

引用格式: 肖晶晶, 冉义军, 许小可. 基于异质模体特征的社交机器人识别 [J]. 电子科技大学学报, 2025, 54(3): 424-431.

XIAO J J, RAN Y J, XU X K. Social robot recognition based on heterogeneous motif features[J]. Journal of University of Electronic Science and Technology of China, 2025, 54(3): 424-431.

## 基于异质模体特征的社交机器人识别



肖晶晶<sup>1</sup>, 冉义军<sup>2,3</sup>, 许小可<sup>1,2,3\*</sup>

(1. 大连民族大学 信息与通信工程学院, 大连 116600; 2. 北京师范大学 计算传播学研究中心, 珠海 519085;

3. 北京师范大学 新闻传播学院, 北京 100875)

**摘要:** 识别社交网络中的机器人账户, 可以保护社交网络运营商及其用户免受各种恶意活动的侵害。现有的基于网络结构的社交机器人识别方法, 忽略了真实网络多数是节点或连边性质具有差异的异质网络, 无法充分利用不同类型节点或连边的拓扑结构信息, 使得在识别社交机器人时存在一定的局限性。基于模体理论融合节点标签信息, 提出一种基于异质模体特征的社交机器人识别方法, 提取更加细节的局部信息来区分人类用户和机器人用户。所提方法与 BotRGCN 方法相比, 在 ACC、Precision、Recall 和 F1 这 4 个指标上均有所提升, 其中 ACC 指标提高了 17.3%, Precision 指标提高了 23.3%, 同时相较于对比方法, 在面对标签噪声时展现出更强的鲁棒性。该识别方法可以更精确地识别社交机器人, 从而有效防止恶意机器人对社交网络平台进行攻击、传播虚假信息或进行欺诈行为。

**关键词:** 社交网络; 异质模体; 社交机器人识别; 社交关系

中图分类号: TP391

文献标志码: A

DOI: 10.12178/1001-0548.2024074

## Social robot recognition based on heterogeneous motif features

XIAO Jingjing<sup>1</sup>, RAN Yijun<sup>2,3</sup>, and XU Xiaoke<sup>1,2,3\*</sup>

(1. College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China;

2. Computational Communication Research Center, Beijing Normal University, Zhuhai 519085, China;

3. School of Journalism and Communication, Beijing Normal University, Beijing 100875, China)

**Abstract:** Identifying bot accounts in social networks can protect social network operators and their users from a variety of malicious activities. The existing social robot recognition methods based on network structure ignore that most of the real networks are heterogeneous networks with different node or edge properties, and cannot make full use of the topological information of different types of nodes or edge, resulting in certain limitations in the identification of social robots. Based on motif theory, a social robot recognition method based on heterogeneous motif features is proposed to extract more detailed local information for distinguishing human users from robot users. Compared with BotRGCN methods, the proposed method has improved in ACC (Accuracy), Precision, Recall and F1, among which ACC has increased by 17.3% and Precision by 23.3%. At the same time, the experimental results show that the proposed method is more robust in the face of label noise than the comparative methods. This identification method can identify social robots more accurately, so as to effectively prevent malicious robots from attacking social network platforms and spreading false information or committing fraud.

**Key words:** social network; heterogeneous motif; social robot recognition; social connections

近年来, 随着社交网络用户数量的增长, 社交网络成为人们获取信息的一个重要渠道。根据 Alexa 的数据, Facebook 已成为全球访问量第三大的网站, 仅次于 Google 和 YouTube<sup>[1]</sup>。然而, 在庞

大的社交平台用户群体中, 机器人账户的涌现逐渐引起人们的关注。这些账户并非由真实的人进行操作, 而是受到程序的操控, 因此被大众称为社交机器人。社交网络上的机器人账号数量之多令人震

收稿日期: 2024-04-02

基金项目: 国家自然科学基金 (62173065); 北京市自然科学基金 (4242040); 辽宁省教育厅项目 (LJKM20220399)

作者简介: 肖晶晶, 博士生, 主要从事社交网络数据挖掘和节点预测方面的研究。

\*通信作者 E-mail: xuxiaoke@foxmail.com

惊。据统计, Twitter 上有 10% 的账号是虚假的<sup>[2]</sup>。机器人已渗透到 Twitter、Facebook、Instagram、问答网站、在线报纸、电子邮件等各种在线对话中, 几乎无处不在<sup>[3]</sup>。这些机器账号并非都是无害的, 部分账号带有恶意性质。它们会散播虚假信息, 误导公众舆论, 进而实现其谋取利益的目的。这些问题不仅严重干扰了社交网络平台的正常运行, 而且对广大用户的权益构成了潜在的威胁。因此, 对这些社交机器人的检测变得至关重要。

基于图论的社交机器人识别方法旨在通过社交网络结构和动力学特征识别机器人账号, 主要利用循环信念传播和随机游走算法进行机器人识别<sup>[4]</sup>。SybilBelief 算法提出一种概率局部规则, 它依据节点邻居的标签信息来评估节点的声誉分数, 该算法将社交网络建模为马尔可夫随机场, 通过一组状态已知的机器人和人类节点, 采用循环信念传播算法, 以获得未知节点是机器人的可能性估计<sup>[5]</sup>。SybilBelief 算法基于一个假设, 即机器人节点在与人类节点建立连接时面临较大的困难, 它们之间的联系存在局限性。SybilFuse 算法提出, 在网络中机器人能够广泛和人类节点建立联系, 于是文献 [6] 在 SybilBelief 算法基础上提出采用集体分类方案, 将局部属性与全局结构结合起来。GANG 算法则是将 SybilBelief 算法应用于有向图中<sup>[7]</sup>。虽然基于循环信念传播方法能够处理节点之间复杂的关系, 但其存在无法保证收敛性且缺乏可扩展性等缺点。

