

基于超图卷积的异质网络半监督节点分类

吴越^{1),2)} 王英^{1),2),3),4)} 王鑫^{2),4),5)} 徐正祥^{1),2)} 李丽娜^{1),2),3)}

¹⁾(吉林大学计算机科学与技术学院, 长春 130012)

²⁾(符号计算与知识工程教育部重点实验室(吉林大学), 长春 130012)

³⁾(吉林大学软件学院, 长春 130012)

⁴⁾(吉林大学人工智能学院, 长春 130012)

⁵⁾(长春工程学院计算机技术与工程学院, 长春 130012)

摘要 近几年, 图神经网络(Graph Neural Network)由于能够较好地提取网络结构信息以获得网络表示, 逐渐成为网络节点分类的主流算法。然而, 与广泛研究的同质信息网络相比, 真实世界中网络往往是由不同类型的对象通过复杂关系相互连接所构成的异质信息网络。异质信息网络包含复杂的结构信息和丰富的语义信息, 这也给网络节点分类提供了新的机遇与挑战。在异质信息网络中, 网络模体(Motif)能够用于理解和探索复杂网络, 其既能描述复杂的语义信息, 又能保存网络中高阶近邻结构信息。因此, 提出基于网络模体的异质超图卷积网络模型 MHGCN (Motif-based HyperGraph Convolutional Network)。首先, 将重复出现的高阶网络模体建模为多个相关节点所构成的超边(hyperedge), 进而将整个异质信息网络转换成由不同超边构成的超图, 以克服同质网络中只能描述节点之间(pair-wise)关系的缺点; 然后, 利用超图的基本性质和谱理论设计超图上的卷积操作, 同时引入超边自注意力机制聚合超图内部不同类型的节点, 并通过在超图网络中加入自环解决在模型的前向传播过程中对异质信息网络覆盖不足的问题; 最后, 通过注意力机制对于不同语义的超图表示进行聚合, 从而使最终的节点表示可以有效保持高阶近邻关系和复杂的语义信息。由于 MHGCN 是端到端的, 最终模型直接学习得到节点的分类标签, 并通过半监督节点分类任务进行验证, 与其它方法相比, MHGCN 在 DBLP-P、DBLP-A 数据集上比最好的基准方法 micro-F1 提高了 0.56%-3.51%, macro-F1 提高了 0.54%-4.37%, 验证了 MHGCN 模型的有效性。

关键词 异质信息网络; 网络模体; 超图; 网络表示学习; 节点分类;

中图法分类号 TP391

Motif-based Hypergraph Convolution Network for Semi-supervised Node Classification on Heterogeneous Graph

WU Yue^{1),2)} WANG Ying^{1),2),3),4)} WANG Xin^{2),4),5)} XU Zheng-Xiang^{1),2)} LI Li-Na^{1),2),3)}

¹⁾(College of Computer Science and Technology, Jilin University, Changchun 130012)

²⁾(Key Laboratory of Symbol Computation and Knowledge Engineering(Jilin University), Ministry of Education, Changchun 130012)

³⁾(College of Software, Jilin University, Changchun 130012)

⁴⁾(School of Artificial Intelligence, Jilin University, Changchun 130012)

⁵⁾(College of Computer Technology and Engineering, Changchun Institute of Technology, Changchun 130012)

Abstract GNN(Graph neural network)-based representation learning methods have been a subject of intense research activities on network representation learning in recent years because of its significant contribution to the extraction of network structure information. Compared with the widely studied homogeneous information network, however, the networks in the real world are often HIN (heterogeneous information networks) composed of different types of objects interconnected by complex relationships. The complex structural information and rich semantic information bring great challenges for network representation learning for HIN. Exploiting the higher-order structural patterns is an effective way of learning the representation for the complex

本课题得到国家自然科学基金项目(61872161,61976103)、吉林省科技发展计划项目(2018101328JC,20200201297JC)、吉林省科技厅优秀青年人才基金项目(20170520059JH)、吉林省发改委项目(2019C053-8)、吉林省教育厅科研项目(JJKH20191257KJ)。吴越, 硕士研究生, 中国计算机学会(CCF) 会员 (B2497G), 主要研究领域为深度学习、数据挖掘、图卷积神经网络、异质信息网络分析与挖掘.E-mail: 13136756616@163.com。王英, 博士, 副教授, 博士生导师, 中国计算机学会(CCF)高级会员, 主要研究领域为机器学习、社交网络、数据挖掘.E-mail:wangying2010@jlu.edu.cn。王鑫(通信作者), 博士, 副教授, 中国计算机学会(CCF) 高级会员, 主要研究领域为机器学习、信息检索、社会计算.E-mail: xinwangjlu@gmail.com。徐正祥, 硕士研究生, 中国计算机学会(CCF) 会员, 主要研究领域为机器学习、数据挖掘、图分类.E-mail: xiang1113@foxmail.com。李丽娜, 博士, 讲师, 主要研究领域为数据挖掘和社交网络分析.E-mail:llnccs@jlu.edu.cn

network, the traditional GCNs are not directly applied to the higher-order heterogeneous network. The motifs have been recognized the most common way for understanding and exploring the complex networks, which can both describe complex semantic information and preserve high-order neighbor structures in the network. We leverage the high-order network patterns in the form of motifs to transform the heterogeneous network into multiple hypergraphs, and then model multiple hypergraphs to learn node representations in a unified way. Compared with the traditional heterogeneous network, the advantages of heterogeneous hypergraph network are that the heterogeneous properties can be fully expressed by introducing high-order structural information and fusing the different types of semantic hyperedges based on various motifs. Therefore, in order to overcome the deficiency that homogeneous network can only describe the pairwise relationship, we proposed a heterogeneous MHGCN (Motif-based HyperGraph Convolutional Network) to model the recurring high-order network patterns as hyperedges formed by multiple related nodes, and transform the entire heterogeneous information network into multiple hypergraphs that composed of different hyperedges. In a homogeneous network, the GCN relies on localized first-order approximations of spectral graph convolutions. However, for a heterogeneous network, the convolutional operation on the motif-based hypergraph should be redefined for preserving all the relevant node features. As convolution operations on hypergraph can be conducted using the basic properties of hypergraph and spectral graph theory, to overcome the issue that the motif-based sampling strategy can not cover all nodes, we integrate multiple hypergraph networks and the original network structural information into hypergraph convolutional network to ensure that all nodes can be aggregated and updated in the process of propagating node features. Inspired by the attention mechanism, we exploit the hypergraph attention network to dynamically aggregate node features based on their importance and semantic roles. Specifically, we conduct two levels of attention networks with a hierarchical structure, namely a hyperedge-level attention to learn the importance among different types of nodes, while a comprehensive semantic-level attention to capture the importance of different types of motif structures. Owing to the proposed MHGCN is in an end to end fashion, it can learn to get the labels of the nodes eventually. The proposed MHGCN improves the micro-F1 by 0.56%-3.51%, improves the macro-F1 by 0.54%-4.37% on two real-world datasets of DBLP-P and DBLP-A in semi-supervised node classification task against the state-of-the-art methods, which also verifies the effectiveness of aggregating different types of nodes based on hyperedge-level attention and comprehensive semantic-level attention.

