitialPartition;

// 图细化

while CoarseGraph \neq 原始图 G **do**└ FineGraph \leftarrow UncoarsenGraph(CoarseGraph); // 粗图映射到上一层细图└ CurrentPartition \leftarrow RefinePartition(FineGraph, CurrentPartition); // 细图划分└ CoarseGraph \leftarrow FineGraph; // 迭代跟新**return** CurrentPartition;

代理模型的构建

- 简化原图模型
- 对原图进行图采样



GraphSaint 采样策略

它的核心思想是利用概率采样策略，在训练过程中动态生成一系列具备统计代表性的子图样本，然后对这些子图及其节点进行加权归一化，确保训练过程对全图的估计是无偏且高效的。

GraphSaint 提供多种子图采样策略，常用的包括：

- **节点采样 (Node Sampling)**

- 从全图中按照某种概率分布随机选择一组节点，然后取出这些节点及其相连的边形成子图。
- 实现简单，对节点特征分布的均匀性有较好的适应性。

- **边采样 (Edge Sampling)**

- 随机选择一批边，将这些边及其关联的节点作为一个子图。
- 对图中边密度的处理更灵活，有时能更好地捕捉图的局部结构特征。

- **随机游走采样 (Random Walk Sampling)**

- 从一个随机选定的节点出发，在图上进行短程随机游走，将访问到的节点和边作为子图。
- 能更好地捕捉图的拓扑结构，适合类社交网络、Web 图等高聚集结构。



GraphSaint+METIS

METIS 专注于图的分割，但它的分割策略更多是基于局部拓扑信息（如节点和边的分布）进行边分割（Edge Cut），从而减少子图之间的通信开销。

然而，METIS 划分的结果往往会忽略一些全局图属性，如图的整体结构信息或节点的特征。

GraphSaint 则通过随机采样子图，可以反复从全图中抽取样本，结合全局的统计信息（如节点重要性、特征聚合）来保留全局图的表示能力。



解释模型的应用

- 解释模型的选取
- 解释子图的生成



GNNExplainer

GNNExplainer 对于解释子图的生成过程是通过优化一个目标函数来实现的，该目标函数旨在最大化 GNN 的预测与子图结构分布之间的互信息。如图 3.1。

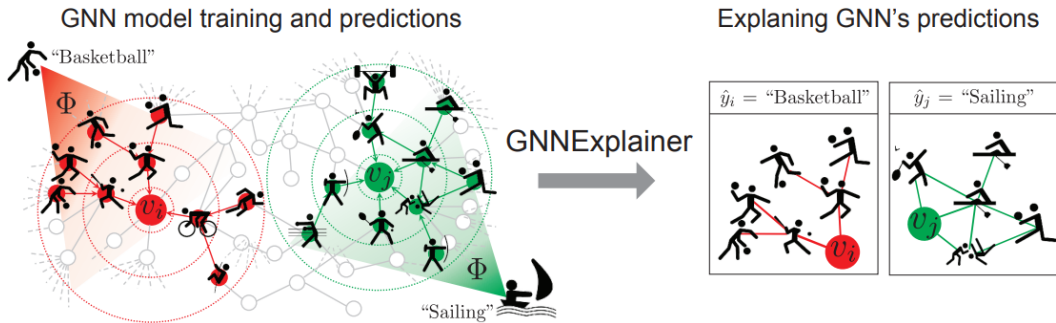


图 3.1: Gnn 生成解释子图示意图



GNNExplainer

GNNExplainer 的目标是通过最大化互信息来优化子图和特征掩码，目标函数可以表示为：

3.1 (GNNExplainer 优化函数)

$$\mathcal{L}(M, F) = -\mathbb{E}_{G_S \sim p(G_S|G, M), X_S^F \sim p(X_S^F|X, F)} \left[\log P_{\Phi}(\hat{Y}|G_S, X_S^F) \right] + \Omega(M, F) \quad (3.1)$$

$$\mathcal{L}_G(M) = -\mathbb{E}_{G_S \sim p(G_S|G, M)} \left[\log P_{\Phi}(\hat{Y}|G_S, X) \right] + \Omega_G(M) \quad (3.2)$$

$$\mathcal{L}_X(F) = -\mathbb{E}_{X_S^F \sim p(X_S^F|X, F)} \left[\log P_{\Phi}(\hat{Y}|G, X_S^F) \right] + \Omega_X(F) \quad (3.3)$$



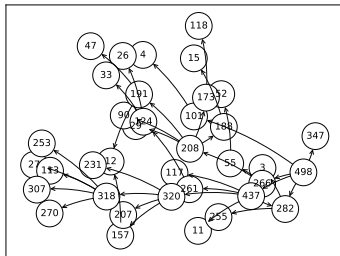
解释模型的应用

- 解释模型的选取
- 解释子图的生成

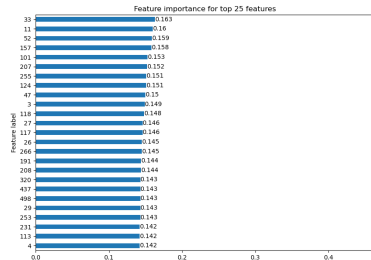


大图生成的解释子图

我们采用两种经典的解释模型，分别在子图和完整图上生成相应的解释子图，并计算解释子图中各节点对目标预测的正确性的影响程度。如图 3.2 为 GNNE 在大图上生成的解释子图对目标节点预测影响度。