基于随机游走的方法利用社交网络的结构特性, 通过随机游走模型为每个用户分配信誉分数, 这些分数可以用来区分用户是否是机器人。SybilRank 将一组标记好的人类用户和机器用户的信誉分数通过随机游走方式分配给其他用户<sup>[8]</sup>。Intergo 首先预测潜在的受害节点, 依据预测结果为社交网络中的连边分配不同的权重; 随后在这个加权网络中执行随机游走, 从而改进 SybilRank<sup>[9]</sup>。SybilWalk 通过额外引入一个标记的人类用户  $l_b$  和一个机器用户  $l_s$ , 将训练集中每个标记的人类节点与  $l_b$  之间创建一条边, 每个标记的机器节点与  $l_s$  之间创建一条连边, 通过随机游走赋给剩余用户信誉分数。该算法解决了随机游走无法同时捕获人类用户和机器人用户标签的问题<sup>[10]</sup>。基于随机游走的方法能够适应不同的网络结构, 但它们对训练数据中的噪声标记缺乏鲁棒性。

除上述利用循环信念传播和随机游走算法外, 许多研究如利用深度学习等也在图论方面展开。如 GAT<sup>[11]</sup> 算法采用了一种基于注意力的架构, 用

于对图结构数据进行节点分类。BotRGCN<sup>[12]</sup> 通过从关注关系中构建异构图, 并应用关系图卷积网络来解决机器人检测问题。文献 [13] 提出了 Ising 算法, 该算法将概率分布放在一个图和它的节点标签上, 使用图模型来表示这种分布, 同时识别图中的所有机器人。尽管上述基于图论的方法在社交机器人的识别方面取得了一定的成效, 但理论上更局部的结构特征在描绘社交网络用户间的复杂关系时会更加合理, 因此本研究旨在探索并应用更加注重局部结构的模体特征来进行机器人识别。

模体是指网络中出现频率较高的子图结构<sup>[8]</sup>, 其不仅能够涵盖网络中的局部特征, 还为理解网络的整体结构演化提供了线索。通过模体进行图分析, 可以获得更为精细的粒度, 从而更准确地把握网络的结构和动态。文献 [14] 基于网络模体和多元统计分析来预测定向生物网络中的重要节点。虽然基于模体结构进行节点预测的研究日益增多, 但大多是在同质网络中进行分析。文献 [15] 提出了一种结合模体滤波和卷积神经网络的分层网络嵌入方法, 可用于进行节点预测。文献 [16] 构建了基于模体的图卷积神经网络 (Motif-GCRNN) 来预测交通速度的低频分量。文献 [17] 提出了一种基于网络模体的复杂预测算法, 用以检测蛋白质复合物。文献 [18] 引入了一种新的 GNN 自监督模体生成框架, 提出了基于模体的图自监督学习和多层次自监督预训练, 用以进行节点预测。尽管以上方法在一些网络中取得了较好的性能, 但它们不能通过区分节点类型来刻画网络的结构特征, 因此无法充分利用节点的异质信息, 影响了预测的准确性。

为了更全面地利用节点的异质信息, 本文提出基于异质模体特征的节点预测方法, 用以识别社交网络中的社交机器人。本文首先将同质模体特征引入社交网络的机器人预测中, 同时结合用户的标签信息将同质模体细化为异质模体, 并提出了基于异质模体特征的社交机器人识别算法。

## 1 研究数据与方法

### 1.1 问题描述

本文使用的社交网络为有向无权网络, 形式为  $G(V, E)$ ,  $V$  为网络中的节点集合,  $E$  为网络中的连边集合, 边用  $e(u, v) \in E$  来表示, 其中  $u, v \in V$ 。其中每个节点  $v \in V$  都有  $f(v) \in A$ , 这表示每个节点的类型映射函数为  $f: V \rightarrow A$ 。每条边  $e \in E$  都有  $r(e) \in R$ , 表示每条边的类型映射函数  $\gamma: E \rightarrow R$ 。当  $R$  和  $A$  满足  $|A| > 1$  或  $|R| > 1$  时, 即边的类型数或者节点的类型

型数大于 1, 则该网络定义为异质网络, 反之为同质网络。从同质网络中提取出的模体, 被定义为同质模体, 反之则是异质模体。对于待测用户节点  $u$  是否为机器人的预测, 是通过与其相连的节点  $v$  的属性信息和节点  $u$  所处的网络结构来进行预测的。本文所研究的社交机器人识别问题, 本质上就是一个针对网络节点的预测问题。

