

一种针对关系不确定性的贝叶斯异质图神经网络

陈冠恒 郭子瑜 梅广旭 刘士军 潘 丽

(山东大学软件学院 济南 250101)

摘 要 由不同类型的节点和边组成的异质图在现实世界中具有广泛的应用场景. 近年来, 针对此类图数据的异质图神经网络研究与应用取得了很好的进展, 但是现有的异质图神经网络在构建时没有考虑到异质图中存在关系的不确定性问题, 因而在面对包含不确定性的关系的对抗性实例干扰时也暴露出鲁棒性较弱的缺点. 针对上述问题, 本文提出贝叶斯异质图神经网络模型 (Bayesian Heterogeneous Neural Network, BHNN), 用于解决异质图中关系不确定性问题并提高模型的鲁棒性. BHNN 首先基于异质图的领域知识预定义不同的元路径, 然后针对表示单一关系的每条元路径构建由相同类型节点构成的元路径邻居图, 每个元路径邻居图被看作是一个随机图参数族的实现并可以使用随机块模型对其建模, 最后结合贝叶斯方法对随机图的参数和节点标签的联合后验进行推理得到节点的预测标签的概率分布. 节点分类实验结果表明: 在 ACM/DBLP 和 IMDB 三个基准数据集上, 相比较于目前最好的图数据训练模型, BHNN 的微 $F1$ 与宏 $F1$ 分别平均提高了 1.59% 与 1.36%, 这验证了本文算法的有效性和优越性. 在节点攻击实验中, 相比较于基准方法中的图神经网络, BHNN 仍保持最好的节点分类效果, 同时节点分类能力下降的幅度小于其他异质图神经网络, 该实验结果验证了本文提出的模型具有更强的鲁棒性.

关键词 异质图; 元路径; 图神经网络; 不确定性; 鲁棒性; 贝叶斯方法

中图法分类号 TP18

DOI 号 10.11897/SP.J.1016.2023.00552

A Bayesian Heterogeneous Graph Neural Network for Relational Uncertainty

CHEN Guan-Heng GUO Zi-Yu MEI Guang-Xu LIU Shi-Jun PAN Li

(School of Software, Shandong University, Jinan 250101)

Abstract Heterogeneous graphs composed of different types of nodes and edges have a wide range of application scenarios in the real world, such as social networks, transportation networks and e-commerce networks, which has attracted the attention to researchers. Although existing heterogeneous graph neural networks achieved excellent performance in the tasks of node classification, node clustering and link prediction, they do not take the uncertainty of relationships in a heterogeneous graph into account. Moreover, the performance of these models is significantly reduced when perturbed by adversarial examples, which exposes the weakness of the model in robustness. The weak robustness limits the application of heterogeneous graph neural networks of the real world, because heterogeneous network is easily changed by the disturbance of uncertain factors. Therefore, it is of great practical significance to improve the robustness of the model. In this paper, we propose a Bayesian Heterogeneous Neural Network (BHNN) to reveal and solve the problem of the uncertainty of relationships in a heterogeneous graph and to improve the robustness of the model. Firstly, BHNN predefines different meta-paths from the domain knowl-

收稿日期: 2021-07-07; 在线发布日期: 2022-12-13. 本课题得到国家自然科学基金(61872222)资助. 陈冠恒, 硕士研究生, 中国计算机学会 (CCF) 会员, 主要研究领域为图神经网络、异质网络计算. E-mail: chenguanheng@mail.sdu.edu.cn. 郭子瑜, 博士研究生, 中国计算机学会 (CCF) 会员, 主要研究领域为图神经网络、时空网络计算. 梅广旭, 博士研究生, 中国计算机学会 (CCF) 会员, 主要研究领域为图神经网络、异质网络计算. 刘士军 (通信作者), 博士, 教授, 中国计算机学会 (CCF) 高级会员, 主要研究领域为服务计算、协同计算、企业关系网络. E-mail: lsj@sdu.edu.cn. 潘 丽 (通信作者), 博士, 副教授, 中国计算机学会 (CCF) 专业会员, 主要研究领域为云计算、云制造、市场导向资源分配. E-mail: panli@sdu.edu.cn.

edge of a heterogeneous graph. Then, it constructs meta-path neighbor graphs through meta-paths collections. Secondly, BHNN uses a random block model to model each meta-path neighbor graph which can be regarded as a realization from a random graph parametric family. Finally, it uses the Bayesian method to infer the joint posterior of the parameters and node labels of the random graph to predict the unknown labels. In this process, the meta-path neighbor graphs are reconstructed to reduce the weak edges and spurious edges caused by the uncertain relationship, so as to get generated graphs that can better reflect the real connection relationship. These generated graphs with different structures from the meta-path neighbor graph can be regarded as adversarial graph data, whose structures are more relevant and contain more abundant information. Furthermore, BHNN samples the weight samples of the generated graphs to further enhance its robustness. Extensive experiments are conducted on three benchmark data sets including ACM, DBLP and IMDB. Compared with the state-of-the-art models, the Micro-F1 and Macro-F1 of BHNN is increased by an average of 1.59% and 1.36% respectively, which illustrate the effectiveness and advancement of BHNN. Meanwhile, in the node attack experiments, BHNN maintains the best performance with less degradation in comparison with other heterogeneous graph neural networks, which also verifies that BHNN achieves a better performance of robustness.

Keywords heterogeneous graph; meta-path; graph neural network; uncertainty; robustness; Bayesian method

1 引言

现实世界中存在的许多网络结构,比如引文网络^[1]、社交网络^[2]、交通网络^[3]等,吸引了研究人员的研究兴趣.由多种类型的节点和边构成的异质信息网络属于其中的一种,此类网络含有丰富的结构和语义信息且在现实世界中广泛存在,备受研究关注.近年来,针对异质信息网络,越来越多的异质图模型被构建出来,应用于异质图中的节点分类^[4]、节点聚类^[5]、链接预测^[6]等任务.

这些模型虽然都表现出良好的性能,但是它们没有考虑到在异质图中存在关系的不确定性问题.结合图 1 来看,现实应用中导致这种不确定性的主要原因有以下几种:

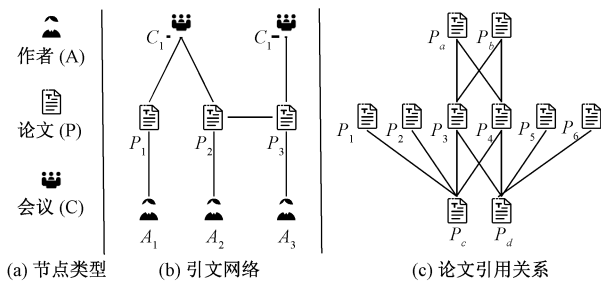


图 1 异质图中的不确定性关系

(1) 异质图关系复杂,构建网络时导入信息缺少

交叉验证,会因为语义歧义或不完备造成关系连接错误.例如因为作者名字相同或名字缩写相同,导致作者节点与论文节点之间存在错误关联,又例如图 1(b)中存在强关联性的两个论文节点 P_1 、 P_2 由于发表时间相同而没有相互引用,网络结构中节点之间缺少连接的边,导致重要关系的缺失.

(2) 数据缺失导致重要关系的丢失或脏数据导致错误的关联关系.例如因为数据的缺失导致两个强关联关系的节点没有边连接,又例如图 1(b)中,由于脏数据的存在,发表于不同会议且无关联性的论文节点 P_2 、 P_3 之间存在错误的引用关系,导致网络结构中毫无联系的节点之间存在连接的边.

(3) 异质图节点类型多样,关系之间的权重不一,过多的次要关系会影响重要关系.例如图 1(c)中同时引用论文 P_3 、 P_4 的论文节点 P_a 与 P_b 拥有较强的关联性,而节点 P_c 与 P_d 之间引用了更多的其他论文导致其包含了过多的次要关系因而对重要关系造成影响.

其他从真实世界获取数据构建的异质信息网络所包含的关系中同样存在类似的不确定性.

目前的异质图神经网络研究除了尚未系统地解决异质图存在关系的不确定性问题,还存在鲁棒性较弱的问题.鲁棒性较弱的模型无法保证稳定的训练效果,因为神经网络容易受到对抗性实例的影响,

数据进行扰动后,模型训练效果容易下降.造成这种现象的原因是在数据集中对抗性实例样本数量太少,而神经网络的高度非线性特性导致模型学习不到这些特殊样本,因而在面对对抗性实例干扰时,模型表现出鲁棒性较弱的现象^[7].这一缺陷严重限制了异质图神经网络在真实世界中的应用,因为真实世界中异质信息网络更容易受到各种不确定性因素干扰而发生改变.

针对上述问题,我们提出一种新的用于节点分类的异质图神经网络,称为贝叶斯异质图神经网络(Bayesian Heterogeneous Neural Network, BHNN).BHNN通过预定义多条元路径以得到多个元路径邻居图,将每个元路径邻居图视为不同的随机图参数族的实现,利用分类混合隶属度随机块模型(as-sortative mixed-membership stochastic block model, a-MMSBM)^[8]对随机图建模,再结合贝叶斯方法构建节点标签的后验推理公式,通过对公式进行推导得到减少关系不确定性的模型.本质上,BHNN通过对异质图的元路径邻居图进行重构以及泛化得到新的图结构,在此过程中,原来异质图中具有强关系的节点之间的边会不断构建,而弱关系或脏数据带来的假边则不断减少,最终得到强关联性的网络结构,以此解决异质图中的关系不确定性问题,基于此,BHNN在面对包含不确定性关系的对抗性实例干扰时,能降低对模型效果的影响,而经过泛化得到的图结构样本因为与原图结构相比拥有不同的网络结构,所以可被视为对抗性实例,因而泛化操作能够增加训练数据中的对抗性实例的数量,进一步增强模型的鲁棒性.

本文主要的贡献如下:(1)提出了一种新的可解决异质图中关系的不确定性问题的异质图神经网络,该异质图神经网络通过对图结构进行重构以减少不确定性关系;(2)提出一种新的增强异质图神经网络鲁棒性的方法,该方法通过泛化重构后的图以增加训练数据中的对抗性实例数量从而达到提高模型鲁棒性的目的;

本文结构如下:第2节介绍相关工作;第3节对本文研究问题涉及的相关概念进行定义;第4节主要介绍我们提出的基于元路径的异质图节点分类方法,并展示构建的细节和公式的推导简化;第5节是实验的结果和分析;最后第6节是对本文内容的总结.

2 相关工作

节点和边类型的多样化使异质信息网络含有丰

富的结构和语义信息,采用何种方式来表达这些信息成为训练异质图数据要解决的首要问题.对此,Yizhou等人^[9]提出元路径的概念来处理异质图的语义信息,此后基于元路径的异质图模型被提出用于训练异质图数据,比如异质图注意力神经网络^[5](Heterogeneous Graph Attention Network, HAN)通过预定义元路径并结合注意力机制构建模型,而在异质图嵌入方法中,Wayne等人^[10]以及Yuxiao等人^[11]均基于在元路径进行随机游走的方式分别提出HERec算法和Metapath2vec算法,以此学习异质图中的节点嵌入表示.这些模型虽然能有效处理异质图的结构和语义信息并在训练任务中取得良好的表现,但是包括上述模型在内的现在已有的异质图模型在构建时均没有考虑到异质图的关系不确定性问题.