Key words heterogeneous information network; network motif; hypergraph; network representation learning; node classification;

1 引言

图神经网络近年来引发了大量的关注,其处理非欧几里得(Non-Euclidean)数据的表示学习能力受到了广泛的认可,其中以GCN^[1], GraphSAGE^[2], GAT^[3]为代表的网络表示学习方法最为突出。这些方法大部分都只聚焦于只有一类节点和边的同质网络,然而,现实世界中大多数网络都含有多种类型的节点和边,也就是广义上的异质信息网络,例如,在MovieLens数据集中有导演、电影、演员等不同类型的节点,这些节点间的关系包括执导、参演等。因此,异质信息网络的复杂性在于其节点和边类型的多样性,以及由此产生的复杂结构和丰富语义。

网络模体^[4](motifs)是复杂网络中的块结构,能够很好地表示网络的局域特性。异质信息网络中的网络模体作为重复出现的子图,连接了不同类型节点,例如,在DBLP中的模体“作者-论文-作者”(APA)代表了两个作者发表同一篇论文的关系。目前,一些异质信息网络表示学习的工作是基于元路径^[5](metapath),然而元路径和网络模体不同,前者是一条路径而后者是网络中的一个子图,子图能够更好地保持网络的结构特征和语义信息。因

此,在异质信息网络节点分类任务中使用网络模体能够:(1)保持网络的异质性,使不同类型节点之间的关系加以区分,而这些复杂的关系在学习异质信息网络表示的过程中是至关重要的;(2)有效探索网络中的高阶邻接关系,并在模体内区分出不同的语义角色,从而得到模体内部节点更完整的语义表示。

另一方面,异质信息网络中节点之间的相互关系复杂,比如附属关系(affiliation relation)和高阶交互关系(interaction relation),都远远超过了同质网络中节点间的点对点关系。在社交网络中,用户大多来自不同的年龄段、地区和群组,若将某一时刻网络中用户之间的联系看作是图中的边,那么可以将所有同一年龄段的用户与某一节点之间的关系考虑成一条边,显然这条边不仅仅只连接了两个节点,C.Berge等人^[6]将这种图上所有节点的非空子集建模为超边,通过超边相互连接的图称为超图。与在传统的异质信息网络上进行表示学习相比,异质超图网络的优势主要体现在网络中异质属性能够通过引入高阶结构信息和融合不同类型网络模体所构成的超边形式来表达。因此,在网络中引入高阶关系结构信息,有助于从超图理论的角度出发研究异质信息网络,这就引出了本文所研究的主要问

题：将异质信息网络建模成超图网络后，利用超图基本性质和谱理论将图神经网络推广到异质超图网络上，从而实现异质信息网络的节点分类。与此同时，面向异质超图网络表示学习的节点分类方法也提出了一系列全新的问题和挑战：

(1) 借助网络模体将异质信息网络建模为异质超图网络，进而保留异质信息网络中高阶近邻网络信息和复杂语义信息，并通过超图卷积网络传播节点特征以学习网络节点表示，为异质信息网络表示学习提供了新视角。

(2) 在模型传播节点特征过程中，为了克服超图网络无法覆盖异质信息网络中所有节点的不足问题，在超图卷积神经网络中引入自环方式融合原始网络结构信息，以保证整个网络都参与节点信息聚合过程。

(3) 提出适用于异质超图网络上的注意力机制，能够动态地依据不同的重要性聚合不同语义角色的节点，使得 MHGCN 即使在较为稀疏的大规模异质信息网络中也能获得较好的节点表示。同时，引入注意力机制为不同语义角色的节点动态分配权重，也为节点的信息聚合过程提供有意义的解释，有助于异质信息网络分析的决策过程变得更加透明。

(4) 在两个异质信息网络数据集上针对 MHGCN 模型实现半监督节点分类进行评估，同现有的基准方法进行比较，从实验结果验证了 MHGCN 模型的有效性。

本文在第 2 节介绍相关性研究工作；第 3 节对本文中所涉及概念给出定义并对研究问题进行形式化描述；第 4 节阐述了基于网络模体的异质超图卷积网络模型；第 5 节为实验数据集、实验设置及结果分析；第 6 节为结论和未来展望。

2 相关工作

近年来，对网络表示学习的研究取得了丰富的成果，其中常见的任务就是对网络中的节点进行划分，通常是通过网络表示学习算法将网络中的节点编码到低维的向量空间，并基于学到的节点特征构建分类器，从而对节点分类。根据网络中包含的节点或边的类型是否相同，可以将网络划分为同质信息网络和异质信息网络。先前的网络表示学习主要是基于谱图理论的降维过程，如等距特征映射、局部线性嵌入等^[7]。随着神经网络和深度学习的兴起，大量的工作致力于同质网络和异质网络上的表示问题，其中根据模型训练方式不同主要分为基于浅层神经网络的方法和基于深度神经网络的方法。

2.1 基于浅层神经网络的节点分类

在同质网络中，Perozzi 等人^[8]提出基于 SkipGram 模型学习节点的拓扑特征，将网络中的节

点看作自然语言模型中的词，将节点序列看作句子，通过在网络上的随机游走产生节点序列并将序列传入 SkipGram 模型中由此获得节点表示。Grover 等人^[9]考虑了两种策略来改进产生节点序列的过程，一种是基于广度优先搜索策略获得节点周围的局部邻接关系，另一种是基于深度优先搜索策略在更大的范围内探索节点的高阶近邻关系，从而在学习到节点表示同时，捕获节点的局部和高阶上下文信息。Tang 等人^[10]重新定义节点间的一阶相似度和二阶相似度，通过对边进行采样更新模型参数，最终在将图中节点嵌入低维空间的同时保留其一阶相似度和二阶相似度。

在异质网络中，Dong 等人^[11]提出通过基于元路径的随机游走构建节点的异质邻居，同时规定节点每次游走的转移概率，将不同类型的节点映射到不同的向量空间，最终通过 SkipGram 模型产生节点表示。Shang 等人^[12]提出的 ESIM 模型考虑了节点间的不同关系，基于元路径进行随机游走，通过将元路径实例的概率最大化来学习出现在该实例中节点的向量表示。Fu 等人^[13]提出将不同类型的节点与节点之间的关系看作为一种二分类问题，在学习网络中节点表示的同时还学到了元路径（关系）的表示。

2.2 基于深度神经网络的节点分类

深度神经网络已经广泛应用于计算机视觉和自然语言处理等领域中，越来越多的研究者开始关注作用于网络结构上的图卷积神经网络。在同质网络上，Bruna 等人^[14]通过引入图的傅里叶变换来定义基于谱图理论的图卷积。Defferrard 等人^[15]在此基础上利用特征值对角矩阵的切比雪夫多项式在谱域中近似的平滑卷积核。Kipf 等人^[1]提出一种卷积神经网络的有效变体用于图结构数据的半监督学习，作为频谱图卷积的局部一阶近似。Hamilton 等人^[2]利用神经网络来汇总邻居的特征信息，通过采样策略使得计算不需在整个图上完成。Velickovic 等人^[3]使用节点自注意力机制来衡量不同邻居的影响，并结合它们的影响以获得节点表示。同时，为了在学习节点表示时融合高阶的结构信息，近年来超图表示学习受到了关注。Feng 等人^[16]提出超图表示学习框架，通过设计超边卷积操作捕获高阶数据之间的关联性。Yadati 等人^[17]基于超图的谱理论提出在超图上训练 GCN 的方法，使相同边上的节点具有相似的表示，用于半监督学习和探索超图上的组合优化问题。Bai 等人^[18]在定义超图卷积的同时，进一步使用注意力机制衡量超边内节点之间的相互影响以获得更好的节点表示。上述这些工作都是只聚焦于同质网络，构建超图的方式主要是通过将节点的一阶近邻或者基于 KNN 算法的 K 个近邻节点建模为超边。

针对异质网络，Sankar 等人^[19]将 CNN 的诸多

特性推广到图结构上,利用网络模体来对节点周围的感受野进行建模以实现模型的局部连通性和平移不变性,同时基于空间域的方法从邻域聚合节点的特征信息并区分其语义角色。Wang 等人^[20]提出一种基于层次注意力的新型异构图神经网络,包括节点级别的注意力和语义级别的注意力,使用节点级注意力学习节点和其异质邻居间的权重,再使用语义级别注意力聚合多条路径之间的权重。此外,其它深度神经网络上的方法也被扩展到图上,Wang 等人^[21]通过在同质网络上采用自编码器的方式抽取节点的非线性特征,提出能够同时描述局部结构和全局结构的目标函数,并使用半监督的方式来拟合优化,最终将自编码器的中间层作为得到的节点表示。Hu 等人^[22]基于生成式对抗网络的思想,提出针对异质信息网络的网络表示学习框架,通过设计生成器和判别器来区分不同关系所连接的节点,从而学习节点分布生成更具鲁棒性的节点表示。

以上的方法中,无监督图编码方式主要是在图上通过学习节点的社区结构来获得节点表示,再通过传统的分类器如 MLP 等来对节点进行分类,但是由于这些方法在训练过程中忽略了节点的标签,缺乏标签信息的监督,学习得到的特征未包含节点类别信息,所以不能取得很好的分类效果。通过引入少量的标注信息,可以有针对性地提高节点表示在后续分类任务中的准确性。本文在上述工作的基础上,采用基于网络模体构建超边的方式,提出了异质网络上的超图卷积模型,在针对异质信息网络的半监督节点分类任务上,通过结合图结构信息、节点特征以及部分节点标签信息,取得更好的分类效果。

