计算机科学
COMPUTER SCIENCE

基于知识图谱的网络空间地理图谱构建方法

吴越, 胡威, 李城龙, 杨家海, 李祉岐, 尹琴, 夏昂, 党芳芳

引用本文

吴越, 胡威, 李城龙, 杨家海, 李祉岐, 尹琴, 夏昂, 党芳芳. 基于知识图谱的网络空间地理图谱构建方法[J]. 计算机科学, 2024, 51(11): 321-328.

WU Yue, HU Wei, LI Chenglong, YANG Jiahai, LI Zhiqi, YIN Qin, XIA Ang, DANG Fangfang. Knowledge Graph Based Approach to Cyberspace Geographic Mapping Construction [J]. Computer Science, 2024, 51(11): 321-328.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.231000127

基于知识图谱的网络空间地理图谱构建方法

吴 越[1] 胡 威[2] 李城龙[1] 杨家海[1] 李祉岐[3] 尹 琴[3] 夏 昂[2] 党芳芳[4]
1 清华大学网络科学与网络空间研究院 北京 100084
2 国家电网有限公司信息通信分公司 北京 100053
3 国网思极网安科技(北京)有限公司 北京 102209
4 国网河南省电力公司信息通信分公司数据中心 郑州 450000
(wuyue23@mails.tsinghua.edu.cn)

摘 要 在互联网快速发展且网络安全愈发重要的数字信息时代,网络空间地理图谱被认为是认知和管理网络空间的新型手段,其通过综合网络空间和地理空间的信息,能够从多个角度更加全面地展示网络空间态势。但目前对于网络空间地理图谱的研究工作缺乏对网络空间模型细粒度的刻画,也缺乏网络空间地理图谱的具体构建方法和应用方式。针对上述问题,以网络空间认知为目标,文中扩展提出了一个带有时间参考轴的四层四级的网络空间分层模型。此外,为了更好地理解复杂的网络空间环境,还结合知识图谱技术,提出了一个构建网络空间地理图谱的具体框架以及构建网络空间本体的方法。基于 Censys 的真实测绘数据,成功构建了一个模拟园区网络的网络空间地理图谱原型。本研究提出了对网络空间分层结构的改进方法,同时也将知识图谱引入网络空间地理学的研究领域。这不仅有助于提高对网络空间的理解,而且在网络安全、资源管理、故障恢复、决策制定等方面具有实际的应用意义。

关键词:网络空间地理学;知识图谱;网络空间层次模型;网络空间本体;网络空间地理图谱
中图分类号 TP393.2

Knowledge Graph Based Approach to Cyberspace Geographic Mapping Construction

WU Yue[1], HU Wei[2], LI Chenglong[1], YANG Jiahai[1], LI Zhiqi[3], YIN Qin[3], XIA Ang[2] and DANG Fangfang[4]
1 Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing 100084, China
2 State Grid Information & Telecommunication Co., Ltd., Beijing 100053, China
3 State Grid Cyber Security Technology(Beijing) Co., Ltd., Beijing 102209, China
4 State Grid Henan Information & Telecommunication Company Data Center, Zhengzhou 450000, China

Abstract In the digital information era of rapid development of the Internet and increasing importance of cybersecurity, cyberspace geographic mapping is regarded as a new type of means of cognition and management of cyberspace. By synthesizing the information of cyberspace and geospatial information, it is able to display the cyberspace situation more comprehensively from multiple perspectives. However, the current research work on cyberspace geographic mapping lacks a fine-grained portrayal of cyberspace model, as well as specific construction methods and application methods of cyberspace geographic mapping. To address the above problems, with the goal of cyberspace cognition, this paper proposes a four-layer, four-level cyberspace hierarchical model with a time reference axis. In addition, in order to better understand the complex cyberspace environment, a specific framework for constructing a cyberspace geographic mapping, as well as a method for constructing a cyberspace ontology, is proposed in conjunction with the knowledge graph technology. Based on real mapping data from Censys, a prototype cyberspace geographic mapping of a simulated park network is successfully constructed. This study proposes an improved approach to the hierarchical structure of cyberspace, and also introduces knowledge mapping into the research field of cyberspace geography, which not only helps to improve the understanding of cyberspace, but also has practical application significance in cybersecurity, resource management, fault recovery, and decision making.

Keywords Cyberspace geography, Knowledge graph, Cyberspace hierarchy model, Cyberspace ontology, Cyberspace geographic mapping

到稿日期:2023-10-20 返修日期:2024-08-26
基金项目:国家电网有限公司科技项目(5700-202252199A-1-1-ZN)
This work was supported by the Science and Technology Project of State Grid Corporation of China(5700-202252199A-1-1-ZN).
通信作者:胡威(whu@sgcc.com.cn)

1 引言

在当今数字时代，网络空间已成为人类社会不可或缺的一部分，成为了与陆、海、空、天四类现实空间并列的第五大战略空间[1]。随着网络的普及和数字技术的迅猛发展，网络安全事件也更加频发。据 Cybersecurity Ventures 发布的 2022 年网络犯罪报告[2]，预计在 2023 年，网络犯罪会在世界范围内造成 8 万亿美元的损失。网络空间的安全问题已经成为必须直面的全球性挑战。网络空间安全成为国家安全的重要基础。

相较于现实世界，网络空间要素体量大、种类杂、分布广、变化快，要较为全面地认知网络空间仍然存在一定难度。目前仍然缺乏一个完善的网络要素体系，已有研究工作[3-7]也未考虑网络空间与现实世界的联系。网络空间地理学是一个关注网络与地理空间交叉与融合的新型学科领域，为认知网络空间，深入理解网络空间与现实地理空间的关系，指导网络空间治理和网络安全防护提供了新的视角。

