

社交知识图谱研究综述

江旭晖^{1),2)} 沈英汉^{1),2)} 李子健^{1),2)} 王元卓^{1),3)} 尹芷仪^{1),2)} 沈华伟^{1),2)}

¹⁾(中国科学院计算技术研究所数据智能系统研究中心 北京 100190)

²⁾(中国科学院大学 北京 100049)

³⁾(中科大数据研究院 郑州 450046)

摘 要 作为通用的知识结构化表示形式,知识图谱被成功应用于医疗、金融、安全等领域.社交知识图谱是一种以人为中心的知识图谱,其融合了动态演化的社交知识.作为知识图谱概念的延伸,社交知识图谱涵盖人、物、事、地等异质信息及其复杂关联;由于其融入了来自社交网络的强时效性知识,能够准确地描述人员的即时状态及其演化趋势,被广泛应用于推荐系统、社交分析等以人为中心的应用中.当前,社交知识图谱的相关工作不断涌现,但缺乏统一的形式化定义以及系统性的分析.基于此,本文首先梳理了社交知识图谱的相关概念,并给出了社交知识图谱的形式化定义.然后从社交知识图谱的定义出发,对其动态性、异质性、情感性、互演化性等性质进行分析.接下来围绕社交知识图谱的生命周期,梳理了社交知识图谱的构建、融合、表示和推理的相关代表性工作.最后介绍了社交知识图谱的相关应用,并展望了社交知识图谱的未来发展蓝图.

关键词 社交知识图谱;社交网络;知识图谱;知识推理;图表示学习;知识抽取;网络融合

中图法分类号 TP18

DOI号 10.11897/SP.J.1016.2023.00304

A Survey of Social Knowledge Graph

JIANG Xu-Hui^{1),2)} SHEN Ying-Han^{1),2)} LI Zi-Jian^{1),2)} WANG Yuan-Zhuo^{1),3)}

YIN Zhi-Yi^{1),2)} SHEN Hua-Wei^{1),2)}

¹⁾(Data Intelligence System Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

²⁾(University of Chinese Academy of Sciences, Beijing 100049)

³⁾(Zhongke Big Data Academy, Zhengzhou 450046)

Abstract Knowledge Graph (KG) has received great attention and has been widely explored since its inception. Recently, there is a trend to introduce social information from Social Network (SN) into KG called Social Knowledge Graph (SKG). SKG is a kind of Knowledge Graph, which is human-centered and dynamic. It contains rich background information from KG and dynamic interaction information about people from SN. Considering that the information complementation of SN and KG can benefit human-oriented applications such as recommendation systems and social analysis. It is essential to propose a unified framework for the integration of SN and KG. However, there is currently no survey focusing on this research trend. In this paper, we introduce the SKG and give a systematic analysis of it. Specifically, we first introduce relevant concepts about SKG, and propose a formal definition of it. Next, we analyze the characteristics of SKG from four perspectives including heterogeneous, dynamic, emotional and coevolutionary. Then, around the life

收稿日期:2021-12-31;在线发布日期:2022-08-18.本课题得到国家自然科学基金(U1836206,U21B2046,62172393)与中原英才计划-中原科技创新领军人才项目(204200510002)资助.江旭晖,博士研究生,中国计算机学会(CCF)学生会会员,主要研究方向为社交知识图谱、动态图表示学习. E-mail: jiangxuhui19g@ict.ac.cn. 沈英汉(共同第一作者),博士研究生,中国计算机学会(CCF)学生会会员,主要研究方向为社交知识图谱、动态图表示学习. E-mail: shenyinghan17s@ict.ac.cn. 李子健,硕士研究生,主要研究方向为社交知识图谱、符号图表示学习. 王元卓(通信作者),博士,研究员,中国计算机学会(CCF)杰出会员、常务理事,主要研究领域为知识图谱、社会计算. E-mail: wangyuanzhuo@ict.ac.cn. 尹芷仪,博士,高级工程师,主要研究方向为智能算法安全治理. 沈华伟,博士,研究员,中国计算机学会(CCF)高级会员,主要研究领域为社交网络分析和社会媒体计算.

cycle of the SKG, we introduce the relevant representative studies of the SKG construction, representation, integration and reasoning. Finally, we introduce the related applications of SKG, and discuss its valuable future direction.

Keywords social knowledge graph; social network; knowledge graph; knowledge reasoning; graph representation learning; knowledge extraction; graph integration

1 引言

大数据与深度学习的兴起与繁荣为互联网领域的诸多应用带来了显著的发展^[1]. 在大数据分析的过程中引入通用或领域知识, 能够为智能系统提供客观世界的认知规律, 从而使其获得解决复杂问题的泛化能力, 这是目前人工智能研究的趋势.

知识图谱(Knowledge Graph)是一种典型的结构化知识表示形式, 由实体、关系和语义描述组成. 其起源于语义网络(Semantic Network)^[2], 经历了长期的发展与演化后, 在 2012 年, 谷歌首次提出了知识图谱的概念^[3]并将通用知识图谱应用至搜索引擎中. 目前, 国内外多家研究团队均构建并发布了相应的通用知识图谱. 著名的英文通用知识图谱有 Freebase^[4]、DBpedia^[5]、YAGO^[6]、ConceptNet^[7]等; 此外, 国内也涌现了许多中文通用知识图谱, 代表性工作如 CN-Probase^[8]和 CN-DBpedia^[9]是基于中文信息构建的大规模知识图谱, OpenKG^[10]是由国内多个学术机构发起的通用可信知识平台, OpenKN^[11]是面向网络大数据中通用知识构建的开放知识网络. 近年来关于知识图谱的研究工作也层出不穷. Yu 等人^[12]对知识图谱领域, 从理论基础、实际应用等方面进行了较为全面地阐述.

知识图谱具有强大的知识建模与表征能力, 被广泛应用于医疗、金融、安全等领域中. 知识图谱仅强调知识的存在性而非时效性. 然而, 在以人为中心的相关应用中, 无时无刻不在产生人与人、人与物的即时互动. 抖音、微信、微博、淘宝等以人为中心的应用需要实时分析用户产生的数据, 为用户提供个性化的服务. 由于知识图谱缺乏对强时效性知识的建模能力, 难以即时准确地分析用户行为和兴趣演化. 因此, 需要围绕以人为中心、动态演化的需求, 引入新的数据源对知识图谱的信息进行补充, 扩展知识图谱的概念. 当前, 融合社交网络信息知识图谱受到了产业界和学术界的广泛关注.

社交网络(Social Network)于 20 世纪 50 年代

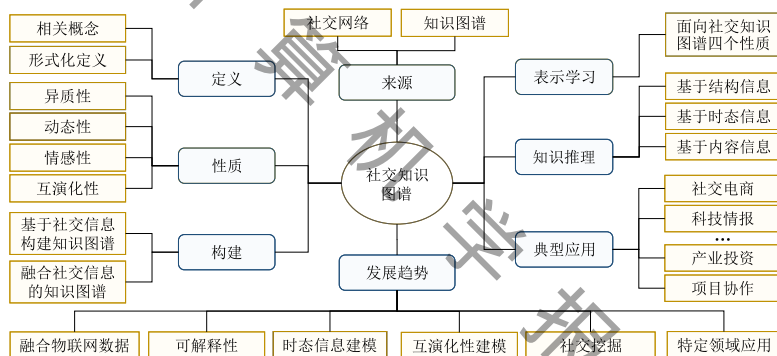
提出^[13], 是另一种典型的图结构数据. 如今所提社交网络通常指以微信、微博、推特、脸书等平台为载体的在线社交网络, 其定义为以社交媒体中的用户为节点, 用户之间的交互或关系为连边而构建的网络^[14]. 社交网络中以人为中心的强时效性知识即时准确地描述了人员的社交行为及其演化趋势, 能补齐传统知识图谱时效性弱的短板.

将社交网络信息融入知识图谱构建以人为中心、动态演化的知识图谱具有充分的理论依据与实践验证, 是建立在定性以及定量地深入剖析社交网络和知识图谱彼此的信息特点基础之上的选择, 且经过了丰富的实践验证. 知识图谱与社交网络均属于图数据的范畴, 但其性质有所不同. 定性地讲, 从建模对象的角度看, 知识图谱与社交网络存在着显著差异. 社交网络侧重建模人员以及人与人之间的交互, 对研究个体与群体的演化特性具有重要的意义. 而知识图谱侧重于建模包括人员在内的异质实体间关联, 蕴含可解释的背景信息. 从知识的演化频率分析, 社交网络的信息来自即时性的社交平台, 时序性强, 知识演化频率高; 知识图谱的信息往往来自开放网页、通用百科或领域文献等, 这些知识来源具有较高的可信度, 侧重于描述实体间存在关联而非强调关联的变化, 时序性弱, 知识演化频率低. 定量地讲, 从网络的同质与异质性分析, 社交网络是由人员组成的网络, 仅包含人员类节点, 是典型的同质图; 知识图谱中实体类型除了人员类节点之外, 还包括其他类型实体, 是典型的异质图. 从数据统计层面分析, 社交网络中存在无标度效应, 即少量节点拥有网络中大多数的链接; 而知识图谱不具有无标度效应, 而在关系类别方面存在长尾效应, 即知识图谱中的大部分关系归属于少数几种关系类型. 从对两者性质分析可以看出, 知识图谱与社交网络可以优势互补, 各取所长. 知识图谱能够为社交网络分析提供人员交互的知识性依据, 而社交网络中的信息又能及时补充或更新知识图谱中的时效性知识. 因此两者的结合潜力巨大, 对于特定目标与群体分析等都有着重要意义.

随着实际应用需求的强劲驱动,社交知识图谱的研究与应用正在被越来越多的科研机构与公司所关注.在学术界,2017 年 Yang 等人^[15]正式提出了社交知识图谱(Social Knowledge Graph)的概念,旨在实现社交交互信息与通用知识的背景信息互补.与社交知识图谱相关的研究也不断涌现^[16-22].值得注意的是,关于社交网络与知识图谱相结合,还有其他类似的概念,如谷歌提出的个人知识图谱^[23]、领英提出的异质职业知识图谱^[24]等概念.在工业界,对于社交知识图谱相关的实践探索也取得了显著成效.如拼多多提出“社交电商”概念,在通过历史浏览记录与购买记录建模用户兴趣的同时,借助社交网络互动分享好友兴趣,为用户推荐其社交好友感兴趣的商品;微信短视频推荐功能也会结合用户自身兴趣主题与用户所在社交好友圈内分享的热点话题为用户做高质量的短视频推荐.此外,社交知识图谱在科技情报、产业投资和项目协作等应用领域皆有

成功实践,证明了其巨大的潜力.