3 相关定义和问题描述

在本节中,首先介绍异质信息网络和超图的一些相关定义,然后给出针对异质超图网络中节点分类问题的形式化描述。

定义 1. 异质信息网络. 信息网络通常被定义为一个图 $G = (V, E)$, 其中 V 是所有节点的集合, E 是所有边的集合。在图上存在一个节点类型映射函数 $\tau()$ 和边类型映射函数 $\phi()$, 对于每一个节点 $v \in V$ 都有 $\tau(v) \in O$, O 是所有节点类型的集合, 同时, 对于每一条边 $e \in E$ 都有 $\phi(e) \in R$, R 是所有边类型的集合。如果在信息网络中, 每个节点都有特定的节点类型, 或者每条边都有特定的边类型, 即 $|O| > 1$ 或 $|R| > 1$, 那么这个网络称为异质信息网络, 否则为同质信息网络。

定义 2. 网络概要. 网络概要是异质信息网络中的元结构, 为异质信息网络中多种类型节点和边以及节点之间的不同关系提供了描述。图 1 给出了

两个典型的异质信息网络中的网络概要, 其中, 图 1(a)是在引文网络中存在的结构, 描述了一个作者和两篇论文之间的关系, 图 1(b)是在社交网络上的一个实例, 针对某一用户发布微博这一行为, 该网络概要可以体现用户对其他用户的关注情况, 发布微博的时间、地点以及微博内容中所包含的情绪等信息。典型的网络概要形式有元路径、网络模体等。

定义 3. 网络模体. 网络模体(Motif)是在复杂网络中频繁出现的一种网络模式, 本质上是子图, 只是该子图结构在真实网络中出现的频率要显著高于其在随机网络中出现的频率。将子图定义为模体 M 在 G 上的实例 $S, S = (V_s, E_s)$, 显然 $V_s \subseteq V$ 并且 $E_s \subseteq E$ 。

在异质信息网络中, 不同于元路径, 网络模体作为子图更能有效地保持网络的结构信息和语义信息。作为引文网络中具有代表性的元路径, 以作者-论文-会议 (A-P-V) 为例, 同模体作者 1-(作者 2)-论文-会议 (A-(A)-P-V) 对比, 前者只能表示一个作者 (A) 在一个会议 (V) 上发表了论文 (P), 路径中仅包含写作关系 (A-P) 和出版关系 (P-V), 而后者除此之外还可以表达两个作者在某一会议发表同一篇论文的协作关系, 以及两个作者在论文写作关系中承担的语义角色。

定义 4. 超图. 超图(Hypergraph)是一种广义上的图, 其边可以和任意数量的顶点连接。在形式化上, V 代表图中所有的节点构成的集合, E_h 是 V 的非空子集称为超边(Hyperedge), 则二元关系 $H = (V, E_h)$ 表示为超图 (对于一个有 N 个节点和 K 条超边的超图来说, 有 $|V| = N$ 且 $|E_h| = K$)。若超图 H 中各超边关联的节点数相同, 则称 H 为一致超图(Uniform Hypergraphs)。

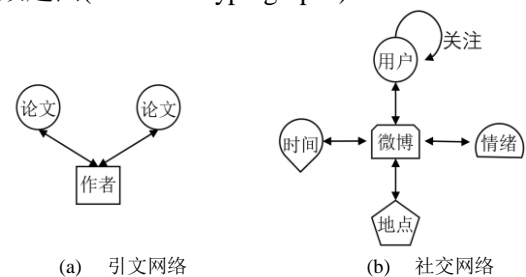


图 1 异质信息网络中的网络概要

目前主要有两种定义超边的方式, 分别是基于属性构建方式和基于邻居构建方式。在基于属性方式构建的超边中, 关联的节点属性都是相同的, 属性可以是节点的特征。然而, 在基于邻居方式构建的超边中, 关联的节点是通过定义节点相似度的阈值, 然后使用 K 近邻算法构建超边。本文将特定类型的网络模体看作是一条超边, 构建的超边大小一致, 因此, 本文基于一致超图进行研究。显然在同

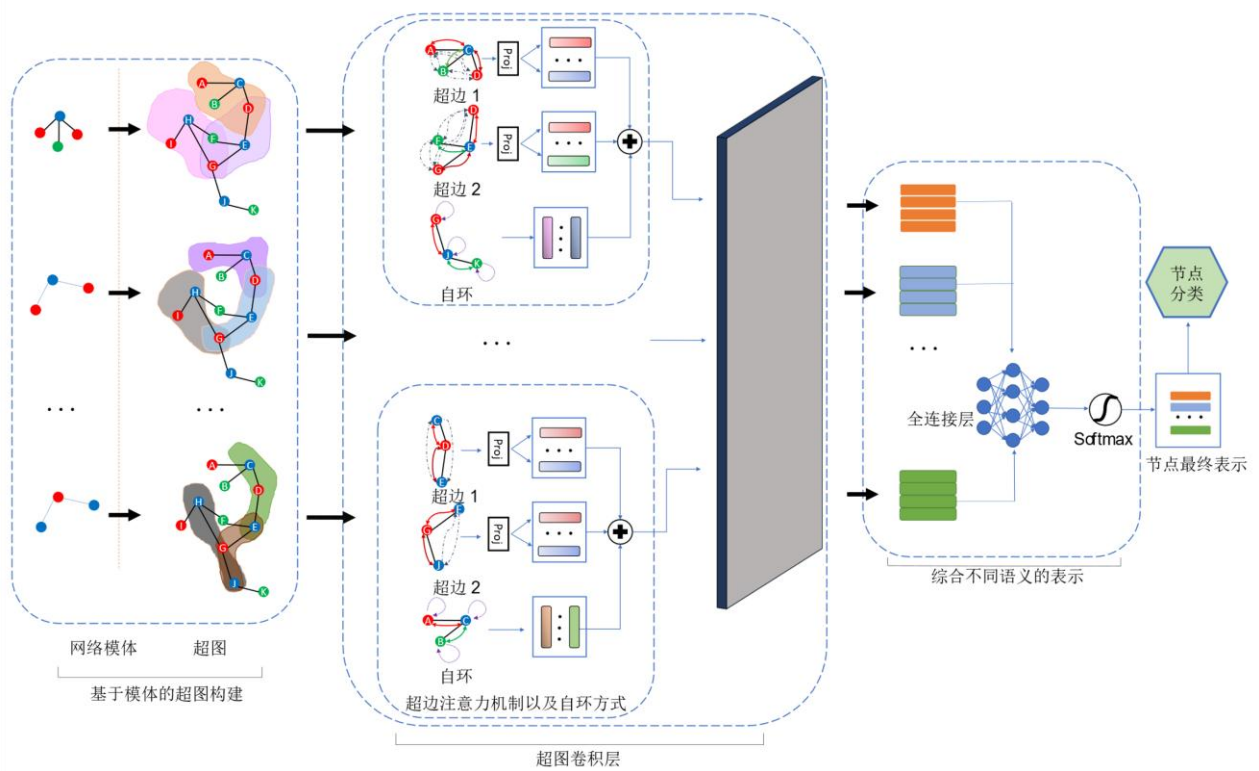


图2 基于网络模体的超图卷积网络模型

一条超边内部的各个节点在原始的空间结构中相互邻近，有着更为相关的语义，因此，这些节点在投影空间中也应有相似的表示，进而捕获异质信息网络复杂的高阶关系。

问题描述. 给定一个异质信息网络 $G=(V, E)$ ，对于多种类型的节点集合 V 和边集合 E （为了简练，本文未考虑边的类型），它们之间的不同语义关系构成了网络模体集合 M 。通过采样 M 在 G 上的实例 S 构建多重超图 H 。设计超图卷积模型 F 分别学习各个节点的表示并在最后使用注意力机制融合，最终得到每个节点的 d 维表示 $Z \in \mathbb{R}^{|V| \times d}$ ， Z 保持了异质网络中的高阶邻接关系和丰富的语义结构信息，并用于半监督分类节点任务。

4 MHGCN 模型

本文提出的用于异质信息网络节点分类的超图卷积模型 MHGCN 包括三个主要模块，首先，基于网络模体进行超图构建，将异质信息网络构建为多重超图；其次，基于超图谱理论设计异质超图网络上的图卷积操作，同时在超图卷积过程中引入超边注意力机制以聚合不同语义角色的节点，并加入自环保证整个网络都参与节点信息聚合过程；最后，通过注意力机制将基于多重超图得到的不同语义的节点表示进行融合，得到最终综合语义的表示。MHGCN 模型框架如图 2 所示。