Chen 等首次提出了网络空间地理图谱的概念[8]。网络空间地理图谱综合了网络空间和地理空间的信息，面对新形势的网络安全威胁，能够从多个角度更加全面地展示网络空间态势。但目前对于网络空间地理图谱的研究工作并不多，文献[8]从理论上提出了一个知识图谱的基础构建框架，文献[9]主要针对网络空间中的情报进行了知识图谱的构建研究。

构建网络空间地理图谱是对网络安全进行综合治理与规划的基础。本文基于现有的研究工作，在网络空间地理学的指导下，对现有的网络空间相关工作进行了总结，进一步完善了网络空间层次模型，细化了网络空间要素的分层分类体系，在 Guo 等提出的四层三级模型[6]的基础上，进行了更细粒度的第四级的划分和补充，并首次加入了网络空间中的时间特性，将网络空间引入三维空间中，提出了一个拥有时间参考轴的四层四级模型。此外，基于网络空间要素分层体系，首次提出了构建网络空间地理图谱中的领域划分方法和本体模型构建方法，并给出了网络空间地理图谱构建的具体技术框架和应用场景示例，以完善对网络空间的认知，为网络建设、网络规划管理和网络安全提供指导和帮助。

本文第 2 章介绍网络地理学的发展，给出目前网络空间的研究进展以及知识图谱的介绍；第 3 章结合已有研究工作，对网络空间层次模型进行改进和优化；第 4 章给出了本文构建的网络空间地理知识图谱框架；第 5 章基于本文提出的分层模型，对不同层次的要素进行本体建模；第 6 章在一个使用真实测绘数据模拟的园区网络上，搭建网络空间地理知识图谱的原型系统；最后总结全文并对网络空间的未来研究方向进行展望。

2 背景和相关工作

2.1 网络空间地理学

网络空间地理学(Cyber Geography)[10]由 Martin Dodge 于 1999 年提出，是一个多学科领域，旨在研究互联网和数字空间的地理分布、结构、演化和影响。目前也有许多工作对网络空间进行分层研究。方滨兴院士从网络安全体系的角度出发，提出了“四横八纵”的网络空间安全体系[4]；Guo 等从网络空间资源测绘的角度出发，为网络空间资源引入了社会形态，定义了新的网络空间资源分类体系[5]；Guo 等梳理了网络空间地理学的基础理论和技术路径，基于“人-地-网”关系提出了网络空间四层分层模型[6]，将地理空间和社会环境纳入对网络空间的认知，弥补了之前相关研究工作的不足，但其只重点关注了网络实体设备以及其间的逻辑链接关系，缺少对网络空间中虚拟资源的关注，而这些虚拟资源是网络空间的重要组成部分之一。

时间特性的缺失导致网络空间知识失去时效性，甚至可能对人们的行为产生误导。因此，本文基于前人的工作，将网络空间的分层模型扩展到第四级，全面考虑网络空间的物理及虚拟属性，引入地理事件属性和时间属性，对网络空间要素体系进行进一步的细化和完善。

2.2 知识图谱

知识图谱是一种将知识结构化并以图形形式表示的技术，目前普遍认可的一种定义为：知识图谱的本质是一种由节点和边组成的语义网络，网络中的节点代表实体或者概念，而边代表实体与概念之间的各种语义关系[11]。知识图谱的真正崛起可以追溯到谷歌在 2012 年推出的知识图谱项目[12]。知识图谱通常由实体、关系和属性组成。实体可以是现实世界中的任何事物，如人、地点、事件、概念等，它们通过关系连接在一起，关系描述了这些实体之间的联系，而属性则为实体提供了更多的信息。知识图谱的构建依赖于自然语言处理、信息抽取、实体识别等技术，可以从大规模文本和网络数据中提取知识。通过知识图谱可以实现对世界万物及其联系的描述和表示[11]。

为了认知网络空间，前人也开展了许多研究工作。Cicalese 等通过地理信息对任播网络基础设施实现定位[13]；Miao 等提出了采用自治域(Autonomous System，AS) 作为网络空间基础坐标向量的网络空间坐标系[3]。这些工作在网络空间资源刻画与表达方面做了大量探索和尝试，也提出了许多适用于特定业务场景或问题的表达方法和相关理论，但仍然存在局限性。1)认知视角的局限性。多数局限于特定的场景或具体问题，相关理论方法或表达模型往往围绕某一类网络空间资源的特定属性或特定关系，无法适用于多维度信息的综合表达。2)表达模型的演进性与时效性问题。相关研究工作提出的理论方法和表达模型大多适用于某一阶段或特定时期的某一问题，并且弱化了数据信息的时效性因素。将知识图谱引入网络空间中，则能够为不同应用和研究领域间的信息共享提供一种有效机制，从而突破认知视角的局限性，实现对网络空间资源更为全面、客观的刻画与表示。Chen 等提出了网络空间地理图谱的概念[8]，但目前对网络空间要素的刻画与表达仍缺乏统一的标准和理论支撑体系，相关表达模型和方法在规范性、可解释性、适用范围和应用场景等方面仍然存在诸多问题。

2.3 可视化技术

可视化技术是随着信息技术的发展而产生的，可视化的过程是通过交互式图形学的技术手段，借助人类的认知能力来提取数据中有意义的信息，以图形的方式呈现复杂数据集

的过程。地理空间可视化研究已经有了长时间的发展和积累，而网络空间具有高度的复杂性，并与地理空间高度关联，构成了人类活动的现实空间。可视化技术的早期形态是二维图形可视化[14]；Cox 等将三维显示技术应用到了网络空间可视化中[15]；美国国安局的“藏宝图计划”对全球互联网进行了全面的测量、探测、分析和可视化[16]。当前国内外开展的网络空间可视化研究大多集中在物理网络层、逻辑网络层，侧重于网络设备的描述和定位、网络运行数据的统计分析等，可视化的效果与应用需求还存在较大差距。