(a) 大图上的解释子图



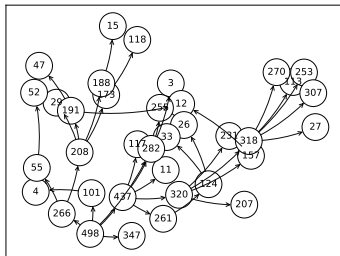
(b) 大图的解释子图对于节点预测影响度排序

图 3.2: 大图的解释子图与其影响程度

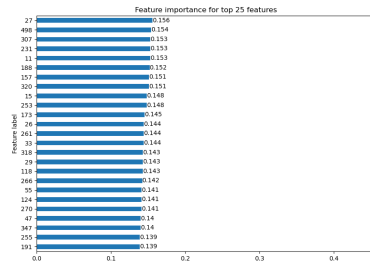


子图生成的解释子图

子图上生成的解释子图对目标节点预测影响度如图 3.3。



(a) 子图上的解释子图



(b) 子图的解释子图对于节点预测影响度排序

图 3.3: 子图的解释子图与其影响程度



解释子图的评价

- 解释子图的评价方法
- 评价指标



评价方法

我们通过从大图中移除由大图生成的解释子图和由子图生成的解释子图，来评估这两种解释子图对大图中目标节点预测概率的影响，从而判断解释子图的质量。如图 4.1。

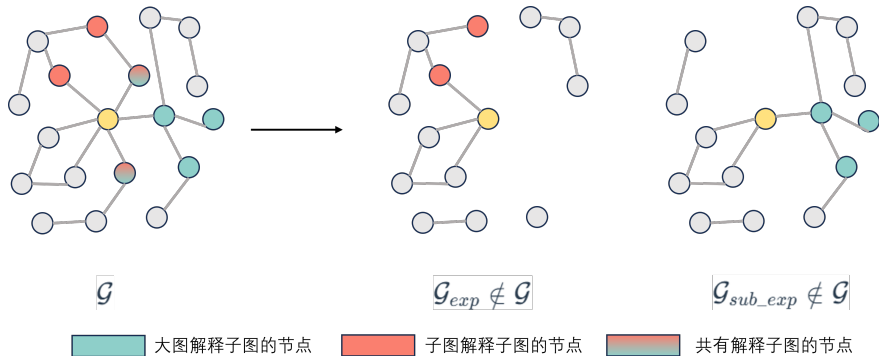


图 4.1: 评价方法示意图



解释子图的评价

- 解释子图的评价方法
- 评价指标



Fidelity 评价指标

Fidelity（忠实度）是衡量解释模型输出与原模型输出一致性的重要指标。它通常用于评估解释模型是否能够准确反映原模型的决策过程。

Fidelity 数学公式

$$\begin{aligned}\text{fid}_+ &= 1 - \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{y}_i^{G_{CS}} = \hat{y}_i) \\ \text{fid}_- &= 1 - \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{y}_i^{G_S} = \hat{y}_i)\end{aligned}\tag{4.1}$$



实验

- 数据集介绍
- 实验结果



数据集选取

表 5.1: 数据集节点数、边数及简介

数据集	节点数	边数	用途	特点
Citeseer	3,327	4,732	节点分类、链接预测	包含论文及其引用关系
Cora	2,708	5,429	节点分类	包含机器学习论文主题分类
PubMed	19,717	44,338	节点分类	生物医学论文与元信息
Reddit	232,965	11,606,919	图分类、社区检测	动态图，建模用户活动
ogb-products	2,449,029	61,859,140	大规模节点分类	OGB 提供的大规模图数据



实验

- 数据集介绍
- 实验结果



GNNE 实验结果

表 5.2: 实验结果对比

数据集	显卡	原始模型				代理模型 (Ours)			
		fid+	fid-	acc	time	fid+	fid-	acc	time
Citeseer	4070ti	0.856	0	0.677	1:12:23	0.723	0	0.691	40:07
Cora	4090	0.817	0	0.819	58:16	0.923	0	0.775	34:20
PubMed	4090	0.798	0	0.793	5:17:23	0.851	0	0.822	1:34:23
Reddit	4090	-	-	0.806	-	0.862	0.016	0.834	3:46:52
ogb-products	4090	-	-	0.736	-	0.846	0.012	0.798	21:23:10



未来计划与安排

- 在更多的图神经网络解释器上进行实验。
- 更多的消融实验。





¹本演示文稿使用 \LaTeX 排版工具制作，若想进一步了解这一强大的开源工具，请移步 一份（不太）简短的 $\text{\LaTeX} 2_{\epsilon}$ 介绍。

²在本演示文稿的制作过程中，作者从 重庆邮电大学 刘立 老师处获得了诸多指导与建议，在此表示十分感谢。

³感谢 重庆邮电大学 李振坤 老师提供的 \LaTeX Beamer 模版。