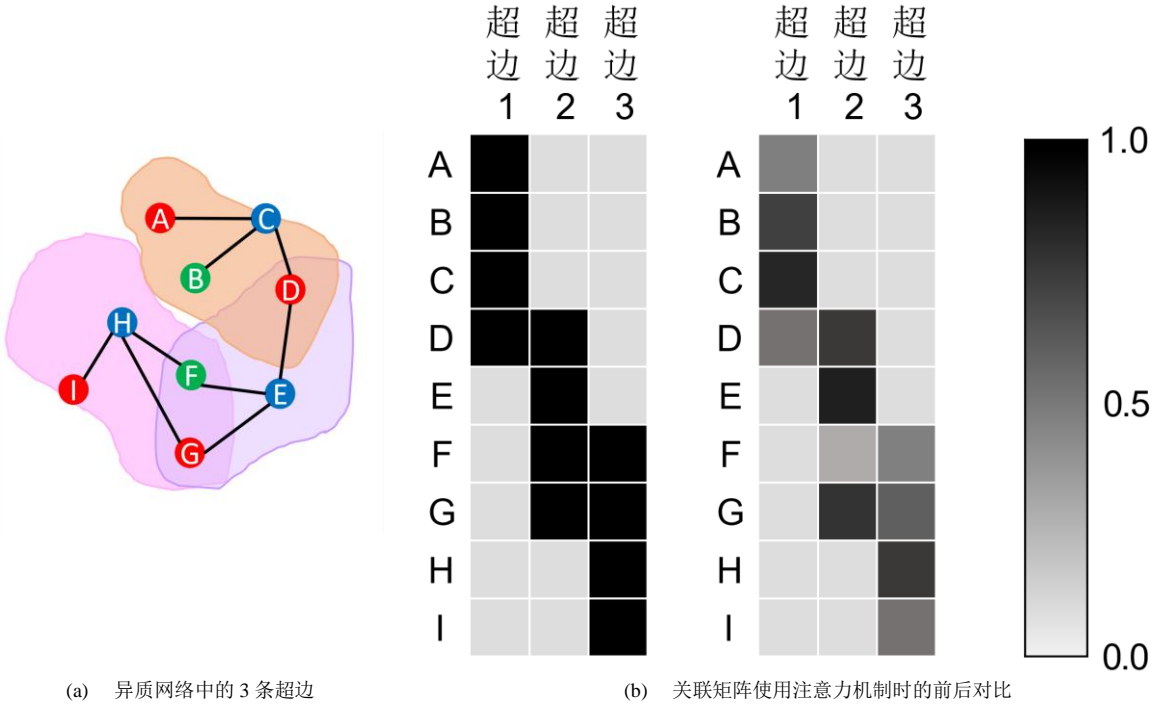
4.1 基于网络模体的超图构建

对于一个给定的异质信息网络 G 和网络模体集合 $M = \{M_1, M_2, \dots, M_T\}$ ，如图 4 所示，给出了在引文网络中定义的模体结构。首先，使用文献^[19]中的网络模体检测算法匹配 M 在 G 上的所有实例 S_M ，若一种模体结构在 G 中出现频率较高，可基于随机采样的策略来缩减其实例的规模。

其次，确定网络模体实例 S_M 。对于一个网络模体 $M_i \in M$ 构建基于 M_i 的超图 $H_{M_i} = (V_{M_i}, S_{M_i})$ ，其中 S_{M_i} 是 M_i 在 G 上的全部实例（假设匹配实例时未经过随机采样）， V_{M_i} 是 S_{M_i} 覆盖的节点集合。由于异质网络含有丰富的语义信息，而单一网络模体只包含一种语义，需要考虑多种模体结构以获得更高质量的网络表示。因此，将异质信息网络 G 转换为多重超图 $\{H_{M_1}, H_{M_2}, \dots, H_{M_T}\}$ 。每个超图 H_{M_i} 可以表示为一个二元关联矩阵 $H_{M_i} \in \mathbb{R}^{N \times |S_{M_i}|}$ ， N 为异质网络 G 中的所有节点数， $|S_{M_i}|$ 为超边个数， H_{M_i} 中的每个元素定义为：

$$h(v, e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{if } v \notin e \end{cases} \quad (1)$$

对于一个节点 v ，当任意超边 $e \in S_{M_i}$ 同节点 v 存在关联时， $h(v, e) = 1$ ，否则为 0。针对每条超边 $e \in S_{M_i}$ 都分配权重 w_e ，所有的权重通过对角矩阵



(a) 异质网络中的3条超边

(b) 关联矩阵使用注意力机制时的前后对比

图3 在超边中使用注意力机制

$\mathbf{W}_{M_i} \in \square^{|S_{M_i}| \times |S_{M_i}|}$ 存储, 本文未考虑不同超边之间的相互影响, 因此, 这里 \mathbf{W}_{M_i} 初始化为单位矩阵, 所有的超边权重相同。同时, 节点 v 的度 $d(v)$ 和超边 e 的度 $\delta(e)$ 分别用对角矩阵 $\mathbf{D}_v \in \square^{N \times N}$ 和 $\mathbf{D}_e \in \square^{|S_{M_i}| \times |S_{M_i}|}$ 来存储。那么任意节点的度定义为:

$$d(v) = \sum_{e \in S_M} w_e \times h(v, e) \quad (2)$$

任意超边 e 所关联的节点数量即超边的度, 定义为:

$$\delta(e) = \sum_{v \in V_M} h(v, e) \quad (3)$$

对于给定的网络模体集合 \mathbf{M} , 构建 T 个不同的基于网络模体的邻接矩阵 $\mathbf{A} = \{\mathbf{A}_{M_1}, \mathbf{A}_{M_2}, \dots, \mathbf{A}_{M_T}\}$ 对应多个超图。对于任意超图 \mathbf{H}_{M_i} 其对应邻接矩阵 \mathbf{A}_{M_i} 定义为:

$$\mathbf{A}_{M_i} = \mathbf{H}_{M_i} \mathbf{W}_{M_i} \mathbf{H}_{M_i}^T \quad (4)$$

4.2 超图卷积网络

4.2.1 超图上的谱卷积

传统的图卷积网络^[11]通过图的拉普拉斯矩阵的特征值和特征向量来研究图的性质, 拉普拉斯矩阵是离散的拉普拉斯算子, 针对一个定义在图节点上的函数, 拉普拉斯算子可以度量该函数的平滑程度是如何相对于图结构而变化的。对于任意超图 \mathbf{H}_{M_i} , 其对应的超图拉普拉斯矩阵^[23] \mathbf{L} 可以定义为:

$$\mathbf{L} = \mathbf{I} - \mathbf{D}_v^{-1/2} \mathbf{H}_{M_i} \mathbf{W}_{M_i} \mathbf{D}_e^{-1} \mathbf{H}_{M_i}^T \mathbf{D}_v^{-1/2} \quad (5)$$

其中, \mathbf{I} 为单位矩阵, 而 \mathbf{L} 则是一个半正定的对称矩阵, 因此经过谱分解有 $\mathbf{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$, \mathbf{U} 中的列向

量为正交的单位特征向量, $\mathbf{\Lambda}$ 中包含对应的非负特征值。基于图信号处理的思路^[14], 超图上的傅里叶变换可以定义为 $\hat{f} = \mathbf{U}^T f$, 傅里叶逆变换定义为 $f = \mathbf{U} \hat{f}$ 。将 \mathbf{U} 看作一组傅里叶基, 而 $\mathbf{\Lambda}$ 为频率, f 为超图上的信号。则 f 和卷积核 g_θ 的卷积操作定义如下:

$$f \hat{\ast} g_\theta = \mathbf{U} \left((\mathbf{U}^T g_\theta) \square (\mathbf{U}^T f) \right) \quad (6)$$

由于公式6描述的超图卷积涉及到矩阵分解和多次矩阵乘法运算, 计算代价较高。因此, 本文使用 K 阶切比雪夫多项式^[15]来近似卷积核 g_θ 展开求解, 这样超图卷积过程可以描述为:

$$f \hat{\ast} g_\theta \approx \sum_{k=0}^K \theta_k T_k(\mathbf{L}) f \quad (7)$$