3　网络空间层次模型

本文围绕地理环境、网络环境和社会环境中不同层次的各种要素对象，结合实战安全需求，将网络地理空间分为地理环境层、网络环境层、社会环境层、业务环境层和一个时间参考轴（见图 1），并在已有工作的基础上，将网络空间模型扩展到由粗到细的 4 个级别，表 1 详细列出了每个层次的各类要素。

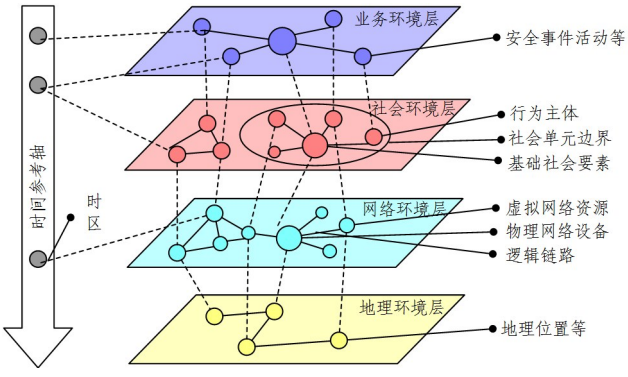

图 1　网络空间层次模型

Fig. 1　Hierarchical model of cyberspace

表 1　网络空间四层模型

Table 1　Four-layer model of cyberspace

一级	二级	三级	四级
地理环境层	基础地理信息	地理位置信息	经纬度等
		地理特征信息	地形、地貌、水系等
	地理环境事件	自然事件	地震、水灾等
		人为事件	战争、开发等
网络环境层	物理网络环境	实体资源	交换机、主机等
	逻辑网络环境	虚拟资源	IP、操作系统、软件、网络服务等
		逻辑属性	逻辑链路等
社会环境层	公共社会环境	基础社会环境	人口、社区、社会单元等
		政治文化环境	安全政策等
	行为主体环境	虚拟角色	网络账号等
		实体角色	身份信息等
		管理人员和组织	网络管理人员、开发商等
业务环境层	安全情报	网络安全事件等	有害程序事件、网络攻击事件等
		动态安全情报	攻击情报等
		静态安全情报	漏洞等
	安全保护	保护对象	重点网络设施等
		安全保护活动	攻防演练等
时间参考轴	绝对时间信息	时间基础属性	绝对时间
		时间社会属性	时区等

地理环境层是网络空间的基础，一般地理环境的信息包括基础地理信息和领域所需要的专题数据[17]。IP 地址是网络空间中的唯一标识，网络环境中的物理网络设备通常拥有一个或多个 IP 地址，而每一个 IP 地址可以定位到一个具体的地理位置。社会环境通常是复杂的，一般拥有各种社会单元边界，如行政边界等，同时，社会环境中拥有对网络环境进行使用、管理、维护或实施攻击等行为的行为主体。行为主体对网络空间中的操作通常可以映射到业务环境层，同时，业务环境层中的事件（如网络攻击行为等）又会直接影响到网络环境的状态，对用户等行为主体直接造成影响。时间参考轴主要为其他层次中的要素提供时间属性。网络空间瞬息万变，时间特征的引入对网络空间中的实时数据分析和网络安全等都至关重要。4 个层次之间相互联系，相互影响，以时间维度作为参考，从多个领域和多个维度构成了网络空间体系。

3.1　地理环境层

地理环境层作为网络空间的物理基础，是各种网络要素的承载者。地理环境层由网络空间所依附的地理环境组成，包括基本的地理信息（如地形、地理位置等）和在地理环境中发生的事件（如自然灾害）。这一层次集合了网络空间要素的地理属性，如网络基础设施和网络参与者的地理位置、空间分布以及区域特性，构成了网络环境层的基础，为上层的网络环境层和业务环境层提供了数据分析的基本框架。

3.2　网络环境层

网络环境层在网络空间扮演着核心角色，包括物理网络和逻辑网络两要素。物理网络环境代表网络的实体属性，由基础设施和相关设备（如互联网和电网系统等）构成，支持信息获取、传输、处理和存储。这一物理层建立在地理环境之上，是网络空间的实体基础，其要素常可映射到具体地理位置，由社会环境中的行为主体进行管理和利用。

网络逻辑环境代表了计算机网络中的虚拟化和抽象层，它规划了网络的逻辑结构、协议和服务，以支持各种通信和应用程序。该层描述了网络的虚拟资源和逻辑拓扑结构（包括节点以及它们之间的连接方式）。这些拓扑关系通常是逻辑的，不一定反映实际的物理布局，呈现了网络空间中节点的逻辑属性和它们之间的逻辑链接关系。在逻辑环境中，每个节点通常由 IP 地址标识。IP 地址代表了节点的逻辑特征，通常对应着地理空间中的某一具体地理位置。此外，逻辑网络还包括基于物理节点的软件、数据等虚拟资源，它们构建了一个信息活动域，其中包含协议、软件和数据等元素。与物理网络相比，逻辑网络环境更加关注网络节点在逻辑处理和数据交换方面的功能。通常情况下，某些物理设备同时具备物理属性以及逻辑属性，用于描述其位置等信息以及记录与其他设备的通信或对设备所包含的虚拟资源的描述。网络环境层是网络空间的核心组成，其中包括各种要素，这些要素通常同时涉及地理特征、社会用户特征和业务安全特征。

3.3　社会环境层

社会环境层与网络环境层以及地理环境层密切相关，由地理属性和网络属性组成。社会环境的地理方面包括人口、城市、国家等元素。人类是社会环境和网络环境中的主要参与者，在网络环境中扮演不同的角色，包括实体和虚拟角色。实体角色指现实社会中的人，包括个人用户、机构单位和网络