## 1.2 数据说明

本文利用 Twitter 中的社交数据, 选取了 2 个公开数据集, 分别为 TwiBot-20<sup>[19]</sup> 和 TwiBot-22<sup>[20]</sup>。TwiBot-20 包含了社交机器人和真实用户, 以便更好地反映现实世界的 Twitter 环境; 而 TwiBot-22 则是一个全面的基于图的 Twitter 机器人检测基准, 提供了迄今为止最大的社交机器人数据集, 其中包含了 Twitter 网络上多样化的实体和关系, 并且注释质量较优。TwiBot-20 和 TwiBot-22 都提供了明确的 Twitter 用户之间的图结构。为了进行社交机器人检测实验, 从原始数据中提取了所有用户是否为机器人的标签信息以及用户之间的网络结构信息, 用以构建异质网络。大规模社交网络的具体信息如表 1 所示。

表 1 网络信息说明

网络信息	节点/个	连边/条	人类节点/个	机器人节点/个
TwiBot-20	205 730	227 477	5 237	6 589
TwiBot-22	693 761	3 711 903	612 329	81 432

## 1.3 评价指标

### 1) Accuracy

Accuracy (准确率) 主要关注整体的预测情况, 是预测正确的结果占总样本的百分比, 即正确预测的人类、机器人数量/预测总数, 其中 TP 和 FP 分别代表实际类别是人类而预测类别是人类和机器人的节点数, TN 和 FN 分别代表实际类别为机器人而预测类别为机器人和人类的节点数。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (1)$$

由该式可知, TP 和 TN 在整体所占的比例越大 Accuracy 值越高, 说明预测得越准确。

### 2) Precision

Precision 是一个衡量模型预测准确性的指标, 它反映了在所有被模型预测为正样本的实例中, 有多少是真正的正样本。Precision 可以定义为:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

Precision 越高, 说明模型预测为正样本的实例

中真正为正样本的比例越高, 也就是模型预测的结果更加准确。

### 3) Recall

Recall 常用于衡量模型找出所有实际为正例的样本的能力。Recall 又称查全率, 本文中其指在实际为人类的样本中被预测为人类的概率, 可以定义为:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

Recall 反映了模型在识别正样本方面的敏感度, 即模型能否将大部分正样本都识别出来。

### 4) F1

F1 是 Precision 和 Recall 的调和平均值, 它综合考虑了 Precision 和 Recall 的性能, 可以看作是两者的平衡。F1 是统计学中用来衡量二分类模型精确度的一种指标。它定义为:

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

F1 的最大值是 1, 最小值是 0, 值越大意味着模型的性能越好。

## 2 基于同质模体的社交机器人识别

### 2.1 基于同质模体的社交网络用户关系

模体一般由几个节点连接构成, 在有向图中, 3 个节点之间可以形成 13 种不同的连通关系。根据待测节点的不同, 这 13 种模体关系可以被进一步细分为 30 种不同的模体结构。这些模体结构可以根据其特征分为 3 类。一类是具有一阶特征的模体, 这类模体的构建起始于待测节点, 只需找出其邻居节点即可形成一个模体, 如图 1a 所示, 其中标注节点为待测节点。另一类是具有闭合特征的模体, 如图 1b 所示, 这类模体的构建起始于待测节点, 只需找出与其相邻的两个节点, 且这两个相邻节点之间存在连边关系。最后一类是具有二阶特征的模体, 如图 1c 所示, 这类模体的构建起始于待测节点, 找出与其相邻的邻居节点后, 需要通过这些节点继续寻找由其邻居节点而构成的模体结构。

基于同质模体特征的社交网络用户关系是在忽略不同类型用户的情况下, 将待检测用户参与的不同结构的模体结构视为该用户参与的不同关系模式。本文利用具有相同属性(人类/机器人)的用户拥有相似的社交关系的普遍特点, 采用模体代表社交网络中的关系模式。在此小节中, 忽略了节点的不同类型, 仅考虑节点之间的连边关系, 以模体的拓扑结构代表网络中的关系模式, 共涉及 30 个三节点模体, 分别代表着 30 种关系模式。

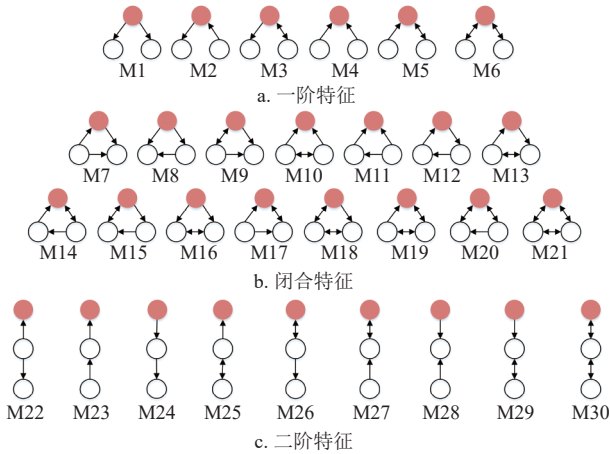


图1 三节点模体结构

### 2.2 基于同质模体特征的识别方法

基于同质模体特征的社交机器人预测中, 将每个预测节点参与的同质模体的数量作为特征值。该特征值用  $Motif_i(n)$  表示, 其中  $Motif_i(a)$  代表节点  $a$  参与第  $i$  种类型的模体的数量。这个提取过程在图2中进行了说明。以图2中的社交网络为例, 分别提取出不同的模体结构, 各模体对应的具体数量即是其特征值用  $Motif_1(a)$ 、 $Motif_7(b)$ 、 $Motif_{22}(a)$  表示。模体  $Motif_1(a)$  的计算方法是寻找待测节点  $a$  指向的互不相邻的2个邻居节点数目。其余特征的计算方法以此类推, 计算出在图2中的社交网络中, 节点  $a$  拥有的模体特征  $Motif_1(a)$  为2,  $Motif_{22}(a)$  为4, 节点  $b$  拥有的模体特征  $Motif_7(b)$  为2。