网络结构的不确定性在同质图中同样存在.Yingxue等人^[12]为了解决因噪声数据导致同质图包含不确定性关系,提出了贝叶斯卷积神经网络(Bayesian graph convolutional neural networks, BGCN),将网络结构视为随机变量并利用a-MMSBM重构同质图,以此解决不确定性问题,同时引入了由Ghahramani等人^[13]提出MC dropout(Monte Carlo dropout)方法,将深度学习中的dropout训练作为深高斯过程中的近似贝叶斯推理,以降低计算难度.Arman等人^[14]为图神经网络的自适应链接采样提出了能概括现有的用于训练图神经网络的随机正则化方法的一个统一框架,以此减轻了深度图神经网络的过度平滑和过度拟合的趋势同时还可以在图分析任务中进行不确定性学习.Zonghan等人^[15]为了捕获远程时间序列信息同时解决时空网络结构的不确定性,提出了一种新的图神经网络架构Graph WaveNet,通过开发一种新的自适应依赖性矩阵并通过节点嵌入来学习该矩阵,以精确地捕获数据中隐藏的空间依赖性,将两个节点的嵌入表示相乘得到它们之间的空间依赖权重再使用激活函数消除弱连接.

作为神经网络最重要的特性之一,鲁棒性同样也是目前研究的一个热点.Shifu等人^[16]为了有效对抗对异质图的对抗性攻击,提出了一种弹性的防御模型,以增强基于异质图的分类器在安卓恶意软件检测中的鲁棒性.Dingyuan等人^[17]将图卷积层中节点表示由普通向量替换为高斯分布,并提出基于方差的注意机制来降低对抗性效应在卷积过程中的传播,以此增强模型应对干扰的能力.Kaidi等

人^[18]提出一种新的基于梯度的攻击方法,对图进行少量的拓扑修改即可有效地降低模型的性能,利用此方法提出基于优化的对抗训练技术,以此提高模型的对抗性鲁棒性.不同于上述方式,本文利用随机图生成模型对图进行重构,以减少网络结构的不确定性,并基于此,通过泛化图结构以增加训练数据中对抗性实例数量实现模型鲁棒性的提高.

3 相关定义

本节首先给出与本文方法有关的几个定义.

定义 1. 异质图^[19] \mathcal{G}_{obs} . 一个异质图定义为 $\mathcal{G}_{\text{obs}} = (V, E)$, 其中 V 表示节点的集合, E 表示边的集合. 节点类型映射函数 $F_v: V \rightarrow \mathcal{A}$, \mathcal{A} 表示节点的类型. 边类型映射函数 $F_e: E \rightarrow \mathcal{R}$, \mathcal{R} 表示边的类型. 当 $|\mathcal{A}| = 1$ 且 $|\mathcal{R}| = 1$ 时, 表示该图是一个同质图, 而在异质图中, $|\mathcal{A}| + |\mathcal{R}| > 2$.

图 2 展示的论文数据集 DBLP 即为一个异质图, 它包含四种节点类型: 作者(A)、论文(P)、会议(C)、以及术语(T), 同时含有三种边类型: 论文-作者(P-A)、论文-术语(P-T)、论文-会议(P-C).

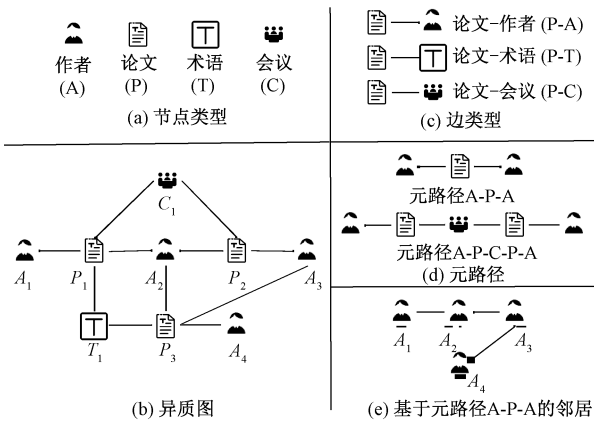


图 2 异质图数据集 (DBLP). (a) 4 种类型节点: 作者、论文、术语、会议. (b) 一个由 4 种类型节点和 3 种边类型组成的一个异质图 DBLP. (c) 3 种边类型: 论文-作者、论文-术语、论文-会议. (d) DBLP 中的两条元路径: A-P-A、A-P-C-P-A. (e) 在异质图 DBLP 中基于元路径 A-P-A 连接的邻居.

定义 2. 元路径^[20] ϕ . 异质图中的一条元路径 ϕ 可描述为 $n_1 \xrightarrow{r_1} n_2 \xrightarrow{r_2} \dots \xrightarrow{r_l} n_{l+1}$, 其中 $n_{i+1} \in \mathcal{A}, r_i \in \mathcal{R}$.

正如图 2(d) 展示, 论文作者之间存在多条元路径, 比如作者-论文-作者 (APA)、作者-论文-会议-论文-作者 (APCPA), 不同的元路径包含不同的语义信息, APA 代表论文的共作关系, APCPA

表示通过同一主题的会议关联不同的作者, 本文通过预定义这些元路径来处理图中不同的结构信息和丰富的语义.

定义 3. 元路径邻居图^[5] \mathcal{G}_ϕ . 每条元路径两端的节点可以理解为通过某种语义连接的邻居, 若它们同属一类节点, 则这些节点构成一个同质图, 本文称之为基于元路径的邻居图 (简称为元路径邻居图), 定义为 $\mathcal{G}_\phi = (V', E')$, 其中 V' 表示由元路径 ϕ 连接的节点集合, E' 表示由元路径 ϕ 连接的边的集合, 且对于任意节点 $v_a, v_b \in V'$, 存在 $F_v(v_a) = F_v(v_b)$.

表 1 符号与解释

符号	含义
\mathcal{G}_{obs}	异质图
V	异质图的节点集合
E	异质图的边集合
F_v	节点类型映射函数
F_e	边类型映射函数
\mathcal{A}	节点的类型
\mathcal{R}	边的类型
ϕ	元路径
\mathcal{G}_ϕ	元路径邻居图
V'	元路径邻居图的节点集合
E'	元路径邻居图的边集合
π_a	节点 a 归属类别概率分布
π_{ak}	节点 a 属于类别 k 的概率
π	所有节点类别概率分布
β_k	类别 k 的类别强度
β	所有类别的类别强度集合
$D_{a \rightarrow b}$	节点对 (a, b) 连接指示变量
L_{ab}	节点对 (a, b) 的连接概率
\mathcal{G}	泛化图样本
φ_a	π_a 的替代参数
θ_{k0}, θ_{k1}	β_k 的替代参数
δ	跨类别链接概率
α, ρ, η	$\phi_a, \theta_{k0}, \theta_{k1}$ 先验的超参数
ϵ_0, τ, κ	ϕ_a, θ_k 迭代更新的超参数
$\mathcal{G}_{n,i,\phi}$	基于 $\mathcal{G}_\phi, \lambda_i$ 得到的第 n 个泛化图样本
$\mathcal{G}_{n,\phi}$	基于 \mathcal{G}_ϕ 得到的第 n 个泛化图样本
X	节点的特征向量
$Y_{\mathcal{L}}$	节点的标签向量
Z	神经网络的输出向量
W	神经网络的权重向量
A	泛化图样本 \mathcal{G} 的连接矩阵
I_N	单位矩阵
D_{ii}	度矩阵
$H^{(l)}$	第 l 层神经网络对应的特征矩阵
$\lambda = \{\pi, \beta\}$	a-MMSBM 建模随机图的随机参数族
U	采样元路径邻居图的样本数量
I	采样生成泛化图的参数样本数量
$N_{\mathcal{G}}$	采样泛化图的样本数量
S	采样权重的样本数量

例如基于异质图 2(b) 中代表共作关系的元路径 APA 可以构建作者邻居图 2(e), 其中节点对之间是否存在边取决于该节点对在异质图 2(b) 中是否位于元路径 APA 的两端。

由于本文使用的符号较多, 所以为更加清晰地展现本文提出的方法, 我们在表 1 中总结了整篇文章所使用的符号。

4 模 型

本文提出的 BHNN 通过重构网络结构来解决异质图中存在的关系不确定性问题, 并结合贝叶斯方法, 将网络结构以及神经网络的权重视为随机变

量, 以节点的标签为推理目标, 利用边缘化方法构建推导公式。

4.1 BHNN 框架

BHNN 框架如图 3 所示, 具体而言, (1) 通过预定义多条元路径的方式构建多个元路径邻居图, 然后把每个元路径邻居图看作是不同随机图参数族实现的随机图; (2) 针对每个随机图, 利用 a-MMSBM 实现随机图的重构并泛化, 从而发现更好的图结构; (3) 通过结合 MC-dropout 方法的图卷积神经网络^[21] (Graph Convolutional Network, GCN) 训练泛化图得到权重参数样本; (4) 利用这些权重参数把不同元路径邻居图的泛化图结合起来, 得到最终的节点标签的概率分布。

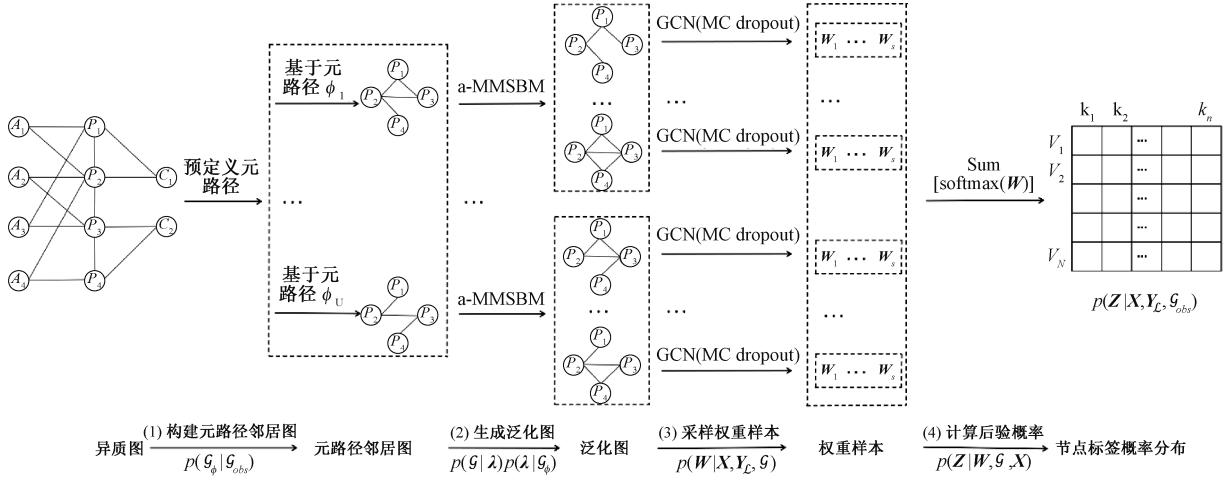


图 3 BHNN 框架图

在 BHNN 中, 元路径邻居图、随机图参数、泛化图的权重参数以及节点标签都被视为随机变量, 而我们最终的目标是获得节点的标签, 所以利用边缘化方法可以得到节点标签的后验概率计算公式:

$$p(\mathbf{Z}, \mathbf{X}, \mathbf{Y}_L, \mathcal{G}_{obs}) = \int \int \int \int p(\mathbf{Z}, \mathbf{W}, \mathcal{G}, \mathbf{X}) p(\mathbf{W}, \mathbf{X}, \mathbf{Y}_L, \mathcal{G}) p(\mathcal{G} | \lambda) p(\lambda | \mathcal{G}_\phi) p(\mathcal{G}_\phi | \mathcal{G}_{obs}) d\mathbf{W} d\mathcal{G} d\lambda d\mathcal{G}_\phi \quad (1)$$