管理员等，他们是网络环境的使用者、维护者和管理者。此外，还存在着黑客和网络犯罪者，他们是网络空间中的不法行为主体，试图入侵、破坏、窃取数据或从事其他恶意活动。虚拟角色是实体角色在网络空间中创建的虚拟个体或身份，包括各种网络账户。虚拟角色与实体角色之间可以通过特定方式进行认证，例如身份验证信息和登录凭据。网络角色和实体角色之间存在多对多的映射关系，一个虚拟账户可能由多个不同的现实实体角色共享，反之亦然，一个实体角色可能拥有多个虚拟身份。

3.4 业务环境层

以网络安全为目标，业务环境层主要包括安全情报和各种安全活动。这一层次根据安全事件的级别分为地理环境层的网络安全事件和网络环境层的安全事件，涵盖了《国家网络安全事件应急预案》中提出的 7 类网络安全事件[18]。网络安全的关键保护对象主要包括要重点保护的网络设备，这些设备包括基础网络设施以及按照特定规则和程序进行信息收集、存储、传输、交换和处理的系统，如云计算平台、物联网设备和工业控制系统等。业务环境层的主要目标是网络安全，它直接影响网络环境层和行为主体层。

3.5 时间参考轴

时间参考轴丰富了网络空间的维度。时间是一维单向的，我们考虑了人类社会因素，将时区概念纳入时间轴，形成了一个二维的时间参考轴。通过一个时间表示和时区可以确定一个具体的时刻。目前的网络空间层次模型大多是二维的，而网络空间中的数据传输、安全攻击等事件都是瞬时性的。时间特性的引入将网络空间扩展到第三个维度，网络空间中的各类要素加上时间戳，为网络空间的信息和知识提供了时效性的保障。

4 网络空间地理知识图谱框架

网络空间地理知识图谱的构建旨在借助知识图谱技术对网络地理空间分散的资源要素进行整合，实现网络地理空间态势感知等目的。与通用知识图谱[12]相比，网络空间地理知识图谱属于领域知识图谱，具有以下特点：网络地理空间中的资源呈现维度多样化、数据信息碎片化、实体间关系动态易变等。结合网络地理空间的特点与知识图谱的构建技术，本文提出了一个构建网络空间地理知识图谱的框架，如图 2 所示。

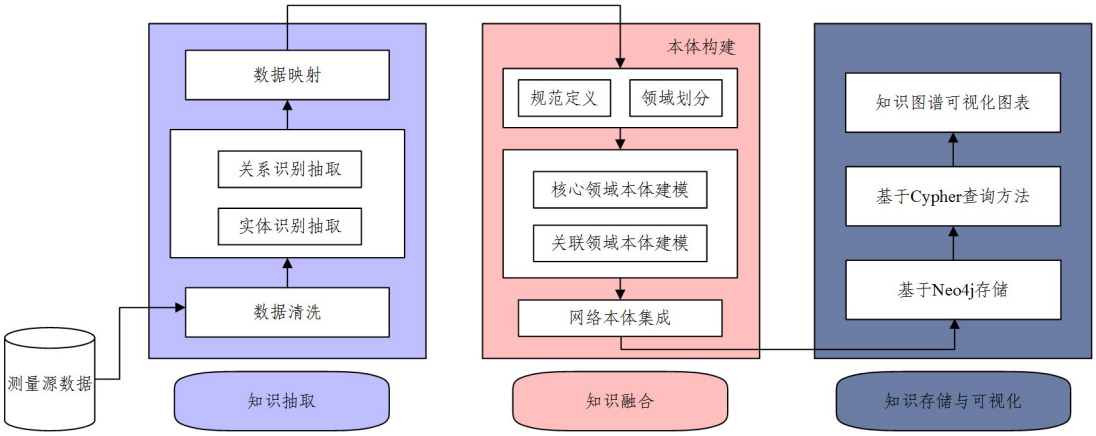

图 2 网络空间地理图谱构建框架

Fig. 2 Construction framework of cyberspace geographic mapping

网络空间地理知识图谱的构建主要包括 3 个层次。1)知识抽取。知识抽取阶段负责将前述数据采集工作获取的结构化网络空间资源数据映射为资源描述框架(Resource Description Framework，RDF)定义的标准 S-P-O(Subject-Predicate-Object，即主体-谓词-宾语)三元组结构知识，包括实体抽取、属性抽取和关系抽取。系统首先通过知识抽取模块，基于网络空间资源本体模型对知识图谱源数据进行知识抽取，为本体模型赋予实例数据。2)知识融合。包括实体链接、本体工程等步骤。知识融合是对源数据进行数据对齐，并对其中包含的相同属性信息或关系信息的数据进行消歧。由于网络空间资源覆盖的领域和范围较大，因此本文首先对网络空间的各个领域特征进行分析，形成网络空间核心领域本体和网络空间关联领域本体，然后对代表性的本体进行建模，最后将各个本体模型进行集成，从而完成本体模型的总体构建。3)知识存储与可视化。在知识存储查询与可视化环节，采用 Neo4j 图数据库[19]对知识数据进行存储，并利用 Cypher 语言[20]构建知识图谱查询方法，形成知识图谱可视化图表。

知识融合的关键在于本体构建。本体[21]是一种形式化的工具，用于清晰而详细地定义共享概念体系。网络空间资源的描述和表达是认知网络空间的基础。为了有效地描绘网络空间，需要使用本体模型来抽象和描述网络空间资源，包括相关概念、属性和关系。网络空间资源本体模型实质上是一种形式化规范，用于详细描述网络空间资源的概念、实体以及它们之间的关系。

5 网络地理空间资源本体模型

5.1 网络地理空间本体建模框架

由于网络空间资源覆盖的领域和范围较大，本文提出了“先主要后次要，先局部再总体”的本体建模思想，如图 3 所示。1)在横向上：首先对网络空间资源核心领域进行本体建模，基于网络空间测量领域知识和本体定义规则对概念、属性、关系进行领域划分，对领域特性模糊和具有争议的部分按照其领域多样性、关联性进行概念细分或重定义，形成网络空间资源核心领域本体；而后对排除在核心领域之外的概念、

