基于异构图的Tor网络关键节点识别方法

王楠楠[1],黄 诚[1*],刘骏以[1],游 畅[1],时金桥[2]
(1. 四川大学网络空间安全学院,四川成都 610064;2. 北京邮电大学网络空间安全学院,北京 100876)

摘 要: 随着匿名通信需求的持续增长,Tor网络因其多条加密路由与去中心化架构,被广泛应用于隐私保护、反审查通信以及敏感信息传输等场景。然而,目前Tor网络中继节点数量增长缓慢、节点负载分布不均以及节点准入机制缺乏严格审查,使得部分高频参与路径构建的关键中继节点逐渐成为攻击者的重点操控目标。一旦这些节点被恶意控制,将显著削弱网络匿名性、破坏路径构建安全性,并对整体网络稳定性产生严重影响。因此,如何精准识别Tor网络中的关键中继节点,已成为提升匿名通信系统安全性与可靠性的核心问题。现有关键节点识别方法多基于静态拓扑结构指标或单一节点行为特征,难以有效刻画节点间复杂的隐式关系与多维语义联系,导致模型在面对真实动态网络环境时泛化能力与鲁棒性不足。针对上述问题,本文提出一种基于异构图建模与关系感知机制的无监督关键节点识别方法。首先,从节点稳定性、链路出现频次、功能标签及资源能力等多维属性出发,构建融合节点特征与隐式关系的多源异构图模型,实现对Tor网络结构的精细化表征;其次,引入关系感知异构注意力网络,对家族关系、自治系统归属关系、地理接近性关系以及路径共现关系等多类型边进行差异化建模,并通过异构注意力机制自适应融合不同关系语义信息,显著提升节点表示的判别性与鲁棒性;最后,在无监督学习框架下设计节点评分机制,实现关键中继节点的重要性排序与自动识别。基于真实Tor共识数据与实际链路构建数据开展系统实验评估,结果表明:在Top-100节点设置下,所提方法实现了85.0%的节点覆盖率,且识别节点的带宽比全网平均节点带宽高约25.9%。进一步实验表明:当移除模型识别出的关键节点后,网络整体带宽与链路覆盖率均显著下降,验证了所识别节点在网络运行中的核心作用。研究结果表明:本文方法能够有效刻画Tor网络中继节点的隐式关联结构,为匿名通信网络关键节点识别提供了一种新的建模范式,对提升Tor基础设施安全性具有重要理论意义与实践价值。

关键词: Tor网络;关键节点;深度学习;异构图;节点评分
基金项目: 国家重点研发计划(No.2023YFB3106600)
中图分类号: TP393.08 **文献标识码:** A **文章编号:** 0372-2112(2026)04-1534-16
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250881

Heterogeneous Graph-Based Identification of Critical Relays in Tor Networks

WANG Nannan[1], HUANG Cheng[1*], LIU Junyi[1], YOU Chang[1], SHI Jinqiao[2]
(1. *School of Cyber Science and Engineering*, *Sichuan University*, *Chengdu*, *Sichuan* 610064, *China*;
2. *School of Cyberspace Security*, *Beijing University of Posts and Telecommunications*, *Beijing* 100876, *China*)

Abstract: With the continuous growth of demand for anonymous communication, the Tor network has been widely adopted in privacy protection, censorship circumvention, and sensitive information transmission scenarios due to its multi-hop encrypted routing and decentralized architecture. However, the slow growth of relay nodes, uneven load distribution, and the lack of strict scrutiny in relay admission mechanisms have caused certain high-frequency relays involved in path construction to gradually become prime targets for adversarial control. Once these relays are maliciously compromised, the anonymity of the network will be significantly weakened, the security of circuit construction will be undermined, and the overall network stability will be severely affected. Therefore, accurately identifying critical relays in the Tor network has become a core issue for enhancing the security and reliability of anonymous communication systems. Existing critical node identification methods mainly rely on static topological metrics or single behavioral features, which fail to effectively capture complex implicit relationships and multidimensional semantic associations among relays, resulting in limited generalization ability and robustness when facing real dynamic network environments. To address these challenges, this paper proposes an unsupervised critical relay identification method based on heterogeneous graph modeling and relation-aware mechanisms. First, a multi-source heterogeneous graph model is constructed by integrating multidimensional attributes such as relay stability, path occurrence frequency, functional labels, and resource capabilities, enabling a fine-grained representation

收稿日期:2025-11-22;录用日期:2026-01-06;责任编辑:宋甲英
*通信作者:黄诚

of the Tor network structure. Then, a relation-aware heterogeneous attention network is introduced to differentially model multiple types of relations, including family relations, autonomous system affiliation, geographic proximity, and path co-occurrence. An heterogeneous attention mechanism is further employed to adaptively fuse diverse relational semantics, significantly enhancing the discriminative power and robustness of relay representations. Finally, an unsupervised scoring mechanism is designed to rank and automatically identify critical relays. Extensive experiments conducted on real Tor consensus data and practical circuit construction data demonstrate that, under the Top-100 relay setting, the proposed method achieves a relay coverage rate of 85.0%, and the bandwidth of the identified relays is approximately 25.9% higher than the network average. Further experiments show that removing the identified critical relays leads to a significant degradation in overall network bandwidth and circuit coverage, validating the core role of these relays in network operation. The results indicate that the proposed method effectively captures the implicit relational structure among Tor relays, providing a novel modeling paradigm for critical relay identification in anonymous communication networks and offering important theoretical and practical implications for enhancing the security of Tor infrastructure.

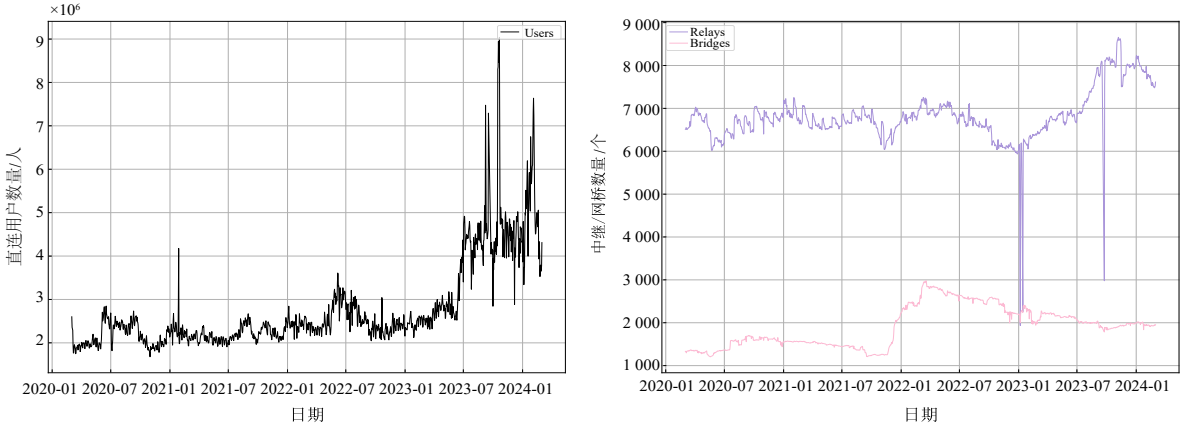
Keywords: Tor network; critical relays; deep learning; heterogeneous graphs; node scoring

Foundation Item(s): National Key Research and Development Program of China (No.2023YFB3106600)

0 引言

在当今高度互联的社会中,互联网已经成为人们获取信息、交流思想和业务处理的重要工具。依据中国互联网网络信息中心(China interNet Network Information Center,CNNIC)发布的《第54次中国互联网络发展状况统计报告》[1]显示(以下简称:《报告》),截至2024年6月,我国网民规模近11亿人,较2023年12月增长742万人,互联网普及率达78.0%。然而,信息化的普及也伴随着隐私泄露和安全威胁的加剧,个人隐私信息成为黑客攻击和数据滥用的主要目标。《报告》显示,遭遇信息泄露的网民比例常年保持在20.0%左右。此外,据计算机犯罪研究中心(Computer Crime Research Center,CCRC)[2]统计,到2025年底,网络犯罪造成的损失将超过12万亿美元,高于Cybersecurity Ventures在2020年所估计[3]的10.5万亿美元。上述趋势表明:网络隐私保护问题已具有全球性与长期性特征,并在学术界与工业界受到持续关注。

为应对这一问题,匿名通信技术应运而生,通过隐藏通信路径和用户IP地址,防止用户通信行为被追踪和分析。其中,Tor[4](The onion router)网络由于其开放性、去中心化以及多跳加密路由的设计,成为当前最流行的匿名通信工具之一。如图1所示,近年来,Tor网络用户数量持续增长,日活跃用户超过4×10^6人[5],但Tor中继节点数量增长缓慢,维持在7 000~8 000个[6],导致部分关键节点负载过重,网络的稳定性与匿名性面临挑战。



(a) 直连用户数量随时间变化　　(b) 中继/网桥数量随时间变化

(a) The number of directly connected users changes over time　　(b) The number of relays/bridges changes over time

图1 Tor Metrics统计信息

Figure 1 Tor Metrics statistics

Tor网络的安全性在很大程度上依赖于分布在全球各地的中继节点。然而,Tor网络的开放式节点加入机制缺乏有效的审查与监管,攻击者可通过Sybil攻击等方式批量部署伪装节点,从而提升在路径选择

中的选中概率。此外,目录服务器对节点带宽和稳定性等属性信息缺乏严格验证[7-9],进一步加剧了被攻击者操控路径构建过程的风险。若关键中继节点被攻击者控制,则将显著破坏网络匿名性,用户的真实身份与通信内容可能遭到泄露。因此,识别和保护关键中继节点,是提升Tor网络安全性的关键问题。

目前,已有研究围绕网络关键节点识别展开探索。一类方法侧重于使用静态拓扑结构信息[10](如基于度中心性[11-12]、介数中心性[13-14]、特征向量中心性[15-16]等图中心性指标)评估节点重要性,但这些方法忽略了节点属性和多维关系的差异,容易受到路径操控攻击影响。另一类研究尝试利用节点行为特征(如带宽、稳定性、在线时间等)结合机器学习模型挖掘关键节点,例如通过监督学习[17-18]框架结合标签特征识别高风险节点,引入强化学习[19]或图卷积网络[20]等深度模型对路径偏好与节点性能进行联合建模。然而,这些方法通常局限于单一视角,难以全面反映节点之间的隐式关系,同时无法直接适用于缺乏明确边关系的Tor网络,导致识别结果在泛化能力与鲁棒性方面仍存在欠缺。