公式(1)的释义为: 在给定异质图 \mathcal{G}_{obs} 、部分节点标签 \mathbf{Y}_L 以及节点特征 \mathbf{X} 的条件下, 预测无标签节点的类别分布 $p(\mathbf{Z}, \mathbf{X}, \mathbf{Y}_L, \mathcal{G}_{obs})$. BHNN 首先获取在已知异质图 \mathcal{G}_{obs} 条件下得到元路径邻居图 \mathcal{G}_ϕ 的概率 $p(\mathcal{G}_\phi | \mathcal{G}_{obs})$, 并获取基于此条件得到随机图参数族 $\lambda = \{\boldsymbol{\pi}, \boldsymbol{\beta}\}$ 的概率 $p(\lambda | \mathcal{G}_\phi)$, 进而获取基于此参数构建泛化图 \mathcal{G} 的概率 $p(\mathcal{G} | \lambda)$, 并获取基于单个泛化图 \mathcal{G} 采样其神经网络权重参数 \mathbf{W} 的概率 $p(\mathbf{W}, \mathbf{X}, \mathbf{Y}_L, \mathcal{G})$, 最终基于节点特征 \mathbf{X} 、泛化图 \mathcal{G} 以及权重样本 \mathbf{W} 得到的节点标签分布 \mathbf{Z} 的概率

$p(\mathbf{Z}, \mathbf{W}, \mathcal{G}, \mathbf{X})$.

4.2 元路径邻居图的构建与建模

BHNN 中元路径邻居图 \mathcal{G}_ϕ 均通过预定义元路径的方式构建, 即基于异质图数据集的领域知识, 通过预定义的方式选择多条符合现实语义的元路径, 其中每条元路径可构建一个元路径邻居图, 利用这些元路径可以构建一定数量的元路径邻居图, 并以此作为元路径邻居图样本集。

此时异质图所包含的不确定性关系被转化为元路径邻居图中的不确定性关系, 为此我们需要为每个元路径邻居图建模以减少其中的不确定性关系, 本文将每个元路径邻居图 \mathcal{G}_ϕ 视为一个随机图参数族的实现, 利用分类混合隶属度随机块模型 (a-MMSBM) 对其建模以获取能反映 \mathcal{G}_ϕ 中节点连接关系的随机图参数族, 基于此实现图的重构以得到减少不确定性关系的泛化图。

具体而言, 对于拥有 N 个节点、 K 种节点类别

的 \mathcal{G}_ϕ , 在利用 a-MMSBM 对其建模以得到泛化图 \mathcal{G} 过程中, 对于任意节点 a , 其归属的类别通过一个维度为 K 的概率分布 $\boldsymbol{\pi}_a$ 表示, 即 $\boldsymbol{\pi}_a = [\pi_{a1}, \dots, \pi_{aK}]^T$, 其中 π_{ak} 表示节点 a 属于类别 k 的概率. 同时每种类别都有其类别强度 $\beta_k \in (0, 1)$, 用于评估该类别的成员之间联系的紧密程度. 对于任意节点对 (a, b) , 存在指示变量 $D_{a \rightarrow b} = k_1$, 表示当节点 a 指向节点 b 时节点 a 所归属的类别为 k_1 , 节点对的连接概率 $L_{ab} \in \{0, 1\}$, 表示节点对之间是否存在连接的边.

总的来说, 泛化图 \mathcal{G} 的节点之间连接与否取决于它们的类别成员的相似性和它们共享类别的强度. 泛化图 \mathcal{G} 生成过程通过下面的描述定义:

1. 对于每个类别 k , 采样其类别强度 $\beta_k \sim \text{Beta}(\eta)$.
2. 对于任意节点 a , 采样其归属的类别分布 $\boldsymbol{\pi}_a \sim \text{Dirichlet}(\alpha)$.
3. 对于任意节点对 (a, b) , 分别采样其指示变量 $D_{a \rightarrow b} \sim \boldsymbol{\pi}_a$ 和 $D_{b \rightarrow a} \sim \boldsymbol{\pi}_b$. 当 $D_{a \rightarrow b} = D_{b \rightarrow a} = k$ 时, 采样它们之间的边 $L_{ab} \sim \text{Bernoulli}(\beta_k)$; 当 $D_{a \rightarrow b} \neq D_{b \rightarrow a}$ 时 $L_{ab} \sim \text{Bernoulli}(\delta)$, 其中 $\delta \in (0, 1)$ 是跨类别链接概率, 通常设为较小值.

其中 $\text{Beta}(\eta)$ 分布表示类别强度 β_k 的先验分布, $\text{Dirichlet}(\alpha)$ 分布表示节点类别的先验分布, η, α 以及 δ 均为超参数.

上述的生成模型过程可以通过下面的联合后验公式描述:

$$p(\mathbf{L}, \mathbf{D}, \boldsymbol{\pi}, \boldsymbol{\beta}, \alpha, \eta, \delta) = \prod_{a=1}^N \prod_{b>a}^N p(L_{ab} D_{a \rightarrow b}, D_{b \rightarrow a}, \boldsymbol{\beta}, \delta) p(D_{a \rightarrow b} \boldsymbol{\pi}_a) p(D_{b \rightarrow a} \boldsymbol{\pi}_b) \prod_{a=1}^N p(\boldsymbol{\pi}_a, \alpha) \prod_{k=1}^K p(\beta_k, \eta) \quad (2)$$

其中 \mathbf{L} 表示所有节点对之间边的集合, \mathbf{D} 表示所有节点对之间指示向量的集合, $\boldsymbol{\pi}$ 表示所有节点归属类别的概率分布的集合, $\boldsymbol{\beta}$ 表示所有类别的类别强度的集合.

4.3 泛化图的生成

联合后验公式(2)实现 a-MMSBM 对元路径邻居图 \mathcal{G}_ϕ 的建模, 基于此公式可实现泛化图的生成, 其中 $p(D_{a \rightarrow b} | \boldsymbol{\pi}_a)$ 与 $p(D_{b \rightarrow a} | \boldsymbol{\pi}_b)$ 分别由 $\text{argmax}(\boldsymbol{\pi}_a)$ 与 $\text{argmax}(\boldsymbol{\pi}_b)$ 定义.

当 $D_{a \rightarrow b} = D_{b \rightarrow a} = k$, 即 $\text{argmax}(\boldsymbol{\pi}_a) = \text{argmax}(\boldsymbol{\pi}_b) = k$ 时, $L_{ab} \sim \text{Bernoulli}(\beta_k)$:

$$p(L_{ab} D_{a \rightarrow b}, D_{b \rightarrow a}, \beta_k, \delta) = (1 - \beta_k)^{1-L_{ab}} \cdot \beta_k^{L_{ab}} \quad (3)$$

当 $D_{a \rightarrow b} \neq D_{b \rightarrow a}$, 即 $\text{argmax}(\boldsymbol{\pi}_a) \neq \text{argmax}(\boldsymbol{\pi}_b)$ 时, $L_{ab} \sim \text{Bernoulli}(\delta)$:

$$p(L_{ab} | D_{a \rightarrow b}, D_{b \rightarrow a}, \boldsymbol{\beta}, \delta) = (1 - \delta)^{1-L_{ab}} \cdot \delta^{L_{ab}} \quad (4)$$

为满足约束条件 $\beta_k, \pi_{ak} \in (0, 1)$ 与 $\sum_{k=1}^K \pi_{ak} = 1$, 本文采用扩大平均参数化方法^[22]实现公式(2)中 $p(\boldsymbol{\pi}_a | \alpha)$ 与 $p(\beta_k | \eta)$ 的计算, 在此过程中引入超参数 ρ .

对于 $p(\boldsymbol{\pi}_a | \alpha)$, 我们引入替代参数 $\varphi_a \in R_+^K$, 采用独立 $\text{Gamma}(\alpha, \rho)$ 分布的乘积作为其先验, 即

$$\begin{aligned} \varphi_a &= \prod_{k=1}^K \text{Gamma}(\varphi_{ak}, \alpha, \rho) \\ &= \prod_{k=1}^K \frac{\rho^\alpha \varphi_{ak}^{\alpha-1} e^{-\rho \varphi_{ak}}}{\Gamma(\alpha)} = \prod_{k=1}^K \frac{\rho^\alpha \varphi_{ak}^{\alpha-1} e^{-\rho \varphi_{ak}}}{\int_0^{+\infty} t^{\alpha-1} e^{-t} dt} \end{aligned} \quad (5)$$

其中伽马函数 $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$, 本文将 $\boldsymbol{\pi}_a$

定义 $\pi_{ak} = \frac{\varphi_{ak}}{\sum_{l=1}^K \varphi_{al}}$, $\varphi_{ak} \geq 0$, 以此结果作为 $\boldsymbol{\pi}_a$ 的

$\text{Dirichlet}(\alpha)$ 先验.

对于 $p(\beta_k | \eta)$, 我们引入替代参数 $\theta_{k0}, \theta_{k1} \geq 0$, 采用独立 $\text{Gamma}(\eta, \rho)$ 分布的乘积作为其先验, 即

$$\begin{aligned} \theta_k &= \prod_{i=0,1} \text{Gamma}(\theta_{ki}, \eta, \rho) \\ &= \prod_{i=0,1} \frac{\rho^\eta \theta_{ki}^{\eta-1} e^{-\rho \theta_{ki}}}{\Gamma(\eta)} = \prod_{i=0,1} \frac{\rho^\eta \theta_{ki}^{\eta-1} e^{-\rho \theta_{ki}}}{\int_0^{+\infty} t^{\eta-1} e^{-t} dt} \end{aligned} \quad (6)$$

将 β_k 定义为 $\beta_k = \frac{\theta_{k1}}{\theta_{k0} + \theta_{k1}}$, 以此结果作为 β_k 的 $\text{Beta}(\eta)$ 先验.