属性、关系进行整理和集成，设置相应的网络空间资源关联领域本体。2)在纵向上：首先从部分概念、属性和关系入手，对具有代表性的局部本体进行建模，而后对各个局部的本体模型进行集成，从而完成本体模型的总体构建。

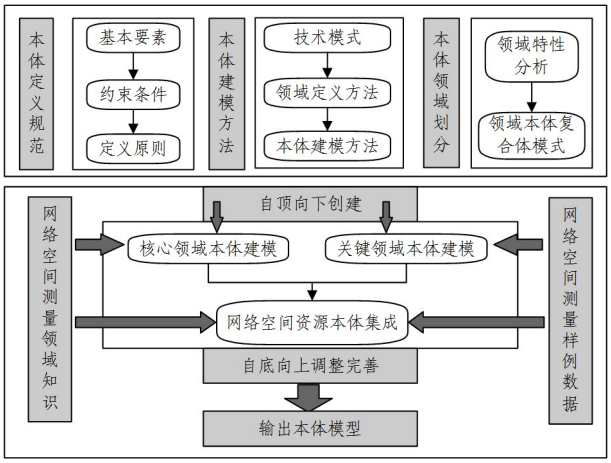

图 3　网络空间地理图谱本体建模框架

Fig. 3　Ontology modeling framework of cyberspace geographic mapping

网络空间资源本体模型采用“核心领域-关联领域”复合体的形式。其中，核心领域本体是网络空间资源领域本体，从属于网络空间本体，主要由 IP 本体、域名本体、网络服务本体等组成。关联领域本体主要包括地理信息空间领域本体、人类社会空间领域本体、时间领域本体。

5.2　本体领域划分

本体的领域划分是本体建模的首要工作。领域本体是对领域内特有概念、属性和关系的形式化描述，因此领域本体具有显著的领域特性。在本体的定义和本体模型的构建过程中，需要首先对网络空间资源本体的领域特性进行分析。

网络空间资源本体兼具基础本体和领域本体的特性。一方面，网络空间资源涉及的领域和维度较为广泛，不同应用和研究领域对网络空间资源的认知不可避免地存在差异，网络空间本体需要具备广泛的、跨领域的语义互操作性(Semantic Interoperability)。另一方面，网络空间本质上是基于网络技术构建的人造空间，网络科学领域知识框架内的概念、关系、属性以及规则和公理等都需要在网络空间资源本体中得到体现。本文基于网络空间测量领域知识和网络测量实际数据，将网络空间资源本体定义为核心领域本体，并将与网络空间资源有密切联系的地理信息空间(Geo Information Space)、人类社会空间(Human Society Space) 和时间领域(Time Domain)定义为关联领域。网络空间资源核心领域定义了 4 个基本要素：IP、域名、网络服务和网络设备。

5.3　本体模型构建方法

本文基于网络测量领域知识，采用自顶向下的方法实施领域本体建模，在核心领域本体的构建方面采用深度优先覆盖的策略，在关联领域本体的设置方面采用广度优先覆盖的策略。在核心领域本体的构建工作中，我们将网络测量领域中的典型要素、重要概念和术语作为本体模型中概念类设置的参考，将网络测量涵盖的重要属性和关系作为本体模型中的数据属性和对象属性的参考。为了补充核心领域之外的知识体系，我们将研究扩展到关联领域本体的构建，对排除在核心领域之外的概念、属性、关系进行整理和集成，设置相应的网络空间资源关联领域本体，并按照实际业务场景和应用需求对领域特性模糊和具有争议的部分进行概念细分或重定义，并划归到相应的领域本体之中。图 4 展示了本体建模流程框架。

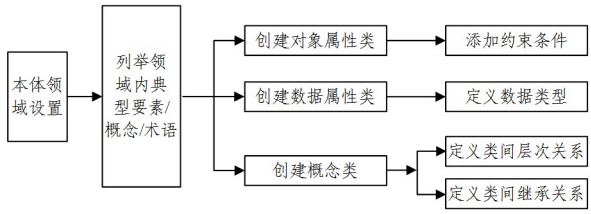

图 4　本体建模流程框架

Fig. 4　Framework of ontology modeling process

5.4　网络核心领域——IP 本体构建

IP 实体是网络空间实体资源的主要组成部分，同时也是网络空间测量领域中网络实体资源识别和探测的主要研究对象，因此我们以 IP 本体构建为例对网络核心领域本体构建方法进行详细的展示。根据网络测量领域知识，在 IP 概念类下设置了两个子类，分别是 IP 概念类和 AS 概念类。由于 IPv4 地址空间与 IPv6 地址空间并存于互联网网络空间，因此 IP 概念类实际上包括 IPv4 概念类、IPv6 概念类、AS 概念类 3 个子类。每一个概念类拥有多个对象实体。IP 本体概念类定义和层级关系如图 3 所示。

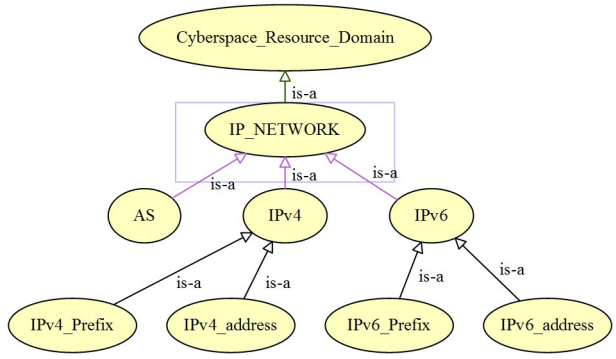

图 5　IP 本体视图

Fig. 5　IP ontology view

创建 IP 本体模型需要明确包含的概念、实体、关系和属性，具体来说，需要定义数据属性和对象属性。这里基于网络测量领域的知识和实际样例数据，对 IP 本体中的典型要素、重要概念和术语进行了整理。同时，以网络测量典型场景为背景，对领域内和跨领域的概念之间的关系进行了总结，对 IP 本体的数据属性和对象属性进行了定义。表 2 列出了详细的 IP 本体数据属性，表 3 列出了详细的 IP 对象属性。以 IP 地址 1.2.3.4 为例，其 IPv4_address 数据属性为 1.2.3.4，属于 AS111111(AS 概念类的实体)，表示为知识图谱的三元组形式，即为(1.2.3.4，Belong_To_AS，AS111111)。三元组中的关系 Belong_To_AS 即为 IP 本体的对象属性。