切比雪夫多项式定义为 $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$, 其中 $T_0(x) = 1$, $T_1(x) = x$ 。 $\theta_k \in \square^K$ 是切比雪夫多项式系数, $T_k(\mathbf{L})$ 是缩放的拉普拉斯矩阵 \mathbf{L} 的 K 阶切比雪夫多项式, 其中 $\mathbf{L} = 2/\lambda_{\max} \times \mathbf{L} - \mathbf{I}$, λ_{\max} 是超图拉普拉斯矩阵 \mathbf{L} 的最大特征值。

公式(7)描述的超图卷积过程相当于使用 g_θ 提取了节点周围的 K 阶近邻信息, 为了减少模型的计算规模, 可以直接使用其一阶近似。当 $K=1$ 时, 超图卷积可以被简化为:

$$f \hat{\ast} g_\theta \approx \theta_0 f + \theta_1 \left(\frac{2}{\lambda_{\max}} \mathbf{L} - \mathbf{I} \right) f \quad (8)$$

此处存在两个卷积核参数 θ_0 和 θ_1 , 为了进一步降低模型在卷积过程中的复杂度并防止过拟合问题, 令

$$\theta_0 = -\frac{1}{2}\theta \left(\frac{2}{\lambda_{\max}}L - \mathbf{I} \right), \theta_1 = -\frac{1}{2}\theta, \text{同时令 } \lambda_{\max} = 2$$

对拉普拉斯矩阵特征值缩放^[1], 则超图卷积过程可表示为:

$$f \hat{\mathbf{a}} g_{\theta} \approx \theta \mathbf{D}_v^{-1/2} \mathbf{H}_{M_t} \mathbf{W}_{M_t} \mathbf{D}_e^{-1} \mathbf{H}_{M_t} \mathbf{D}_v^{-1/2} f \quad (9)$$

改写上式为矩阵形式, 则超图卷积的逐层传播规则为:

$$\mathbf{X}^{(l+1)} = \sigma(\mathbf{D}_v^{-1/2} \mathbf{H}_{M_t} \mathbf{W}_{M_t} \mathbf{D}_e^{-1} \mathbf{H}_{M_t} \mathbf{D}_v^{-1/2} \mathbf{X}^{(l)} \Theta) \quad (10)$$

其中, $\mathbf{X}^{(l)} \in \mathbb{R}^{N \times C^{(l)}}$ 和 $\mathbf{X}^{(l+1)} \in \mathbb{R}^{N \times C^{(l+1)}}$ 分别为第 l 层和第 $l+1$ 层的超图信号, C 为节点的特征维度。 Θ 作为卷积核其值在训练过程中被不断更新, 从而达到提取超图中节点特征的效果。

4.2.2 超边注意力机制

在本节中, 主要讨论如何度量超边内部不同类型的节点对目标节点表示的影响。对于图结构上的半监督学习, 实际上是把具有标签的节点看作是“信息源”, 图卷积过程就是邻近节点之间发生的信息交换过程, 最终的信息分布在几轮迭代之后达到平衡。在这个过程中, 信息在节点间的转移概率是由邻接矩阵和度矩阵所确定的^[1]。公式(10)表明, 超图卷积也是通过每个节点的转移概率与特征矩阵的乘积构造新的节点表示, 而转移概率同样是由超图的邻接矩阵 \mathbf{A}_{M_t} 和度矩阵 \mathbf{D}_v 、 \mathbf{D}_e 所确定的。

由于该转移概率是固定的, 使得在信息传播过程中每个节点对其邻居节点的影响也不会发生变化, 为此引入超边内部自注意力机制使转移概率变成可以训练的参数, 从而动态地聚合不同类型的节点。

由于无法直接将转移概率参数化, 因此, 本文通过采用注意力机制, 将关联矩阵 \mathbf{H}_{M_t} 中的值变成可以在优化模型过程中学习得到, 以克服不同类型节点和语义角色由于其特征空间不同对目标节点表示所带来的影响, 并间接使模型拥有可训练的转移概率。新的关联矩阵用 $\hat{\mathbf{H}}_{M_t}$ 表示, α_{ij} 对应 $\hat{\mathbf{H}}_{M_t}$ 中的一个元素。超边中的注意力机制如图 3 所示, 图 3(a) 是引文网络上的三个超边实例, 其中蓝色为待分类的论文节点, 红色为作者节点, 绿色为会议节点。对于未使用注意力机制的超边, 不同类型的节点之间连接概率只能为 0 或 1, 显然无法提取不同语义角色的节点信息。而超边注意力机制可以为不同类型的节点动态分配权重, 如图 3(b) 所示, 例如, D 节点同时存在于超边 1 和超边 2 中, 在以论文节点为分类目标节点的情况下, D 节点在超边 2 中有更重要的影响力。

对于第 i 个节点, 我们希望通过基于概率分布模型来得到其所在第 j 条超边中对应的注意力分数 $\alpha_{i,j}$ 。假设 $\alpha_{i,j}$ 满足某一概率分布, 则只有在节点集来自同域时才能对节点特征的相似度进行统一衡

量。因此, 为了使该计算可行, 需要对超边 e^j 内各个节点特征进行累加, 来计算超边的特征 x_{e^j} 。这里考虑了四种累加方式: concat, mean, 1-norm, 2-norm, 其中 concat 是通过拼接 e^j 内各个节点的特征 $\{x_{v^1}, x_{v^2}, \dots, x_{v^m}\}$, m 是超边内的节点个数, 另外三种方式可以通过以下方式计算:

$$x_{e^j} = \begin{cases} \frac{1}{m}(x_{v^1} + x_{v^2} + \dots + x_{v^m}) & \text{mean} \\ \sqrt[p]{|x_{v^1}|^p + |x_{v^2}|^p + \dots + |x_{v^m}|^p} & \text{p-norm} \end{cases} \quad (11)$$

其中, p 等于 1 或 2, x_{v^i} 和 x_{e^j} 分别代表节点和超边特征向量的第 n 列元素。因此, 参考同质网络中计算注意力的方式^[3], 引入超边自注意力机制, 对于第 i 个节点和其所在的第 j 条超边, α_{ij} 的计算公式为:

$$\alpha_{i,j} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}\left[x_{v^i} \mathbf{P}_{M_t} \parallel x_{e^j} \mathbf{P}_{M_t}\right]\right)\right)}{\sum_{k \in \mathbf{N}^j(i)} \exp\left(\text{LeakyReLU}\left(\mathbf{a}\left[x_{v^k} \mathbf{P}_{M_t} \parallel x_{e^j} \mathbf{P}_{M_t}\right]\right)\right)} \quad (12)$$

其中, LeakyReLU 是非线性激活函数。为了计算异质节点之间的相似度, 需要保证不同类型的节点在相同的向量空间, 因此权重矩阵 \mathbf{P}_{M_t} 负责将特征向量 x_{v^i} 和 x_{e^j} 映射到统一隐空间。 \parallel 代表拼接操作, \mathbf{a} 是可训练的注意力向量, 在不同的超图之间共享。 $\mathbf{N}^j(i)$ 是节点 v^i 在超边 e^j 中的邻居集合(包括 v^i)。最终, 通过学习超边内自注意力, 超边 e^j 内不同类型节点之间的相互影响得到更精确的描述。

4.2.3 超图中加入自环

基于单一网络模体构建的超图无法覆盖原始网络中所有的节点和边, 因此部分节点未参与公式(10)中的节点信息传播过程, 并且其表示也不会被其他节点所更新。因此, 为了保证所有的节点信息都能被聚合和更新, 将超图邻接矩阵和原始邻接矩阵+相融合, 并加入自环, 得到最终的超图卷积传播规则:

$$\mathbf{X}^{(l+1)} = \sigma((\mathbf{D}_v^{-1/2} \hat{\mathbf{H}}_{M_t} \mathbf{W}_{M_t} \mathbf{D}_e^{-1} \hat{\mathbf{H}}_{M_t} \mathbf{D}_v^{-1/2} + \mathbf{A}) \mathbf{X}^{(l)} \Theta) \quad (13)$$

其中, $\mathbf{A} = \mathbf{A} + \mathbf{I}$ 是加入自环后的邻接矩阵。与公式(10)中采用的固定转移概率不同, 上述公式融合了可训练的超边内节点转移概率以及全局的节点转移概率, 使得原本就存在于超边中的节点邻居在每一次信息聚合的过程中有更高的权重。因此在超边内部, 任意一个节点的一阶近邻总是要比超边中的其它邻居对该节点具有更高的影响力。同时,