然而,当前社交知识图谱的概念尚未统一,且缺少对社交知识图谱性质及主要研究问题的总结和分析,无法形成系统性的理论体系.基于此,本文的目标是聚焦于社交知识图谱进行综述研究,为社交知识图谱研究提供系统性的理论支撑,从而推动社交知识图谱的实现与应用.考虑到目前未有研究工作提出社交知识图谱的形式化定义,且缺少社交知识图谱特性的深入分析.本文首次给出了社交知识图谱的形式化定义.其次,针对社交知识图谱当前研究碎片化、缺少完整体系的问题,全面梳理与介绍了社交知识图谱相关的研究工作,并从动态性、异质性、情感性和互演化性等角度进行详尽的分析,最后指出了未来应用与研究方向的展望.除此之外,本文全面整理了社交知识图谱的文献以及数据集等资源并构建了开源项目^①,旨在促进该领域发展.社交知识图谱的整体研究框架如图 1 所示.



本文第 2 节梳理已有的社交知识图谱相关定义,并且给出本文对于社交知识图谱的形式化定义;第 3 节论述并分析社交知识图谱的特性;第 4 节梳理社交知识图谱的构建与融合相关工作;第 5 节阐述社交知识图谱表示学习相关的工作;第 6 节整理基于社交知识图谱的推理相关工作;第 7 节介绍现有社交知识图谱相关应用;第 8 节展望社交知识图谱具有前景的研究方向;最后,在第 9 节对全文进行归纳总结.

2 社交知识图谱的定义

当前社交知识图谱的研究涌现,但对社交知识图谱概念并未形成统一的定义.本文梳理了社交知识图谱相关概念,归纳了社交知识图谱的研究特征,并在此基础上提出社交知识图谱形式化定义.

2.1 社交知识图谱相关概念

自知识图谱的概念提出之后,自然地,考虑到社交网络与知识图谱的丰富信息与广袤的应用,围绕两者结合的探索在学术界和工业界都始终存在.

学术界聚焦于社交知识图谱的理论研究.其中 Yang 等人^[15]提出了社交知识图谱的概念,旨在将学术社交网络中的学者节点与通用知识库中的若干研究领域概念关联,以此实现两个网络的融合,并用于学者职业档案匹配与异常研究领域检测等任务.这篇工作首次提出社交知识图谱,但其缺乏对于社交知识图谱定义以及性质的深入剖析.在此之后,围绕社交知识图谱进行的理论研究^[16-20]也在不断地丰富着社交知识图谱概念的内涵与外延.其中 Tweeki^[16]和 SocialLink^[17]探索基于社交网络与通用知识图谱

① <https://github.com/jxh4945777/Social-Knowledge-Graph-Papers>

融合来构建社交知识图谱的方法, SocialScope^[18-20] 聚焦基于社交信息的信息获取, 用于构建社交知识图谱并服务于事件分析等应用。

工业界则聚焦于社交知识图谱的业务落地, 微软 Alonso 等人^[21] 认为社交知识图谱是从推特、脸书等在线社交平台的文本信息中抽取用户、链接、主题、交互文本等知识构建而成的知识图谱, 用于服务微软必应搜索、新闻推荐等应用。谷歌 Balog 等人^[23] 从用户的视角审视知识图谱, 并提出个人知识图谱 (Personal Knowledge Graph) 的概念, 以用户为中心, 融合用户社交行为、通用知识图谱、领域知识图谱等信息构建知识图谱, 为用户提供定制的个性化服务。领英 Shi 等人^[22,24] 聚焦于知识图谱在领英业务中的应用, 利用领英的求职者、职业技能、职业头衔、公司等数据构建社交知识图谱, 指出了社交知识图谱具有异质性和动态性并加以利用, 为领英的用户职业生涯规划等业务提供支持。

总体来讲, 以上工作在社交知识图谱的理论基础、应用落地等方面进行了探索, 也从侧面印证了社交知识图谱正在受到学术界和工业界的广泛关注。然而, 这些工作没有明确给出社交知识图谱形式化定义, 缺乏统一的定义以及研究框架, 无益于后续相关研究工作的开展。本文在总结上述工作的前提下, 归纳了社交知识图谱的两点主要研究目标: (1) “以人为中心”, 即社交知识图谱是以人为中心展开的; 需要建立人员为中心, 多类型实体关联的知识图谱, 展示人-事-物-时间-地点之间的复杂关联, 用于分析人物画像、解释人员行为等, 服务于智能推荐、风险分析和情报挖掘等应用; (2) “动态演化”, 即人员的状态、行为与关联并非一成不变, 而是动态演化的, 状态之间既包含共现也存在因果联系, 需要从时间维度出发挖掘潜在的因果联系并分析其演化规律。

综上所述, 社交知识图谱的本质是以人为中心, 动态演化的知识图谱, 是社交网络与知识图谱在概念、方法、模型方面的交叉融合, 最终目标是服务于以人为中心的应用。

2.2 社交知识图谱定义

基于上文所提社交知识图谱相关概念与研究目标, 本小节给出社交知识图谱形式化定义如下:

给定社交知识图谱 $G=(V, E, P, R, I, T)$, V 是社交知识图谱中节点集; E 是社交知识图谱中的边集; P 是节点的类型, 包括人员类节点 P_{person} 和其他与人员类节点直接相关的节点类型集合 P_{other} 。社交知识图谱中, 依据边所关联实体类型, 边可分为〈人-人〉、

〈人-其他实体〉和〈其他实体-其他实体〉三种; 依据边的时态特性, 边的类型分为稳态关系 (Relation) R 与瞬态交互 (Interaction) I 。稳态关系 R 是相对稳定的关联方式, 瞬态交互 I 是相对瞬时的关联方式; T 是边的时间范围集合。特别地, 存在节点类型映射函数 $\phi(\cdot)$, 对于节点 v , 有 $\phi(v)=P_u$, 其中 $P_u \in P$; 存在边的类型映射函数 $\psi(\cdot)$, 对于关系类型的边 e_r , 有 $\psi(e_r)=R_e$, 其中 $R_e \in R$; 对于交互类型的边 e_i 有 $\psi(e_i)=I_e$, 其中 $I_e \in I$; 对于边 e , 存在时间属性映射 $\eta(e_r)=[t_{e_r}^{\text{start}}, t_{e_r}^{\text{end}}]$, 其中 $t_{e_r}^{\text{start}}, t_{e_r}^{\text{end}} \in T$ 分别对应边的开始时间和结束时间, 对于瞬时交互 e_i , 有 $\eta(e_i)=t_{e_i}^{\text{occur}}$, 其中 $t_{e_i}^{\text{occur}} \in T$ 是交互发生时间。

图 2 中展示了以美职篮球员为例的社交知识图谱, 基于“以人为中心”和“动态演化”的研究目标, 社交知识图谱的信息围绕人物展开, 包含稳态关系和瞬态交互, 能够建模网络的动态演化。具体地, 该示意包含源自社交网络 (如推特、脸书) 的球员信息, 和知识图谱 (如维基百科) 中与人相关的其他类型实体信息, 如球队、毕业院校、国籍、城市等, 具有异质性; 图中存在人员效力球队、人员所属球队位置以及国籍等稳态关系, 以及人员在社交平台中的点赞、评论等瞬态交互, 具有动态性; 人与人之间关系具有情感性, 如队友、竞争对手等; 网络中多种异质类型数据与时态知识相互作用, 具有互演化性。下节将对以上性质展开详细分析。

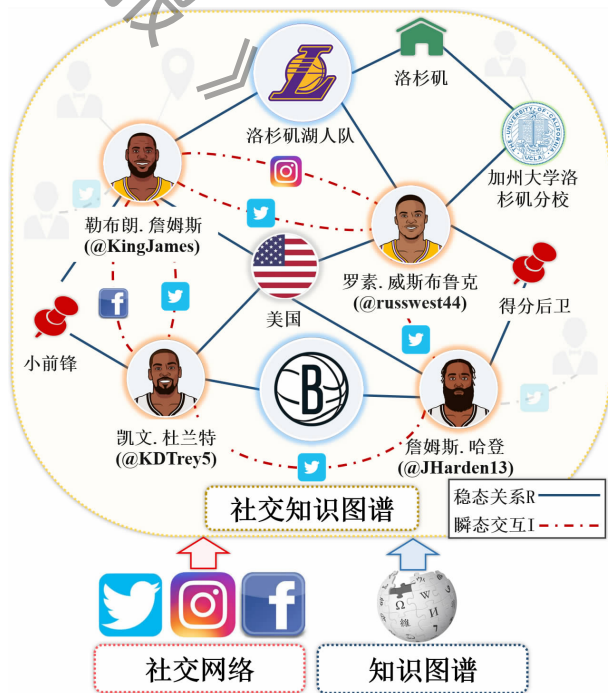


图 2 社交知识图谱示意

3 社交知识图谱的性质

根据定义,可以发现社交知识图谱在继承了知识图谱的特性基础上,由于融合了社交网络中以人为中心且动态的信息,丰富了知识图谱的内核,并带来全新且富有价值的特性。

具体来讲,不同于在线社交网络仅仅关注于建模用户节点以及用户之间的关系与交互,社交知识图谱中源自于知识图谱的能够代表与用户相关的背景知识信息并服务于应用大量的异质节点,具有异质(Heterogeneous)性;不同于通用知识图谱和领域知识图谱仅利用三元组建模知识信息,社交知识图谱在继承知识图谱中大量稳态知识的同时,能够保留人与人之间大量的瞬态交互信息,具有动态(Dynamic)性;社交知识图谱以人为中心,聚焦于从知识图谱和社交网络两个角度建模人物,在更全面地描述人物本身的同时,也能够描述人物之间的情感,具有情感(Emotional)性;社交知识图谱中人物与相关实体的变化会影响人物的行为,而行为又会反作用于人物与相关实体的变化,具有互演化(Coevolutionary)性。本节将从以上4个角度深入地剖析社交知识图谱的性质。

3.1 社交知识图谱的异质性

社交知识图谱的异质性包括节点异质和数据源异质两个层次。其中节点异质继承自知识图谱中丰富的实体类型。传统社交网络往往只考虑同质用户节点之间的交互,而忽略了丰富的异质信息,而这些异质信息作为人物节点的有效补充,对于社交信息的充分挖掘同样十分重要。传统的开放知识图谱具有十分丰富的节点类型,难以找寻其背后的模式规律,不同于此,通过对于社交知识图谱相关概念^[16-22]以及本文所给出的定义可以发现,社交知识图谱以人为中心,从知识图谱中引入与人相关的异质信息而非将通用知识图谱整体引入,便于通过用户节点与其他异质节点的交互找寻其背后用户的行为模式。因此,需重点关注如何建模和有效利用社交知识图谱的异质节点信息。