为解决上述问题,本文提出一种基于异构图建模与关系感知机制的Tor关键节点识别方法。具体而言,本文首先对Tor网络中继节点的多维特征进行系统性建模,包括节点稳定性、链路出现频次、标签类型等多源属性,并通过统计分析与标签编码构建反映语义联系的异构图结构;其次,借鉴异构图神经网络(Graphic Neural Network,GNN)的设计思想,构建关系感知异构注意力网络模型,通过引入关系感知层对不同类型边进行建模,并利用异构注意力机制提升节点表示学习的区分性与鲁棒性;最后,结合无监督评分机制对节点的重要性进行量化识别,从而发现Tor网络中的关键中继节点。

本文的主要贡献包括:

(1)构建融合节点属性与隐式关系的Tor异构网络模型,全面刻画中继节点之间的关联关系;

(2)设计融合关系感知与异构注意力机制的无监督关键节点识别模型,提升节点表征质量并进行有效评分;

(3)在真实Tor共识与链路数据集上开展验证实验,从节点覆盖率、带宽影响度评估模型有效性。

综上所述,本文围绕Tor网络中继节点关键性识别问题展开研究,提出融合异构建模与关系感知机制的无监督识别方法,为匿名通信网络的结构优化与安全增强提供了新的建模视角与技术路径。

1 相关工作

关键节点[21]是指在网络中对整体功能、通连性或信息流动产生重大影响的部分节点,因此也常被称为影响力节点[22]或重要节点[23]。这些节点虽然数量相对较少,但其往往决定网络的稳定性和安全性。若关键节点遭受攻击,则网络的信息传输将受到严重影响,甚至导致网络瘫痪。因此,识别关键节点不仅有助于提高网络的鲁棒性和安全性,还能有效降低监控成本,增强防御策略。

关键节点识别技术作为复杂网络研究的重要方向,已广泛应用于社交网络、通信网络、交通网络及匿名通信网络等多个领域,其核心在于如何有效量化节点的重要程度。由于不同类型的网络在结构和信息传播机制上各具特点,可能涉及不同的节点类型及节点间关系。因此,依据不同的网络模型,关键节点识别方法可分为单层网络和多层网络两大主要模式。

1.1 单层网络关键节点识别

在单层网络中,由于其仅包含单一的节点类型和关系,关键节点识别方法通常集中在基于网络物理结构的中心性算法上,如度中心性(degree centrality)、介数中心性(betweenness centrality)、*K*-Shell算法和PageRank算法等。除上述传统方法外,近年来研究者们提出了诸多新兴算法以提升识别精度。例如,文献[24]提出的局部模糊信息中心度方法,通过聚类计算局部节点的信息量并进行重要性排序。文献[25]则通过分析不同阶数邻居数量的变化来评估节点的传播能力和重要程度。文献[23]综合考虑网络的局部与全局拓扑结构,进一步提升了节点识别精度和鲁棒性。文献[26]利用卷积神经网络对节点的影响力进行训练与预测,从而有效识别关键节点。这些创新方法为单层网络中的关键节点识别提供了更加灵活和高效的解决方案。

1.2 多层网络关键节点识别

与单层网络相比,多层网络的结构更加复杂,具有多维性和信息多样性,因此只有使用更为复杂的分析方法才能准确描述其拓扑结构和特性,常见分析方法包括网络中心性统计、多层网络特性和网络聚合等。例如,文献[27-28]提出了两种适用于多层网络的PageRank中心性算法,分别结合静态概率分布计算[27]和改进的重心度(gravity centrality)[28]算法,以识别网络中的关键节点。文献[29]则通过对不同节点间连接进行加权,改进了传统的重心度算法,并成功将其应用于多层网络。此外,部分研究引入了社区信息以提升关键节点识别的准确性。文献[30]提出了基于社区的影响力最大化(Community-based Influence Maximization,CIM)算法,通过融合节点所在社区的重要性信息,得到一组重要节点序列。然而,CIM算法假设网络的社区结构是已知的,这在某些网络中可能无法适用。文献[31]提出了一种基于社区结构的中

心性算法,在网络所有层中对节点重要性进行加权,但该方法需假设每一层使用相同的社区结果,显然存在局限性。总体来看,与成熟的单层网络方法相比,当前多层网络关键节点识别的研究仍较为薄弱,需要进一步深化与改进。

1.3　匿名通信网络关键节点识别

匿名通信技术近年来受到广泛关注,除常见的流量分析[32-36]、指纹分析[37-38]研究外,已有多项相关工作[39-40]对匿名通信技术的发展现状、典型体系架构及安全挑战进行了系统综述,为关键节点识别等安全防护研究奠定了理论基础。进一步地,文献[41]提出了Tor节点可靠性分析方案(Tor Nodes Reliability Analysis Scheme, TNRAS)。该方法从节点行为特征入手,选取节点稳定性、带宽等六方面指标训练模型,有效预测并剔除Tor网络中的Sybil恶意节点,筛选出高可靠节点用于实际通信。此外,文献[42]提出了基于节点匿名度的Tor网络匿名性评估方法,其通过监测每个节点行为的波动来量化节点的匿名性贡献,可在遭受分布式拒绝服务攻击(Distributed Denial of Service,DDoS)攻击时及时发现异常节点并评估全网匿名度的变化,这对于理解哪些节点被攻击时对匿名通信影响最大有重要意义。值得一提的是,近年来也有研究将匿名通信节点识别方法拓展到其他系统。有工作针对Tor网络的流量关联攻击,引入机器学习改进关键路径节点的发现。例如,文献[43]提出的流量关联图卷积网络(Flow Correlation Graph Convolutional Network,FlowCorr-GCN)方法,结合图卷积神经网络(Graph Convolutional Networks,GCN)与三元组网络,实现对洋葱服务流量的高效关联分析,为匿名网络关键节点和链路识别提供了新的技术手段。

1.4　GNN在关键节点识别中的应用

随着深度学习的发展,GNN已经被广泛用于复杂网络的关键节点识别与网络安全分析。文献[44]将图卷积网络引入节点影响力评估,其将关键节点识别视作回归预测任务,利用GCN提取节点的局部和全局特征来预测节点的重要度,从而更准确地挖掘网络中隐含的高影响力节点。文献[45]亦提出了基于GNN的机会网络节点重要度评估方法,在动态网络场景中取得了优异表现。此外,文献[46]提出融合自动编码器与GNN的深度模型自动编码器图神经网络(Autoencoder Graph Neural Network,AGNN)。该模型先利用GCN自动编码器学习节点的潜在表示,再结合传播模型和排序学习策略列表级最大似然估计(listwise Maximum Likelihood Estimation,listMLE)来优化关键节点排序,这类工作证明,将结构表示学习与节点重要性评估相结合是提升识别性能的有效途径。在网络安全领域,GNN同样展示出强大潜力。例如,文献[47]将图卷积与长短期记忆网络结合,针对动态时序网络预测每一时刻传播能力最强的关键节点。该动态GNN模型在时变网络中捕捉传播模式,用于识别在不同时间快照中最易引发大规模传播的节点。这对具有演化特性的通信网络(包括匿名网络中节点上下线变化)具有借鉴意义。

2　方法

在本研究中,“关键节点”是指在Tor网络运行过程中,对链路构建稳定性、带宽承载能力以及匿名性保障具有显著影响的中继节点。具体而言,关键节点通常具备以下特征:(1)在真实链路中出现频次较高,其被选中构成通信路径的概率显著高于普通节点;(2)在网络流量传输中承担重要角色,如高速出口节点或高可用入口节点;(3)其失效会对网络连通性、带宽资源或匿名性造成客观影响。因此,本文将节点的重要性视为节点在网络结构、属性特征与隐式关系三者综合影响下的全局贡献度,并以此作为模型识别的目标。

本节将介绍面向Tor网络关键节点识别方法,整体研究框架如图2所示。首先,通过对节点信息的统计分析,从节点稳定性、节点被选择率等多方面对节点进行重要性分析,该分析补全了现有关键节点识别在刻画维度上的不足,同时为后续节点识别提供依据。其次,通过综合考虑节点属性和隐式关系,构建一个异构图模型,全面刻画Tor网络的潜在结构特性。最后,通过结合关系感知层和异构注意力机制,实现对节点重要性评分的自动计算,并采用Top-K策略识别不同角色中的关键节点。

2.1　节点关键性分析

2.1.1　在线时长分布分析

Tor网络的核心由多个路由节点组成,其节点在线时长和上下线频率直接影响网络的性能与稳定性。本文通过分析Tor系统每小时发布的微描述共识数据,统计节点的在线状态。具体来说,若一个节点出现在某个小时发布的微描述共识中,本文则认为其在该小时内为在线状态(记为“1”),否则为离线状态(记为“0”)。

基于上述规则,本文即可构建起节点在线时间状态矩阵,该矩阵能够反映一段时间内节点的上下线情况。2024年11月微描述共识共出现的公共节点数为9 647个,发布共识文件712份,因此该矩阵维度为712 × 9 647。同时,为了更直观地反映节点的在线时长,本文参考位图[48](bitmap),通过绘制图像实现了矩阵数据可视化,具体如图3所示。其中每行代表一个

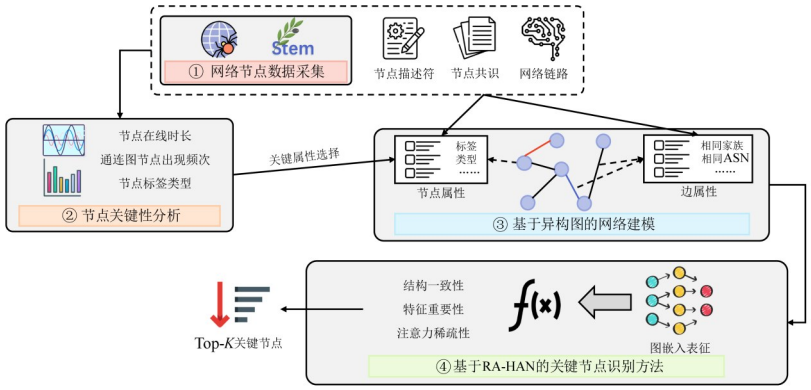
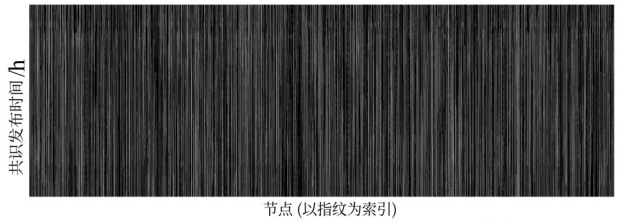