参考文献[12]利用 φ 与 θ 对角矩阵的逆矩阵得到预处理矩阵 $G(\varphi) = \text{diag}(\varphi)^{-1}$ 与 $G(\theta) = \text{diag}(\theta)^{-1}$ 以此得到 φ, θ 的迭代公式:

$$\varphi_{ak}^{(t+1)} = \left| \varphi_{ak}^{(t)} + \varepsilon_t (\alpha - 1 - \rho \varphi_{ak}^{(t)} \sum_{b=1, b \neq a}^N g_{ab}(\varphi_{ak}^{(t)})) \right| \quad (7)$$

$$\theta_{ki}^{(t+1)} = \left| \theta_{ki}^{(t)} + \varepsilon_t (\eta - 1 - \rho \theta_{ki}^{(t)} + \theta_{ki}^{(t)} \sum_{a=1}^N \sum_{b=a+1}^N g_{ab}(\theta_{ki}^{(t)})) \right| \quad (8)$$

其中 $\varepsilon_t = \varepsilon_0 (t + \tau)^{-\kappa}$ 为缩小步长, ε_0, τ 以及 κ 均为超参数, $g_{ab}(\varphi_{ak})$ 与 $g_{ab}(\theta_{ki})$ 分别是 $\log p(L_{ab}, \boldsymbol{\pi}_a, \boldsymbol{\pi}_b, \boldsymbol{\beta})$ 中关于 φ_{ak} 与 θ_{ki} 的偏导数, 参考文献[23]可得

$$g_{ab}(\varphi_{ak}) = \frac{\pi_{ak} [\beta_k^{L_{ab}} (1 - \beta_k)^{1-L_{ab}} \pi_{bk} + \delta^{L_{ab}} (1 - \delta)^{1-L_{ab}} (1 - \pi_{bk})]}{Z_{ab}^{L_{ab}} \varphi_{ak}} - \frac{1}{\varphi_a} \quad (9)$$

$$g_{ab}(\theta_{ki}) = \frac{\beta_k^{L_{ab}} (1 - \beta_k)^{1-L_{ab}} \pi_{ak} \pi_{bk} \left(\frac{1 - i - L_{ab}}{\theta_{ki}} - \frac{1}{\theta_k} \right)}{Z_{ab}^{L_{ab}}} \quad (10)$$

其中,

$$Z_{ab}^{L_{ab}} = \delta^{L_{ab}} (1 - \delta)^{1-L_{ab}} + \sum_{k=1}^K [\beta_k^{L_{ab}} (1 - \beta_k)^{1-L_{ab}} - \delta^{L_{ab}} (1 - \delta)^{1-L_{ab}}] \pi_{ak} \pi_{bk} \quad (11)$$

经过参数的不断更新迭代,所生成的泛化图中节点之间的边更能反映真实连接关系,不确定性关系导致的弱边和假边越来越少,由此得到减少不确定性关系的网络结构.

4.4 使用 MC-dropout 采样权重样本

在节点分类任务中,对于单个元路径邻居图的泛化图样本 \mathcal{G} ,输入数据表示为节点特征 X 、节点标签 Y_L 与之相对应的输出表示为 $Z = \{z_1, \dots, z_n\}$,目标是通过神经网络训练得到一个能反映输入和输出之间关系的函数 $z = f(x)$,为此,可以使用贝叶斯方法,将权重参数 W 建模为随机变量,引入它们的先验分布,同时因为 W 是不确定的,所以神经网络的输出也是随机变量.对于新输入 x 的预测,可以根据相应 W 的后验分布,通过以下积分函数得到

$$p(zx, X, Y_L, \mathcal{G}) = \int p(zx, W, \mathcal{G}) p(WX, Y_L, \mathcal{G}) dW \quad (12)$$

公式(12)的积分函数是很难直接计算出来的,但是可以通过变分方法和马尔科夫链蒙特卡洛方法得到该积分的近似结果.在分类任务中,通过将 softmax 函数应用于神经网络的输出可以得到节点的分类分布,以此分布对 $p(zx, W, \mathcal{G})$ 进行建模,此外我们可以使用文献[13,24-25]中的变分推理或文献[26,28]的 MCMC 方法来近似权重的后验 $p(WX, Y_L, \mathcal{G})$,其中文献[13]通过变分推理得到权重后验的变分近似,并将 MC-dropout 方法等价于从近似的后验中采样 W 的样本.由此可以通过蒙特卡洛近似得到公式(12)的近似公式:

$$p(zx, X, Y_L, \mathcal{G}) \approx \frac{1}{S} \sum_{s=1}^S p(zx, W_s, \mathcal{G}) \quad (13)$$

其中 S 个权重样本 W_s 通过结合 MC-dropout 方法的 GCN 得到.

相比于元路径邻居图,4.3 节中所生成的泛化图样本减少了不确定性关系,拥有与之不同的网络结构,可被视为对抗性图数据,这些对抗性实例的关系关联性更强且蕴含更加丰富的结构信息,利用结合了 MC-dropout 的 GCN 采样其权重样本能进一步增强其泛化能力,因而本文提出的模型拥有更强的鲁棒性.

对于任意权重样本 W 及其对应节点分类分布 Z 可以通过 GCN 传播规则(15)得到:

$$\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \quad (14)$$

$$Z = \text{softmax}(\hat{A} \text{ReLU}(\hat{A} X W^{(0)})) W^{(1)} \quad (15)$$

上式中 A 为泛化图 \mathcal{G} 的邻接矩阵, $\tilde{A} = A + I_N$, I_N 为单位矩阵,度矩阵 $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$,其中 $X \in R^{N \times C}$, $W^{(0)} \in R^{C \times H}$, $W^{(1)} \in R^{H \times K}$, N, C, H, K 分别表示节点数量、节点特征维度、神经网络隐藏单元数量以及节点的类别数量.

GCN 在训练图的过程中聚合了节点的特征信息,因而所得到的节点类别概率分布实质上包含了节点的特征信息,实验中我们使用该节点类别概率分布初始化 α -MMSBM 的参数,所以基于此参数构建的泛化图中节点之间连接与否会受到节点特征信息的影响,由于弱关联关系的节点之间特征区别较大,而强关联关系的节点之间特征相近,所以所生成的泛化图中含有的弱关联性的边会减少,同时强关联性的边会增加,由此减少了图结构所包含的不确定性关系

4.5 近似求解及公式简化

基于上述推导,我们进一步给出公式(1)的蒙特卡洛近似表达.对于 $p(\mathcal{G}_\phi | \mathcal{G}_{obs})$ 的实现方式,本文针对不同数据集预定义多种元路径,基于这些元路径构建元路径邻居图样本集,再通过均匀采样的方式实现从异质图中采样元路径邻居图.对于 $p(\lambda | \mathcal{G}_\phi)$ 以及 $p(\mathcal{G} | \lambda)$,则通过随机图生成模型 α -MMSBM 对其建模来推理实现,其中推理时可以使用变分推理^[8]或 MCMC^[23].对于权重后验 $p(W | X, Y_L, \mathcal{G})$,我们使用变分推理^[13]得到近似后验,通过结合 MC-dropout 方法的 GCN 实现权重样本 W 采样.最后通过 GCN 的 softmax 函数的结果 $p(Z | W, \mathcal{G}, X)$ 建模.由此可以得到公式(1)的蒙特卡洛近似:

$$p(\mathbf{Z}\mathbf{X}, \mathbf{Y}_{\mathcal{L}}, \mathcal{G}_{obs}) \approx \frac{1}{U} \sum_{\phi} \frac{1}{I} \sum_i \frac{1}{N_{\mathcal{G}}} \sum_{n=1}^{N_{\mathcal{G}}} \sum_{s=1}^S p(\mathbf{Z}\mathbf{W}_{s,n,i,\phi}, \mathcal{G}_{n,i,\phi}, \mathbf{X}) \quad (16)$$

在近似公式(16)中,从 $p(\mathcal{G}_{\phi} | \mathcal{G}_{obs})$ 采样 U 个元路径邻居图样本 \mathcal{G}_{ϕ} ,对于每个 \mathcal{G}_{ϕ} ,通过随机图生成模型,从 $p(\lambda | \mathcal{G}_{\phi})$ 获取 I 个参数族样本 λ_i ,从 $p(\mathcal{G} | \lambda_i)$ 采样出 $N_{\mathcal{G}}$ 个泛化图样本 $\mathcal{G}_{n,i,\phi}$. 这些泛化图样本的准确度取决于对元路径邻居图样本所构建的随机图生成模型. 对于权重样本的采样 $p(\mathbf{W} | \mathbf{X}, \mathbf{Y}_{\mathcal{L}}, \mathcal{G}_{n,i,\phi})$,我们通过结合 MC-dropout 方法的 GCN 对每个 $\mathcal{G}_{n,i,\phi}$ 采样 S 个权重样本 $\mathbf{W}_{s,n,i,\phi}$,并对 $\mathbf{W}_{s,n,i,\phi}$ 使用 $softmax$ 函数得到节点标签的概率分布 $p(\mathbf{Z} | \mathbf{W}_{s,n,i,\phi}, \mathcal{G}_{n,i,\phi}, \mathbf{X})$,最后通过累加这些标签分布得到节点最终的标签分布 $p(\mathbf{Z} | \mathbf{X}, \mathbf{Y}_{\mathcal{L}}, \mathcal{G}_{obs})$.

在本文中,我们通过随机优化方法学习 a -MMSBM 的随机参数 $\boldsymbol{\pi}, \boldsymbol{\beta}$. 但是由于 a -MMSBM 的后验维度过高,对其参数采用随机初始化的方式会影响训练的效果,所以本文使用 GCN 预训练元路径邻居图 \mathcal{G}_{ϕ} ,利用 $softmax$ 函数的输出来初始化参数 $\boldsymbol{\pi}$ 和 $\boldsymbol{\beta}$. 同时为了避免因为参数 $\boldsymbol{\pi}$ 和 $\boldsymbol{\beta}$ 取值范围过大导致 a -MMSBM 生成的泛化图样本 $\mathcal{G}_{n,i,\phi}$ 与 \mathcal{G}_{ϕ} 差异过大,我们使用最大后验估计来代替 $\boldsymbol{\pi}$ 和 $\boldsymbol{\beta}$ 的积分,利用合适的 $\boldsymbol{\pi}$ 和 $\boldsymbol{\beta}$ 的先验得到近似公式:

$$\{\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\beta}}\} = \underset{\boldsymbol{\pi}, \boldsymbol{\beta}}{\operatorname{argmax}} p(\boldsymbol{\pi}, \boldsymbol{\beta} | \mathcal{G}_{\phi}) \quad (17)$$

由此 $\mathcal{G}_{n,i,\phi}$ 改写为 $\mathcal{G}_{n,\phi}$, $\mathbf{W}_{s,n,i,\phi}$ 改写为 $\mathbf{W}_{s,n,\phi}$, 并可以进一步简化公式(16)得到:

$$p(\mathbf{Z}\mathbf{X}, \mathbf{Y}_{\mathcal{L}}, \mathcal{G}_{obs}) \approx \frac{1}{U} \sum_{\phi} \frac{1}{N_{\mathcal{G}}} \sum_{n=1}^{N_{\mathcal{G}}} \sum_{s=1}^S p(\mathbf{Z}\mathbf{W}_{s,n,\phi}, \mathcal{G}_{n,\phi}, \mathbf{X}) \quad (18)$$

在近似公式(18)中,利用 a -MMSBM 从 $p(\mathcal{G}_{n,\phi} | \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\beta}})$ 采样得到 $\mathcal{G}_{n,\phi}$,结合 MC-dropout 方法对 $\mathcal{G}_{n,\phi}$ 使用 GCN 从 $p(\mathbf{W}\mathbf{X}, \mathbf{Y}_{\mathcal{L}}, \mathcal{G}_{n,\phi})$ 采样权重样

本 $\mathbf{W}_{s,n,\phi}$,并使用 $softmax$ 函数得到节点标签分布,最后采用累加的方式计算节点的标签分布.