社交知识图谱源自于知识图谱和社交网络的多源数据的融合^[22],其数据源异质,具体表现为社交网络和知识图谱的网络结构特点以及节点的属性特点都存在差异性,其中社交网络的交互数据还包含大量的多模态内容信息,在实现信息互补的同时,也

为社交知识图谱的融合带来了挑战。构建连接社交网络与知识图谱的桥梁,跨越两个网络的语义鸿沟实现深度融合,继而对其中的多模态数据进行利用,是社交知识图谱面临的关键问题之一。

3.2 社交知识图谱的动态性

社交知识图谱具有动态性,这是由于知识图谱的关系与社交网络的交互均带有时态信息^[25-26],这类知识被称为时态知识。社交知识图谱中的时态知识具有多粒度、多类型、时序性的特点。时态知识的多粒度是指,社交知识图谱中知识的时态信息精确范围可以从年到秒,不同的时态粒度反映了知识中时态信息的精确程度。时态知识的多类型是指,依据知识的延续性进行划分,社交知识图谱中有描述交互的瞬态知识,其时态信息使用时间点表示;也有描述相对稳定关系的稳态知识,其时态信息使用时间区间表示,这两类知识的时态信息表示是截然不同的^[27]。时态知识的时序性是指,时态知识间并非毫无关联,不同时态知识间存在一定的规律性与因果性,按照时间先后形成了时态知识的序列,即时序信息。时序信息描述了知识间先后顺序关系,是研究知识演化规律重要依据。

建模社交知识图谱中的时态信息与时序信息并研究知识的演化规律是一项具有挑战性的任务。首先,考虑到时态知识多粒度,需研究如何将时态信息整合成统一的时序信息表示,并且在此基础上实现不同时间粒度、瞬态和稳态知识的表示与推理;其次,考虑到时态知识多类型,需研究如何建模瞬态交互以及稳态关系,并挖掘瞬态交互以及稳态关系之间的关联;最后,基于前两步的基础上,考虑到多粒度、多类型时态知识之间复杂的时序关系,需研究时态知识之间的先后序中蕴藏的因果规律,挖掘可解释的时态知识链,并以此进行时态知识的推理和预测。

3.3 社交知识图谱的情感性

社交知识图谱中,人与人之间除包含交互和关系信息外,还包含丰富的情感信息,这些情感信息反映了深层的人际关系^[28]。例如,在同一公司工作的员工,因此他们间的关系都是同事关系,但是人与人之间的关系有远有近,其间可能只是普通的同事关系,甚至彼此不认识,而也可能是亲密的同事关系,彼此经常合作互助。同样,在不同的关系下,相似的交互模式也会反映出不同的情感特性。例如,两个用户之间是同事关系,那么他们之间每天规律的通信交互可能只是工作原因,而两个同学之间规律的交

互更有可能代表他们之间的亲密关系。

社交知识图谱中的情感信息在以人为中心的应用中具有重要的意义。以用户聚类为例,在不考虑情感信息时,用户聚类往往只能依靠用户间的交互频率等内容信息进行聚类,这样聚类所得的簇仅能代表用户间交互密集,但是这种交互密集很容易被各种因素破坏。而在考虑了情感信息后,聚类算法能够将真实关系更加亲密的用户聚为一类,簇内更为紧密,且簇间的边界更为明显,这样聚类所得的簇更为稳定。考虑到实际数据中对于情感信息缺乏标注,如何在低资源条件下挖掘情感信息并对情感信息进行利用是社交知识图谱的重要研究方向。

3.4 社交知识图谱的互演化性

社交知识图谱的互演化建立在社交网络与知识图谱两个数据源深度融合的基础上,其整体的表现为人物的属性以及人物与其他相关实体之间关系的变化会影响人物的行为,而人物的行为又反过来会影响人物的属性以及人物与其他相关实体之间关系的变化。以美国职业篮球运动员为例,霍顿-塔克加入湖人队在知识图谱中体现为其与湖人队和其他湖人队的球员建立起稳态关系,这些关系会影响到霍顿-塔克的人物属性及其与其他球员之间在社交平台的交互,而社交网络中频繁的交互也会反过来作用于霍顿-塔克与湖人队的其他球员之间的关系变化,例如队友关系变化为挚友关系,以及与其他球员相关的人物或地点建立起联系。

社交知识图谱的互演化性与前文社交知识图谱的异质特性、动态特性和情感特性紧密联系。从节点的异质角度考虑,人物与其他类型节点的关系变化会影响人物的行为;从数据源的异质角度考虑,源自于社交网络数据的变化与源自于知识图谱数据的变化相互影响、相互映照^[22];从动态的角度考虑,稳态关系与瞬态交互之间也存在互相影响,对于社交知识图谱中的人物,稳态关系决定了其瞬态交互对象,而瞬态交互也改变着其稳态关系^[29];从情感角度考虑,人物之间关系背后的情感信息会影响其交互的情感态度,反之亦然^[28]。相较于原有基于单一网络或简单网络融合的分析方式,社交知识图谱的出现为挖掘互演化现象潜在的成因提供了优质的数据基础以及有潜力的研究方向。

异质特性、动态特性、情感特性以及互演化特性是社交知识图谱的重要特性,随着对社交知识图谱的研究深入,其涵盖范围会更广,内核也更为丰富,

将有更多的特性需要在未来的工作中探索。

4 社交知识图谱的构建

现有社交知识图谱的产生方式主要分为两大类,即通过社交网络中的信息构建社交知识图谱,以及对社交网络与通用知识图谱进行融合所得的社交知识图谱。社交知识图谱的构建是后续工作的基础,但通过第2节对定义和示例的分析可知,社交知识图谱包含多样的节点、边和时态信息,如何整合源自多源异构数据的丰富的节点和边,建模不同类型时态信息使该工作充满了挑战。本节将对于两种社交知识图谱的构建方式的工作进行系统性地梳理,并且对于两种方式的优缺点进行分析,在此基础上提出同时基于社交信息知识获取和跨网络融合的社交知识图谱构建。

4.1 基于社交信息构建知识图谱

在线社交平台中时刻产生着大量动态且形式多样的内容,其中蕴含了丰富的知识信息,包括人与人之间动态关系和与人有关的事件,如何利用丰富的社交信息构建社交知识图谱是该研究方向的重点。基于社交信息构建知识图谱示意如图3所示。

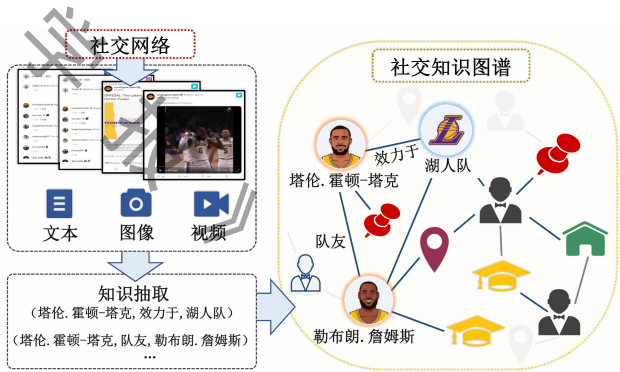


图3 基于社交信息构建知识图谱示意

与通用知识图谱的构建方式相似,社交知识图谱的构建是建立在对于大量来自于社交网络的非结构化数据(如用户或媒体发布的文本内容)和半结构化数据(如用户的简历信息)利用自然语言处理的方式进行知识抽取的基础之上,因此知识抽取的准确程度关乎所构建的社交知识图谱质量。通用知识图谱构建过程中所需的知识抽取技术主要包括两大部分:命名实体识别(Named Entity Recognition)和关系抽取(Relation Extraction)^[22]。工作^[21]聚焦于社交知识图谱本体层概念及其关系的设计,且给出了

基于推特社交数据所构建的社交知识图谱示例。系列工作^[18-20]设计并提出了一套完整的社交知识图谱构建框架——SocialScope。其中,工作^[18-19]考虑到对于在线社交平台所产生的大量社交信息有效利用,提出了 SocialScope 框架且给出了详细的构建流程,旨在通过大量、动态的社交信息构建动态的社交知识图谱,并且探讨了在社交知识图谱上进行用户推断的相关应用。工作^[20]则是从利用大量动态社交数据进行事件抽取的角度对于 SocialScope 框架进行了扩展。工作^[22]以领英社交平台的数据为例,全面阐述了社交知识图谱的构建生命周期,指出命名实体识别和关系抽取是社交知识图谱构建过程中知识抽取的重要手段,这篇工作同时也探讨了社交知识图谱构建过程的难点,即在缺乏标注,社交数据包含噪音等受限条件下,如何构建高质量的社交知识图谱。

以上工作皆是从社交知识图谱构建整体上进行了研究,而有关社交知识图谱中的知识抽取具体研究工作也层出不穷。关于社交知识图谱构建过程中的命名实体识别,研究^[30]面向于社交数据中的命名实体识别任务,使用双向长短时记忆神经网络,结合条件随机场提出了一套多通道的社交文本命名实体识别框架;工作^[31]考虑到社交网络中用户所发布的信息类型多样性,提出了面向社交文本的多模态内容命名实体识别模型;工作^[32]则是考虑到社交文本中的噪音对于命名实体识别任务的影响,因此将多任务的方式引入社交数据命名实体识别任务,通过多个有关命名实体分割的次级任务辅助命名实体分类的首要任务,以此来实现多任务协同的社交文本命名实体识别;工作^[33]聚焦于中文社交平台,考虑到中文文本需要分词等特点,针对性地设计了面向中文社交文本的命名实体识别模型。

关于社交关系抽取,TransNet^[34]考虑到其不同于传统关系抽取的任务,往往社交关系隐含表示在文本当中,且难以用单一的标签描述,因此,这篇工作首先定义了社交关系抽取(Social Relation Extraction)任务,其中提出用多个标签表示社交关系,并且针对性地设计了社交关系抽取模型 TransNet,通过自编码器结构结合 Trans 机制学习社交关系的嵌入式表示;LTE^[35]关注到社交关系的动态变化,提出了动态关系抽取框架;FL-MSRE^[36]聚焦于小样本条件下多模态数据中的社交关系抽取任务,在 BERT 模型的基础上针对性地设计了 FL-