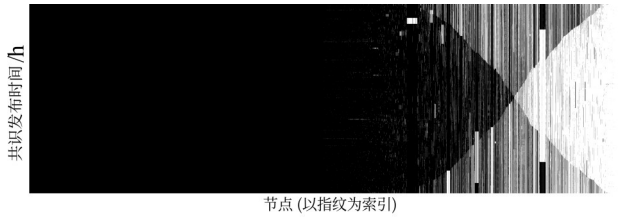
图2 Tor网络关键节点识别整体设计方案

Figure 2 Overall design for Tor critical relay identification



(a) 节点在线时长分布(未排序)

(a) Node online time distribution (unsorted)



(b) 节点在线时长分布(按在线时长排序)

(b) Node online time distribution (sorted by online time)

图3 Tor节点在线时长分布

Figure 3 Distribution of Tor relays online hours

小时,每列代表唯一节点,行列交叉处的黑色像素代表该节点在该小时内为在线状态,反之,白色像素代表该节点在该小时内为离线状态。

由图3可知,大部分节点在线状态稳定,然而,仍有近半数节点在不同时间段存在波动,这表明系统性事件或区域性网络故障可能同时影响多个节点。同时,图中白色区域呈现对角线模式且出现黑色柱状图形,可能因区域性网络故障、系统维护或主机群控等原因导致节点周期性上下线。因此,若网络中的关键节点频繁波动,则可能严重影响Tor网络的匿名性、可用性和路径选择稳定性,尤其当流量较多的节点频繁变化时,更易引发网络性能问题。

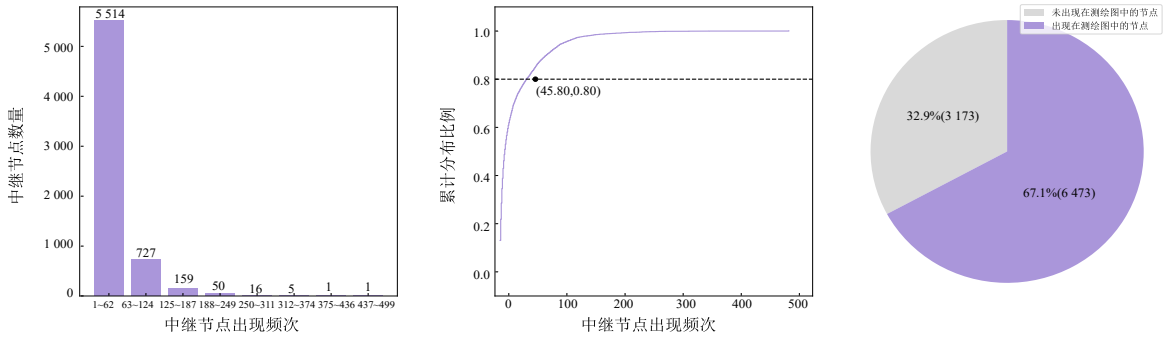
2.1.2 通连图出现频次分布分析

Tor网络依靠节点链路实现数据传输,节点的安全性和链路选中频率直接影响网络稳定性。本文使用Tor客户端自动构建链路,并以GETINFO指令获取链路信息,从而获取一定时间范围内不同时刻的链路信息,并将其存储于文件中。具体来说,若一个节点出现在某个时刻Tor的链路信息中,本文则认为其在该时刻被Tor客户端选中作为链路的节点之一。此外,由于节点在链路中的不同位置将扮演不同的角色,因此本文将依据节点的出现位置不同分别对节点出现频次进行统计。

为了获得真实且可复现的链路选取数据,本文部署了一个位于香港地区的Tor客户端节点(Ubuntu Server 22.04 LTS,Tor 0.4.8.10),使用Stem控制接口生成链路,具体可参考4.1.2节。基于上述规则,本文统计了每个节点在2024年11月中分别于中间节点和出口节点出现的次数,并分别绘制了对应的出现频次分布图、出现频次累积分布图以及出现节点占比图,从而获得更清晰的展示。出现频次分布图能够反映不同出现频次区间的节点数目情况,本文将区间[0,该类节点最大出现频次]进行n(n=8)等分,以作为该图的横坐标区间;出现频次累积分布图能够反映完整描述节点出现频次的概率分布,本文绘制了累积概率分布占比为m(m=80%)所对应的直线,以了解大部分节点出现的次数范围;出现节点占比图能够反映现在通连图中的节点占全部该类型节点的比例。

对于中间节点,具体统计数据如图4所示。统计结果显示,节点最多出现499次,其中5 514个节点出现频次低于62次,占比达57.1%;80%的节点出现次数少于45.8次。整体来看,仅67.1%的节点作为中间节点被使用过,这说明节点的选用概率存在明显差异,部分节点因被频繁选用,成为影响网络稳定性的关键因素,这些节点一旦遭到攻击,将对网络性能造成显著影响。

对于出口节点,具体统计数据如图5所示,统计结果显示,节点出现频次最高达614次,其中1 441个节点的出现次数不超过76次,占总数的53.5%;80%的节点

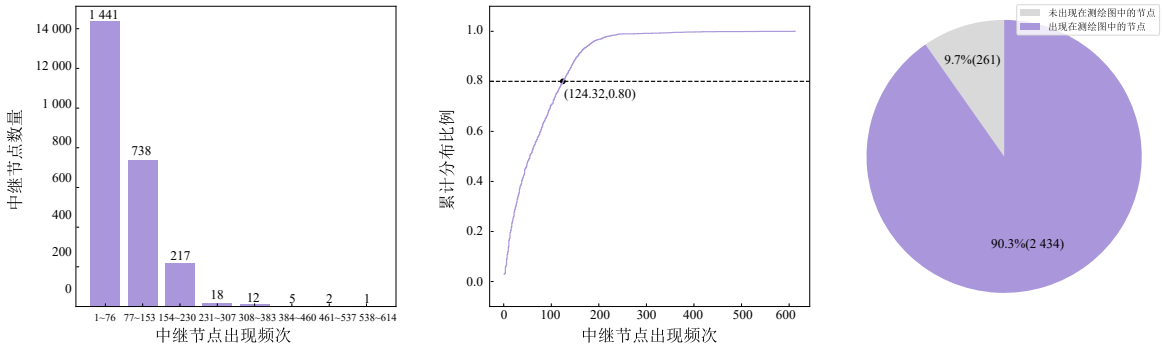


(a) Middle 节点出现频次分布　(b) Middle 节点出现频次累计分布　(c) Middle 节点在测绘图中的分布

(a) Frequency distribution of Middle　(b) Cumulative frequency distribution of Middle　(c) Distribution of Middle in mapping

图 4　Middle 节点出现频次统计

Figure 4　Frequency statistics of Middle relays



(a) Exit 节点出现频次分布　(b) Exit 节点出现频次累计分布　(c) Exit 节点在测绘图中的分布

(a) Frequency distribution of Exit　(b) Cumulative frequency distribution of Exit　(c) Distribution of Exit in mapping

图 5　Exit 节点出现频次统计

Figure 5　Frequency statistics of Exit relays

出现次数小于 124.3 次,但超过 90% 的出口节点曾被使用过。尽管出口节点的数量较少,但在 Tor 网络中频繁被选中,这提高了出口节点被攻击的风险和影响网络稳定性的可能性。尽管 Tor 官方已建议个人用户避免部署出口节点,但由于出口节点在匿名通信的重要性,仍有攻击者持续部署此类节点,构成安全隐患。

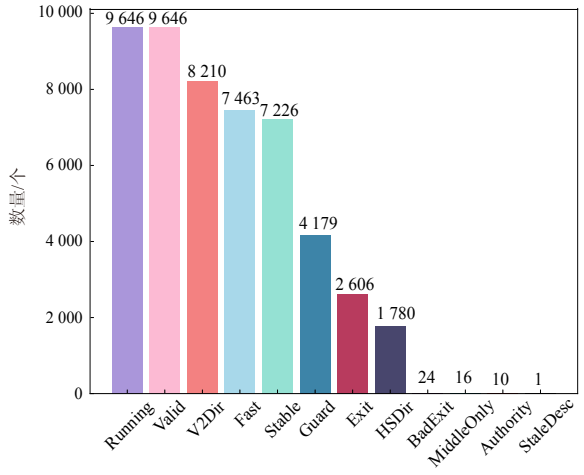
对于入口节点,由于其具有较强的持久性和稳定性,Tor 客户端通常在较长时间内固定使用极少数 Guard 节点。在本实验中,由于 Stem 所生成的链路在入口位置高度集中于个别固定 Guard 节点,因此并未对其进行统计工作。

2.1.3　标签类型分布分析

Tor 网络中的节点在运行过程中将被赋予不同的标签,以表示节点的具体功能、性能和稳定性。本文对 9 647 个节点的标签类型进行了详细统计分析,并绘制了数据集中的标签数量分布,如图 6(a)所示。统计结果显示,所有节点均被赋予 Running 和 Valid 标签,这表明 Tor 网络及时剔除了不合规的节点,网络整体运行状态保持稳定。此外,具有 V2Dir、Fast、Stable、Guard 标签的节点也较多,这些标签分别反映节点在目录服务、带宽能力和稳定性方面表现优异。而具有 Exit、HSDir、BadExit 标签的节点数量则明显较少,这表明能够承担出口功能的节点数量有限,且部分节点可能存在安全问题,需要重点关注。

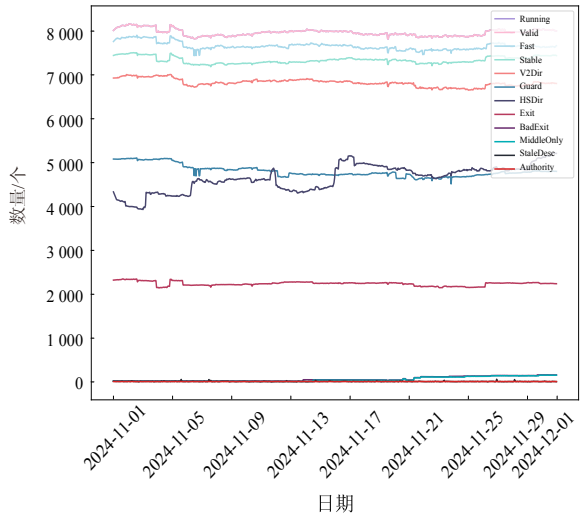
为了进一步分析不同标签的动态变化,本文统计了各类标签随时间的数量变化趋势,如图 6(b)所示。由变化趋势可知,部分标签(如 Running、Valid)的节点数量较为稳定,而另一部分标签(如 Guard、Exit)的数量存在一定的波动。这种波动可能是因受到节点上下线行为、带宽调整以及 Tor 网络自身运行机制的影响。通过对标签数量的统计和变化分析,可更好地理解 Tor 网络中继节点的分布特征,为后续研究关键节点的识别提供数据支持。