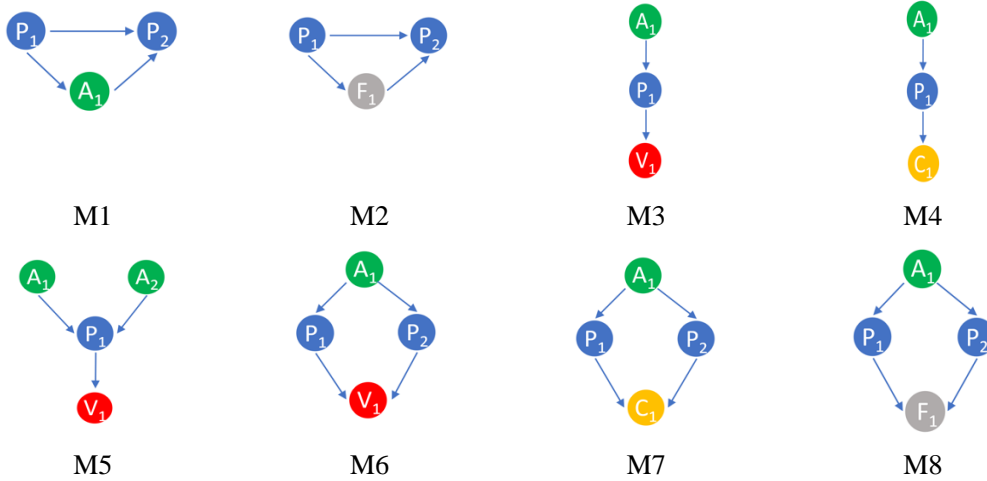


图4 数据集中使用的模体

对于没有被超图覆盖的节点，自环的加入保证了所有的节点信息随着超图卷积的过程得到更新。

因此，在卷积过程中，对于超边中的待分类节点，由于在加入了邻接矩阵以及自环，引入注意力机制可以使得不同类型的节点信息相互传播，图2中用不同颜色的箭头加以区分。如果是两个节点之间原本就存在链接，信息流可以直接沿着链接传播，传播过程在图2中用实心双向箭头表示。对于原本不存在链接但是在同一超边内的，节点信息可以基于超图谱卷积进行传播，图2中用虚线单项箭头表示。对于没有被超边覆盖到的节点，邻接矩阵以及自环的存在能保证其节点信息也能被更新。

4.3 综合语义的超图卷积网络

单一超图 H_{M_i} 的表示可以基于上述的超图卷积网络获得。而对于由网络模体集合 M 构建的不同超图 $\{H_{M_1}, H_{M_2}, \dots, H_{M_T}\}$ ，可以根据公式(13)学习得到特定语义的网络表示 $\{Z_{M_1}, Z_{M_2}, \dots, Z_{M_T}\}$ 。由于不同的超图对生成最终的异质网络表示有不同程度的影响，因此，使用基于注意力机制计算的系数 C_{M_i} 融合不同语义的网络表示：

$$C_{M_i} = \text{Softmax}(\tanh(\mathbf{W} \cdot \mathbf{Z}_{M_i} + \mathbf{b}) \cdot \bar{\mathbf{a}}^T) \quad (14)$$

其中， $\bar{\mathbf{a}}$ 是用于综合语义的注意力向量， \mathbf{W} 和 \mathbf{b} 是全连接层的权重和偏置，全连接层用于把不同语义的节点表示投影到相同的向量空间，再使用 $\text{Softmax}(\cdot)$ 进行对数归一化，得到融合系数 C_{M_i} 。最

后对各个特定语义的网络表示进行融合，从而得到异质网络 G 的表示：

$$\mathbf{Z} = \sum_{i=1}^T C_{M_i} \mathbf{Z}_{M_i} \quad (15)$$

对于多分类任务而言，其中 $\mathbf{Z} \in \mathbb{R}^{N \times q}$ ， q 为节点的类别。通过最小化带标签节点的真实标签和预测标

签的交叉熵，定义半监督节点分类的损失函数：

$$L = - \sum_{i \in V_L} \sum_{j=1}^q Y_{ij} \ln Z_{ij} \quad (16)$$

其中， V_L 是所有带标签节点的集合， Y 为节点的真实标签。在标签信息的指引下，可以基于反向传播和梯度下降训练模型中的参数，最终得到所有节点的表示并预测出其标签。

5 实验

在本节中，将 MHGCN 在两个异质数据集上与先进基准方法进行比较，验证其在半监督节点分类上的有效性。

表1 不同数据集的统计信息

数据集	节点数	链接数	使用的模体种类
DBLP-P	35570	131636	M1、M3、M5、M6、M8
DBLP-A	37791	170794	M1、M2、M4、M7

5.1 数据集

本文使用了两个公开的引文网络数据集如表1所示，对于每个数据集都设计了多个模体用于模型抽取语义，使用的网络模体如图4所示。最后在每个网络上进行半监督节点分类，分类时所用到的初始节点特征是稀疏的词袋特征向量。

DBLP-A 是从 Tang Jie 等人^[24]提出的 DBLP 引文关系网络中抽取的对作者节点分类的数据集。其中有 14475 个作者，14376 篇论文，20 个会议。根据作者提交的会议，将其研究领域划分为：数据库、数据挖掘、机器学习和信息检索，并以此作为作者节点的分类标签。

DBLP-P 和 DBLP-A 类似，不同的是每一篇论文根据 Cora 数据集^[25]的论文类别划分为 10 类，包括信息检索、人工智能、操作系统、数据库等，并以此作为论文节点的分类标签。DBLP-P 包含 17411 个作者、18059 篇论文和 300 个会议。

表 2 不同分类方法的 micro-F1 和 macro-F1

Datasets	Metrics	Training-size	LINE	MetaGNN	GCN	HAN	MHGCN-noAtt	MHGCN
DBLP-P	micro-F1	20%	0.7059	0.7649	0.7252	0.6684	0.7713	0.7815
		40%	0.7297	0.7723	0.7345	0.6495	0.7675	0.7918
		60%	0.7509	0.7765	0.7431	0.6520	0.7809	0.7931
		80%	0.7605	0.7819	0.7697	0.6754	0.7683	0.8017
	macro-F1	20%	0.6602	0.6907	0.6642	0.4527	0.7201	0.7288
		40%	0.6805	0.7171	0.6792	0.4390	0.7309	0.7456
		60%	0.6957	0.7286	0.6826	0.4491	0.7421	0.7449
		80%	0.7109	0.7315	0.7115	0.4601	0.7014	0.7574
DBLP-A	micro-F1	20%	0.8472	0.8532	0.8589	0.8742	0.8921	0.9143
		40%	0.8805	0.8914	0.8927	0.8915	0.9115	0.9258
		60%	0.8903	0.8927	0.9062	0.9003	0.9159	0.9274
		80%	0.8914	0.8938	0.8962	0.9110	0.9145	0.9166
	macro-F1	20%	0.8422	0.8457	0.8496	0.8650	0.8837	0.9135
		40%	0.8752	0.8812	0.8856	0.8830	0.9045	0.9197
		60%	0.8871	0.8896	0.9039	0.8936	0.9022	0.9243
		80%	0.8886	0.8921	0.8940	0.9144	0.8998	0.9121

5.2 基准方法

现有的用于半监督节点分类任务的方法，主要分为浅层的基于游走模型和深层的深度神经网络。对于其中仅能适用于同质网络的方法，将原始的异质信息网络作为输入，在已标注的节点限制下训练其它节点。这里选取五种网络表示方法作为基准方法与 MHGCN 进行性能比较，其中 LINE 是浅层模型，而 MetaGNN、GCN、HAN 是深层模型：

LINE: 该方法^[10]能够将大规模的节点映射到低维空间，通过定义节点间的一阶相似度和二阶相似度保持局部和全局的网络拓扑结构信息，是一种通用、高效的网络表示方法，但并未考虑网络的异质性。

MetaGNN: 由文献^[19]提出的异质卷积神经网络模型，通过采样网络模体来保持高阶结构信息，并在此基础上定义目标节点的感受野使模型能够识别不同的语义角色，从而捕获异质信息网络的复杂语义信息。

GCN: 由文献^[1]提出的图卷积神经网络模型，基于谱图卷积的一阶近似推导将图像处理中的卷积操作成功推广到了图结构上，模型将节点特征信息和图的邻接矩阵作为输入，最后将输出的节点表示运用于同质网络上的半监督节点分类任务。

HAN: 在文献^[20]中提出，是目前最先进的异质信息网络表示方法。HAN 使用分层注意力机制：使用节点级注意力学习节点和其基于元路径的邻居节点之间的权重，再使用语义级注意力来聚合多条元路径之间的不同权重。