MSRE 模型;考虑到以往工作关注于从微博等公开发布的内容中抽取社交关系,而缺少从对话中抽取社交关系,工作^[37]提出了基于多轮对话的社交关系抽取数据集 DialogRE,旨在研究通过多轮对话抽取包括对话人以及其中提到的任务之间的社交关系;基于此数据集,RAEGCN^[38]提出了基于图神经网络的对话社交关系抽取方法;工作^[39]则利用异质图神经网络,实现同时考虑包括文档和人物等节点之间联系的社交关系抽取。除传统的实体与关系抽取,考虑到社交信息的动态性,在用户所发布的内容中还包括当下热点和前沿事件。

从社交信息中有效地对于事件进行抽取长期以来受到研究者关注^[40],相关的研究工作也层出不穷。EventKG^[41]面向动态的事件信息,提出了事件为中心的知识图谱。工作^[42]从事件抽取、事件补全、事件推断和事件预测 4 个角度阐述包含事件信息知识图谱构建过程,全面梳理了相关研究工作。

通过社交信息构建社交知识图谱不仅通过知识抽取的方式,实现了社交内容信息与结构信息的有效利用,同时能够实现社交知识图谱动态信息及时地更新。然而,考虑到社交信息本身的局限,在知识图谱构建时所使用的信息来源于推特、微博等公开发布的动态内容,缺少对相对稳定的知识(如用户的背景信息、亲属关系等)进行有效建模,具有一定的片面性。此外,社交平台所发布的内容信息往往包含着大量的噪音,且用户的倾向性会导致部分内容存在偏差甚至冲突,社交平台快速迭代的内容中还包含大量虚假信息。因此,在后续研究工作中需要考虑如何利用高质量的通用知识验证从社交信息中抽取的知识真实性,以及考虑如何准确、高效地处理存在冲突的知识,甄别虚假信息,实现基于社交信息的高质量社交知识图谱构建。

4.2 融合社交信息的知识图谱

在线社交平台中有以用户为中心,用户之间蕴含大量动态交互的社交网络。同时,诸如 Freebase^[4]、DBpedia^[5]等高质量的知识图谱以三元组的形式描述现实世界中的通用知识。将现有的社交网络与通用知识图谱进行融合,是社交知识图谱的另一种重要构建方式。

如图 4 所示,给定社交网络与通用知识图谱两个图,融合的目标在于将两个图融合成一个同时包含社交信息和通用知识信息的社交知识图谱。两者的融合有利于实现信息互补和增强信息的跨网络流

动,结合各自网络的特点与优势,实现信息互补,与基于社交信息构建的方式不同,通过网络融合的方式构建社交知识图谱重点不是通过知识抽取的方式将半结构或无结构的内容转化成结构化知识,其重点以及难点在于如何对现有的社交网络与知识图谱构建链接的桥梁并实现网络信息融合。

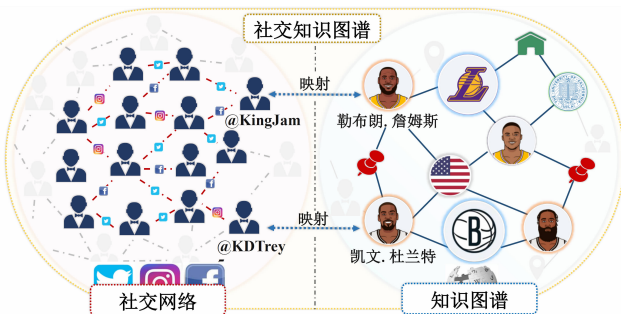


图4 融合社交信息的知识图谱示意

本节从社交网络和知识图谱两个角度出发对社交知识图谱网络信息融合的相关研究工作进行梳理。从社交网络的角度出发,工作^[43]指出用户简介数据不对称、用户产生的内容异构以及网络的稀疏性和噪声等社交网络跨网络融合所存在的挑战,对于社交网络间的融合进行了详细地定义,并且在此基础上提出了通用框架,即先从包括用户介绍、用户发布内容、网络结构特征等信息中进行特征抽取,然后在此基础上进行模型预测。在此框架下,这篇文章对于相关的研究工作、数据集和评价指标进行了梳理。近年来,有工作^[44]关注到社交网络的动态性并设计融合模型DNA,旨在通过用户之间的动态交互信息辅助社交网络融合。

从知识图谱的角度出发,工作^[45]根据现有的实体对齐方法的技术特点进行了分类,并在此基础上构建了开源的跨网络实体对齐库OpenEA;工作^[46]进行了大量实验分析,对于跨网络实体对齐的组件包括嵌入式表示模块、对齐模块以及其他信息建模模块进行了功能性分析,并且提出了新的数据集DBP-FR;BootEA^[47]提出一种基于BootStrapping的方式进行半监督实体对齐方法;MultiKE^[48]考虑到不同类型的信息对于实体对齐任务的增益,提出了一种新的实体对齐框架,从属性、实体名、关系三个角度来学习实体的嵌入式表示,并且给出了多种结合策略;BERT-INT^[49]基于BERT设计了只利用属性信息的跨网络实体对齐方法。考虑到图神经网络能有效建模图结构信息并且识别同构子图,因此近年来有研究考虑通过图神经网络实现网络融合。GCN-

Align^[50]首次基于图卷积神经网络设计了网络融合模型;RDGCN^[51]引入对偶关系图用于建模网络融合过程中知识图谱所包含的复杂关系信息,并且提出了关系感知的对偶图神经网络;R-GCN^[52]通过权重矩阵建模不同类型的关系;AVR-GCN^[53]将R-GCN与Trans方法结合,对于R-GCN进行了改进;HGCN^[54]考虑到以往的实体对齐方法没有利用好实体之间的关系来辅助实体对齐,因此该工作基于图神经网络,提出一种新的实体和关系的联合学习框架用于实体对齐;MuGNN^[55]则考虑到了网络融合过程中可能存在图本身不完整的情况,因此提出了先进行网络内关系补全再进行网络融合的操作,并且通过联合学习的方式进行优化;AliNet^[56]考虑到两个融合的网络中相对应实体之间可能存在不同的子图结构,因此引入注意力机制和门控机制用于缓解邻居非同构的问题。

以上网络融合的研究从社交网络或知识图谱的单一角度出发,融合目标也是社交网络-社交网络或知识图谱-知识图谱的组合,在进行社交网络与知识图谱的融合时存在一定的局限性。社交网络与知识图谱本身结构特点不同,具有异构性,因此难以直接依靠结构信息进行融合,且两个网络的内容信息也各有所侧重。因此,需要同时从社交网络和知识图谱两个角度出发,明确网络融合时连接两者的桥梁,并且针对性地设计模型。SOCINST^[57]利用社交节点信息以及通用知识库辅助社交文本中的实例识别。该工作较早考虑到在具体任务中,将社交网络中的信息与知识图谱的信息进行结合实现信息互补,继而服务于相关应用。工作^[15]首次明确社交网络与知识图谱融合,并明确提出社交知识图谱概念,其面向于学术社交网络,设计方法GenVector,旨在将学者与其研究领域相关概念的链接作为桥梁,融合社交网络与知识图谱。谷歌所提出的个人知识图谱^[23]也是社交网络与知识图谱融合的产物,其旨在围绕特定用户个体,将知识图谱和社交网络等图中与用户相关的信息进行结合;与之相似的还有领英的工作^[24],利用与用户职业相关的异质知识对于社交网络进行扩充,服务于诸如职业推荐等具体业务。在社交电商领域,工作^[58-60]以用户-物品-关联实体作为桥梁,将社交网络与知识图谱进行结合,对于知识图谱而言能够进行知识的补全^[59],对于社交网络而言能够精准刻画用户兴趣继而服务于内容推荐^[58,60]。TIMME^[61]则将部分精英人群作为跨越社交网络与

知识图谱的桥梁,服务于倾向性预测。

有工作^[16-17]关注于社交知识图谱数据集的构建,Tweeki^[16]项目^①旨在将推特发布的内容所提到的实体作为桥梁,连接到知识图谱中对应的知识,提出了 Tweeki 数据集用于后续的社交知识图谱研究;SocialLink^[17]项目^②旨在将推特中的用户与 DBpedia 知识库中的实体相连,搭建连接社交网络与知识图谱的桥梁.具体来讲,SocialLink 通过推特提供的接口,对于知识库出现的人名进行检索,然后再依据社交网络的用户表示学习所得到的嵌入式表示对候选的用户进行排序,最终进行连接。

通过社交网络与知识图谱网络融合构建社交知识图谱的方式,实现了这两种图数据的信息互补,且能够作为一种有效的纠错手段^[62],通过两个图中存在关联的信息交叉验证,实现对于原图中存在的错误信息进行纠正。

虽然以融合的方式构建的知识图谱已经在相关应用中取得一定的进展,但仍存在许多挑战需要后续进行研究探索.其一,现有社交网络与知识图谱融合的方法都仅仅利用了部分信息,在融合时没有充分考虑社交网络的动态性特点,例如没有有效地解决如何将社交网络中的动态演化特点融入社交知识图谱当中.其二,缺乏研究工作能够充分地考虑社交网络与知识图谱结构的异构性,以及内容的差异性,导致两个网络的融合不充分且网络之间的桥梁存在噪音或缺失,影响后续基于社交知识图谱的应用.其三,社交知识图谱在融合后结构并不是一成不变的,还具有互演化的特性,即考虑到社交网络动态产生的新内容也会更新社交知识图谱^[63],需要将知识获取与网络融合有机结合,即在网络融合基础上,设计方法挖掘社交网络或开放网页所产生的新内容信息,通过知识获取的方式将新的内容转化为结构化知识,并对于社交知识图谱进行更新,如图 5 所示。

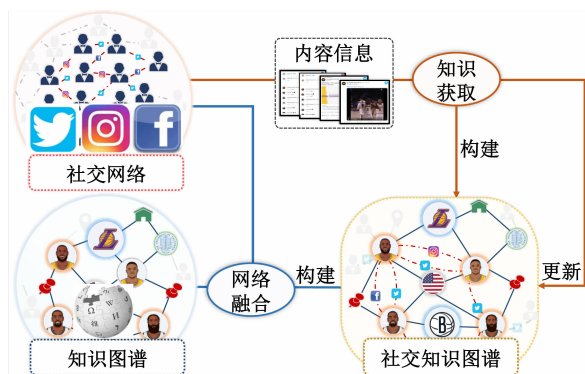


图 5 结合知识获取与网络融合的社交知识图谱构建示意

5 面向社交知识图谱的表示学习

面向社交知识图谱的表示学习是挖掘社交知识图谱中数据潜在规律的主流手段,也是面向下游任务的通用解决方案,具有广泛的应用基础.本节聚焦于分析如何在充分考虑社交知识图谱特性的基础上,依据社交知识图谱的信息进行表示学习方法的设计.具体地,根据社交知识图谱数据特点,以及从前文定义中归纳而出社交知识图谱的特性,本节将从异质性、动态性、情感性和互演化性四个角度展开分析,并对于相关的研究方法进行梳理。