为了确保对多标签节点的统计不会产生重复计数或语义冲突,本文在处理 Tor 节点标签时采取了如下方法:首先,Tor 中继节点可能同时拥有多个功能性标签,如同时具备 Guard、Exit、Fast 等属性,因此本文将每个节点的标签视为一个标签集合(label set),并对其进行结构化存储,而不是将多标签节点拆分为多个独立样本。其次,在统计标签数量时采用“一节点



(a) 数据集标签数量统计

(a) Dataset label count statistics



(b) 各类标签数量变化

(b) Variation in the number of labels per category

图6 节点标签类型分布

Figure 6 Distribution of relay label types

多计数”策略,即每个节点在拥有某一标签时仅在该标签对应类别下计为一次,而不会在整体节点数量中被重复添加。

2.2 基于异构图的Tor网络建模

在Tor网络研究中,其匿名通信机制导致网络拓扑结构无法被外界直接获取,进而无法大量获取节点之间真实的连接关系,只能借助节点公开信息和隐含关联来重构网络表示。为解决这一问题,本文构建了一个合理的异构图模型来描述Tor网络结构,图中节点对应Tor网络中的中继节点,边表示节点间的隐式关系。为系统刻画Tor网络中继节点及其多维关系结构,本文将构建的网络表示为一个异构图,计算式为

$$G=(\mathcal{V},\mathcal{E},\mathcal{A},\mathcal{R}) \tag{1}$$

式中,$\mathcal{V}$为节点集合,每个节点对应一个Tor中继;$\mathcal{A}$为节点属性集合,包括带宽、运行时间、标签向量、地理位置等多维特征;$\mathcal{R}$为关系类型集合,包含家族关系、自治系统(Autonomous System,AS)归属关系、地理接近性关系以及路径共现关系四类边。对于任意关系类型$r\in\mathcal{R}$,其对应的边集合定义为

$$\mathcal{E}_r=\left\{\left(v_i,v_j\right)\in\mathcal{V}\times\mathcal{V}\middle|\left(v_i,v_j,r\right)\in\mathcal{E}\right\} \tag{2}$$

并为每个关系类型构建独立的邻接矩阵$\boldsymbol{A}^{(r)}$。此外,节点特征向量表示为

$$\boldsymbol{X}_i=\left[X_i^{\text{role}},X_i^{\text{resource}},X_i^{\text{stability}},X_i^{\text{geo}}\right] \tag{3}$$

用于统一描述节点的类别属性、资源能力、稳定性指标与地理信息。上述形式化定义为后续关系感知异构注意力网络提供了严格的数学输入结构,使模型能够在多关系环境中进行有效学习。

下面将详细介绍基于节点属性的图节点构建方法和基于隐式关系的图边构建方法。

2.2.1 基于节点属性的图节点构建

Tor网络中的节点均具有特定的属性信息,这些信息不仅能够刻画节点的基本属性,还可能影响其在通信链路中的作用。因此,在构建图的过程中,首先需定义节点的特征,以便后续的结构分析和关键节点识别。本节将从节点属性选择和特征嵌入两个方面展开介绍。

(1)节点属性选择

Tor网络的节点属性多种多样,需要确定能够有效反映Tor网络中的关键特征,并剔除冗余或无关属性。综合考虑网络拓扑结构、节点功能及匿名性影响,本文将节点属性划分为基础信息、角色信息、资源能力和稳定性信息四个类别,具体如表1所示。

表1 关键节点属性及表征方法

Table 1 Critical relay attributes and attribute representations

特征类别	具体特征名称	表征方法
基础信息	监听端口	One-Hot编码
	AS归属	One-Hot编码
	地理位置	One-Hot编码
角色信息	节点类型	One-Hot编码
	节点标签	One-Hot编码
资源能力	声明带宽	Min-Max归一化
	IPv6兼容性	0/1二值编码
稳定性信息	运行时间	Min-Max归一化
	BadExit标签	0/1二值编码

角色信息反映了节点的具体功能及其在路径选择中的重要性。入口节点是用户进入Tor网络的第一跳,稳定性高的入口节点更易被选中,但长期使用相同入口可能增加流量分析风险。中继节点负责数据转发,影响网络连通性。出口节点是流量的最终出

口,可观察明文流量,若被攻击者控制,则可能导致匿名性降低。此外,节点标签影响路径选择。例如,Fast和Stable标签表示高带宽、长期在线的节点,通常更易被选为路径组成部分。

资源能力决定了节点在路径选择中的竞争力,主要包括声明带宽、协议支持。声明带宽是指节点自报的最大可用带宽,高带宽节点更易被选为路径组成部分。而是否支持IPv6关系到其在新型网络环境下的可达性,支持IPv6的节点能够为更多终端提供连接能力,在路径选择中具备更强的适配性和优先级,从而进一步增强其竞争优势。

(2)节点特征向量化

为了有效利用Tor网络的节点属性进行分析,需将不同类型的特征转换为统一的数值表示,使其适用于机器学习和图模型分析。本节采用类别变量嵌入、数值变量归一化两种方式,以提高模型对不同类型数据的适应性,并确保信息表达完整。各类特征对应嵌入方式如表1所示。

(a)类别变量嵌入

类别变量通常是离散的、无序的值,如节点类型、监听端口、节点标签、AS归属等,不能直接用于计算,因此需转换为数值形式。本文直接采用0/1二值编码和One-Hot编码进行转换。

(b)数值变量归一化

数值变量(如带宽、运行时间等)具有不同的取值范围,若直接输入模型,可能导致不同变量的影响权重不均衡。因此,对于数值分布均匀的变量,采用Min-Max归一化将特征值缩放至[0,1],具体计算式为

$$x'=\frac{x-\min(x)}{\max(x)-\min(x)} \tag{4}$$

此外,对于具有长尾分布的变量,Min-Max归一化可能仍导致极端值影响较大,因此先进行对数变换,具体计算式为

$$x'=\ln(x+1) \tag{5}$$

2.2.2　基于隐式关系的图边构建

在Tor网络分析中,仅依靠节点属性无法完整描述其运行特性,节点之间的关系(边)同样是关键因素。由于缺乏Tor网络直接的拓扑边信息,因此,本文从Tor节点之间的隐式关联出发,定义多种类型的关系边来丰富图结构。具体而言,本文引入了家族关系边、AS归属边、地理接近性边和路径共现边四种边来描述节点间潜在关联,并依据其对匿名通信的影响设定不同的权重,以区分正负关系。各类关系边的构建方法和含义如下。

(1)家族关系边

若两个节点在Tor目录中声明其属于同一个家族(Family),则在该两个节点之间连接一条家族关系边。在Tor网络中,当客户端进行路径选择时会避免同一家族的节点出现在同一线路中,因为这些家族节点通常由同一运营者控制,同时失陷或被控制会削弱整体网络匿名性。因此,本文将家族边视作一种负相关关系,其权重设为-1,表示两节点不应同时出现在同一路径内,强调这种节点组合所带来的安全风险。该负权重在图模型中促使算法在聚合邻居信息时降低家族关联节点对彼此的重要性贡献。

(2)AS归属边

AS归属关系用于描述两个节点是否处于同一AS管辖下。若两个中继的IP地址属于同一AS,则本文在它们之间添加AS归属边。该关系反映了潜在的网络间路径依赖,若入口和出口中继处于同一AS,则该AS有能力监测到流量的两端,从而导致匿名流量面临被单一实体对象同时观察的风险。鉴于此安全隐患,本文将AS归属边同样视作负面关系,其权重赋值为-1,表示应降低同一AS内节点组合在路径中的机会。通过在图中引入AS关联,从而帮助识别受到AS级别攻击威胁的潜在节点对。

(3)地理接近性边

地理位置接近的中继节点可能具有相似的管辖和监控环境。例如,位于同一国家或地区的节点可能受到同一管制机构影响,致使某些国家级攻击者可同时对该区域内的多个节点实施流量分析。因此,本文通过定义地理接近性边以连接地理位置邻近的节点,例如,在处于相同国家/地区的两节点间添加边。与家族和AS不同,地理接近性本质上是一种中性关系,其并不必然意味着恶意或协同作用,代表了一种潜在的相关性。本文为此类边赋予中性权重0.5,表示适度相关但非强关联。该权重设定使模型在聚合时能够感知地缘因素对路径选择的影响(例如,某些客户端可能倾向于选择地理位置更近的节点以降低延迟),同时不会过度强化这种关系以免影响匿名性。

(4)路径共现边

路径共现关系描述了两个节点在通信链路中频繁连续出现的情况。若某对节点(如Guard和Exit)在模拟电路或Tor流量记录中的共现频率显著高于随机期望,则在它们之间构建路径共现边。该关系能够反映Tor路径选择的偏好,例如,高带宽的入口和出口节点常被同时选用组成高速通道。本文将该关系视为正向关系,赋予权重1,使GNN聚合时增强此类节点的影响力,从而识别高频服务用户的关键节点组合。此外,路径共现关系在一定程度上弥补了拓扑信息的缺失,帮助模型推测对匿名流量影响较大的节点对。

通过引入隐式关系边,异构图模型能够更全面地刻画Tor网络中节点之间的潜在联系。模型能够捕捉Tor路径选择的显式规则[49]和业界建议的改进策

略[50]；同时，融入了网络使用过程中的统计特征，从多个角度揭示潜在的安全隐患和性能因素。

2.3 基于RA-HAN的关键节点识别方法

虽然GNN在复杂网络分析中被广泛应用，但其在Tor网络场景下并不能直接适用。原因主要包括三个方面：首先，传统同质图GNN[如GCN、图采样与聚合(Graph Sample and AggreGatE，GraphSAGE)]假设图中仅包含单一节点类型与边类型，而Tor网络节点具有多维异构属性，且节点之间存在多种隐式关系(如家族关系、AS归属、地理接近性与路径共现)，使用同质图模型会导致语义混淆，难以准确刻画节点间的结构差异；其次，现有面向属性图的模型[如异构图注意力网络(Heterogeneous graph Attention Network，HAN)、R-GCN]通常依赖显式关系或已知拓扑，而Tor网络缺乏公开的真实拓扑结构，仅能从统计行为中推断隐式关系，这需模型能够对每类关系分开建模并进行加权融合；最后，Tor网络缺乏关键节点的标注数据，使得监督式GNN无法直接训练。因此，仅采用通用GNN模型难以满足Tor网络的异构性、弱拓扑性及无监督特性需求。