表 2 IP 本体数据属性

Table 2 IP ontology data attributes

属性名称	注释	主语	数据类型
IP Description	IP 地址描述信息	IPv4_address IPv6_address	xsd:string
IPv4 address	IPv4 标准格式地址	IPv4_address	scoket. AF_INET4
IPv6 address	IPv6 标准格式地址	IPv6_address	scoket. AF_INET6
ASN	自治系统编号	AS	xsd:short
AS Name	自治系统名	AS	xsd:string

表 3 IP 本体对象属性

Table 3 IP ontology object properties

属性名称	注释	主语	宾语
Double_Stack	双栈关系	IPv4_address IPv6_address	IPv4_address IPv6_address
Has_Prefix_IPv4	IPv4 地址所属前缀	IPv4_address	IPv4_Prefix
Has_Prefix_IPv6	IPv6 地址所属前缀	IPv6_address	IPv6_Prefix
Belong_To_AS	地址所属 AS	IPv4_address IPv6_address AS	AS
Location_City	所在城市	IPv4_address IPv6_address AS	City
Location_Continent	所在大陆	IPv4_address IPv6_address AS	Continent
Location_Coordinates	所在经纬度	IPv4_address IPv6_address AS	Coordinates
Location_Country	所在国家名	IPv4_address IPv6_address AS	Country_Name
Location_District	所在行政区	IPv4_address IPv6_address AS	District
Location_Postal_Code	所在地邮政编码	IPv4_address IPv6_address AS	Postal_Code
Location_Street	所在地街道	IPv4_address IPv6_address AS	Street
Location_Timezone	所在地时区	IPv4_address IPv6_address AS	Time_Zone
Belong_To_Org	所属组织	IPv4_address IPv6_address AS	Origanization
Mnt_by	所属管理机构	AS	Organization
tech_c	技术联系人	AS	Organization
Updated_Time	信息更新时间	IPv4_address IPv6_address AS	Time_Stamp

根据“领域本体复合体”模式定义，将排除在网络空间资源核心领域之外的地理位置信息(Location)和组织机构信息(Organization)分别归入对应的地理信息空间(Geo Information Space)关联领域本体、人类社会空间(Human Society Space)关联领域本体和时间领域(Time Domain)本体。

6 网络空间地理图谱原型系统

6.1 系统实现

本文在一个模拟的电力园区网络上进行网络地理知识图谱构建。本文采用开源的 D2RQ 工具[22]和 D2RQ Mapping 映射语言[23]将从测绘平台 Censys[24]获得的部分网络空间资源真实测绘数据映射为资源描述框架定义的标准 S-P-O 三元组结构知识，并存入 Neo4j 图数据库。由于知识图谱的内容包括了实体、属性、关系等，因此普通的关系数据库不能很好地体现出这个数据的特点。Neo4j 图形数据库还提供了 Neo4j Browser 的可视化界面，因此本文选用 Neo4j 图关系数据库作为知识图谱的存储和可视化方案。

6.2 模拟园区网络要素定义

我们将可构建电力系统网络空间的最小独立元素视为网络空间资产，具体包括但不限于 Web 服务器、路由交换设备、网络摄像头、物联网设备、网络打印机、操作系统、应用服务等物理设备和 IP、Mac、端口、服务、域名、组件等逻辑要素资产。在本次模拟园区网络中，引入了交换机、路由器、服务器、防火墙、客户终端、继电器、控制器等网络物理设备以及在各种物理设备上应用的操作系统等软件、地理环境层要素和社会环境层的要素。各类要素及其所属的层次如表 4 所列。

表 4 模拟园区网络中的资产要素

Table 4 Asset elements in simulated park network

所属层次	要素名称	数量	所属层次	要素名称	数量
物理网络环境	交换机	1	逻辑网络环境	IP	6
	路由器	1		操作系统、应用软件等	5
	服务器	1	时间领域	时间戳等	5
	防火墙	1	地理环境	国家	1
	客户终端	1		城市	1
	继电器	1	社会环境	开发商	4
业务环境	安全漏洞	6		组织	2

6.3 知识图谱表达方法

本文采用 Neo4j 作为存储和可视化工具来展示网络地理空间。每一类本体都存在本体数据属性和本体对象属性。在我们模拟的图 7 中的网络空间地理知识图谱中，节点表示本体，节点之间的边表示两个节点间的关系。以图 6 为例，device(设备)拥有若干个数据属性，如 device_type(设备类型)等设备属性，同时也拥有一个 HAS_IPv4(存在 IPv4 地址)的对象属性，与它存在这个关系的节点为另一个 IP 节点。两个节点和其之间的连线等同于三元组关系(device, has_ipv4, ip)。

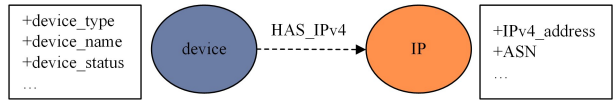

图 6 知识图谱节点与关系

Fig. 6 Knowledge graph nodes and their relationships

6.4 模拟园区网络空间地理知识图谱

该模拟园区采用从 Censys 获取的测量数据，包含位于北京的交换器、路由器、防火墙、继电器、服务器和终端主机等，构建一个小型网络空间地理知识图谱。如图 7 所示，该可视化结果展示了模拟园区的所有要素资产情况以及各个要素之间的联系。