5.1 面向社交知识图谱异质性的表示学习

社交知识图谱的异质性包括节点异质和数据源异质两个层次.节点异质指社交知识图谱中除了人物节点之外,还包括大量与人相关的异质节点.为深入探究图中节点的异质性,Sun 等人在 2012 年提出了异质图(Heterogeneous Information Networks)的概念^[64-65],其规定节点类型与边的类型之和大于二的图为异质图,包括社交知识图谱在内的图数据都属于异质图的范畴.因此,异质图的相关研究是社交知识图谱节点异质性挖掘的重要参照. Shi 等人分析了图数据结构的异质性并撰写了中文综述^[66],文中首先明确了异质信息网络的中文概念并给出了形式化定义,接着梳理了异质图的经典与前沿工作,然后对于异质信息网络的应用从安全、社交和生物制药等角度进行了叙述,最后对异质信息网络的发展提出了展望. Yang 等人的工作^[67]是另一篇关于异质图的代表性综述文章,不同于上述 Shi 的综述,这篇文章侧重于提出统一的框架用于描述已有的异质图表示学习方法,并对于代表性的方法进行详尽的评测.上述综述工作对于异质图的梳理方式与性质剖析值得社交知识图谱的研究借鉴。

异质图表示学习是异质图数据挖掘的主流方式,近年来,随着深度学习技术的发展,利用深度学习方法进行异质图表示学习的方法涌现,其主要分为四类:基于随机游走机制的表示学习、基于关系三元组的表示学习、基于异质图神经网络的表示学习以及神经-符号结合的表示学习方法。

基于随机游走机制的异质图表示学习的核心思想是通过随机游走机制对于图中节点进行采样,考

① <https://ucinlp.github.io/tweeki/>

② <http://sociallink.futuro.media/>

考虑采样路径上节点的共现关系,并通过结合 Skip-Gram 机制^[68]更新节点表示。模型 Metapath2Vec^[69]和 HIN2Vec^[70]首先将随机游走机制应用于异质图中,通过预先定义的元路径对于随机游走的采样进行约束,通过不同元路径采样得到的节点序列最大化共现概率,实现考虑异质信息的图表示学习。Star2vec^[24]是针对社交知识图谱设计的表示学习方法,其围绕人物及相关背景信息设计随机游走模型进行人物节点的表示学习。基于随机游走机制的表示学习能够有效利用异质图的结构信息,但对社交知识图谱中节点属性信息和内容信息的利用有限。

对于诸如 DBpedia、Wikidata 和 Probase 等通用知识图谱,其节点和边的异质类型十分丰富,因此难以通过元路径对于复杂的节点间关系进行枚举。对于该类异质图数据的表示学习广泛采用基于关系三元组转移(Translation)机制的一系列方法,后简称 Trans 方法,其核心思想是对包含〈头实体,关系,尾实体〉的关系三元组数据中元素的向量表示进行约束,使得头实体与关系的向量之和尽可能地接近尾实体的向量表示,在此约束下实现考虑丰富节点和边类型的异质图表示学习^[12]。TransE^[71]首次提出上文所提到的基于关系三元组的约束关系用于知识图谱的表示学习。TransH^[72]考虑到知识之间一对多/多对一/多对多的复杂关系,因此提出通过超平面对于头实体和尾实体进行投影,用于区分复杂关系。TransR^[73]考虑到不同关系类型包含的不同语义信息,因此提出依据关系类型将实体映射到不同的语义空间再结合三元组约束进行表示学习。TransA^[74]考虑到知识图谱的不同规模与结构,提出了包含局部自适应机制的知识图谱表示学习方法。上述 Trans 系列的方法能够在异质信息十分丰富的图中对于节点和边的表示同时进行有效地建模,但是其仅仅从微观层面对于每个三元组进行表示,缺乏从整个图的宏观角度进行考虑,且难以适用于社交知识图谱中异质节点类型有限的情况。

将基于随机游走的方法与基于关系三元组的方法结合,能够实现多角度地对于异质信息进行深入挖掘,RHINE^[75]在随机游走时将元路径预先分成了两类,即从属关系和交互关系,对于从属关系,其考虑通过欧氏距离度量相似性,对于交互关系,其考虑通过 TransE 类似的关系三元组约束进行建模。PRE^[76]则是关注于传统的 Trans 系列表示学习方法没法很好地捕获到高阶关系之间节点的相似性,因此设计了一种联合学习的方式,首先通过 Trans

系列方法学习节点的表示,并且通过关系三元组约束的打分函数进行打分,用于指导图上的随机游走概率,然后通过 Skip-Gram 机制再进行节点表示的优化,通过联合学习的方式能够对于 Trans 系列方法没有充分学习的节点表示进行节点表示的更新。

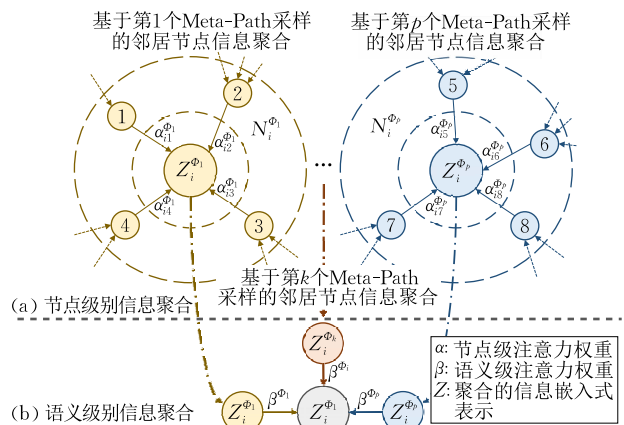


图 6 异质图神经网络模型 HAN 示意图

图神经网络将神经网络结构以及诸如卷积、循环等广泛使用的神经网络机制^[77-78]迁移到图结构数据上,其核心思想是通过神经网络建模的消息传递机制和节点信息聚合机制迭代式地对于节点的表示进行更新^[79-80]。基于图神经网络的异质图表示学习工作在消息传递机制和节点信息聚合机制中考虑到了节点和边的异质性,并且建模了异质节点之间的复杂关联^[81],适用于节点属性信息和内容信息丰富的社交知识图谱。其中,R-GCN^[52]考虑到异质图中边的类型丰富,但缺乏有效的图神经网络建模方式,因此设计模型在信息传递时对于不同类型的边使用不同的权值矩阵,且采用共享权值矩阵的方式,将不同类型边视作多个权值矩阵的带权加和,以此缩小训练的参数量。HAN^[82]是将异质图与图神经网络结合的一篇经典工作,不同于其他图神经网络直接聚合邻居信息,HAN 通过 Meta-Path 采集到多跳的邻居并据此将异质图同质化再聚合邻居节点,以实现聚合元路径上节点的信息;HAN 同时提出分层注意力机制,用于衡量不同邻居的权重,以及不同语义(元路径)信息的权重,其模型示意如图 6 所示。HGT^[83]考虑到以往方法人工定义元路径的局限性,提出了 Meta-Relation 的概念,并结合注意力机制建模异质节点之间的关系,提出了相对时间编码用于建模图中的动态信息,且设计了异质图采样机制使模型能够应用于大规模异质图中。除以上工作之外,还有许多工作针对于异质信息的挖掘提出了模型的改进,HetGNN^[84]引入带重启的随机游

走机制为每个节点采样异质邻居用于信息聚合; Het-GAT^[85] 利用异质图采样过程中,元路径上节点的信息优化节点的表示;KG-GCN^[86] 在考虑边的建模时,利用了与该边相关的实体信息进行聚合,更新边的表示.以上异质图神经网络学习到节点的表示在诸如节点分类、链接预测的部分下游任务中表现优于基于随机游走的表示学习方法,证明了基于图神经网络的异质图表示学习方法在社交知识图谱上应用的潜力.

知识图谱是符号(Symbolic)信息的离散表示形式^[87],大量的符号信息体现着异质节点之间的内在逻辑关系.目前,有研究工作将符号信息引入表示学习(Neural-Symbolic)用于增强知识表示能力^[88-90].其中,KALE^[88]在Trans系列方法对实体和关系表示进行约束的基础上,利用预定义的知识之间的逻辑规则将多个三元组拼接进行打分,实现二次约束,用于优化实体和关系的表示.在此基础上,RUGE^[89]将二次约束方式优化为依据逻辑规则迭代式插入未标注的三元组并判别正确与否的方式,实现更好地模型性能.然而,KALE和RUGE都依赖于预定义的逻辑规则,鉴于此,工作^[90]设计了知识图谱表示学习框架IterE,能够迭代式地发现新规则,并将新规则用于增强实体和关系的表示.上述符号-神经结合的方法聚焦于单一数据源,且缺少对于图谱动态演化的建模,难以建模社交知识图谱数据源异质且动态的情形,限制了其表征能力.

上述表示学习工作适用于社交知识图谱的节点异质性挖掘,而数据源异质层次的挖掘同样十分重要.其中,代表性工作UniSKGRep^[91]聚焦于将社交网络和知识图谱进行统一表示,并通过实验验证了多数据源统一表示模型相较于多个单一数据源表示模型的性能优越性.鉴于社交网络和知识图谱的结构、内容差异,以及动态、互演化的特性,亟需针对性地设计社交知识图谱的统一表征方法.

综上所述,通过本文从节点异质和数据源异质两个层次对面向社交知识图谱异质性的表示学习相关研究工作的梳理,可以发现,从节点异质层次来讲,已有的方法聚焦于在广义的异质图中,缺乏对于社交知识图谱这种以人为中心的异质图建模方法,且考虑到异质信息之间的符号逻辑能够为可解释性带来的增益,将符号-神经结合的表示学习方法研究有限但潜力巨大;从数据源异质层次来讲,融合多源异质数据的社交知识图谱缺乏统一表示方法,继而会影响社交网络和知识图谱信息互补的效果,因此

需要在后续的研究工作中立足于社交网络和知识图谱的性质,针对性地设计表示模型.

5.2 面向社交知识图谱动态性的表示学习

动态图表示学习作为社交知识图谱中动态演化数据挖掘的重要手段,相关的研究受到了广泛关注^[25-26].动态图的建模方式主要分为两大类:离散时态动态图(Discrete Temporal Dynamic Graph, DTDG, 简称为离散动态图)与连续时态动态图(Continuous Temporal Dynamic Graph, CTDG, 简称为连续动态图),如图7所示.