本节基于上述声明的Tor网络异构图模型，提出一种基于关系感知异构注意力网络(Relation-Aware Heterogeneous Attention Network，RA-HAN)的关键节点识别模型。该模型结合关系感知层(Relation-Aware Layer，RAL)和异构注意力机制(Heterogeneous Attention Mechanism，HAM)，能够在无监督学习框架下，自适应地计算节点在网络中的重要性评分，进一步通过Top-K选择策略识别不同角色中的关键节点。

2.3.1 关系感知层

在异构图中，不同类型的节点关系(边)承载着多样化的语义信息。若简单地将所有关系的邻居节点信息混合聚合，则易导致语义混淆和信息干扰，从而影响模型对不同粒度关系模式的捕捉。例如，Tor网络中的“家族关系”和“地理接近性关系”分别表示节点间的路由冲突和地域上的安全风险，由于这两类关系的语义差异显著，模型需在聚合阶段对其进行差异化处理。

为了解决这一问题，本文借鉴R-GCN[51]的消息传递思想，为RA-HAN模型引入了关系感知层(relation-aware layer)，即通过为每种关系类型 $r\in\mathcal{R}$ 定义独立的线性变换矩阵 $\boldsymbol{W}_r$，使模型能够分别聚合不同关系的邻居信息，从而确保独立计算不同关系类型对节点表示的贡献。此外，在模型实现中，本文还引入了自环连接(self-loop)的权重矩阵 $\boldsymbol{W}_0$，即引入节点自身信息，使节点表示不仅依赖邻居节点，还能包含自身原始特征，从而增强模型的表达能力。

具体而言，对于目标节点 i，模型首先将属于关系 r 的每一个邻居节点 j 的特征向量 $\boldsymbol{h}_j$ 乘以对应的 $\boldsymbol{W}_r$，实现特征空间的变换。其次，在同一关系类型内部，对经过变换的邻居特征进行聚合，得到节点 i 在关系 r 下的中间表示。此外，在进行所有关系的聚合时，模型还会通过归一化因子 $c_{i,r}$ 平衡不同邻居数量对特征聚合的影响。最后，节点 i 在第 $l+1$ 层的表示更新式为

$$\boldsymbol{h}_i^{(l+1)}=\sigma\left(\boldsymbol{W}_0\boldsymbol{h}_i^{(l)}+\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_i^r}\frac{1}{c_{i,r}}\boldsymbol{W}_r\boldsymbol{h}_j^{(l)}\right)\tag{6}$$

式中，$\mathcal{N}_i^r$ 为节点 i 通过关系 r 连接的邻居集合；$\sigma(\cdot)$ 为非线性激活函数(如ReLU)；$c_{i,r}$ 为归一化因子(通常设置为 $\sqrt{\left|\mathcal{N}_i^r\right|}$)，以平衡不同邻居数量的影响；$\boldsymbol{h}_j^{(l)}$ 为节点 j 在第 l 层的特征表。

这种按关系分类聚合的机制，使模型在节点表示中既保留了节点自身信息，又能够从多种关系类型的邻居中提取语义特异性的信息，为后续的注意力机制提供了多样化的特征输入。

2.3.2 异构注意力机制

虽然关系感知层能够在模型中区分不同关系类型的语义，但在具体的关系内部，不同邻居节点的贡献度也可能存在差异。此外，从全局视角来看，不同关系类型对节点表示的重要性也可能不尽相同。为解决这一问题，在完成关系感知层的基础表示后，本文进一步通过异构注意力机制[52]进行更高级的特征聚合。该机制通过为不同的邻居和关系分配可学习的权重，实现对重要信息的自适应聚合，本文主要包含两种注意力。

(1)节点-邻居级注意力(Node-Neighbor Level)。在每种关系 r 下，模型通过注意力机制计算节点 i 及其邻居 j 之间的注意力权重 α_{ij}^r，从而识别在特定关系语境中最重要的邻居节点。具体计算式为

$$\boldsymbol{\alpha}_{ij}^r=\frac{\exp\left(\mathrm{LeakyReLU}\left(\boldsymbol{a}_r^{\mathrm{T}}\left[\boldsymbol{W}_r\boldsymbol{h}_i\|\boldsymbol{W}_r\boldsymbol{h}_j\right]\right)\right)}{\sum_{k\in\mathcal{N}_i^r}\exp\left(\mathrm{LeakyReLU}\left(\boldsymbol{a}_r^{\mathrm{T}}\left[\boldsymbol{W}_r\boldsymbol{h}_i\|\boldsymbol{W}_r\boldsymbol{h}_k\right]\right)\right)}\tag{7}$$

式中，$\boldsymbol{a}_r$ 为关系特定的可学习向量，用于计算节点 i 和邻居 j 在关系 r 下的注意力分数。

(2)关系类型级注意力(relation-type level)。模型在节点 i 已经得到各关系类型 r 下的邻居聚合表示，随后计算不同关系的重要性权重 β_i^r，从全局视角平衡不同关系类型的影响力。具体计算式为

$$\beta_i^r=\frac{\exp\left(\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}_i^r\right)}{\sum_{t\in\mathcal{R}}\exp\left(\boldsymbol{q}^{\mathrm{T}}\boldsymbol{z}_i^t\right)}\tag{8}$$

式中，$\boldsymbol{q}$ 为全局关系的可学习权重向量，用于评估关系类型 r 对节点 i 的影响力；z_i^r 为节点 i 在关系 r

下的特征聚合表示,通常通过节点-邻居级注意力聚合得到的特征。

聚合表示。模型通过节点-邻居级和关系类型级的注意力权重相结合,得到节点 i 的最终表示。具体计算式为

$$\boldsymbol{h}_i^{(l+1)}=\sigma\left(\sum_{r\in\mathcal{R}}\beta_i^r\sum_{j\in\mathcal{N}_i^r}\alpha_{ij}^r\boldsymbol{W}_r\boldsymbol{h}_j^{(l)}\right) \tag{9}$$

这种双重加权聚合的方式,使RA-HAN在处理多关系异构图时,能够精准识别关键的邻居节点和关系类型,从而显著提升模型的准确性和鲁棒性。

2.3.3　损失函数与节点评分

在完成对节点信息的有效聚合后,本文需设计合适的机制来实现关键节点的识别。然而,由于缺乏明确的监督信号(即关键节点标签),本文将采用无监督学习框架,因此损失函数不仅要优化节点表示,还需确保模型在无标签数据的情况下,仍能有效区分关键节点与普通节点。为此,本文设计了三种损失函数,确保节点评分 S 合理且符合网络结构特性,具体内容如下。

(1)结构一致性损失。该损失约束相连节点的评分,使其在局部网络结构上保持连续性,具体计算式为

$$L_{\text{struct}}=\sum_{(i,j)\in E}\left|S_i-S_j\right| \tag{10}$$

(2)特征重要性损失。该损失鼓励模型关注高影响力的特征,使关键节点的评分更具区分度。具体计算式为

$$L_{\text{feature}}=-\ln\left(\sigma\left(\text{mean}(S)\right)+\epsilon\right) \tag{11}$$

(3)注意力稀疏性损失。该损失约束注意力权重的分布,使其更具选择性,减少对不重要关系的关注。具体计算式为

$$L_{att}=\left\|\boldsymbol{A}_{\text{relation}}\right\|_1 \tag{12}$$

式中,$\boldsymbol{A}_{\text{relation}}$ 为所有关系级别注意力权重矩阵。

最终,在优化损失的基础上,模型使用多层感知机(Multi-Layer Perceptron,MLP)作为评分层,计算节点的重要性评分,并根据Top-K排序以识别Tor网络中的关键节点。

3　实验

本节将先介绍数据集,再从关键节点识别模型评估和模型有效性验证两大方面进行实验设置,并对实验结果进行分析。

3.1　数据集

3.1.1　Tor官方数据

由于实时的Tor网络研究存在诸多局限性,因此,与大部分现有研究[53-54]类似,本文主要基于Tor的历史数据开展研究[8,55]。Tor官方项目组提供了一个Tor生态系统历史数据存储与发布平台Tor Metric,其根据数据功能的不同,将历史数据分为Tor中继描述符、Tor网桥描述符、Tor Web服务器记录等多种类别。本研究主要基于2024年11月的Tor中继服务器描述符和微描述共识。

3.1.2　自生成数据

除与节点相关的描述符合共识外,本文通过在香港服务器上部署Tor客户端,使用Python脚本实现与客户端的自动化交互,从而在真实Tor网络中重复构建与更新网络链路,获取有关Tor链路的相关数据。

Stem库[56]是Tor官方构建的一个Python库,用户可结合Stem和Tor控制协议来编写控制Tor进程的Python脚本,或者构建诸如Nyx(Tor命令行监视器)之类的工具。本研究主要通过使用GETINFO函数获取三跳路由作为Tor链路信息,同时使用SIGNAL函数向客户端定期发送NEWNYM信号,控制Tor切换到干净链路,从而保证链路数据多样性。具体链路数据主要由链路ID、链路创建时间以及三跳路由的昵称、IP和指纹构成,包含2024年11月生成的链路数据176 805条,以及12月生成的链路数据39 676条。

3.2　实验设置

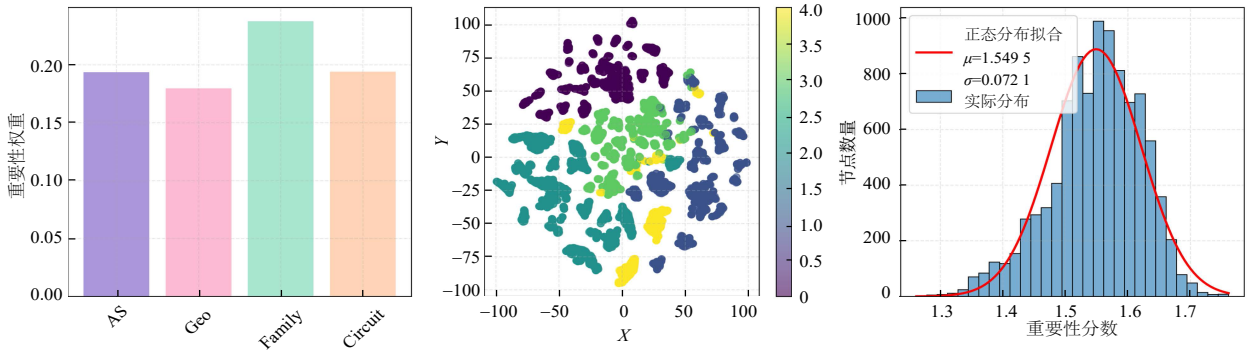
本次实验主要分为两部分。第一部分为关键节点识别模型的性能评估实验,旨在分析本文提出的关键节点识别方法的整体效果及Top-K节点的识别能力。整体效果评估包括对模型输出的节点重要性分数分布以及各类关联关系重要性的深入分析,以验证模型在理解Tor网络整体结构方面的有效性。Top-K节点分析则具体针对模型识别出的排名靠前的节点,观察其特征属性(如带宽)的分布情况,以进一步验证关键节点的识别质量。第二部分为模型有效性验证实验,旨在从两个维度验证本文提出的关键节点识别方法的实际有效性。首先,通过在真实Tor网络环境中构建大量电路,分析模型所识别的Top-K关键节点对网络链路的覆盖程度。若Top-K节点能够覆盖绝大部分网络链路,则说明模型能够准确反映Tor网络的实际路径选择偏好,进一步验证模型的合理性。其次,模拟攻击场景,通过移除模型识别出的Top-K关键节点,观察网络整体带宽性能的变化。若移除这些节点后网络的带宽明显下降,则充分证明模型所识别的节点对网络性能具有重要影响,从而证实模型的有效性和识别结果的可靠性。