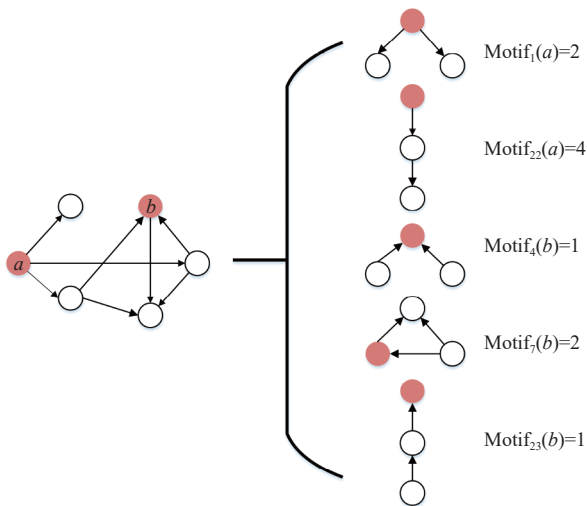


图2 同质模体特征提取

采用同质模体特征为基础的识别手段, 将节点参与各类同质模体的数量作为数值特征, 用于预测用户是否为机器人, 识别模型框架如图3所示。在实证研究阶段, 将数据集分割成训练集与测试集, 其中训练集与测试集的比例为9:1, 同时确保测

试集中人类节点与机器人节点的比例为1:1。随后, 构建使用随机森林模型进行训练和测试, 并采用了十折交叉验证的方法来全面评估模型的性能。随机森林模型与决策树、逻辑回归、支持向量机以及朴素贝叶斯等模型相比, 具有更高的稳定性及准确性, 因为它采用集成学习的思想, 结合了多个决策树的预测结果。这种“森林”式的结构使得模型能够捕捉到数据中的多种模式和特征, 从而降低了单一模型可能产生的偏差和过拟合问题。同时, 随机森林中的每棵树都是基于随机抽样和特征随机选择构建的, 这种随机性进一步增加了模型的多样性和稳定性, 使得整体预测结果更为准确可靠。此外, 在存在异常值的情况下, 随机森林构建的多个决策树以投票方式进行预测, 这有助于减轻异常值对模型的影响, 更适合解决本文提出的多模体预测问题。

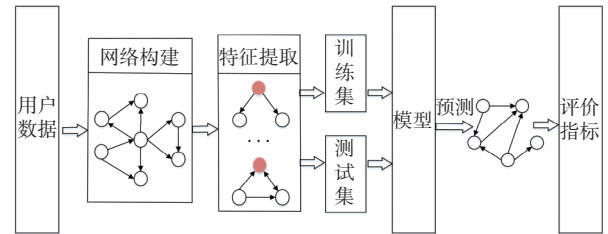


图3 模型预测框架示意图

在实验过程中, 对数据集进行了合理切分后, 首先提取了每个节点参与的同质模体的数量。接着, 将这些单个模体数量特征作为模型的输入, 构建了多个专门针对不同模体的分类器。随后, 将所有同质模体的数量特征集合在一起, 形成了一个多模体特征向量。这个特征向量被用来构建一个多模体分类器, 该分类器能够全面利用各种模体的信息, 以提高预测结果的准确性和稳定性。

本文针对单个同质模体特征以及多个同质模体特征进行预测, 并计算出评价指标包括准确率 (ACC)、精确率 (Precision)、查全率 (Recall) 以及 F1 分数的值, 结果如表2所示, 单个模体特征的较优预测性能和多模体特征的预测效果加粗标出。从表2可以看出, 在利用单个同质模体进行社交机器人预测时, Twibot-20 网络中的模体特征 M3 以及 Twibot-22 网络中的模体特征 M8 的表现较为出色。然而, 当综合运用多个同质模体特征进行预测时, 其预测效果 ACC 相较于使用单个特征较优效果提升了 1.0%~2.9%, Precision 相较于使用单个特征较优效果提升了 2.3%~4.6%, 而 Recall 和 F1 的值保持在可接受的范围, 这表明综