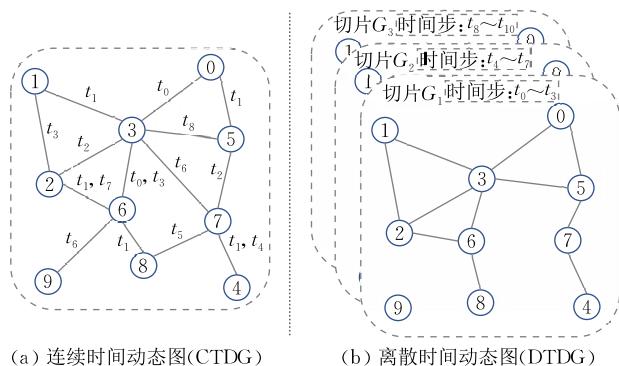


图7 离散时间动态图与连续时间动态图示意

离散动态图表示为 T 个离散的时间点的静态图切片集合 $\{G_0, G_1, \dots, G_{T-1}\}$, 即将动态图划分为多个静态图切片序列, 其中每个切片包含多个时间步 t . 通过离散时间动态图的建模, 能够将原本复杂的动态图转化为多个静态图的方式, 继而利用静态图的方法结合序列建模的方法进行分析. 离散动态图的表示学习的核心思路是, 通过设计编码器对于每一个时间切片的动态图进行表示, 再基于序列模型挖掘静态图表示的序列, 学习节点在不同时刻的表示并研究其演化规律. 工作^[92]针对于离散时间动态图中的四种演化行为(增加/删除节点、增加/删除边)设计了动态图上的随机游走算法. DynGEM^[93]在节点表示过程中同时保留结构信息以及动态信息, 利用深度自编码器建模历史信息(前 $t-1$ 个时间步)的演化规律, 用于得到当前时间步 t 时节点的表示. dyngraph2vec^[94]提出了一种能够捕捉动态图演化的动力学特征, 生成动态图表示的方法, 其采用自编码器、循环神经网络以及自编码器结合循环神经网络三种方式, 依据动态图的前 t 个时间步的快照, 预测 $t+1$ 时刻的图中节点的嵌入式表达, 应用于诸如链接预测、属性预测等下游任务. DynamicTriad^[95]关注到动态图中三元闭合的过程, 即当前时刻两个节点并未关联, 但都存在一个共同关联的节点, 通过

建立模型学习节点的表示并且度量节点之间的相似性,用于预测下一时刻是否能够建立关联实现三元闭合.为关注到时间序列中重要的时间节点,DANE^[96]将注意力机制引入动态图表示学习,DySAT^[97]和TemporalGAT^[98]将图神经网络 GAT^[80]迁移到动态图中,设计动态图神经网络的注意力机制,从图结构的角度和时间序列的角度计算注意力权重,用于更好地学习节点的表示. EvolveGCN^[99]同样是使用图神经网络学习每个时间步节点的表示,但不同的是,以往工作用 RNN 编码各个时间步节点表示的演化,而 EvolveGCN 通过使用 RNN 去编码图神经网络参数演化,以此来捕获图的动态演化.离散时间动态图的表示学习方法建模方式简单直接,但时间快照序列的动态图表示方式需要对于每一个时间段内的信息进行压缩,继而损失了时间段内的动态交互信息,且对于变长时间步或时间跨度较长的动态图建模缺乏灵活性.且难以对社交知识图谱中多粒度、多类型和时序的信息准确建模.

连续动态图将不同时刻节点之间的交互建模在统一的一张图中,其动态演化的过程能够通过一个事件序列 $L = \{l_0, l_1, l_2, \dots, l_n\}$ 来表示. 其中 $l_i \in L$ 代表一次事件,表现为四元组的形式^[26]: $l_i = \{u_i, v_i, t_i, \eta_i\}$, 其中 u_i, v_i 代表发生事件的两个节点, t_i 代表发生事件的时间, η_i 代表事件类型例如增加边、删除边. 连续时间动态图完整保留了节点之间的动态交互信息,通过设定时间间隔进行快照,能够将连续时间动态图转化为离散时间动态图,但由于丢失了具体的时间信息,无法通过离散时间动态图还原成连续时间动态图. CTDNE^[100]将静态图表示学习中广泛使用的随机游走机制迁移到连续时间动态图上,设计了带时间约束的随机游走机制,即在一条随机游走路径上,从起始节点到终止节点要尽量满足连边的时间信息依次递增,在得到采样序列后通过 Skip-Gram 机制学习节点的表示. TDIG^[101]与 MMDNE^[102]则通过时序点过程的方式对于连续时间动态图进行建模,并从宏观(Macro)和微观(Micro)两个角度建模图的动态演化规律. TagGen^[103]从图生成的角度对于动态图的演化进行建模. JODIE^[104]聚焦于通过表示学习的方式建模用户-物品二部图之间的动态演化,继而服务于推荐系统等应用,除此之外,考虑到商品推荐场景动态图的规模巨大, JODIE 提出了 t -Batch 算法,依据时间一致性(time-constant)来对训练的数据进行划分,且能够实现动态图表示学习并行计算,将训练效率提升了九倍.

TGN^[105]同样关注于动态图表示学习效率,提出了一套通用的连续时间动态图表示框架,并且配套提出了能够并行加速训练的算法. 工作^[106]考虑到动态图中日益增长的节点需求难以用直推式(Transductive)的方式进行有效学习,若出现新的节点,现有的模型无法很好地学习,因此提出了归纳式(Inductive)的连续时间动态图表示学习算法 TGAT. 为建模连续时间动态图中多个节点的复杂演化,工作^[107-110]将 Motif 的概念引入动态图表示学习中,并且通过匿名路径游走(Anonymous Walk)^[111]机制来对于动态图中的 Motif 实现近似采样,以此来研究动态图的演化规律. 连续时间动态图的表示学习方式由于保留了完整的动态信息,且事件序列的方式能够建模不同类型的时间信息,相较于离散动态图的表示方式,能够更好地适用于建模社交知识图谱的动态性.

上述工作皆聚焦于社交知识图谱的单一特性进行表示学习,近期也有工作^[112-115]同时关注到图的动态信息和异质信息. Know-Evolve^[116]与 EKG^[117],提出了能够在动态知识图谱上学习实体表示随时间动态演化的框架. TDLF^[112]提出了时间差路径的概念,将异质关系的时间信息融入到网络上的关系路径中,将时间信息和结构信息整合用于随机游走的采样,基于采样的路径序列训练回归模型用于做时序关系预测. DHNE^[113]通过构建历史-当前(Historical-Current)图将中心节点的历史邻居信息与当前邻居信息进行拼接,并在此基础上进行随机游走采样学习节点的表示. DyHATR^[114]利用异质图神经网络学习离散时间动态图节点的表示,并且利用循环神经网络建模节点表示的演化. DyHNE^[115]设计了动态异质图的增量式更新方法,将图的演化转化成特征值和特征向量的变化.

综上所述,围绕社交知识图谱动态性,通过对动态图表示学习相关工作从离散动态图和连续动态图两个角度进行梳理,可以发现社交知识图谱的动态性研究仍充满挑战. 如何处理真实数据中存在缺失或表述不一致的时态信息、如何设计统一的表示框架建模瞬态交互与稳态关系及其间的关联、如何从时序信息中挖掘知识的演化规律都是该领域方向亟待解决的问题.

5.3 面向社交知识图谱情感性的表示学习

社交知识图谱的引入能够挖掘人物之间关系与交互背后潜在的情感信息. 网络中的情感信息通过符号图的方式建模,具体地,符号图中的边带有正或负两种对立的符号属性,正边可以表示朋友、赞同、

喜欢等积极的关系或交互,负边可以表示敌人、反对、厌恶等消极的关系或交互^[28].通过边的符号属性,符号图可以表示人与人、人与物之间的情感好恶,包含更加丰富的情感信息.

符号图的相关模型主要使用两个社会学理论:结构平衡理论和社会地位理论^[118-119].结构平衡理论基于几个基本的直觉,即“朋友的朋友是我的朋友”、“朋友的敌人是我的敌人”.这个理论通常用于无向符号图^[120].工作^[121]对平衡理论进行了扩展,符号图的结构应该使用户与朋友更相似,与敌人更有区分度.社会地位理论一般用于有向符号图,图中每个人都有相应的社会地位,人会向社会地位更高的人产生正边,向社会地位低的人产生负边^[122].

符号图表示学习是挖掘社交知识图谱中情感信息的主流方法,其将社会学理论一般化,引入模型结构和目标函数中,在建模思路上借鉴了无符号图表示学习工作,并在信息聚合等过程中对正负边进行了区分和融合. SiNE^[123]使用查询表编码节点向量表示,依据扩展后的平衡理论设计目标函数,目标函数采用负采样的间隔损失函数,将正样本替换为与朋友邻居的相似度,将负样本替换为与敌人邻居的相似度. SNE^[124]借鉴 Skip-gram 模型,在已知路径前置节点的情况下,使用对数双线性模型预测目标节点,用于学习节点的向量表示. SIDE^[125]将随机游走算法引入有向符号图中,将图结构转化为序列结构,通过节点在随机游走中的共现频率学习符号图中的邻域信息.以上三个模型中, SiNE 和 SIDE 根据结构平衡理论的一般化假设设计目标函数,考虑了符号图的社会学特性,但是没有区分正负边,丢失了符号图中边的信息. SNE 模型区分了正负边,但只是简单地把边的符号看作两种不同类型的关系,丢失了符号图的社会学特性.

SGCN^[126]将图神经网络引入符号图研究中,并且根据结构平衡理论,定义了符号图中的聚合操作和目标函数.具体地,SGCN 模型定义了平衡集和非平衡集,并使用两个独立的聚合器,分别聚合目标节点平衡集中的节点信息和非平衡集中的节点信息.在目标函数中,模型约束正节点对间的距离小于无连接节点对,无连接节点对的距离小于负节点对,以此实现节点表示的优化.在 SGCN 基础上, SiGAT^[127]将图注意力机制引入符号图用于度量邻居节点的重要性,其根据连边的符号和方向定义了 38 种三角模体 (Motif),并对不同模体使用独立的聚合器进行邻居节点的信息聚合.而预定义模体的方式会使模型灵

活性下降.如果数据中包含中立连边或更细粒度符号分类,就需要重新定义模体.