3.3　实验结果与分析

3.3.1　关键节点识别模型评估实验结果

为验证所提的基于RA-HAN的关键节点识别方法的可行性,实验对模型的输出结果进行了详细评估。第一个实验通过分析关系类型重要性、节点嵌入聚类以及节点重要性分布来评估模型对Tor网络关

键节点的识别能力，整体效果评估如图7所示。在关系类型重要性分析中，模型对Tor网络中的不同隐式关系赋予了不同的权重。从结果来看，家族关系（Family）具有最高的权重，而地理位置（Geo）权重相对较低，这表明家族成员之间的关联性在关键节点识别中起到了较大的作用。



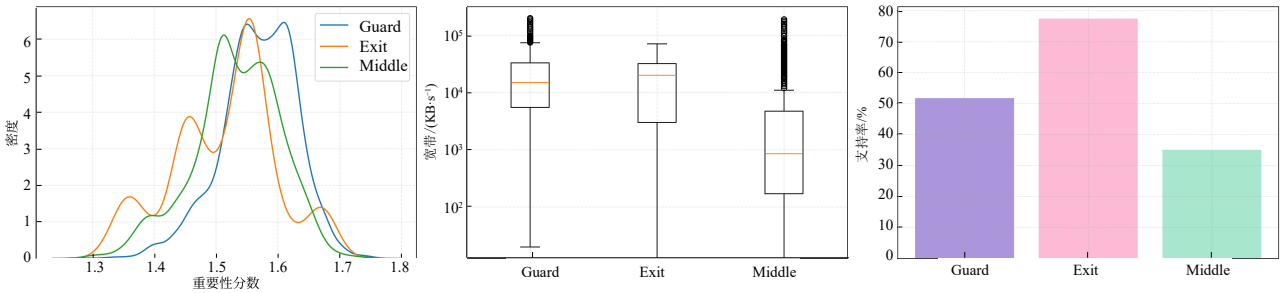
(a) 关系类型重要性分布　(b) 节点嵌入聚类分布　(c) 节点重要性分数分布

(a) Relation type importance distribution　(b) Node embedding clustering distribution　(c) Node importance score distribution

图7 整体效果评估

Figure 7 Overall evaluation of the model

节点嵌入的可视化结果表明模型在低维空间中形成了明显的聚类结构。这说明模型能够有效捕捉节点之间的特征相似性。此外，该结构使得不同类别的节点在嵌入空间中呈现出更好的区分性。与此同时，节点重要性分数的统计分布结果表明：节点评分整体服从正态分布，其均值约为1.55，标准差为0.07。这说明模型生成的重要性评分具有较好的稳定性，不存在极端值或评分偏移现象，进一步验证了评分的合理性。在此前分析的基础上，本文进一步统计不同类型节点的特征差异，结果如图8所示。从评分分布来看，入口节点（Guard）的评分整体较高，表明其在Tor网络中的重要性较大，这与其承担流量入口、保障匿名性的功能相符。相比之下，出口（Exit）和中间（Middle）节点的评分分布较为接近，但仍存在一定的区分度。在带宽方面，出口节点的带宽中位数最高，且分布较为分散，说明出口节点由于其稀缺性往往需要有高带宽支持，而入口和中间节点的带宽较为稳定。此外，在IPv6支持率上，出口节点支持IPv6的比例最高，超过70%，而入口和中间节点相对较低，这表明出口节点更倾向于适配更广泛的网络环境。



(a) 不同类型节点重要性分数分布　(b) 不同类型节点带宽分布　(c) 不同类型节点IPv6支持率

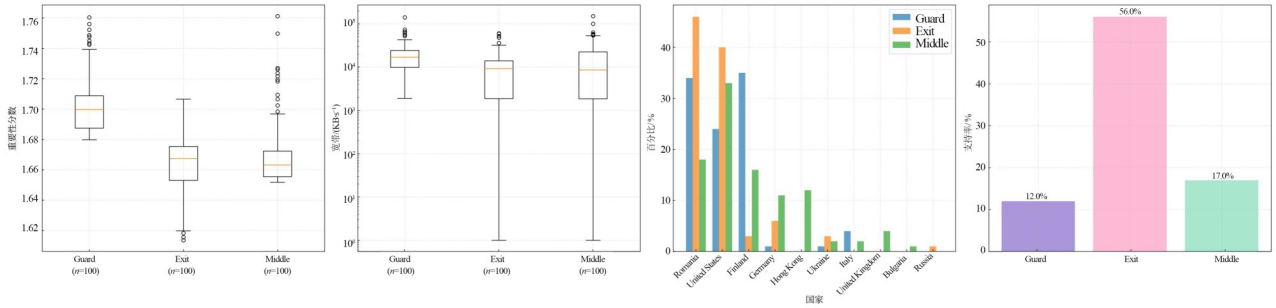
(a) Importance score distribution by node type　(b) Bandwidth distribution by node type　(c) IPv6 support rate by node type

图8 不同节点的部分属性分布情况

Figure 8 Distribution of some attributes for different relays

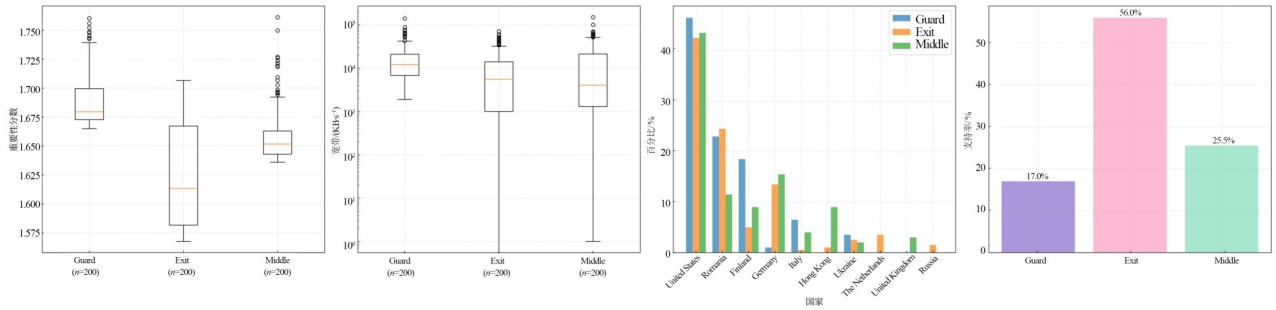
第二个实验主要分析不同Top-K关键节点的特征变化，关注重要性评分、带宽、地理归属和IPv6支持率，以验证模型在不同K取值（50、100、200、500）下的适用性和稳定性。实验结果如图9所示。

从重要性评分分布来看，入口节点的评分整体最高，且随着K值的增加，评分范围逐渐扩大，模型纳入了更多次级节点。然而，不同类型节点的重要性排名整体保持不变，说明模型在不同Top-K取值下对节点关键性判断较为稳定，能够保持一致的节点分类标准。在带宽分布方面，节点带宽分析开始相对较小，随着K增大，节点带宽范围也在不断增大，这说明模型在较小K取值时优先选择高带宽节点，而在较大



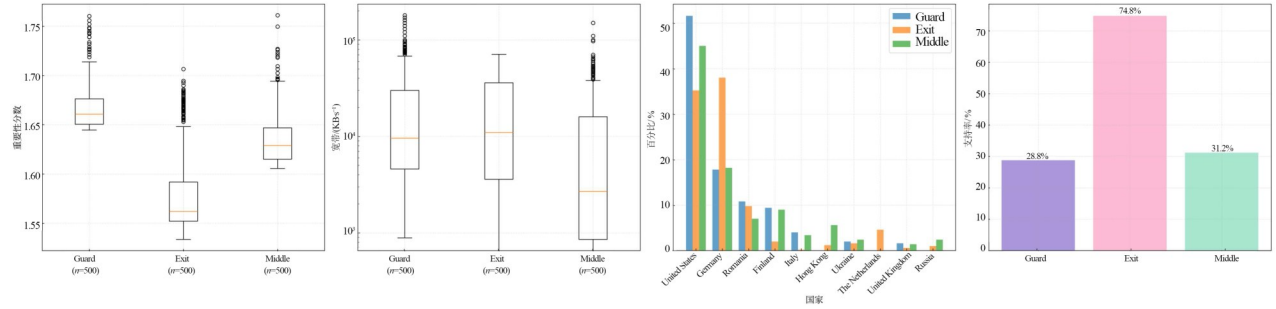
(a) Top-100节点重要性分数分布　(b) Top-100节点带宽分布　(c) Top-100节点地理分布(前10国家)　(d) Top-100节点IPv6支持率

(a) Importance scores of Top-100 nodes　(b) Bandwidth of Top-100 nodes　(c) Geography of Top-100 nodes (top 10 countries)　(d) IPv6 support of Top-100 nodes