MHGcn-noAtt: 本文提出的 MHGCN 的变体，但是由于未使用超边注意力机制，所以节点信息传

播过程中转移概率是固定的。

MHGcn: 本文提出的基于超图卷积的模型，引入超边注意力机制。

5.3 实验设置

本文基于 Pytorch 框架实现了 MHGCN 模型，模型迭代一轮代表一次优化过程，实验结果均基于 5 次实验的平均值。在实验中，首先使用文献^[19]中的网络模体匹配算法来抽取两个数据集中的模体，对于链接较为密集的数据集如 DBLP-A 则随机采样 30% 的实例作为超边，并使用它们构成超图。每个数据集的初始节点特征维度为 1500（或通过多层感知机将节点初始特征投影到 1500 维的特征空间），然后使用两层超图卷积层来提取节点特征。使用 Adam 优化器将学习率设置为 0.005，采用早停策略，假如验证集的损失在连续 50 轮迭代时下降或者上升的变化小于 0.002，则终止模型的训练。其它方法的参数设置均依据原文中的最佳组合。最后，将两个数据集中带标签数据按照 20%、40%、60% 和 80% 的划分比例设置为训练集，验证集和测试集均为 20%。

在机器学习中，micro-F1 和 macro-F1 是分类问题中重要的评价指标，主要基于混淆矩阵计算查准率和查全率，然后分别通过微平均和宏平均的方式对分类结果进行综合考察，因此，本文选用 micro-F1 和 macro-F1 作为评价指标。

5.4 实验结果分析

在两个数据集上，本文提出的方法与其它模型相比都取得了更好的实验结果，对比结果如表 2 所示。可以看出，MHGCN 模型的综合效果最好，

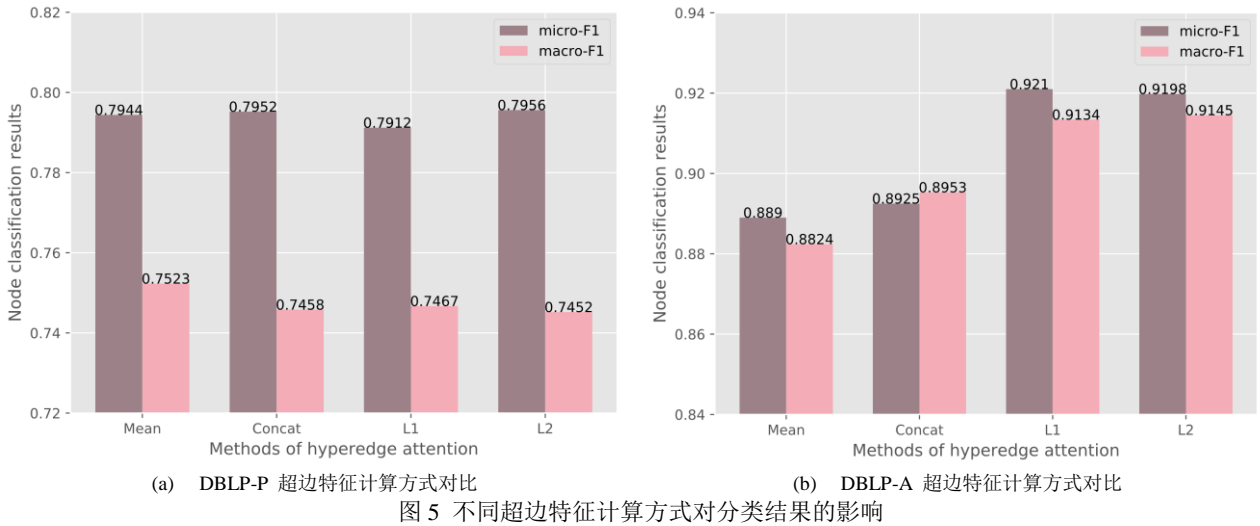


图5 不同超边特征计算方式对分类结果的影响

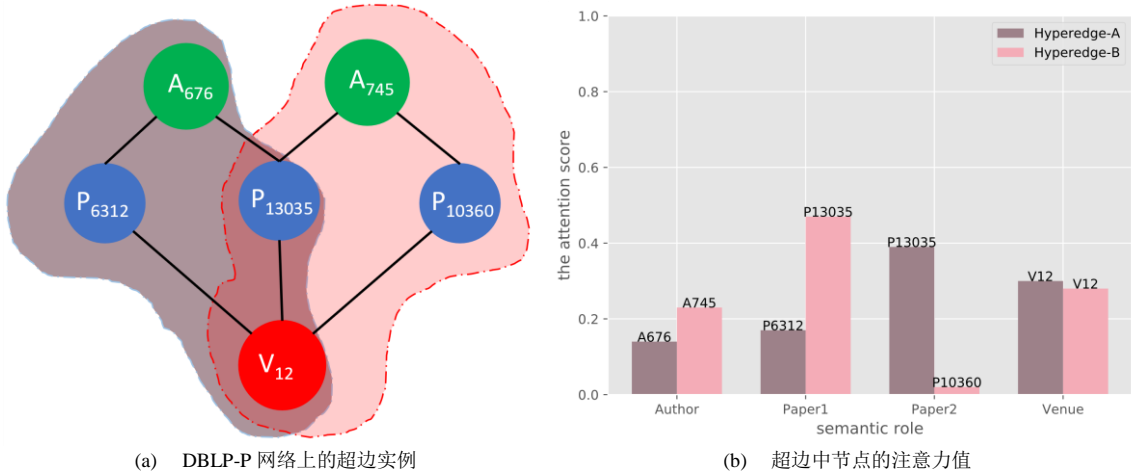


图6 超边注意力机制在引文网络上的验证

micro-F1 较最好的基准方法平均提高了 0.56%-3.51%，macro-F1 提高了 0.54%-4.37%。对于基于游走的 LINE 方法，由于它们均是通过游走产生序列再使用 skip-gram 来产生节点的表示，忽略了异质图的结构信息，因此，实验结果不如深层模型。作为深层模型，GCN 在两个数据集上取得了较好的效果，但是前提是本文通过抽取同质图的方式提升了 GCN 异质节点的分类表现，而这种方式无法获得综合的语义信息。MetaGNN 考虑了一种基于模体的空间域卷积操作，并利用注意力结合多重模体来获得目标节点表示，然而 MetaGNN 是通过增加模体类型来提高对网络的覆盖率，其表现依赖于人为设定的模体质量。HAN 使用节点级注意力聚合通过元路径相邻的节点，再使用语义级注意力进一步聚合多条元路径的表示，然而，HAN 基于元路径获得异质信息网络表示的特性使其受限于元路径对高阶结构和语义信息建模能力的不足，因此，无法获得更高质量的表示。本文提出的方法使用超边来保持异质信息网络中的高阶关系，使用注意力机制区分复杂的语义信息，因此，在两个数据集上均取得了较好的分类结果。同时，

MHGCN-noAtt 只有在 DBLP-P 上当训练数据比例为 20%时取得了较好的结果，其余效果均不如使用了注意力机制的 MHGCN，可见引入超边注意力是十分必要的。

针对本文提出的四种超边特征计算方式，在 60% 的训练数据比例下，其在两个数据集上的分类结果如图 5 所示。其中，使用 mean 方式计算超边特征时间复杂度低，且在 DBLP-P 上取得了较好的效果，然而在 DBLP-A 上分类效果最差。concat 方式涉及到多个向量的拼接，占用大量内存。通过对比分析可以看出，1-norm 方式和 2-norm 方式在两个数据集上都取得了较好的效果。本文选取 2-norm 方式作为 MHGCN 模型的最终超边聚合方式。

5.5 模型性能分析

本文提出的 MHGCN 模型除了在理论上基于超图谱理论之外，最显著的不同就是使用了超边内部自注意力机制并加入了自环方式，最后综合各个超图得到综合语义的节点表示。以 DBLP-P 数据集的论文节点分类结果为例，对超边注意力机制如何衡量节点在超边中的重要程度进行了分析。实验结果如图 6 所示，图 6(a)展示了在 DBLP-P 数据