4.6 算法

结合公式(18)展示的最终推导结果,BHNN 的算法过程可以通过算法 1 展示.

算法 1. 贝叶斯异质图神经网络

输入: 异质图 $\mathcal{G}_{obs} = (V, E)$; 节点特征 \mathbf{X} ; 节点标签 $\mathbf{Y}_{\mathcal{L}}$;

输出: 节点标签分布 $p(\mathbf{Z} | \mathbf{X}, \mathbf{Y}_{\mathcal{L}}, \mathcal{G}_{obs})$;

1: 预定义元路径 $\phi \in \{\phi_0, \phi_1, \dots, \phi_U\}$,基于 $p(\mathcal{G}_{\phi} | \mathcal{G}_{obs})$ 从 \mathcal{G}_{obs} 中获取元路径邻居图集合 $\{\mathcal{G}_0, \mathcal{G}_1, \dots, \mathcal{G}_U\}$;

2: FOR $\mathcal{G}_{\phi} \in \{\mathcal{G}_0, \mathcal{G}_1, \dots, \mathcal{G}_U\}$ DO

3: 使用 GCN 训练随机图 \mathcal{G}_{ϕ} 以初始化 a -MMSBM 参数和权重参数;

4: FOR $n = 1; N_{\mathcal{G}}$ DO

5: 执行 N 次 a -MMSBM 的迭代推理,结合公式(17)以

获得 $(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\beta}})$;

6: 利用 a -MMSBM 生成图样本 $\mathcal{G}_{n,\phi} \sim$

$p(\mathcal{G} | \hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\beta}})$;

7: FOR $s = 1; S$ DO

8: 通过 MC-dropout 方法,使用 GCN 训练 $\mathcal{G}_{n,\phi}$ 来采样权重样本 $\mathbf{W}_{s,n,\phi}$;

9: END FOR

10: END FOR

11: END FOR

12: 使用公式(18)得到 $p(\mathbf{Z} | \mathbf{X}, \mathbf{Y}_{\mathcal{L}}, \mathcal{G}_{obs})$ 的近似结果;

算法 1 以异质图的节点特征、节点标签以及网络结构作为输入,得到节点的预测标签,其中算法第 1 行表示基于预定义的元路径从输入的异质图中构建元路径邻居图样本集,第 2,3 行表示遍历元路径邻居图样本集,对于每个元路径邻居图使用 a -MMSBM 建模并通过 GCN 的 $softmax$ 函数输出对参数进行初始化,第 4~6 行表示泛化图的生成过程,包括利用最大后验估计处理 a -MMSBM 参数以及图样本的生成,第 7,8 行表示使用结合 MC-dropout 方法的 GCN 采样权重样本,第 12 行表示累加这些权重样本的 $softmax$ 输出得到最终的节点标签.

表 2 数据集基本信息

数据集	节点的数量	边的数量	节点的类型	边的类型	节点特征	训练集	验证集	测试集
DBLP	27194	122393	4	3	334	800	400	2857
ACM	8916	12769	3	2	1830	600	300	2125
IMDB	12890	19120	3	2	1232	658	329	2300

算法 1 中对采样 U 个元路径邻居图的计算复杂度是 $\mathcal{O}(U)$,采样 $N_{\mathcal{G}}$ 个泛化图样本的计算复杂度是 $\mathcal{O}(N_{\mathcal{G}} |V'|^2 K)$,其中 $|V'|^2, K$ 分别表示节点

数量、节点的类别数量,采样 S 个权重样本的计算复杂度是 $\mathcal{O}(S |E'| CHK)$,其中 $\mathcal{O}(|E'|)$ 是 GCN 中邻接矩阵的稀疏表示所需内存,其大小与边的数

量 $|E'|$ 成线性关系, C, H 分别表示节点特征维度以及神经网络隐藏单元数量, 上述过程中 $U, N_G, K, S, C, H \ll |V'|^2, |E'|$, 所以算法 1 的计算复杂度是 $O(|V'|^2 |E'|)$.

5 实验与分析

在本节中, 所有实验都是基于三个基准异质图数据集进行的, 在各个实验中, 通过与基准方法对比, 展示我们提出的方法的先进性.

5.1 数据集

本文所使用的异质图数据集的详细描述都展示在表 2 中, 其中 DBLP 和 ACM 数据集都是论文引用网络数据, 而 IMDB 数据集是电影网络数据.

DBLP^①. 该数据集中的论文数据来源于四个存在关联的领域: 数据库、数据挖掘、机器学习、信息检索, 该网络中包含四种类型的节点, 分别是 14328 个论文节点(P)、4057 个作者节点(A)、20 个会议节点(C)和 8789 个术语节点(T), 拥有 122393 条边其中包含三种边类型: 19645 条论文—作者(P-A)、14328 条论文—会议(P-C)、88420 条论文—术语(P-T). 在实验中我们选择由多种边类型构成的两条元路径 APA、APCPA 作为输入数据.

ACM^②. 该数据集中的论文数据来源于数据库、无线通讯、数据挖掘方向的国际会议, 此异质图由 3025 个论文节点(P)、5835 个作者节点(A)、56 个主题节点(L)、9744 条论文—作者边(P-A)、3025 条论文—主题边(P-L)构成, 论文节点的特征由论文内容的词袋模型的关键词构成, 不同类型的边可以构建不同的元路径, 我们选择其中的两条元路径 PAP、PLP 进行实验.

IMDB^③. 该数据集属于电影网络, 其中包含 4780 个电影节点(M)、5841 个演员节点(A)、2269 个导演节点(D)、14340 条电影—演员(M-A)边和 4780 条电影—导演边(M-D), 电影类型可以分为三类: 动作片、喜剧片和戏剧片, 每个电影节点的特征同样由词袋模型得到. 在实验中, 我们选择两条元路径 MAM、MDM 作为模型的输入.

5.2 基准方法

为了验证本文提出的模型的有效性和先进性, 在基准试验中, 我们不仅从目前已有的异质图节点嵌入方法以及异质图神经网络中选择最先进的方法模型进行对比, 而且还对比了同质图中最先进的方法模型.

DeepWalk^[29]: 一个用于学习同质图网络嵌入的模型, 先利用随机游走策略构建节点的序列, 再将这些节点序列作为 word2vec 的输入, 从而得到节点的嵌入表示, 实验中我们不考虑图中节点存在类型上的差异, 直接在整个异质图上执行 DeepWalk.

HERec^[10]: 一个学习异质图节点嵌入的方法, 基于元路径进行随机游走, 同时结合过滤机制得到有效的节点序列, 再利用 skip-gram 学习节点的嵌入. 实验中我们分别测试每条元路径并以最好的结果作为最终结果.

Metapath2vec^[11]: 一个学习异质图节点嵌入的方法, 通过在元路径上随机游走并利用 skip-gram 学习节点的嵌入, 与 HERec 相比, 缺少对节点序列的过滤策略. 实验中我们分别测试每条元路径并以最好的结果作为最终结果.

GCN^[21]: 一个训练同质图的图神经网络, 将卷积神经网络应用到图数据中, 通过图拉普拉斯算子的特征分解, 在傅里叶域实现图的卷积运算. 实验中我们分别测试每条元路径并以最好的结果作为最终结果.

GAT^[30]: 一个应用于同质图的图神经网络, 利用注意力机制, 以邻居节点的特征来更新中心节点特征. 实验中我们分别测试每条元路径并以最好的结果作为最终结果.

HAN^[5]: 一个应用于异质图的图神经网络, 通过节点层次和语义层次的注意力机制分别学习基于元路径的节点之间的重要性以及不同元路径的重要性, 以此更新中心节点的特征. 实验中我们将已经预定义元路径的基准异质图数据集作为输入, 以训练得到结果作为最终结果.

BGCN^[12]: 一个应用于同质图的图神经网络, 结合贝叶斯方法以及随机块模型, 处理图结构的不确定性. 实验中我们分别测试每条元路径并以最好的结果作为最终结果.

DGMI^[31]: 一种应用于异质图的图神经网络, 用于多重属性网络的简单有效的无监督网络嵌入方法, 可最大化图的局部和整个图的全局表示之间的互信息. 实验中我们将已经预定义元路径的基准异质图数据集作为输入, 以训练得到结果作为最终结果.

LGCN^[32]: 一种应用于同质图的图神经网络,

① 1 <https://dblp.uni-trier.de>

② 2 <https://dl.acm.org/>

③ 3 <https://www.imdb.com>

通过洛伦兹重建双曲 GCN 的图操作来严格保证学习到的节点特征遵循双曲几何的一种新型双曲 GCN 模型. 实验中我们分别测试每条元路径并以最好的结果作为最终结果.

BHNN: 本文提出的模型, 基于异质图的元路径, 将元路径邻居图视为随机图, 结合贝叶斯方法以及随机块模型, 用于解决异质图结构的不确定性同时提高模型的鲁棒性. 实验中我们将已经预定义元路径的基准异质图数据集作为输入, 以训练得到结果作为最终结果.

5.3 实验参数设置和评价指标

(1) 实验平台参数. 本文实验运行于版本为 18.04 的 Ubuntu 系统, 使用 GeForce RTX 2080 Ti 显卡, CUDA 版本为 10.2, 使用版本为 1.14 的深度学习框架 tensorflow-gpu 搭建算法代码.

(2) 实验参数. 本文的基准实验中, 为了能准确对比各个模型的性能, 本文将数据集划分为训练集、验证集以及测试集, 每个模型使用同样的数据. 对于各个模型相同的训练参数, 我们设置相同的值, 其中模型训练迭代 epoch 数量设为 600, 优化器使用 Adam^[33], 模型输出 embedding 维度设为 64, 学习率设为 0.01, L2 正则化参数设为 5×10^{-4} , dropout 率设为 0.5, 每层神经网络隐藏单元数量设为 16 个. 对于 HAN 与 GAT 模型的 attention head 的数量设为 8, attention 的 dropout 值设为 0.5, HAN 模型语义层次的 attention 向量维度设为 128, GCN 和本文提出的 BHNN 中的切比雪夫多项式的最大度均设为 3, 而 BGCN 与 BHNN 中的 a-MMSBM 的推理参数设置一致, 设定 $\eta = 1, \alpha = \frac{1}{K}, K$ 为节点类别数量, $\delta = 10^{-4}, \rho = 1, \epsilon_0 = 1, \tau = 1024$ 以及 $\kappa = 0.5$. BHNN、DGMI、BGCN、GCN 模型均使用 ReLU 作为激活函数而 HAN、GAT 与 LGCN 的激活函数是 ELU. 对于基于随机游走的图嵌入模型 DeepWalk、HERec 以及 Metapath2vec, 我们将窗口大小设为 5, 每个节点的序列数量设为 40, 游走步长设为 100, 负采样样本数设为 5.