综上所述,本节聚焦社交知识图谱的情感性挖掘,以符号图为主对于相关表示学习工作进行梳理.可以发现,目前大部分的工作根据符号图的社会学假设,对无符号图模型进行了改进,但是这些模型大都只考虑正负情感,没有考虑社交知识图谱中无情感偏向的交互和客观的关系,往往会有数据稀疏的问题.另外,现有工作往往将情感分为正负极性,缺少更加细粒度的情感表示.情感也具有动态性,人在不同阶段会有不同的情感倾向,情感也会受到交互等因素的影响而发生改变,现有的工作无法建模情感的演化和溯源.而社交知识图谱的异质性和动态性可以有效地建模情感,可以通过人物间的关系和交互信息辅助情感信息的推理,解决数据稀疏问题.另外可以根据用户间交互类型不同,推理更加细粒度情感,例如喜欢和热爱、厌恶和痛恨等.同时可以根据社交知识图谱中用户动态交互的内容和频率变化,推理用户间情感演化,并为情感的变化提供依据.

5.4 面向社交知识图谱互演化性的表示学习

社交知识图谱的互演化性与其异质性、动态性和情感性紧密关联,现有关于互演化性的研究有限.工作^[29,128]关注到了长期的社交关系和瞬时的交互信息两种不同的动态信息,其中 DyREP^[29]提出了一套基于图神经网络的连续时间动态图表示学习框架,该框架能很好地建模网络的动态演化特征.对于动态图结构中节点的交互行为,作者将其分为关系 (Association) 与交互 (Communication) 两种,前者代表长期稳定的关系,后者代表短暂、临时的交互.在此基础上, DyREP 提出条件强度的概念来建模节点之间动态的相互影响程度,以及用邻接矩阵的概念来建模节点之间的关联,并设计动态更新机制建模两者之间互相作用.在节点的信息传播方面, DyREP 将节点的信息传播分为为局部信息传播 (Localized Embedding Propagation)、自传播 (Self-Propagation) 和外因驱动的信息传播 (Exogenous Drive),从三个不同地角度建模节点的演化; MRINF^[128]分别从宏观结构和微观交互两个角度对于人物的行为进行建模,并且引入图注意力神经网络学习人物的嵌入式表示.

综上所述,可以发现围绕社交知识图谱互演化性的研究工作相对有限,上述研究对于互演化性的建模方式仍较为简单,未考虑交互信息与长期稳定社交关系之间的内在联系、不同时间细粒度的动态

信息,以及未考虑到交互中所蕴含的大量内容信息,且上述方法建模的对象是仅包括人员节点的社交网络,而在社交知识图谱中,稳态关系不仅存在于人员之间,也存在于人员与机构、组织、地点等其他类型实体间,考虑到社交知识图谱的异质性,如何设计知识表示算法适用于社交知识图谱中的异质节点,以及建模异质节点之间动态的相互作用也是需要考虑的问题。

6 基于社交知识图谱的知识推理

在实现社交知识图谱构建的基础上,由于数据源的局限性以及隐含知识的存在,构建完毕的社交知识图谱中仍然存在大量缺失的知识,为服务于实际以人为中心的应用,需要对社交知识图谱中缺失的属性或关系,以及对难以直接在社交网络或知识图谱推出的深层次属性或关系进行推理补全。不同于上一节介绍的通过表示学习的方式寻找下游任务的通用解决方案,本节所述的知识推理方法则是立足于传统方法或表示学习方法,聚焦于知识推理这一任务并进行模型方法专门的设计与优化。

知识推理方法内部也存在丰富的类型区分,根据目标的类型不同,可以划分为特定人、特定模式集合和特定群体推理;根据应用而又可以划分为人物节点分类、链接预测、关系分类和关系补全。有关社交网络^[129]或知识图谱^[130]推理的研究已有详尽的梳理工作。不同于此,本文从社交知识图谱的研究目标出发,聚焦于以人为中心的知识推理,探讨社交知识图谱在融合了社交网络与知识图谱信息之后,如何充分利用社交知识图谱的特点,在提升知识推理性能的同时,对于时态知识推理、可解释知识推理等实际需求上,发挥社交知识图谱推理的优越性。从社交知识图谱的定义出发,根据前文所述社交知识图谱中信息的形式,本节将从结构信息、时态信息和内容信息三个角度展开进行相关工作的梳理,并且就社交知识图谱在动态可解释的知识推理上的优越性进行分析。

6.1 基于社交知识图谱结构信息的知识推理

社交知识图谱在融合了社交网络与知识图谱之后,既包含了人与其他异质类型节点的拓扑结构信息,也包含了人与人之间的拓扑结构信息。本小节关注于如何有效利用以上结构信息进行社交知识图谱中以人为中心的知识推理。

结合前文对于社交知识图谱异质特性的分析,

能够发现社交知识图谱中包含大量源自于知识图谱的与人相关的异质节点信息(如学校、地点、公司和专业技能等),这些节点描述了人物相关的背景信息,能够用于推理人物的属性以及人物之间的关系。工作^[131]是利用异质信息进行知识推理的开创性工作,该工作将元路径(Meta-path)的概念引入包含异质信息的网络中,并且提出了设计元路径并通过图搜索以及模式匹配的方式进行学术合作关系推理。在此基础上,工作^[131]将基于元路径的推理从学术社交网络扩展到通用的异质信息网络上进行人物关系推理(如元路径“人-学校-人”能够用于推理同学关系,“人-公司-人”能够推理同事关系);工作^[132]则提出将元路径信息用于人物的属性推理。以上三篇工作能够充分利用异质节点信息进行知识推理,但基于元路径的模式匹配需要消耗大量的时间,难以保证效率,且在数据质量存在噪音或缺失的情况下推理效果难以保障。基于异质图神经网络的知识推理工作^[133-134]将元路径的思想引入图神经网络的框架中,相较于传统基于元路径的图搜索方法无论是性能还是效率都有提升。除元路径外,有工作GraIL^[135]关注到异质节点之间更为复杂的关联,其基于头、尾实体进行子图采样并求交集,在此基础上通过基于注意力机制的多关系图神经网络实现从图的层面进行归纳式的知识推理。

人与人之间关系或交互的拓扑结构信息能够反应人物的社会地位或人与人之间的关系,能够用于以人为中心的知识推理。其中,ConPI^[136]利用人物之间过往的交互行为作为背景信息进行人物之间的链接预测以及关系语义挖掘;工作^[137]则考虑到不同来源的社交网络交互行为数据能够互相作为背景信息,并且援引三个经典社会学理论:社交平衡(Social Balance)、社会地位(Social Status)和社交结构洞(Social Hole),用于从社会学的角度建模社交关系,并且在此基础上设计模型,实现跨网络的社交关系推理;SHINE^[138]通过符号图的方式将人物之间的情感背景信息引入知识推理中,具体地,SHINE使用三个独立的自编码器分别提取符号图、关注社交网络以及人物属性网络的人物向量,然后将三个向量拼接为人物的嵌入向量,使用向量内积进行社交关系推理。工作^[139-141]关注到不同类型的网络拓扑结构蕴含着不同的社交关系信息(例如层级结构反映上下级公司关系,全连接结构反映家庭关系),如图8所示。其中,GraphSTONE^[140]通过匿名路径机制以序列的方式建模中心人物节点周围的子图结

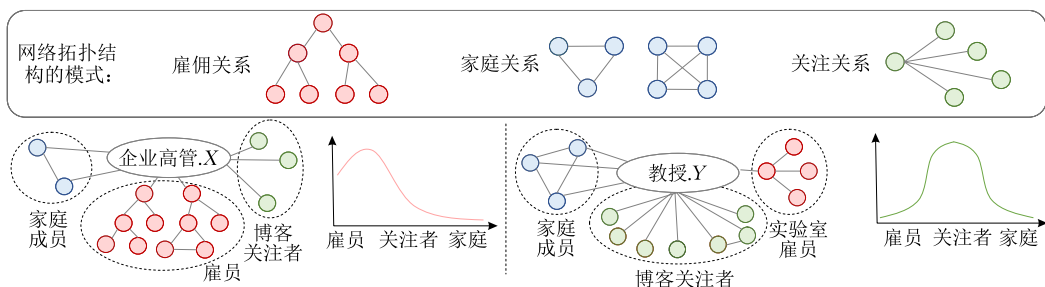


图 8 基于结构信息的社交关系推理示意

构,并通过主题聚类的方式挖掘序列背后所反映的社交关系;MHCN^[141]则利用超图建模社交网络中结构信息所反映的社交关系。

本小节从人与人、人与其他节点交互的拓扑结构两个角度对基于社交知识图谱结构信息的推理工作进行了梳理。然而,根据前文对于社交知识图谱互演化性的分析,人与人、人与其他节点之间交互之间存在潜在的关联,因此如何建模这种潜在的关联值得在后续的社交知识图谱研究工作中探索。

6.2 基于社交知识图谱时态信息的知识推理

社交知识图谱中融合了来自社交网络大量强时态性、动态变化的社交知识,相较更新缓慢、不考虑时间信息的传统知识图谱数据,能够有效建模人员之间关系随时间的演化轨迹。结合社交知识图谱定义和前文 3.2 节所述,在统一的时态表述体系下,本文按时态信息类型将图谱中的时态知识分成两种:瞬态知识与稳态知识^[26],如图 9(a)、(b)所示。本小节首先从瞬态和稳态两个角度展开介绍社交知识图谱的时态知识推理;考虑到时态知识间序列关系对研究知识演化的作用,在瞬态和稳态知识推理基础上提出时序关系推理的概念,如图 9(c)所示。

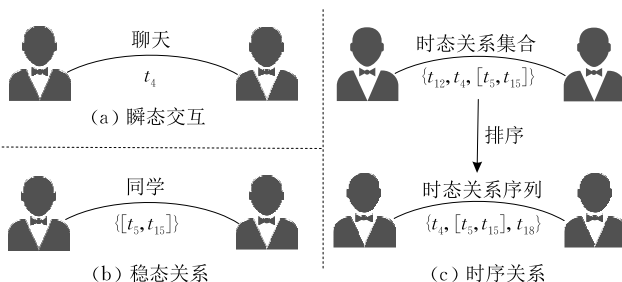


图 9 时态关系示意

瞬态知识的概念由文献^[26]提出,瞬态知识的延续性在时间表述体系下短暂到可被忽略不计,可被视为无延续性(如上述文献中提到的邮件通讯关系);或考虑其延续性无意义(如文献中提到的论文引用关系)。社交知识图谱中的瞬态知识表示为 (u, r, v, t) 的形式,代表 t 时刻 u 节点与 v 节点发生了一次类