合利用多个模体特征进行预测能取得更好的效果。然而, 仅仅利用同质模体特征无法全面反映异质网络中的社交关系, 还可能引入一些噪声信息。因

此, 本文提出将同质模体根据节点用户是否为机器人的标签进行细化, 形成异质模体, 以便更全面地提取社交关系信息。

表 2 基于同质模体特征的社交机器人识别结果

结构特征	模体名称	TwiBot-20				TwiBot-22			
		ACC	Precision	Recall	F1	ACC	Precision	Recall	F1
闭合	M7	0.482	0.484	1.000	0.650	0.507	0.513	0.231	0.316
	M8	0.490	0.487	0.992	0.651	<b>0.565</b>	<b>0.548</b>	<b>0.712</b>	<b>0.623</b>
	M9	0.483	0.483	1.000	0.651	0.550	0.531	0.831	0.649
	M11	0.484	0.484	1.000	0.652	0.562	0.571	0.497	0.532
	M17	0.482	0.482	0.998	0.650	0.538	0.523	0.869	0.653
	M1	0.560	0.589	0.296	0.394	0.522	0.522	0.511	0.517
一阶	M2	0.523	0.505	0.939	0.655	0.539	0.529	0.731	0.612
	M3	<b>0.627</b>	<b>0.564</b>	<b>0.961</b>	<b>0.715</b>	0.543	0.569	0.355	0.437
	M4	0.483	0.483	1.000	0.651	0.544	0.545	0.527	0.536
	M5	0.483	0.483	1.000	0.651	0.524	0.557	0.229	0.325
	M23	0.521	0.504	0.944	0.654	0.534	0.585	0.243	0.344
	M25	0.488	0.485	1.000	0.652	0.502	0.505	0.257	0.341
二阶	M26	0.483	0.483	1.000	0.651	0.546	0.580	0.330	0.421
	M28	0.573	0.573	0.995	0.652	0.547	0.530	0.778	0.632
	混合结构	同质模体	0.637	0.587	0.831	0.687	0.594	0.616	0.544

### 3 基于异质模体的社交机器人检测

#### 3.1 基于异质模体的社交网络用户关系

异质模体特征主要针对异质网络, 即网络中不止存在一种节点类型。根据异质网络的拓扑结构, 构建区分节点类型的模体结构特征, 将其定义为异质模体。本小节将利用节点属性信息来细化同质模体, 从而获得更加精确的异质模体来代表节点的社交关系。在预处理网络数据后, 根据用户是否为机器人将同质模体细分, 具体细分过程如图 4 所示, 其中蓝色节点代表人类, 紫色节点代表机器人。

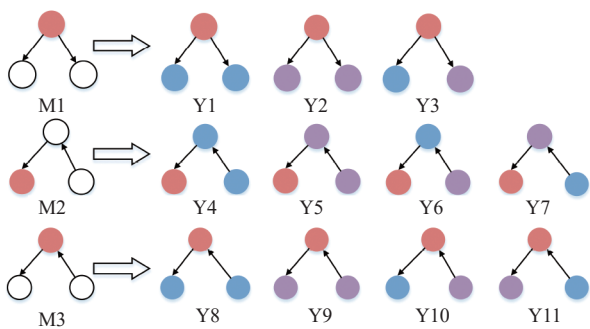


图 4 模体细化过程示意图

基于异质模体的社交关系是将用户属性纳入到社交用户关系构建框架中。本节主要使用三节点模体, 以用户标签 (人类/机器人) 作为细分同质模体的属性, 同质模体结构参照图 1。同质模体特征采用 30 种同质模体代表用户关系模式, 相比之下, 异质模体特征在此基础上采用 114 种三节点异质模体表示 114 种关系模式, 相比同质模体特征,

异质模体特征可以更充分地利用社交关系信息。

本文列举了所有异质模体, 这样做有两方面原因。首先, 全面准确地涵盖所有模体特征可以最大程度地保证局部结构信息的完整性。其次, 这种方法可以最大程度地避免网络中出现的特殊情况对结果的影响。考虑到所有的关系模式和结构, 可以在一定程度上减少特例情况对分类和预测的干扰, 提高模型的泛化能力。

#### 3.2 基于异质模体特征的社交机器人预测

基于异质模体特征的社交机器人预测中, 将每个预测节点参与的异质模体的数量作为特征值。该特征值用  $YMotif_i(n)$  表示, 其中  $YMotif_i(n)$  代表节点  $n$  参与第  $i$  种类型的异质模体的数量。这个提取过程在图 5 中进行了说明。

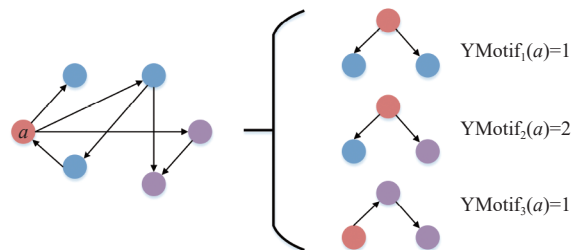


图 5 异质模体特征的提取过程

以图 5 中的社交网络为例, 分别提取出异质模体, 异质模体对应的具体个数即是其特征值用  $YMotif_1(a)$ 、 $YMotif_2(a)$ 、 $YMotif_3(a)$  表示。异质模体  $YMotif_1(a)$  的计算方法是寻找待测节点  $a$  和邻居节点中互不相邻的 2 个具有人类属性的节点构成的

模体数目。其余特征的计算方法以此类推, 计算出图5中的社交网络中, 节点  $a$  拥有的异质模体特征  $YMotif_1(a)$  为 1,  $YMotif_2(a)$  为 1,  $YMotif_3(a)$  为 3。