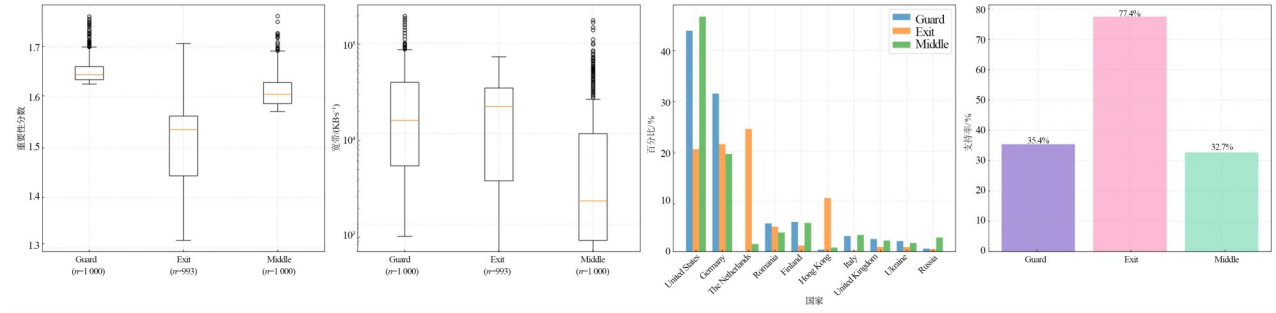
(e) Top-200节点重要性分数分布　(f) Top-200节点带宽分布　(g) Top-200节点地理分布(前10国家)　(h) Top-200节点IPv6支持率

(e) Importance scores of Top-200 nodes　(f) Bandwidth of Top-200 nodes　(g) Geography of Top-200 nodes (top 10 countries)　(h) IPv6 support of Top-200 nodes



(i) Top-500节点重要性分数分布　(j) Top-500节点带宽分布　(k) Top-500节点地理分布(前10国家)　(l) Top-500节点IPv6支持率

(i) Importance scores of Top-500 nodes　(j) Bandwidth of Top-500 nodes　(k) Geography of Top-500 nodes (top 10 countries)　(l) IPv6 support of Top-500 nodes



(m) Top-1 000节点重要性分数分布　(n) Top-1 000节点带宽分布　(o) Top-1 000节点地理分布(前10国家)　(p) Top-1 000节点IPv6支持率

(m) Importance scores of Top-1 000 nodes　(n) Bandwidth of Top-1 000 nodes　(o) Geography of Top-1 000 nodes (top 10 countries)　(p) IPv6 support of Top-1 000 nodes

图9　不同Top-*K*关键节点的部分属性分布情况

Figure 9　Distribution of some attributes for different Top-*K* critical relays

规模任务下，则能够识别出更多带宽中等或较低但仍具有一定重要性的节点，从而提高关键节点的覆盖度。

在地理分布方面，首先能够看出关键节点主要集中在少数国家，这说明这部分国家对Tor的控制能力较强。其次，随着K值增大，地区分布存在一定变化，这说明不同国家对部署节点的类型存在一定偏好性。在IPv6支持率方面，出口节点的IPv6支持率始终最高，超过70%，而入口和中间节点的支持率相对较低。随着K值的增加，IPv6支持率整体保持稳定，这说明是否支持IPv6尽管并不是关键节点的必要特征，但确实对节点关键性存在一定影响。

综合上述两个实验来看，实验结果验证了本方法能够有效建模Tor网络中的隐式关系，从而准确识别在网络当中发挥重要作用的节点。此外，通过实验也揭示了不同类型关键节点在带宽和IPv6支持等特性上的差异，进一步证明了方法的有效性和适用性。

3.3.2 模型有效性验证实验结果

为了验证所提出的基于RA-HAN的关键节点识别方法的有效性，实验对从链路覆盖率和网络带宽两方面进行了验证。在网络覆盖率方面，由于用于模型训练的链路数据来源于香港服务器，因此，为了更好地保持地域性，本实验所使用的验证链路数据也来自同一服务器，采集时间为2024年12月1日至2024年12月31日，共39 676条。在此数据集上，本文进行了计算相应的评价指标。具体结果如表2所示。

表2 覆盖率实验结果

Table 2 Coverage experiment results

总链路数	节点类型	出现节点数	Top-K	节点重叠率/%	链路覆盖率/%
39 676	中间节点	4 643	100	85.0	2.79
			200	80.5	4.87
			500	79.2	11.98
	出口节点	2 016	100	89.0	6.73
			200	80.0	11.29
			500	81.0	25.52

由于入口节点的切换较慢（只有4个），在短时间内其变动较小，对其进行统计分析的意义有限，因此实验仅对中间节点和出口节点进行评估。从整体结果来看，模型识别出的关键节点大多能够在新生成的电路中出现，覆盖率达到80%左右，这表明模型能有效地识别Tor网络中的关键节点，同时能够较好地匹配Tor网络的路径构建规律。此外，实验结果显示，随着Top-K取值的增加，节点覆盖率稍显下降，电路覆盖率显著提升，这说明尽管模型识别的准确率有所下降，但这些关键节点在电路中的整体参与度相对稳定，进一步验证了模型在Tor网络关键节点识别中的有效性。

在网络带宽方面，实验主要评估移除关键节点后对Tor网络整体带宽的影响，以衡量模型识别的关键节点对网络流量的贡献程度。实验统计了总带宽、总平均带宽，并分析了不同Top-K关键节点被移除后，网络损失的带宽情况，包括移除带宽和平均移除带宽。实验结果如表3所示。

表3 带宽影响实验结果

Table 3 Bandwidth impact experimental results

总带宽/(KB·s^{-1})	总平均带宽/(KB·s^{-1})	Top-K	移除带宽/(KB·s^{-1})	平均移除带宽/(KB·s^{-1})
162 129.75	16.81	100	2 118.36	21.18
		200	3 502.82	17.51
		500	8 432.32	16.87

从实验结果来看，模型识别出的前100个关键节点带宽贡献较为突出，移除后对网络带宽的影响最大。随着Top-K取值的增加，被移除节点的整体影响增强，但单个节点的带宽贡献相对减少。表明被识别出的关键节点对Tor网络带宽有较大的贡献，因此对这些节点的保护和优化对于维持Tor网络的稳定性和高效性至关重要。

4 结论

本文提出一种基于异构图建模与关系感知机制的无监督关键节点识别方法。该方法从节点稳定性、链路频次与标签特征等维度出发，构建多源异构图结构；引入关系感知异构注意力网络模型，融合多类型边关系与节点属性信息，提升节点表示的判别性；结合无监督评分机制实现关键性节点排序与识别。在真实Tor共识与链路数据上开展实验，结果显示所提方法在Top-100节点设置下节点覆盖率达85.0%，节点带宽比全网平均节点带宽高约25.9%。该研究为Tor网络基础设施安全建模提供了新路径，具有重要的理论价值与实用意义。

此外，本研究提出的关键中继节点识别方法不仅能够刻画Tor网络结构中继节点的重要性，还具有显著的安全管理应用价值。首先，该方法能够辅助目录授权方（Directory Authorities，DA）提升中继筛选与信誉评估的能力，通过识别在链路构建中起重要作用的高关键性中继，可在节点加入审查机制中优先关注此

类节点的带宽真实性、稳定性与历史行为,从而降低恶意节点通过伪造性能参数混入关键路径的风险。在路径选择策略优化方面,识别出的关键中继节点可用于构建更加稳健的路由决策模型,例如在保持匿名性要求的前提下,为客户端提供具备高稳定性和低攻击面暴露的候选节点集合,从而提升全网路由效率与抗攻击能力。此外,该方法在应急处置场景中也具有实际价值。当监测系统检测到部分节点存在异常行为或潜在攻击迹象时,可通过本方法确定哪些关键中继的失陷会导致整体性能或匿名性显著下降,并据此制定针对性的防护或隔离策略,提高Tor网络对突发事件的响应效率。总的来说,该方法能够为Tor生态中的管理者、运维者和研究人员提供精细化的节点分析依据,为提升匿名通信基础设施的可信度和稳健性提供技术支撑。

参考文献

[1] 中国互联网络信息中心. 第54次中国互联网络发展状况统计报告[R/OL]. (2024-09-19) [2025-11-10]. https://www.199it.com/archives/1718242.html.
China Internet Network Information Center (CNNIC). The 54th statistical report on China's internet development[R/OL]. (2024-09-19)[2025-11-10]. https://www.199it.com/archives/1718242.html. (in Chinese)

[2] Computer Crime Research Center. The cost of cybercrime to reach over $12tn by 2025[EB/OL]. (2024-01-24) [2025-11-10]. https://www.crime-research.org/news/24.01.2024/4132/.

[3] Ventures Cybersecurity. Cybercrime to cost the world $10.5 trillion annually by 2025[EB/OL]. (2020-11-13)[2025-11-10]. https://cybersecurityventures.com/cybercrime-damage-costs-10-trillion-by-2025/.

[4] Dingledine R, Mathewson N, Syverson P. Tor: The second-generation onion router[C]//Proceedings of the 13th Conference on USENIX Security Symposium. Berkeley: USENIX Association, 2004: 21.

[5] The Tor Project. Tor metrics[EB/OL]. (2010-11-30)[2025-11-10]. https://metrics.torproject.org/userstats-relay-country.html.

[6] The Tor Project. Tor metrics[EB/OL]. (2010-11-30)[2024-12-10]. https://metrics.torproject.org/networksize.html.

[7] Bauer K, McCoy D, Grunwald D, et al. Low-resource routing attacks against tor[C]//Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society. New York: ACM, 2007: 11-20.

[8] Greubel A, Pohl S, Kounev S. Quantifying measurement quality and load distribution in tor[C]//Proceedings of the 36th Annual Computer Security Applications Conference. New York: ACM, 2020: 129-140.

[9] Jansen R, Johnson A. On the accuracy of tor bandwidth estimation[C]//Proceedings of the 22nd International Conference on Passive and Active Measurement. Heidelberg: Springer, 2021: 481-498.

[10] Zhang Junlong, Luo Yu. Degree centrality, betweenness centrality, and closeness centrality in social network[C]//Proceedings of the 2nd International Conference on Modelling, Simulation and Applied Mathematics. [S.l.]: Atlantis Press, 2017: 300-303.

[11] Al-Nabki M W, Fidalgo E, Alegre E, et al. ToRank: Identifying the most influential suspicious domains in the Tor network[J]. Expert Systems with Applications, 2019, 123: 212-226.

[12] Srinivas A, Velusamy R L. Identification of influential nodes from social networks based on Enhanced Degree Centrality Measure[C]//2015 IEEE International Advance Computing Conference. Piscataway: IEEE, 2015: 1179-1184.

[13] Song Zhichao, Duan Hong, Ge Yuanzheng, et al. A novel measure of centrality based on betweenness[C]//2015 Chinese Automation Congress. Piscataway: IEEE, 2016: 174-178.

[14] Kianian S, Rostamnia M. An efficient path-based approach for influence maximization in social networks[J]. Expert Systems with Applications, 2021, 167: 114168.

[15] Zhong Linfeng, Shang Mingsheng, Chen Xiaolong, et al. Identifying the influential nodes via eigen-centrality from the differences and similarities of structure[J]. Physica A: Statistical Mechanics and its Applications, 2018, 510: 77-82.