从最内核的地理位置节点到最外层的相关安全漏洞，分别对应地理环境层、网络环境层(逻辑网络环境层和物理网络

环境层）、社会环境层和业务环境层。其中，地理环境层包括国家和城市节点；社会环境层包括设备开发商和设备所属组织两类社会角色；业务环境层包括物理设备或虚拟设备存在的相关漏洞。该园区网络模拟位置位于北京，是网络空间到地理空间的映射。时间参考轴为网络空间中可发现的网络设备记录了更新情况，可为当前的网络状况提供时间参考。

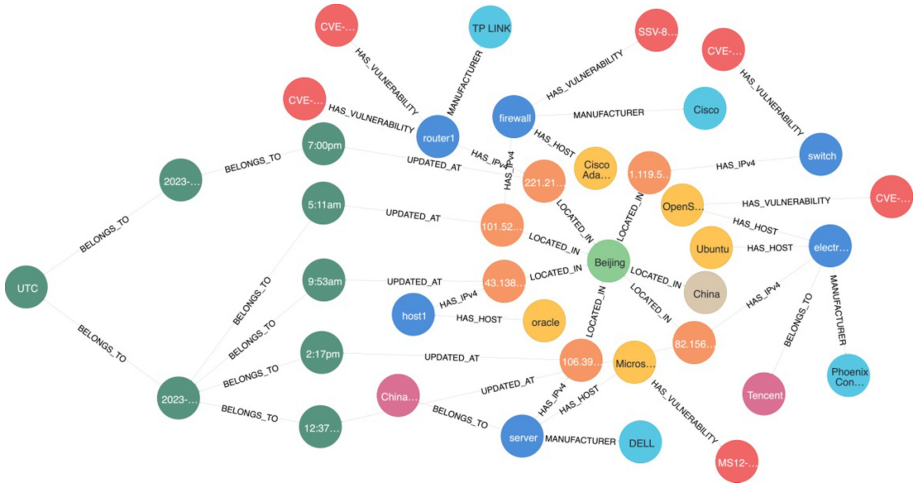

图 7　可视化的模拟园区知识图谱示意图

Fig. 7　Visualization of knowledge map diagram of simulated park

真实世界的网络-地理空间会更加复杂，存在多种地理环境、多种社会角色、多时刻发生的各类网络事件。该模拟园区网络在一定程度上展示了利用知识图谱去认知“地理-网络-社会”空间的可行性。

6.5　可视化知识图谱应用

网络空间资源信息数据规模极为庞大，本文设计实现的原型系统可以基于网络空间本体和知识图谱源数据成功抽取搭建知识图谱，实现了高效的知识存储和知识查询。此外，构建网络空间地理图谱对于加强网络空间的安全性具有重要意义。通过结构化表示网络空间和地理信息，知识图谱可以帮助识别潜在的安全威胁，如分析攻击路径、识别脆弱的网络节点和通信链路。以攻击路径分析为例，知识图谱可以存储和展示网络中各个节点和连接的详细信息，包括它们的配置、服务和存在的安全漏洞。通过分析这些数据，安全专家可以识别和预测潜在的攻击路径，帮助识别通过一系列脆弱节点和服务进行的攻击可能性。同时，网络空间地理图谱将地理信息与网络信息相结合，这有助于识别地理位置上的安全风险，如特定区域内频繁发生的网络攻击或自然灾害对网络的潜在影响。

6.6　系统评估

我们从两个关键方面对本文工作进行了评估：网络空间的层次模型以及网络空间地理图谱的原型系统。

对于网络空间的层次模型，我们从准确性、可扩展性、数据整合能力和可视化能力等方面与已有工作进行了比较，结果如表 5 所列。其中圆圈颜色深浅代表对每一项指标的满足程度。

表 5　IP 本体对象属性

Table 5　IP ontology object attributes

	准确性	可扩展性	数据整合	可视化
文献[5]	○	●	◉	○
文献[6]	◉	●	●	○
Ours	●	●	●	●

这些工作从不同角度推进了网络空间要素体系的完善，但仍然存在一些尚未解决的问题，导致对网络空间层次的概括仍显不足。具体表现为：1）忽略了地理空间事件对网络空间的影响。地理空间的灾害事件，如地震、洪水、飓风、火灾等，对网络安全的影响是多方面的。2）缺少网络空间的时间特性。除此以外，本文工作还给出了相应的可视化工具，并建立了一个小型电网的原型系统。

对于网络空间地理图谱的原型系统，目前仍然缺少将知识图谱实际应用到网络地理空间的相关工作。传统的可视化表达中，主要采用基于地理信息（如 CAIDA[25]，ZoomEye[26] 等兼容地理信息维度的网络空间综合信息平台）、网络拓扑或管理视图进行表达。随着网络技术的不断发展，这些工作也面临着挑战：1）单纯依靠地理位置信息与网络空间资源映射关系构建的网络空间信息模型难以匹配网络空间的发展现状；2）网络拓扑结构复杂易变，基于网络拓扑刻画存在时间维度上的局限性；3）网络管理[27] 的方式类似于知识图谱，但主要是一种基于协作关系的网络空间中的特殊场景，是较为成熟的网络空间资源表达方式。我们利用网络空间地理图谱将传统意义上零散、孤立的网络空间资源数据进行标准化、规范化和结构化的组织与呈现，提供了一种更为系统的知识表达方式，为网络空间的多维信息表达、关联分析和可视化建立了基础。但不可否认的是，本文工作仍然存在局限性。目前，原型系统仅在一个规模非常有限的模拟园区网络中得以实现，在未来工作中，我们计划将这一图谱应用于更广泛的网络范围，以进一步验证和扩展其有效性和适用性。