在利用异质模体特征进行社交网络用户节点标签预测时, 计算了类似图5中的异质模体特征。然后将训练集中节点参与的每种异质模体特征和训练集中的节点标签作为机器学习模型的输入进行训练, 生成为具有客观性的随机森林模型。接着将测试集中节点参与的每种异质模体特征作为输入, 利用训练好的模型进行打分并添加标签。最后将测试集中真实的打分和标签进行对比, 从而进行分类并得出评价指标。在确保训练集和测试集一致的前提下, 将提出的异质模体方法与 Ising 算法、BGSRD<sup>[21]</sup> 算法、Abreu<sup>[22]</sup> 算法、BotRGCN<sup>[12]</sup> 算法、Botometer<sup>[23]</sup> 算法以及 T5<sup>[24]</sup> 算

法等方法进行了对比。

比较结果如表3所示, 从表3中可以看出, 在 TwiBot-20 网络中, 异质模体方法的准确率相较于其他对比算法并非最优。这是由于该网络中已知标签的节点数量较少, 在提取异质模体特征时, 大量未知邻居被忽略, 因此所提取出的异质模体无法准确且完整地涵盖网络的微观结构信息, 从而影响了机器人的识别精度。通过实验, 随机选取一定比例的节点并移除其标签, 查看不同标签缺失比例下的模型性能, 可发现随着网络中标签缺失比例的增大, 模型的性能呈现出一定程度的下降趋势。从而验证了上述结论。相对而言, BotRGCN 算法展现出了较优性能, 这主要归功于其不仅利用了图结构信息, 还结合了用户的词条描述、推文内容等多模态用户信息, 从而更精准地识别出社交机器人。

表3 多种方法的社交机器人预测结果

对比算法	TwiBot-20				TwiBot-22			
	ACC	Precision	Recall	F1	ACC	Precision	Recall	F1
Ising <sup>[13]</sup>	0.573 0	0.587 0	0.857 0	0.697 0	0.493 0	0.491 0	0.823 0	0.615 0
BGSRD <sup>[23]</sup>	0.663 6	0.676 4	0.731 9	0.700 5	0.718 8	0.225 5	0.199 0	0.211 4
Abreu <sup>[24]</sup>	0.734 5	0.722 0	0.828 1	0.771 4	0.706 6	0.509 2	0.117 3	0.534 4
BotRGCN <sup>[12]</sup>	<b>0.858 0</b>	<b>0.845 0</b>	<b>0.902 0</b>	<b>0.873 0</b>	0.797 0	0.748 0	0.468 0	0.575 0
Botometer <sup>[25]</sup>	0.530 9	0.531 3	0.556 7	0.508 2	0.498 7	0.427 5	0.308 1	0.698 0
T5 <sup>[26]</sup>	0.735 7	0.721 9	0.690 5	0.705 7	0.720 5	0.632 7	0.120 9	0.202 7
同质模体	0.637 0	0.587 0	0.831 0	0.687 0	0.594 0	0.616 0	0.544 0	0.579 0
异质模体	0.816 0	0.778 0	0.895 0	0.836 0	<b>0.970 0</b>	<b>0.982 0</b>	<b>0.957 0</b>	<b>0.969 0</b>

而在 TwiBot-22 网络中, 异质模体方法在机器人识别方面表现优于其他方法。相较于 BotRGCN 方法, 该方法在 ACC 指标上提升了 17.3%, Precision 指标提高了 23.3%, Recall 指标以及 F1 指标均有提升。这得益于该网络中完整的节点标签, 使得提取出的异质模体特征能够准确捕捉到图结构中的局部信息。这些信息充分反映了节点之间的相互作用和关系, 从而能够更准确地识别出社交机器人和人类用户的行为模式, 进而提高了社交机器人检测的准确率。此外, 还评估了随机森林识别模型中 114 种异质模体特征的重要性, 发现 Y12、Y8 和 Y4 等模体具有显著的重要性。当仅使用这些高重要性特征组合作为输入时, 模型的 ACC 高达 0.95。然而, 当使用重要性较低的特征组合时, 模型的准确率大幅下降至 0.56。这表示, 在本文提出的识别方法中, 异质模体特征起到了关键作用。

ROC 曲线是评估模型性能的一种图形工具。ROC 曲线通过调整分类阈值, 绘制出真正例率与假正例率之间的关系, 曲线越接近左上角, 表明模

型性能越优异。AUC 是 ROC 曲线下与坐标轴围成的面积, 用于量化模型的性能。AUC 越接近 1, 说明模型在不同阈值下的性能越好。图6展示了各分类模型下的 ROC 曲线, 由图可看出, 基于异质模体特征的分类模型的 ROC 曲线更偏向左上角, 且 AUC 均明显大于其余的对比算法, 充分证明了其出色的分类能力。

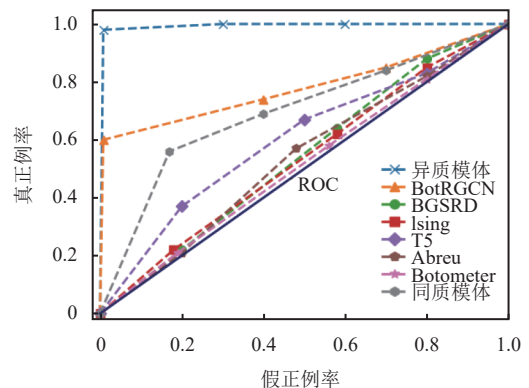


图6 各分类模型下的 ROC 曲线图

本文不仅深入比较了多种机器人识别方法在实际应用中的性能表现, 还进一步利用 SHAP 概要图

对 TwiBot-22 网络中涉及的异质模体结构特征进行了深入分析。SHAP 概要图作为一种综合展示特征重要性与特征效应的图形工具,不仅能够直观展现各特征在识别过程中的关键作用,还能精准揭示每个样本中特征影响力的具体大小和正负效应。