集上的两个超边实例：Gareth Tyson(A745)和Matthew Farrens(A676)两位作者在MICRO(V12)期刊上发表过的三篇文章P6312^①、P10360^②、P13035^③，三篇文章的领域相同。首先，通过图4中模体M6来构建超边A和超边B，两条超边包含的每个节点都有着不同的语义角色，通过节点对应的注意力分数来衡量其在分类任务下对其它节点的影响。如图6(b)所示，在超边A中，作者节点A676和论文节点P6312的注意力分数较低，说明在以论文分类节点为目标的模型传播过程中，二者只提供了较少的信息。同样在超边B中，作者节点A745和论文节点P10360的注意力分数偏低，其中P10360是所有节点中得分最低的，经过验证，我们认为论文节点P10360的另一作者发表的另外两篇论文同P10360具有不同的标签信息，在端到端的训练过程中，这两篇论文作为P10360的二阶邻居，也会向P10360传播自身的节点特征和标签信息，因此，降低了P10360在超边中的影响力。最后，对于两条超边重叠部分中的节点P13035和V12，由于它们所包含的节点信息对于各自所在的超边内其它论文节点的分类有着更为重要的影响力，因此，在所有节点中注意力分数最高。基于以上分析，在模型传播过程中，超边注意力能够对承担不同语义角色的节点动态分配权重，从而最大程度上保留对待分类目标节点有意义的信息。

6 结论与展望

半监督节点分类是网络分析中的基本问题，然而在现实世界中的网络大多是异质信息网络，有着复杂的结构和丰富的语义。本文通过抽取异质网络中的模体来保存高阶复杂语义信息，并将网络模体建模成超边，构建超图异质信息网络，在此基础上提出异质超图卷积网络模型MHGCN。为了解决不同类型节点因其语义角色不同导致其特征空间不同的问题，使用超边自注意力机制动态聚合不同类型节点。通过在两个异质信息网络数据集上实验表明，相比目前最先进的异质信息网络表示方法，自注意力机制的加入使得MHGCN在节点分类效果上提升显著，同时，使节点信息的聚合过程变得透明，提供了良好的可解释性。因为不同超边之间信息重要程度并不相同，未来我们将进一步考虑不一致超图、超边的自动发现以及不同超边间的注意力，进一步提升异质信息网络节点分类的性能。

^① Gareth Tyson,Todd M.Austin.Improving the accuracy and performance of memory communication through renaming,MICRO'97

^② Gareth Tyson,Matthew Farrens,John Matthews.A modified approach to data cache management,MICRO'97

^③ Gareth Tyson,Matthew Farrens.Techniques for extracting instruction level parallelism on MIMD architectures,MICRO'97

参考文献

- [1] Thomas N. Kipf, Max Welling. Semi-Supervised Classification with Graph Convolutional Networks//Proceedings of the International Conference on Learning Representations. Toulon, France, 2017
- [2] Hamilton W, Ying R, Leskovec J. Inductive representation learning on large graphs//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 1024-1034
- [3] Velivckovic P, Cucurull G, Casanova A, et al. Graph attention networks//Proceedings of the International Conference on Learning Representations. Vancouver, Canada, 2018
- [4] Milo R, Shen-O S, et al. Network motifs: simple building blocks of complex networks. Science, 2002, 298(5594): 824-827
- [5] Sun Y, Han J, et al. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks//Proceedings of the VLDB Endowment. Seattle, USA, 2011: 992-1003
- [6] Berge C. Graphs and Hypergraphs. Amsterdam: North-Holland, 1973
- [7] Yan S, Xu D, et al. Graph embedding and extensions: A general framework for dimensionality reduction. IEEE transactions on pattern analysis and machine intelligence, 2006, 29(1): 40-51
- [8] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online learning of social representations// Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA, 2014: 701-710
- [9] Grover A, Leskovec J. node2vec: Scalable feature learning for networks//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016: 855-864
- [10] Tang Jian, et al. Line: Large-scale information network embedding//Proceedings of the 24th International Conference on World Wide Web. Florence, Italy, 2015: 1067-1077
- [11] Dong Y, Chawla N V, Swami A. metapath2vec: Scalable representation learning for heterogeneous networks//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Cambridge, UK, 2017: 295-304
- [12] Shang J, Qu M, Liu J, Kaplan L M, Han J, Jian P. Meta-path guided embedding for similarity search in large-scale heterogeneous information networks. arXiv preprint arXiv:1610.09769, 2016
- [13] Fu T, Lee W-C, Lei Z. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning// Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Singapore, 2017: 1797-1806
- [14] Bruna J, Zaremba W, Szlam A, Lecun Y. Spectral Networks and Locally Connected Networks on Graphs//Proceedings of the International Conference on Learning Representations. Banff, Canada, 2014
- [15] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering//Proceedings of the 30th Conference on Neural Information Processing Systems. Barcelona, Spain, 2016: 3844-3852
- [16] Feng Y, You H, Zhang Z, Ji R, Gao Y. Hypergraph neural networks//Proceedings of the 33rd AAAI Conference on Artificial Intelligence, Hawaii, USA, 2019:3558-3565
- [17] Yadati N, Nimishakavi M, Yadav P, et al. HyperGCN: A New Method of Training Graph Convolutional Networks on Hypergraphs// Proceedings of the 33rd Conference on Neural Information Processing Systems, Vancouver, Canada, 2019: 1511-1522
- [18] Bai S, Zhang F, et al. Hypergraph convolution and hypergraph attention. arXiv preprint arXiv:1901.08150, 2019
- [19] Sankar A, Zhang X, et al. Motif-based convolutional neural network on graphs. arXiv preprint arXiv:1711.05697, 2017
- [20] Wang X, Ji H, Shi C, et al. Heterogeneous Graph Attention Network//Proceedings of the World Wide Web Conference. San Francisco, USA, 2019: 2022-2032
- [21] Wang D, Cui P, Zhu W. Structural Deep Network Embedding//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016: 1225-1234
- [22] Hu B, Fang Y, Shi C. Adversarial Learning on Heterogeneous Information Networks// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Anchorage, USA, 2019: 120-129
- [23] Agarwal S, Branson K, Belongie S. Higher order learning with graphs//Proceedings of the 23rd international conference on Machine learning. Pittsburgh, USA, 2006: 17-24

- [24] Tang J, Zhang J, Yao L, Li J, et al. Armetminer: extraction and mining of academic social networks//Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data mining. Las Vegas, USA, 2008: 990-998
- [25] Sen P, Namata G, Bilgic M, et al. Collective classification in network data. *AI magazine*, 2008, 29(3): 93-93



Wu Yue Master candidate. His main research interests include deep learning, data mining, and analysis and mining of heterogeneous information networks.

Wang Ying PhD, associate professor. Senior member of CCF. Her main research interests include machine learning, social network and data mining.

Wang Xin PhD, associate professor. Senior member of CCF. His main research interests include machine learning, information retrieval and social computing.

Xu Zheng-Xiang Master candidate. His main research interests include machine learning, data mining and graph classification.

Li Li-Na PhD, lecturer. Her main research interests include data mining and social network.

Background

This paper focuses on heterogeneous information network analysis and applications. Compared to homogeneous information network, the heterogeneous information network can effectively fuse more information and contain richer semantics in objects and links, and thus it forms a new development of data mining. Driven by the success of deep learning, researchers have borrowed ideas from convolution networks to design the architecture of graph neural networks. Recall that heterogeneous networks have multiple classes of nodes or edges, the analysis of heterogeneous information network still has challenges, especially for network representation.

The metapath-based methods have been applied in heterogeneous graph mining in recent years because it can describe semantic information in the network. The latest motif patterns can both describe complex semantic relationships and preserve high-order neighbor structure in heterogeneous information network. This paper model the recurring high-order network patterns as hyperedges formed by multiple related nodes, and transform the entire heterogeneous information network into a hypergraph that composed of different

hyperedges. which can conduct convolution operations on hypergraph using the basic properties of hypergraph and spectral graph theory. This paper chiefly solves the following two problems: (1)Using the motifs and hyperedges to solve the problem of homogenous network can only describe the pairwise relationship; (2)proposed hyperedge self-attention mechanism aggregate different types of nodes come from different projection space, so as to generate the labels of nodes effectively and apply to sem-supervised node classification task.

This work is support by the National Natural Science Foundation of China (61872161,61976103) , the Project of Science and Technology Development Plan of Jilin Province (2018101328JC,20200201297JC) , the Science and Technology Department Excellent Youth Talent Foundation of Jilin Province (20170520059JH), the Project of Development and Reform of Jilin Province (2019C053-8) , and the Scientific Research Item of Jilin Provincial Education Department (JJKH20191257KJ) .