[16] Harooni A, Lotfi-Shahreza M, Shams M, et al. Centrality in multilayer networks: Accurate measurements with MultiNet-Py[J]. The Journal of Supercomputing, 2025, 81(5): 689.

[17] Stolarski M, Piróg A, Bródka P. Identifying key nodes for the influence spread using a machine learning approach[J]. Entropy, 2024, 26(11): 955.

[18] Asgharian Rezaei A, Munoz J, Jalili M, et al. A machine learning-based approach for vital node identification in complex networks[J]. Expert Systems with Applications, 2023, 214: 119086.

[19] Zhang Ying, Hou Xuhang. Node importance assessment method based on double deep Q-network[C]//Proceedings of the 5th International Conference on Frontiers Technology of Information and Computer. Piscataway: IEEE, 2023: 870-873.

[20] Rashid Y, Bhat J I. OlapGN: A multi-layered graph convolution network-based model for locating influential nodes in graph networks[J]. Knowledge-Based Systems, 2024, 283: 111163.

[21] Lalou M, Tahraoui M A, Kheddouci H. The critical node detection problem in networks: A survey[J]. Computer

Science Review, 2018, 28: 92-117.

[22] Curado M, Tortosa L, Vicent J F. A novel measure to identify influential nodes: Return random walk gravity centrality[J]. Information Sciences, 2023, 628: 177-195.

[23] Ullah A, Wang Bin, Sheng Jinfang, et al. Identifying vital nodes from local and global perspectives in complex networks[J]. Expert Systems with Applications, 2021, 186: 115778.

[24] Zhang Haotian, Zhong Shen, Deng Yong, et al. LFIC: Identifying influential nodes in complex networks by local fuzzy information centrality[J]. IEEE Transactions on Fuzzy Systems, 2022, 30(8): 3284-3296.

[25] Zhong Shen, Zhang Haotian, Deng Yong. Identification of influential nodes in complex networks: A local degree dimension approach[J]. Information Sciences, 2022, 610: 994-1009.

[26] Yu Enyu, Wang Yueping, Fu Yan, et al. Identifying critical nodes in complex networks via graph convolutional networks[J]. Knowledge-Based Systems, 2020, 198: 105893.

[27] Lv Laishui, Zhang Kun, Bardou D, et al. A new centrality measure based on random walks for multilayer networks under the framework of tensor computation[J]. Physica A: Statistical Mechanics and its Applications, 2019, 526: 121000.

[28] Lv Laishui, Zhang Ting, Hu Peng, et al. An improved gravity centrality for finding important nodes in multi-layer networks based on multi-PageRank[J]. Expert Systems with Applications, 2024, 238: 122171.

[29] Ni Chengzhang, Yang Jun, Pang Zezhao, et al. Seeding strategy based on weighted gravity centrality in multiplex networks[J]. IEEE Transactions on Network Science and Engineering, 2023, 10(1): 331-345.

[30] Venkatakrishna Rao K, Katukuri M, Jagarapu M. CIM: Clique-based heuristic for finding influential nodes in multilayer networks[J]. Applied Intelligence, 2022, 52(5): 5173-5184.

[31] Lv Laishui, Hu Peng, Zheng Zijun, et al. A community-based centrality measure for identifying key nodes in multilayer networks[J]. IEEE Transactions on Computational Social Systems, 2024, 11(2): 2448-2463.

[32] 梅汉涛, 程光, 朱怡霖, 等. Tor被动流量分析综述[J]. 软件学报, 2025, 36(1): 253-288.
Mei Hantao, Cheng Guang, Zhu Yilin, et al. Survey on Tor passive traffic analysis[J]. Journal of Software, 2025, 36(1): 253-288. (in Chinese)

[33] Shen Meng, Wu Jinhe, Ai Junyu, et al. Swallow: A transfer-robust website fingerprinting attack via consistent feature learning[C]//Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2025: 1574-1588.

[34] Shen Meng, Wu Jinhe, Ye Ke, et al. Robust detection of malicious encrypted traffic via contrastive learning[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 4228-4242.

[35] Qing Yuqi, Yin Qilei, Deng Xinhao, et al. Training robust classifiers for classifying encrypted traffic under dynamic network conditions[C]//Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2025: 3564-3578.

[36] Fu Chuanpu, Li Qi, Bertino E, et al. Training with only 1.0‰ samples: Malicious traffic detection via cross-modality feature fusion[C]//Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2025: 3930-3944.

[37] 郑献春, 王瑞, 闫皓楠, 等. 基于分布式爬虫的高性能Tor网络内容监控系统[J]. 信息安全学报, 2023, 8(1): 144-153.
Zheng Xianchun, Wang Rui, Yan Haonan, et al. A high performance Tor web content monitoring system based on distributed crawlers[J]. Journal of Cyber Security, 2023, 8(1): 144-153. (in Chinese)

[38] 孙学良, 黄安欣, 罗夏朴, 等. 针对Tor的网页指纹识别研究综述[J]. 计算机研究与发展, 2021, 58(8): 1773-1788.
Sun Xueliang, Huang Anxin, Luo Xiaopu, et al. Webpage fingerprinting identification on Tor: A survey[J]. Journal of Computer Research and Development, 2021, 58(8): 1773-1788. (in Chinese)

[39] 罗军舟, 杨明, 凌振, 等. 匿名通信与暗网研究综述[J]. 计算机研究与发展, 2019, 56(1): 103-130.
Luo Junzhou, Yang Ming, Ling Zhen, et al. Anonymous communication and darknet: A survey[J]. Journal of Computer Research and Development, 2019, 56(1): 103-130. (in Chinese)

[40] 马传旺, 张宇, 方滨兴, 等. 匿名网络综述[J]. 软件学报, 2023, 34(1): 404-420.
Ma Chuanwang, Zhang Yu, Fang Binxing, et al. Survey on anonymous networks[J]. Journal of Software, 2023, 34(1): 404-420. (in Chinese)

[41] Zhang Wenzhen, Lu Tianbo, Du Zeyu. TNRAS: Tor nodes reliability analysis scheme[C]//Proceedings of the 11th International Conference on Communication and Network Security. New York: ACM, 2021: 21-26.

[42] Cui Jun, Huang Changqi, Meng Huan, et al. Tor network anonymity evaluation based on node anonymity[J]. Cybersecurity, 2023, 6(1): 55.

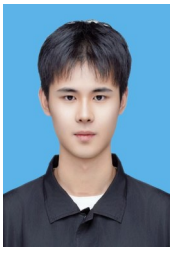
[43] Zhai Jiangtao, Zhang Kaijie, Zeng Xiaolong, et al. FlowCorrGCN: Enhancing flow correlation through graph convolutional networks and triplet networks[J]. International

Journal of Intelligent Systems, 2024, 2024: 8823511.
[44] Qiu Qinjun, Liu Jiandong, Hao Mengqi, et al. DCKH-CNN: A multimetric graph-based convolutional neural network for identifying key influential nodes in earth surface data linked networks[J]. Transactions in GIS, 2025, 29(2): e70016.
[45] 刘琳岚, 谭镇阳, 舒坚. 基于图神经网络的机会网络节点重要度评估方法[J]. 计算机研究与发展, 2022, 59(4): 834-851.
Liu Linlan, Tan Zhenyang, Shu Jian. Node importance estimation method for opportunistic network based on graph neural networks[J]. Journal of Computer Research and Development, 2022, 59(4): 834-851. (in Chinese)
[46] Xiong You, Hu Zheng, Su Chang, et al. Vital node identification in complex networks based on autoencoder and graph neural network[J]. Applied Soft Computing, 2024, 163: 111895.
[47] Yu Enyu, Fu Yan, Zhou Junlin, et al. Predicting critical nodes in temporal networks by dynamic graph convolutional networks[J]. Applied Sciences, 2023, 13(12): 7272.
[48] Chan C Y, Ioannidis Y E. Bitmap index design and evaluation[C]//Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data. New York: ACM, 1998: 355-366.
[49] The Tor Project. Path selection and constraints[EB/OL]. (2024-07-16)[2024-12-10]. https://spec.torproject.org/path-spec/path-selection-constraints.html.
[50] Edman M, Syverson P. As-awareness in tor path selection[C]//Proceedings of the 16th ACM Conference on Computer and Communications Security. New York: ACM, 2009: 380-389.
[51] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks[C]//Proceedings of the 15th International Conference on the Semantic Web. Heidelberg: Springer, 2018: 593-607.
[52] Wang Xiao, Ji Houye, Shi Chuan, et al. Heterogeneous graph attention network[C]//The World Wide Web Conference. New York: ACM, 2019: 2022-2032.
[53] Wang Chunmian, Luo Junzhou, Ling Zhen, et al. A comprehensive and long-term evaluation of tor V3 onion services[C]//IEEE INFOCOM 2023 - IEEE Conference on Computer Communications. Piscataway: IEEE, 2023: 1-10.
[54] Sendner C, Stang J, Dmitrienko A, et al. MirageFlow: A new bandwidth inflation attack on tor[C]//Proceedings of the 31st Annual Network and Distributed System Security Symposium. Rosten: Internet Society, 2024.
[55] Döpmann C, Tschorsch F. Modeling tor network growth by extrapolating consensus data[C]//Proceedings of the 18th International Conference on Availability, Reliability and Security. New York: ACM, 2023: 29.
[56] The Tor Project. Stem[EB/OL]. (2013-03-26) [2024-12-10]. https://github.com/torproject/stem.

作者简介

王楠楠　男,2001年2月生,山西晋城人。2025年硕士毕业于四川大学网络空间安全学院,现为成都市公安局高新分局民警。主要研究方向为网络公害治理与网络安全。
E-mail: ning000121@gmail.com

黄　诚　男,1987年9月生,重庆云阳人。博士,现为四川大学网络空间安全学院教授,博士生导师。中国计算机学会高级会员。主要研究方向为网络空间安全,尤其是威胁情报、供应链检测和黑灰产治理。
E-mail: codesec@scu.edu.cn



刘骏以　男,2002年5月生,重庆云阳人。现为四川大学网络空间安全学院在读硕士研究生。主要研究方向为匿名通信网络安全。
E-mail: 1325389290@qq.com

游　畅　男,2001年4月生,四川成都人。四川大学网络空间安全学院硕士研究生。主要研究方向为匿名通信与网络安全。
E-mail: 1162242952@qq.com

时金桥　男,1978年1月生,黑龙江哈尔滨人。博士,现为北京邮电大学网络空间安全学院教授、博士生导师。中国计算机学会会员。主要研究方向为网络与信息安全,尤其是隐私增强技术和数据泄露检测。
E-mail: shijingiao@bupt.edu.cn