(3) 评价指标. 本文的所有实验均使用微 $F1$ 、宏 $F1$ 作为评价模型分类效果的指标, 且每个实验重复 10 次, 取微 $F1$ 与宏 $F1$ 的平均值作为实验结果. 微 $F1$ 与宏 $F1$ 可通过以下定义得到. 其中 TP_k (True Positive) 表示类别为 k 的样本被正确预测的数量, FN_k (False Negative) 表示类别为 k 的样本被错误预测的数量, FP_k (False Positive) 表示类别不为 k

的样本被预测为 k 的数量.

微 $F1$ ^[34] 的计算公式如式(19)所示.

微 $F1 =$

$$2 \times \frac{\sum_k^K TP_k}{\sum_k^K TP_k + \sum_k^K FP_k} \times \frac{\sum_k^K TP_k}{\sum_k^K TP_k + \sum_k^K FN_k} \\ \frac{\sum_k^K TP_k}{\sum_k^K TP_k + \sum_k^K FP_k} + \frac{\sum_k^K TP_k}{\sum_k^K TP_k + \sum_k^K FN_k} \quad (19)$$

宏 $F1$ ^[34] 的计算公式如式(20)所示.

$$\text{宏 } F1 = \frac{1}{K} \sum_k^K \frac{2 \times \frac{TP_k}{TP_k + FP_k} \times \frac{TP_k}{TP_k + FN_k}}{\frac{TP_k}{TP_k + FP_k} + \frac{TP_k}{TP_k + FN_k}} \quad (20)$$

5.4 节点分类实验结果和分析

表 3 展示了本文提出的方法 BHNN 以及基准方法在基准异质图数据集 ACM、DBLP 和 IMDB 上的节点分类结果, 模型的训练效果通过评价指标微 $F1$ 与宏 $F1$ 展示.

从结果来看, 在异质图的节点嵌入方法中, 加入对节点序列过滤机制的 HERec 在三个数据集上的微 $F1$ 与宏 $F1$ 均高于仅使用随机游走构建节点序列的 Metapath2vec, 而原本适用于训练同质图数据的 DeepWalk 在大部分数据集中表现优于其他节点嵌入方法, 这是因为 DeepWalk 是在整个异质图上进行游走的而不是基于元路径邻居图, 由此能学习到图上更多的信息. 与节点嵌入的方法相比, 同时结合图的结构和节点特征信息的图神经网络方法 GCN、GAT、HAN、BGCN、DMGI 以及 LGCN 在节点分类的效果上有很大的提升.

对比图神经网络方法可以发现, 直接将同质图训练模型应用于异质图时, 节点分类的效果并不一定差于异质图神经网络, 这是因为同质图模型展现的结果仅基于表现最好的单个元路径邻居图, 相对而言数据所包含的不确定性关系以及数据噪声较少, 而考虑到数据噪声问题的同质图模型 BGCN, 相比较于其他同质图模型中也有更好的表现.

本文提出的 BHNN 不仅能得到类似同质图中的最佳的结果, 还基于预定义元路径并结合贝叶斯方法和随机块模型解决了异质图的关系不确定性问题以及数据噪声问题, 从而得到目前最好的结果. 对比于图嵌入和其他图神经网络方法, BHNN 在三

个数据集中均取得最优的性能,节点分类结果的微 $F1$ 与宏 $F1$ 高于其他方法,且微 $F1$ 平均提高了

1.59%,宏 $F1$ 平均提高了 1.36%,验证了我们提出的方法的有效性和先进性.

表 3 节点分类实验结果

数据集	评价指标	模型									
		DeepWalk	HERec	Metapath2vec	GCN	GAT	HAN	BGCN	DMGI	LGCN	BHNN
ACM	宏 $F1$	77.25	66.17	65.09	85.36	85.45	86.50	86.10	83.98	84.96	87.81
	微 $F1$	76.92	66.03	65.00	85.45	85.57	86.58	85.97	84.79	85.08	88.00
DBLP	宏 $F1$	77.43	91.68	90.16	89.29	87.34	90.68	90.00	91.47	92.11	92.39
	微 $F1$	79.37	92.69	91.53	90.51	89.04	91.79	91.14	92.23	93.07	93.27
IMDB	宏 $F1$	40.72	41.65	41.16	57.48	52.09	58.36	59.29	50.53	56.99	60.34
	微 $F1$	46.38	45.81	45.65	59.26	59.30	59.47	60.95	54.86	58.70	62.82

5.5 消融实验结果和分析

为了更好地展现本文提出的异质图神经网络各部分效果,本节实验将 BHNN 的部分模块剥离后与基准方法中的其他异质图神经网络进行比较.我们测试了将 BHNN 中的随机块模型剥离,仅使用 GCN 进行训练,我们简称为 HGCN,训练时设置的参数与 BHNN 完全一致,通过节点分类的结果评价模块的有效性.

表 4 消融实验结果 (单位:%)

数据集	评价指标	模型			
		HAN	DMGI	HGCN	BHNN
ACM	宏 $F1$	86.50	83.98	88.26	87.81
	微 $F1$	86.58	84.79	88.28	88.00
DBLP	宏 $F1$	90.68	91.47	91.56	92.39
	微 $F1$	91.79	92.23	92.43	93.27
IMDB	宏 $F1$	58.36	50.53	58.39	60.34
	微 $F1$	59.47	54.86	60.52	62.82

从表 4 展示的结果分析,使用贝叶斯方法的 HGCN 在 ACM、DBLP、IMDB 三个数据集上的节点分类效果均优于 HAN 与 DMGI,这是因为通过贝叶斯方法,HGCN 能涵盖更多的可能性,得到更多的信息.而 BHNN 相比较于 HGCN,在 DBLP、IMDB 数据集上微 $F1$ 分别提高了 0.84%、2.3%,宏 $F1$ 分别提高了 0.83%、1.95%,在 ACM 数据集上 BHNN 的微 $F1$ 与宏 $F1$ 分别比 HGCN 略低 0.28%、0.45%,这是因为 ACM 数据集中包含的节点和边类型较少,且规模结构较小,因而整体数据噪声较少,所以图结构中关系的不确定性问题不突出,而这导致 BHNN 在使用随机块模型构建泛化图时生成的图样本相似度过高,所以在训练 ACM 数据集时,BHNN 出现过拟合现象,因而分类效果稍差于 HGCN.

上述的实验结果验证了 BHNN 中使用的贝叶斯方法以及随机块模型的有效性,能够解决异质图

上的关系不确定性问题并提高节点分类的效果.

5.6 节点攻击下的节点分类实验结果和分析

为了对比 BHNN 与基准方法中图神经网络的鲁棒性强弱,本节实验采用 Daniel 等人^[35]设计的针对图数据的对抗攻击的方法,即在模型训练前对数据进行干扰,利用修改后的数据进行训练,这种机制称为中毒攻击.在实验中我们分别对 10 个、50 个以及 100 个节点进行攻击,对于每个节点都限定扰动数量为 $\Delta = d_{v_0} + 2$,以避免攻击后的图结构与原图差异过大,其中 d_{v_0} 表示目标节点 v_0 的度,中毒攻击后,随机删除 $(d_{v_0} + 2)/2$ 条目标节点与原邻居之间的边,同时会选择 $(d_{v_0} + 2)/2$ 个与目标节点类别不相同的节点,并在目标节点与这些节点之间构建连接的边.

表 5 攻击 10 个节点后节点分类结果 (单位:%)

数据集	评价指标	模型						
		GCN	GAT	HAN	BGCN	DMGI	LGCN	BHNN
ACM	宏 $F1$	84.86	85.26	82.60	86.47	83.14	83.57	87.99
	微 $F1$	84.96	85.37	82.93	86.54	82.99	83.72	88.18
DBLP	宏 $F1$	89.34	89.91	90.27	89.95	91.08	92.16	92.25
	微 $F1$	90.56	91.12	91.43	91.14	91.96	93.12	93.17
IMDB	宏 $F1$	56.92	53.06	56.96	59.27	49.32	56.02	59.80
	微 $F1$	58.84	59.28	58.00	60.43	53.78	57.83	62.17

表 6 攻击 50 个节点后节点分类结果 (单位:%)

数据集	评价指标	模型						
		GCN	GAT	HAN	BGCN	DMGI	LGCN	BHNN
ACM	宏 $F1$	83.06	83.14	81.31	84.47	81.42	81.06	86.67
	微 $F1$	83.15	83.28	81.56	84.56	81.15	81.13	86.82
DBLP	宏 $F1$	89.20	88.83	89.83	90.01	90.55	91.02	91.90
	微 $F1$	90.43	90.25	91.01	91.21	91.45	92.61	92.89
IMDB	宏 $F1$	55.98	52.25	54.91	58.41	47.99	53.22	59.11
	微 $F1$	57.90	58.77	55.92	59.86	52.12	54.00	61.17

表 5、6、7 分别展示了攻击 10 个、50 个、100 个节点后,BHNN 和其他图神经网络的节点分类实验结果.从分类结果分析,虽然随着攻击节点数量的增加,各模型的训练效果大体上呈现逐步下降的趋势,

但是相比较于其他 GNN, BHNN 仍能得到最好的分类结果,同时注意到,存在个别数据出现上升的情况,这是因为在节点攻击的过程中,新边的构建具有一定的随机性,相对于被删除的边,新边的关联性可能会更强,因而节点攻击后新构建的图中强关联性的边可能会增加,模型的节点分类效果有所提高.

表 7 攻击 100 个节点后节点分类结果 (单位:%)

数据集	评价指标	模型						
		GCN	GAT	HAN	BGCN	DMGI	LGCN	BHNN
ACM	宏 F1	82.22	81.71	78.95	83.59	80.34	80.66	86.85
	微 F1	82.33	81.90	79.39	83.67	80.01	80.89	86.96
DBLP	宏 F1	89.26	87.62	89.97	89.94	90.19	90.84	91.55
	微 F1	90.48	89.15	91.12	91.10	91.15	91.94	92.54
IMDB	宏 F1	54.16	50.31	53.79	58.06	45.94	52.46	58.22
	微 F1	56.23	57.35	55.04	59.86	50.33	56.43	60.65

为了更直观地对比这些模型受到节点攻击后节点分类结果的变化趋势,我们通过折线图进行比较.

图 4 与图 5 分别展示 ACM 数据集的节点受到攻击后,不同图神经网络的节点分类结果中微 F1 与宏 F1 的变化趋势.图中显示在被攻击的节点数量增加时每个模型都受到不同程度的影响,微 F1 与宏 F1 整体呈现下降趋势,但相比较而言,BHNN 在保持最高的微 F1 与宏 F1 的同时,微 F1 与宏 F1 的最大降幅为 1.36% 与 1.32%,明显小于 BGCN 的 2.87% 与 2.88%、GCN 的 3.12% 与 3.14%、LGCN 的 4.19% 与 4.3%、GAT 的 3.67% 与 3.74% 以及 HAN 的 7.19% 与 7.55%、DMGI 的 4.78% 与 3.64%,这说明 BHNN 在处理扰动数据时具有更强的抗干扰能力.