本文分析了异质模体结构特征的 SHAP 概要图,其结果如图 7 所示。图中纵轴列出了部分特征重要程度较高的异质模体结构,特征的重要性从下往上递增。横轴代表每个特征的 SHAP 值,SHAP 值是基于 Shapley 值的概念,通过计算特征在模型预测中的边际贡献来评估其重要性。横轴的正值意味着该特征对预测结果有积极影响,而负值意味着该特征对预测结果有消极影响。图中每个点代表一个样本,点的颜色代表该样本的特征值,其中点的颜色由蓝到红表示该点所代表的样本的特征值由低到高(即异质模体数量由小到大)。从图 7 可以看出,异质模体结构特征 Y12 在机器人识别过程中具有较高的重要性。该特征中,蓝色点广泛分布在正负值区域,表明这个特征对于大多数样本都具有影响,无论是正面还是负面,这意味着该特征具有较强的预测能力。此外,一些具有二阶特征的模体结构如 Y13、Y15 也具有相对较高的重要性,这表明具备二阶特征的异质模体结构在模型预测中也发挥着一定作用。

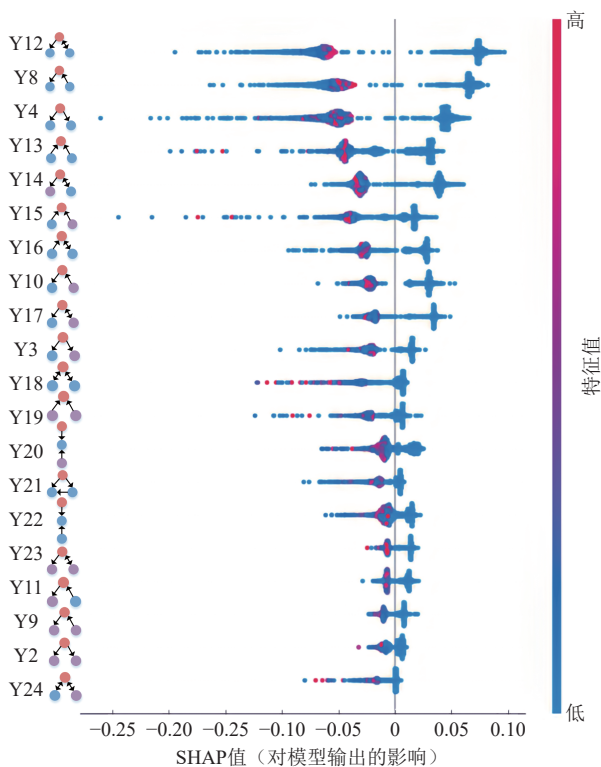


图 7 SHAP 概要图

### 3.3 对标签噪声的鲁棒性

在实际应用中,训练数据集可能存在噪声,即一些标记为人类的用户实际上可能是机器人,而一些标记为机器人的用户实际上可能是人类,从而对识别方法的准确性造成一定的干扰。本小节将探究所提出的算法对这种标签噪声的容忍能力,以评估该算法的鲁棒性。如果提供的用户的标签信息与实际标签不符,则该节点具有噪声标签。具体实验时,在 TwiBot-22 数据集上随机选取了一定比例被标记为人类的节点,并将其标记为机器人节点,同时随机选取了一定比例被标记为机器人的节点,并将其标记为人类节点。图 8 显示了当增加一定比例的标签噪声时,比较不同方法在数据集上的 ACC。

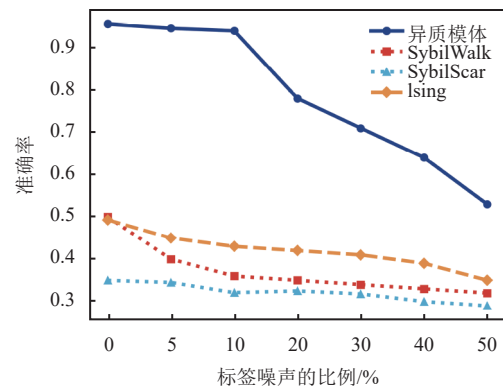


图 8 不同算法对标签噪声的鲁棒性

由图 8 可知,当标签噪声比例高达 10% 时,基于异质模体的检测算法的 ACC 接近 1。而其他算法在不到 10% 的标签噪声下,其预测表现比随机猜测更差,因此基于异质模体特征的算法在标记噪声方面更健壮。当 50% 的标记节点有噪声标记时,对比方法的 ACC 都小于 0.5,而本文方法具有明显优势。

## 4 结束语

本文以 Twitter 数据为研究对象,探讨了模体作为图结构特征在社交机器人检测领域的应用。首先,提出了一种基于同质模体特征的检测方法,该方法用同质模体代表社交网络用户群关系。实验结果表明,多个同质模体特征组合的检测效果优于单个模体特征。其次,由于异质模体特征可以更精细地描述社交关系,因此将同质模体特征利用社交网络用户标签信息进行细化成异质模体,提出了基于异质模体特征的检测方法。相比于对比方法,该方法在 4 个指标上均有所提升,其中在 ACC 指标上提高了 17.3%, Precision 指标提高了 23.3%。同