结束语　随着互联网的普及与发展，网络空间已经成为人类活动的新型空间形态。网络空间安全不仅关乎个人和组织的利益，也关系到国家的安全和社会的发展。本文基于对网络地理空间的已有研究工作，对网络空间的分层模型进行了完善，首次提出了一个带有时间参考轴的四层四级分层模型。时间特性的引入给网络空间注入了新的维度。此外，本文引入知识图谱技术，提出了一种基于网络空间测量领域

知识和实际测量数据构建网络空间资源本体的方法，系统性地提出了网络空间资源本体的基本构成元素、定义规范和建模方法。在模拟的小型园区网络中，完成了网络空间地理知识图谱的建立和可视化。网络空间地理图谱的建立能够为网络安全提供有价值的洞察和分析，加强网络防御，建立更安全和包容的网络环境。在未来，我们可以进一步拓展网络空间资源知识图谱的内涵，构建囊括网络空间安全、网络空间行为感知等多维度的，实现实用化、大规模，具备演进性和实时更新能力的网络空间地理信息图谱。

参 考 文 献

[1] FENG D G, LIAN Y F. Challenges and Countermeasures to Cyberspace Security[J]. Proceedings of the Chinese Academy of Sciences, 2021, 36(10): 1239-1245.

[2] MORGAN S. 2022 Official Cybercrime Report[EB/OL]. (2022-10-17) [2023-09-01]. https://s3. ca-central-1. amazonaws. com/esentire-dot-com-assets/assets/resourcefiles/2022-Official-Cybercrime-Report. pdf.

[3] MIAO C, WANG J, ZHUANG S, et al. A Coordinated View of Cyberspace[J]. arXiv: 1910. 09787, 2019.

[4] FANG B X. Coverage Areas of Cyberspace Security Technologies From a Hierarchical Perspective[J]. Journal of Network and Information Security, 2015, 1(1): 2-7.

[5] GUO L, CAO Y N, SU M J, et al. Cyberspace Resource Mapping: Concepts and Techniques[J]. Journal of Information Security, 2018, 3(4): 1-14.

[6] GUO Q Q, GAO C D, SUN K F, et al. Construction of a Hierarchical System of Cyberspace Elements based on the "People-Place-Network" relationship [J]. Geographical Studies, 2021, 40(1): 109-118.

[7] ZHANG L, ZHOU Y, SHI Q S, et al. Cyberspace Map Model with Tight Geospatial Correlation[J]. Journal of Information Security, 2018, 3(4): 63-72.

[8] CHEN S, GUO Q Q, GAO C D, et al. Concepts and Methods of Cyberspace Geographic Mapping[J]. Science & Technology Review, 2023, 41(13): 14-22.

[9] DONG C, JIANG B, LU Z G, et al. A Review of Knowledge Graphs for Cyberspace Security Intelligence[J]. Journal of Information Security, 2020, 5(5): 56-76.

[10] BATTY M. The geography of cyberspace [J]. Environment and Planning B: Planning and Design, 1993, 20(6): 615-616.

[11] WANG H F, DING J, HU F H, et al. An Overview of Large-Scale Enterprise-Class Knowledge Graph Practices[J]. Computer Engineering, 2020, 46(7): 1-13.

[12] SINGHAL A. Introducing the knowledge graph: Things, not strings. official blog[EB/OL]. (2012-05-16) [2023-09-01]. https://blog. google/products/search/introducing-knowledge-graph-things-not/.

[13] CICALESE D, AUGE J, JOUMBLATT D, et al. Characterizing IPv4 Anycast Adoption and Deployment [C]// Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies. 2015: 1-13.

[14] YANG J H, HE L, LI C L. Cyberspace Mapping-Principles, Techniques and Applications[M]// Posts and Telecom Press, 2023.

[15] COX K C, EICK S G, HE T. 3D Geographic Network Displays [J]. ACM SIGMOD Record, 1996, 25(4): 50-54.

[16] NSA/CSS Threat Operations Center. Bad Guys Are Everywhere, Good Guys Are Somewhere [EB/OL]. (2014-09-14) [2023-09-01]. https://nsarchive. gwu. edu/sites/default/files/documents/3469146/Document-01-NSA-CSS-Threat-Operations-Center. pdf.

[17] HE J B, LI X T. Awareness and Reflection on Geographic Information Classification and Coding[J]. Geography and Geographic Information Science, 2002, 18(3): 1-7.

[18] Cyberspace of the Central Cyberspace Affairs Commission. National cybersecurity incident response plan[EB/OL]. (2017-06-27) [2023-09-01]. http://www. cac. gov. cn/2017-06/27/c_1121220113. htm.

[19] Neo4j. Neo4j Graph Database [EB/OL]. (2023-09-01) [2023-09-01]. https://neo4j. com/.

[20] Neo4j. Cypher Manual [EB/OL]. (2023-09-01) [2023-09-01]. https://neo4j. com/docs/cypher-manual/current/.

[21] GRUBER T R. A Translation Approach to Portable Ontology Specifications[J]. Knowledge Acquisition, 1993, 5(2): 199-220.

[22] BIZER C. The D2RQ Platform [EB/OL]. (2010-03-04) [2023-09-01]. http://d2rq. org/.

[23] The D2RQ Mapping Language[EB/OL]. (2012-06-22) [2023-09-01]. http://d2rq. org/d2rq-language.

[24] CENSYS. The Censys Internet Map [EB/OL]. (2023-09-01) [2023-09-01]. https://about. censys. io/.

[25] CAIDA. Caida[EB/OL]. 2023. https://www. caida. org. /.

[26] ZoomEye. Zoomeye[EB/OL]. https://www. zoomeye. org/.

[27] BEN-ARTZI A, CHANDNA A, WARRIER U. Network management of tcp/ip networks: present and future[J]. IEEE Network, 1990, 4(4): 35-43.

WU Yue, born in 2002, Ph.D. Her main research interests include network measurement and security and so on.

HU Wei, born in 1977, professor. His main research interests include network information security and situational awareness.

（责任编辑：何杨）