表 8 在不同的预训练次数条件下节点分类结果

(单位:%)

数据集	评价指标	模型							
		50	100	150	200	250	300	350	400
ACM	宏 F1	87.50	87.53	87.73	87.43	87.67	87.39	87.58	87.81
	微 F1	87.71	87.71	87.90	87.62	87.85	87.57	87.76	88.00
DBLP	宏 F1	91.65	91.69	91.76	91.75	91.63	91.83	92.20	92.39
	微 F1	92.64	92.75	92.78	92.68	92.57	92.78	93.10	93.27
IMDB	宏 F1	60.31	60.45	60.16	60.14	60.13	60.09	60.60	60.34
	微 F1	62.95	62.86	62.73	62.43	62.43	62.39	63.08	62.82

5.7 模型参数敏感度分析与计算效率比较

在这一部分,我们将评估 BHNN 对超参数预训练轮数、权重样本数量以及 α -MMSBM 超参数设置的敏感度,同时对比各个图神经网络的计算效率,并对实验结果进行分析.

在评估 BHNN 对超参数预训练轮数和权重样本数量设置的敏感度的实验中,我们设置每个元路径邻居图的训练迭代轮数为 $N_{train} = 600$ 的实验条

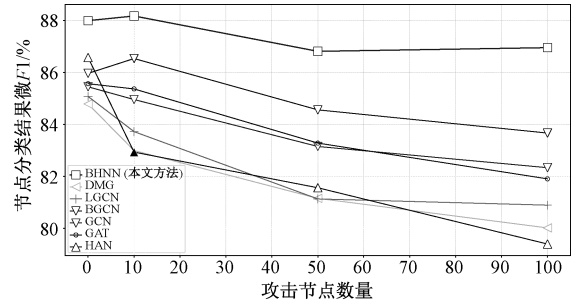


图 4 节点攻击下模型分类任务的微 F1 变化

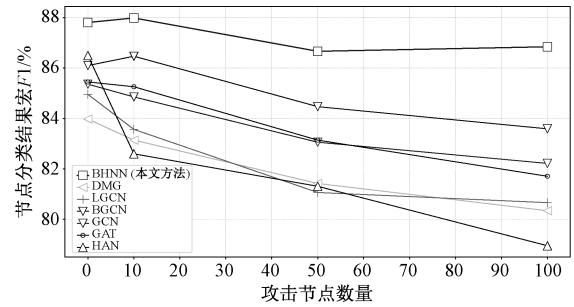


图 5 节点攻击下模型分类任务的宏 F1 变化

相比较于基准方法中的异质图神经网络, BHNN 在 ACM、DBLP、IMDB 三个基准异质图数据集上的微 F1 与宏 F1 平均仅下降了 1.42% 与 1.43%,而异质图神经网络 HAN 与 DMGI 分别平均下降了 4.13% 与 3.46%、而宏 F1 分别平均下降了 3.17% 与 4.32%,这充分说明 BHNN 具有更强的鲁棒性.这一特性在真实世界中具有重要意义,例如在社交网络中,添加或删除好友的操作比较频繁,网络结构变动较大,使用一个具有更强适应性的模型能保证的相对稳定的训练结果.

件下,模型的训练过程可以分为两个阶段,第一个阶段是预训练阶段,该阶段使用 GCN 预训练 N_{pre} 轮后利用 $softmax$ 函数的输出对 α -MMSBM 参数和权重参数进行初始化,第二个阶段是训练阶段,该阶段通过结合 MC-dropout 方法的 GCN 采 N_w 个权重样本,其中 $N_{pre} + N_w \leq N_{train}$,所以本节实验分别对影响模型性能的两个参数,即预训练轮数 N_{pre} 和权重样本数量 N_w 进行测试,并分析它们在

取何值时,模型达到最佳性能.

在分析预训练轮数对模型性能影响的实验中,我们设置权重的样本数量为 $N_w = 200$ 个,则预训练轮数 $N_{pre} \leq N_{train} - N_w = 400$,在此基础上分别测试不同的预训练轮数,得到模型在进行节点分类实验时训练效果的变化趋势.表 8 展示了 BHNN 在 ACM、DBLP 以及 IMDB 三个数据集上设置预训练轮数为 50 轮、100 轮、150 轮、200 轮、250 轮、300 轮、350 轮、400 轮的条件,节点分类实验的结果并标注了其中最值.分析发现,随着预训练轮数的增加,BHNN 在各个数据集上的节点分类效果并没有明显的提升或下降,其中在 ACM、DBLP、IMDB 数据集上微 $F1$ 与宏 $F1$ 的最大变化幅度分别是 0.43% 与 0.42%、0.63% 与 0.76%、0.69% 与 0.51%,微 $F1$ 与宏 $F1$ 平均最大变化幅度分别是 0.58% 与 0.56%,这说明预训练轮数对 BHNN 的性能影响较小,但是在各个数据集中,当预训练轮数取最大值时,分类结果的微 $F1$ 与宏 $F1$ 达到最大值或相当接近最大值,所以在将预训练轮数设置为 $N_{pre} = N_{train} - N_w$ 时,模型能达到较好的性能.

根据上述得出的结论,在分析权重样本数量对模型性能影响的实验时,对于不同的权重样本数量取值 N_w ,预训练轮数设置为 $N_{pre} = N_{train} - N_w$.

图 6 展示了在 DBLP、ACM 和 IMDB 三个数据集上采样 50 个、100 个、150 个、200 个、250 个、300 个、350 个权重样本的条件下,模型节点分类效果的变化趋势.从折线图的变化趋势可以看出,在数据集 DBLP、ACM、IMDB 的权重样本分别增加到 200 个、100 个、250 个之前,模型的节点分类结果的微 $F1$ 与宏 $F1$ 呈现递增趋势,而在此之后,随着权重样本数量的增加,各个数据集的微 $F1$ 与宏 $F1$ 均大体呈现下降趋势,这是因为适度地提高权重样本数量有助于模型学习到更多信息,而过度地增加权重样本的数量会导致过拟合问题,进而造成模型性能降低.

在评估 BHNN 对 α -MMSBM 的超参数 α 、 η 、 ρ 以及 δ 设置的敏感度实验中,我们设置每个元路径邻居图的训练迭代轮数为 $N_{train} = 600$ 、预训练轮数为 $N_{pre} = 400$ 、权重的样本数量为 $N_w = 200$,在此实验条件下,分析 BHNN 进行节点分类实验的性能变化.图 7(a)、(b)、(c)、(d) 分别展示超参数 α 、 η 、 ρ 以及 δ 在取值范围 $\{10^{-4}, 10^{-3}, 0.01, 0.1, 0.5, 1\}$ 内, BHNN 分类数据集 DBLP、ACM 以及 IMDB 的节点类别的结果.

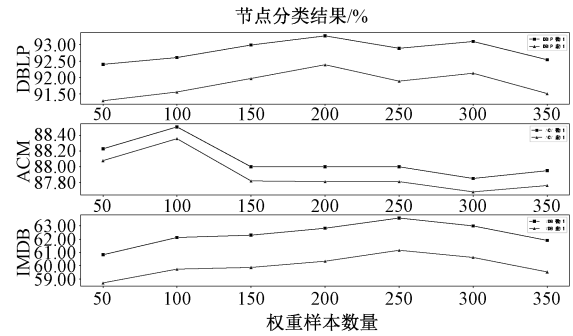


图 6 权重样本数量对分类结果的影响

分析发现,在 $1 \geq \alpha \geq 0.01, 1 \geq \eta \geq 0.01, \rho = \{10^{-4}, 10^{-3}, 1\}$ 以及 $\delta = 10^{-4}$ 的情况下,节点分类结果的微 $F1$ 与宏 $F1$ 取得最高或较高的值,将超参数设置为其他值时, α -MMSBM 模型所构建的泛化图与元路径邻居图结构上存在较大的差异,导致在训练时引入过多的噪声,因而模型的节点分类效果较差.本文所有实验中 α -MMSBM 超参数的设置主要参考文献[12]以及文献[23],将 α 设置为 $\frac{1}{K}$, K 为节点类别数量, η 设置为 1, ρ 设置为 1, δ 设置为 10^{-4} .

图神经网络的计算效率的高低可以通过模型训练时间的长短体现出来,所以本节实验通过统计每个模型训练数据所需的平均时间实现模型计算效率的对比.

表 9 图神经网络训练时间 (单位:s)

数据集	模型						
	GCN	GAT	HAN	DMGI	LGCN	BGCN	BHNN
ACM	308	885	478	392	277	1243	1280
DBLP	483	917	492	405	452	1829	1996
IMDB	190	661	349	311	139	956	958

表 9 中展示出了基准方法中图神经网络在训练 ACM、DBLP 以及 IMDB 数据集所需的时间.从展示的时间数据分析, BHNN 所需要的训练时间多于其他图神经网络模型,这是因为 BHNN 在训练过程中需要使用结合 MC-dropout 的 GCN 对神经网络权重进行采样,并多次使用 α -MMSBM 模型对图进行重构,耗费较多时间,所以 BHNN 的平均训练时间 1411 秒长于 GCN 的 327 秒、GAT 的 821 秒、LGCN 的 289 秒、HAN 的 439 秒与 DGMI 的 369 秒,但接近于同样需要采样权重样本和进行图重构的 BGCN 的 1343 秒.