时, 为评估该方法对于标签噪声的鲁棒性, 本研究采用了随机选择一定比例节点的方式, 将这些节点的标签赋予与原始标签相反类别, 然后对各算法进行对比实验。实验结果显示, 基于异质模体特征的检测方法相较于其他方法展现出了更强的鲁棒性。本研究有助于对社交网络中的用户进行属性预测, 在用户行为分析、异常行为检测以及用户隐私保护等方面具有广阔的应用前景。

### 参考文献

- [1] FERRARA E, VAROL O, DAVIS C, et al. The rise of social bots[J]. *Communications of the ACM*, 2016, 59(7): 96-104.
- [2] YU Z H, BAI L X, YE O, et al. Social robot detection method with improved graph neural networks[J]. *Computers, Materials & Continua*, 2024, 78(2): 1773-1795.
- [3] DANIEL F, CAPIELLO C, BENATALLAH B. Bots acting like humans: Understanding and preventing harm[J]. *IEEE Internet Computing*, 2019, 23(2): 40-49.
- [4] BEBENSEE B, NAZAROV N, ZHANG B T. Leveraging node neighborhoods and egograph topology for better bot detection in social graphs[J]. *Social Network Analysis and Mining*, 2021, 11(1): 10.
- [5] GONG N Z, FRANK M, MITTAL P. SybilBelief: A semi-supervised learning approach for structure-based sybil detection[J]. *IEEE Transactions on Information Forensics and Security*, 2014, 9(6): 976-987.
- [6] GAO P, WANG B H, GONG N Z, et al. SYBILFUSE: Combining local attributes with global structure to perform robust sybil detection[C]//*Proceedings of the IEEE Conference on Communications and Network Security*. New York: IEEE, 2018: 1-9.
- [7] WANG B H, GONG N Z, FU H. GANG: Detecting fraudulent users in online social networks via guilt-by-association on directed graphs[C]//*Proceedings of the IEEE International Conference on Data Mining*. New York: IEEE, 2017: 465-474.
- [8] CAO Q, SIRIVIANOS M, YANG X W, et al. Aiding the detection of fake accounts in large scale social online services[C]//*Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation*. San Jose: [s.n.], 2012: 197-210.
- [9] BOSHMAF Y, LOGOTHETIS D, SIGANOS G, et al. Íntegro: Leveraging victim prediction for robust fake account detection in large scale OSNs[J]. *Computers & Security*, 2016, 61: 142-168.
- [10] JIA J Y, WANG B H, GONG N Z. Random walk based fake account detection in online social networks[C]//*Proceedings of the 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*. New York: IEEE, 2017: 273-284.
- [11] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. *Stat*, 2017, 1050(20): 10-22.
- [12] FENG S, WAN H, WANG N, et al. BotRGCN: Twitter bot detection with relational graph convolutional networks[C]//*Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. Piscataway: IEEE, 2021: 236-239.
- [13] DES MESNARDS N G, HUNTER D S, EL HJOUI Z, et al. Detecting bots and assessing their impact in social networks[J]. *Operations Research*, 2022, 70(1): 1-22.
- [14] WANG P, LÜ J H, YU X H. Identification of important nodes in directed biological networks: A network motif approach[J]. *PLoS One*, 2014, 9(8): e106132.
- [15] YANG C, LIU M X, ZHENG V W, et al. Node, motif and subgraph: Leveraging network functional blocks through structural convolution[C]//*Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. New York: IEEE, 2018: 47-52.
- [16] ZHANG N, GUAN X, CAO J, et al. A hybrid traffic speed forecasting approach integrating wavelet transform and motif-based graph convolutional recurrent neural network[EB/OL]. [2024-03-15]. <https://doi.org/10.48550/arXiv.1904.06656>.
- [17] PATRA S, MOHAPATRA A. Protein complex prediction in interaction network based on network motif[J]. *Computational Biology and Chemistry*, 2020, 89: 107399.
- [18] ZHANG Z, LIU Q, WANG H, et al. Motif-based graph self-supervised learning for molecular property prediction[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 15870-15882.
- [19] FENG S B, WAN H R, WANG N N, et al. TwiBot-20: A comprehensive twitter bot detection benchmark[C]//*Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. New York: ACM, 2021: 4485-4494.
- [20] FENG S, TAN Z, WAN H, et al. Twibot-22: Towards graph-based twitter bot detection[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 35254-35269.
- [21] GUO Q, XIE H, LI Y, et al. Social bots detection via fusing bert and graph convolutional networks[J]. *Symmetry*, 2021, 14(1): 30.
- [22] FONSECA A J V, GHEDINI R C, COSTA G J J. Twitter bot detection with reduced feature set[C]//*Proceedings of the IEEE International Conference on Intelligence and Security Informatics*. New York: IEEE, 2020: 1-6.
- [23] YANG K C, FERRARA E, MENCZER F. Botometer 101: Social bot practicum for computational social scientists[J]. *Journal of Computational Social Science*, 2022, 5(2): 1511-1528.
- [24] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. *Journal of Machine Learning Research*, 2020, 21(140): 1-67.