6 总 结

针对异质图网络结构中存在的关系不确定性问

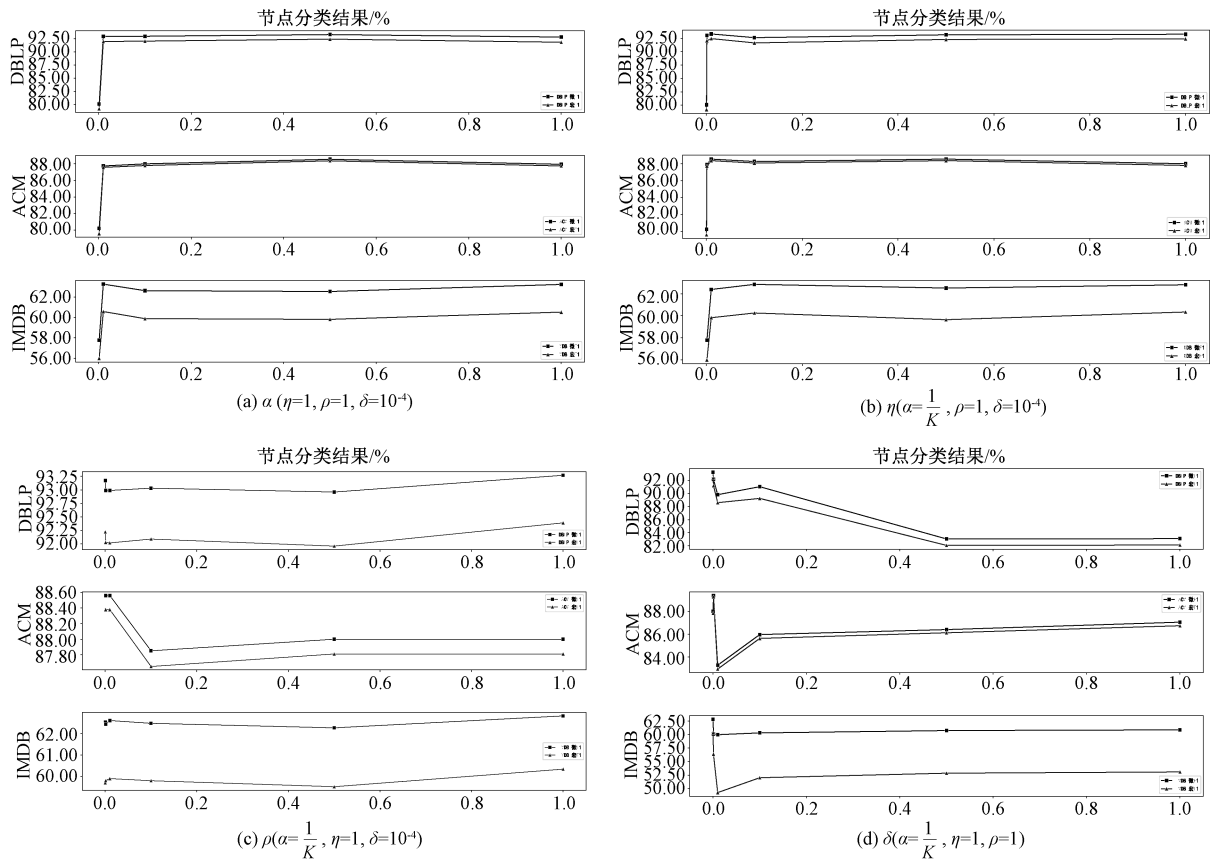


图7 α -MMSBM 超参数对分类结果的影响

题以及现有的异质图神经网络鲁棒性较弱问题,本文结合贝叶斯方法提出一种基于元路径的异质图神经网络 BHNN,给出了模型的详细设计和有关公式的推导,并进一步给出节点标签后验计算公式及其近似求解表达.该模型利用随机块模型重构异质图网络结构以此解决关系不确定性问题,同时通过提高训练数据中的对抗性实例数量增强模型的鲁棒性.在 ACM、DBLP、IMDB 三个基准异质图数据集上进行的节点分类实验以及节点攻击实验的结果表明,BHNN 优于现有的异质图神经网络,具有更强的节点分类能力以及鲁棒性.

在实验中我们也发现 BHNN 的训练时间长于现有的异质图神经网络,同时训练时需要更多的计算资源的问题.因此进一步缩短训练时间以及减少计算资源的需求是今后工作的研究重点.

参 考 文 献

[1] Liu H W, Kou H Z, Chao Y, Qi L Y. Link prediction in paper citation network to construct paper correlation graph. EURASIP Journal on Wireless Communications and Networking,

2019, 2019(1): 1-12
 [2] Wang Z Q, Tan Y W, Zhang M. Graph-based recommendation on social networks//Proceedings of the 12th Asia-Pacific Web-Conference. Busan, Korea, 2010: 116-122
 [3] Sun B, Zhao D, Shi X, et al. Modeling global spatial-temporal graph attention network for traffic prediction. IEEE Access, 2021, 9(99): 8581-8594
 [4] Yun S, Jeong M, Kim R, et al. Graph transformer networks//Proceedings of the 33rd Annual Conference on Neural Information Processing Systems. Vancouver, Canada, 2019: 11960-11970
 [5] Wang X, Ji H Y, Shi C, et al. Heterogeneous graph attention network//Proceedings of the 19th: The Web Conference on The World Wide Web Conference. San Francisco, USA, 2019: 2022-2032
 [6] Cai X C, Bain M, Krzywicki A, et al. Reciprocal and heterogeneous link prediction in social networks//Proceedings of the 16th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining. Kuala Lumpur, Malaysia, 2012: 193-204
 [7] Szegedy C, Zaremba W, Sutskever I, et al. Intriguing properties of neural networks//Proceedings of the 2nd International Conference on Learning Representations. Banff, Canada, 2014: 1-10

- [8] Prem G, David M, Mimno, et al. Scalable inference of overlapping communities//Proceedings of the 26th Annual Conference on Neural Information Processing Systems. Lake Tahoe, USA, 2012; 2258-2266
- [9] Sun Y, Han J, Yan X, et al. PathSim: Meta path-based top-K similarity search in heterogeneous information networks. *Vldb Endowment*, 2011, 4(11): 992-1003
- [10] Shi C, Hu B B, Philip S. Y. Heterogeneous information network embedding for recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2019, 31(2): 357-370
- [11] Dong Y, Chawla N V, Swami A. Metapath2vec: Scalable representation learning for heterogeneous networks//Proceedings of the 23rd International Conference on Knowledge Discovery and Data Mining. Halifax, Canada, 2017; 135-144
- [12] Zhang Y, Pal S, Coates M, et al. Bayesian graph convolutional neural networks for semi-supervised classification//Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu, USA, 2019; 5829-5836
- [13] Yarín G, Zoubin G. Dropout as a bayesian approximation: Representing model uncertainty in deep learning//Proceedings of the 33rd International Conference on Machine Learning. New York, USA, 2016; 1050-1059
- [14] Hasanzadeh A, Hajiramezani E, Boluki S, et al. Bayesian graph neural networks with adaptive connection sampling//Proceedings of the 37th International Conference on Machine Learning. Vienna, Austria, 2020; 4094-4104
- [15] Wu ZH, Pan S R, Long G D, et al. Graph wavenet for deep spatial-temporal graph modeling//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, China, 2019; 1907-1913
- [16] Hou S F, Fan Y J, Zhang Y M, et al. aCyber: Enhancing robustness of android malware detection system against adversarial attacks on heterogeneous graph based model//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China, 2019; 609-618
- [17] Ioannidis V N, Giannakis G B. Defending graph convolutional networks against adversarial attacks//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Barcelona, Spain, 2019; 1399-1407
- [18] Xu K, Chen H, Liu S, et al. Topology attack and defense for graph neural networks: An optimization perspective//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, China, 2019; 3961-3967
- [19] Huang L W, Li D Y, Ma Y T, et al. A meta path-based link prediction model for heterogeneous information networks. *Chinese Journal of Computers*, 2014, 37(4): 848-858
(黄立威, 李德毅, 马于涛, 郑思仪, 张海粟, 付鹰. 一种基于元路径的异质信息网络链路预测模型. *计算机学报*, 2014, 37(4): 848-858)
- [20] Zhou L H, Wang J L, Wang H M, Kong B. Heterogeneous information network representation learning: A survey. *Chinese Journal of Computers*, 2022, 45(01): 160-189
(周丽华, 王家龙, 王丽珍, 陈红梅, 孔兵. 异质信息网络表征学习综述. *计算机学报*, 2022, 45(01): 160-189)
- [21] Kip F T N, Welling M. Semi-supervised classification with graph convolutional networks//Proceedings of the 5th International Conference on Learning Representations. Toulon, France, 2017; 1-14
- [22] Sam P, Yee Whye Teh. Stochastic gradient Riemannian langevin dynamics on the probability simplex//Proceedings of the Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems. Lake Tahoe, USA, 2013; 3102-3110
- [23] Li W, Ahn S, Welling M. Scalable MCMC for mixed membership stochastic blockmodels//Proceedings of the 19th International Conference on Artificial Intelligence and Statistics. Cadiz, Spain, 2016; 723-731
- [24] Sun S, Chen C, Carin L. Learning structured weight uncertainty in Bayesian neural networks//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, USA, 2017; 1283-1292
- [25] Christos Louizos and Max Welling. Multiplicative normalizing flows for variational bayesian neural networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia, 2017; 2218-2227
- [26] Radford M, Neal. Bayesian learning via stochastic dynamics//Proceedings of the Advances in Neural Information Processing Systems. Denver, USA, 1992; 475-482
- [27] Anoop K B, Vivek R, Kevin P Murphy, Max Welling. Bayesian dark knowledge//Proceedings of the Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems. Montreal, Canada, 2015; 3438-3446
- [28] Li C, Chen C, Carlson D, et al. Preconditioned stochastic gradient langevin dynamics for deep neural networks//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Phoenix, USA, 2016; 1788-1794
- [29] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online learning of social representations//Proceedings of the The 20th ACM-SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA, 2014; 701-710
- [30] Velickovi P, Cucurull G, Casanova A, et al. Graph attention networks//Proceedings of the International Conference on Learning Representations. Vancouver, Canada, 2018; 1-12
- [31] Park C, Kim D, Han J, et al. Unsupervised attributed multiplex network embedding//Proceedings of the Thirty-fourth AAAI Conference On Artificial Intelligence. New York, USA, 2020; 5371-5378
- [32] Zhang Y D, Wang X, Shi C, et al. Lorentzian graph convolutional networks//Proceedings of the WWW '21: The Web Conference 2021. Ljubljana, Slovenia, 2021; 1249-1261
- [33] Diederik P Kingma, Jimmy Ba. Adam: A method for stochastic optimization//Proceedings of the 3rd International

Conference on Learning Representations. San Diego, USA, 2015:1-15

- [34] Van Rijsbergen, C. Information retrieval: theory and practice//Proceedings of the Joint IBM/University of Newcastle upon Tyne Seminar on Data Base Systems. Callaghan, Aus-

tralia,1979:1-14

- [35] Daniel Zügner, Akbarnejad A, Stephan Günnemann. Adversarial Attacks on Neural Networks for Graph Data//Proceedings of the 24th International Conference on Knowledge Discovery Data Mining. London, UK, 2018:2847-2856



CHEN Guan-Heng, M. S. candidate. His research interests include graph neural network, heterogeneous network computing.

GUO Zi-Yu, Ph. D. candidate. Her research interests include graph neural network, spatio-temporal network computing.

puting.

MEI Guang-Xu, Ph. D. candidate. His research interests include graph neural network, heterogeneous network computing.

LIU Shi-Jun, Ph. D. , professor. His research interests include services computing, cooperative computing, enterprise relationship network.

PAN Li, Ph. D. , associate professor. Her research interests include service computing, cloud computing

Background

Relation uncertainty is an important problem in the field of the graph. Many practical problems can be tackled by abstracting as a graph structure, but the constructed graph may contain uncertainty for some reasons in this process. However, a heterogeneous graph containing complex structure and semantic information lacks corresponding research while existing graph neural networks that are related to the problem of graph uncertainty are mostly homogeneous graph models. Furthermore, for the problem of relationship uncertainty in a heterogeneous graph, there are few works of literature that propose solutions before. The weak robustness of the model is a common problem in neural networks, thereby many methods have been proposed to improve the robustness of the model, but only a few methods are suitable for a heterogeneous graph.

In this paper, we propose BHNN to solve the problem of relationship uncertainty in heterogeneous graphs and to provide a novel method for enhancing the robustness of BHNN. We construct multiple meta path based neighbor graphs from a heterogeneous graph by predefining meta paths and use a MMSBM to reconstruct and generalize the meta path based neighbor graphs. In this way, we tackle the problem of relationship uncertainty and improve the robustness of BHNN. To predict node labels, we build a posterior inference formula of node labels based on the Bayesian method, and finally, we obtain the probability distribution of node labels through formula derivation. Our experimental results verify the effectiveness and advancement of our method.

This work is supported by the National Natural Science Foundation of China under Grant No. 61872222.