别为 r 的交互。在研究瞬态知识时,仅考虑其发生时间,并不考虑结束时间,这是由于瞬态知识的持续时间极短,可以近似地将发生时间视作结束时间。瞬态知识推理的任务被定义为给定四元组中 (u, r, v, t) 中任意三元的情况下,预测未给定的元素。按照实现方法区分,瞬态知识推理可以分为基于矩阵分解的方法、基于转移机制的方法以及基于深度学习的方法三种。

基于矩阵分解的瞬态知识推理方法应用于表示为多个静态图快照序列的离散动态图中。在仅考虑单一关系类型的前提下,每一张静态图快照可以视为一个二维矩阵,基于矩阵分解的时态知识关系推理方法将时间建模为矩阵的第三个维度,并在三维矩阵上进行推理。其中,工作^[142]首先提出了使用离散动态图建模演化的社交网络数据的方法,工作^[143]通过三维矩阵建模动态社交网络,在此基础上使用成分矩阵分解方式实现社交交互的推理。考虑到在时态知识图谱中关系的丰富类型,工作^[144]引入矩阵的第四维用于代表不同的关系类型,并通过四阶张量分解的方法推理不同类型关系在不同时刻存在与否。基于矩阵分解的方法时间复杂度高,且难以推理社交知识图谱中新加入的知识。

基于转移机制的瞬态知识推理工作借鉴了静态知识图谱中 Trans 系列方法的思想,并将转移模型扩展至时序知识图谱中,其思路是将实体、关系与时间投影至超平面中,并通过打分函数建立三者之间的联系。由于时序知识图谱被表示为离散动态图的形式,因此基于 Trans 模型的瞬态知识推理工作仅能用于离散动态图中。HyTE^[145]将动态图中的每一个时间点 t 看成隐平面 w_t ,通过将实体和关系类型投影到每一个隐平面上,从而得到实体和关系类型在不同时刻的表示。TTransE^[146]在静态知识图谱的 Trans 模型基础上赋予每一个时间点一个表示,将时间表示作为系数影响关系表示。上述两种方法只能应用于离散动态图中,而基于深度学习方法实现

的瞬态知识推理方法既可应用于离散动态图,也可应用于连续动态图中. 在离散动态图中,TA-TransE^[147]提出时间词序列的概念,例如“出生于”的发生时间为1985年,其可以看成词序列(*born_in*,1y,9y,8y,5y),*born_in*为序列中的事件类型词,剩余的看成时间词,并在此基础上使用循环神经网络建模时间词序列,并用最后一个时刻的循环神经网络的输出去替代静态知识图谱表示模型的知识表示. 工作^[148-150]使用了循环神经网络建模观测到的静态图快照序列,以此挖掘实体在不同时刻的演化规律,训练不同时刻的实体表示进行知识推理. T-GAP^[150]建模了时间差信息,并且提出了一种基于注意力传播的时序消息传播算法.

在连续动态图中,Know-Evolve^[116]首次将时序点过程与深度学习结合,提出了深度时序点过程概念,并将其用于瞬态知识推理. 其将时态知识视作事件,在给定时间的情况下,通过实体在时刻的表示建模条件强度函数,从而推理关系的发生概率. 其他基于时序点过程的工作^[151-152]使用Hawkes过程或其他动态点过程方法结合节点在特定时刻的表示,建模条件强度函数. 基于时序点过程的方法很好地解决了时间插值问题,能够预测任意时间点的交互. 然而,其仅在交互发生时才更新节点表示,会存在冷启动问题;此外,条件密度函数计算复杂,在大规模数据场景中计算效率低.

瞬态交互在真实社交网络中更新频繁,而现有的瞬态知识推理模型在推理增量式、实时更新的瞬态知识时仍存在计算复杂度高,训练效率低下的问题,难以适配大规模社交知识图谱的应用需求.

稳态知识是时态知识的另一种形式,其具有可被研究的延续性^[26]. 稳态知识中的关系,如在读关系、就职关系、亲属关系等. 社交知识图谱中的稳态知识表示为 $(u, r, v, [t_{\text{start}}, t_{\text{end}}])$ 的形式,代表 t_{start} 到 t_{end} 时间段内 u 节点与 v 节点存在类别为 r 的关系. 在推理稳态知识时,需要考虑到稳态知识的发生时间与结束时间(或持续时长). 本文定义稳态知识推理的任务为:给定稳态知识 $(u, r, v, [t_{\text{start}}, t_{\text{end}}])$ 其中四元的情况下,预测另一缺失的元素. 传统的动态知识图谱(Temporal knowledge graph)或事件图谱(Event graph)使用单一时间戳表示知识中的动态信息,这种建模方式无法表示稳态知识中的时间区间信息. 然而,许多时态知识中的时间信息并不能简单地被单一时间戳描述,如(奥巴马,是…总统,美国)这条稳态知识的时间信息为[2009,2017]这个区

间,意味着这条知识在该时间区间内有效,而传统的时态知识图谱不能描述这种能够在一定时间区间内有效的稳态知识. TIMEPLEX^[153]首次聚焦于稳态知识推理任务,并将稳态知识推理任务分为链接预测和时间预测两步. 在此基础上,该工作提出了一种基于四阶张量分解表示的方法,将实体、关系以及动态信息表示为向量,考虑四元组共现概率、稳态知识的周期性以及稳态知识之间的时序依赖性这三项,计算四元组的打分函数. TIMEPLEX考虑到了动态关系的有效时间区间,首次聚焦于稳态知识推理任务;然而TIMEPLEX需要为每个时间点训练表示,在社交知识图谱中时间点数量多、粒度细的情况下,学习时间点的表示训练代价大,且不能为训练集中未出现的时间点生成表示,泛化推理能力有限. TComplex^[154]和TeLM^[155]同样注意到了动态知识图谱中稳态知识的存在. 其中,TComplex提出了基于复数域的四阶张量分解法. 不同于其他使用实数向量或复数向量表示实体的方法,TeLM使用线性动态约束器用于使相邻时间的表示更加接近,并使用多重向量(multivector)嵌入式表示的方式用于建模不同知识在不同时间步的表示. 从实验效果上看,TeLM效果超过了使用实数或复数向量表示的稳态知识推理工作;然而,与TIMEPLEX的方式相同,TComplex和TeLM仍基于张量分解方法实现,需要对每个时间点计算时间表示,训练效率低,且缺少对未出现在训练集中时间点的泛化能力.

综上所述,通过从瞬态知识推理和稳态知识推理两个角度对社交知识图谱时态知识推理相关工作进行梳理,可以发现该研究方向仍面临着巨大挑战. 其一,构建社交知识图谱的数据来源于多通道的社交网络与知识图谱数据,而这些数据中包含的时态信息并不保证是完整的,也不能保证时态信息的表述方式是完全统一的,如“1960年4月16日”,“今天下午3点”分别表达了两种不同粒度、不同表述形式的时态信息. 其二,社交知识图谱中的两种时态知识并不互斥,两个实体在具有稳态知识的同时也可以具有瞬态知识或具有其他稳态知识,即知识间的时序关系并非仅有先序关系与后序关系,还存在时序包含关系. 这里本文定义时序知识推理的任务为:给定节点之间的时态知识集合,推理出时态关系集合中关系的时序关系序列,包括前后关系与包含关系. 目前对于多尺度动态关系中的时序关系推理研究较少,现有工作中,DyREP^[29]关注到了不同尺度的时态知识,但仍未考虑时序信息以及建模知识的

演化. 在统一时态表述体系的基础上, 依据给定的时态关系集, 梳理时态关系之间存在的时序脉络, 并推理不同类型时态关系组成的时序链, 是分析社交知识图谱关系时序演化的关键. 推理出的时序链可以为以人为中心的应用中分析用户行为演化提供依据, 具有丰富的应用价值.

6.3 基于社交知识图谱内容信息的知识推理

社交知识图谱中的内容信息包括人物本身背景描述, 以及人物之间的交互所包含的丰富信息, 并以文本、图像和视频等丰富的形式呈现, 这些内容信息与个人的属性偏好以及人与人之间的关系存在潜在的联系, 因此内容信息对于社交知识图谱中的知识推理也十分重要. 本小节从文本内容和多模态内容两个角度入手进行相关工作的梳理与分析.

社交知识图谱中文本内容包括人物本身的背景描述(如个人简历信息), 以及用户交互(如对话)中的文本信息. 对于人物本身的背景描述信息, 工作 CANE^[156] 通过深度学习的方式学习社交网络中个人背景信息(如社交主页的自我介绍)的嵌入式表示, 并用于增强节点的嵌入式表示继而服务于知识推理, 后续的知识推理研究工作中将该方式作为包含背景描述信息的图数据中用户节点初始化的通用范式. 对于对话中的文本信息, 工作^[37-39] 关注到对话中蕴含的社交关系信息, 设计抽取式的方法推理对话中所提到的人物之间的社交关系, 并在多轮对话数据集上取得了良好地表现, 证明了利用交互内容信息推理社交关系的可行性, 但以上工作缺乏对于隐含以及深层次社交关系的挖掘.

工作^[157-161] 致力于推理交互内容中所隐含的社交关系, 其通过交互内容信息进行人物之间社交关系的表示学习并用于推理社交关系的深层次语义信息. 其中 TransConv^[157] 首先提出了对话相似度指数, 用于建模用户之间的互相交流是否相似, 再通过交流频率指数用于建模用户之间经常对话的主题的权重, 并结合 Trans 模型以及通过不同的关系映射到不同的超平面用于建模用户节点和社交关系的嵌入式表示, 继而服务于知识推理任务; Topic-GCN^[158] 关注到对话中隐含着多种话题, 且由于社交关系不同, 话题的占比不同, 因此这篇工作通过多头注意力机制建模对话中话题的分布, 并且使用变分自编码器学习边的表示; MERL^[159] 关注到人物之间的关系不对称性; BaKGraSTeC^[160] 通过引入知识图谱信息作为背景知识辅助知识推理; RELEARN^[161] 同时利用社交结构信息、交互内容信息以及用户本身的属

性信息, 并且考虑到交互信息的不完整性以及噪音, 设计了基于图神经网络和变分自编码器的图表示学习框架, 用于推理人物之间的关系语义信息.

随着在线社交网络平台的发展, 用户交互的方式也不仅仅局限于文本的形式, 包括视频、语音、Emoji 表情等多样化的用户交互行为愈发常见, 因此对于以上多模态的交互内容进行深入挖掘对于社交知识图谱的知识推理至关重要. 例如图像和视频中出现的人与物等都能够反映相关人物潜在的社交关系, 如图 10 所示.

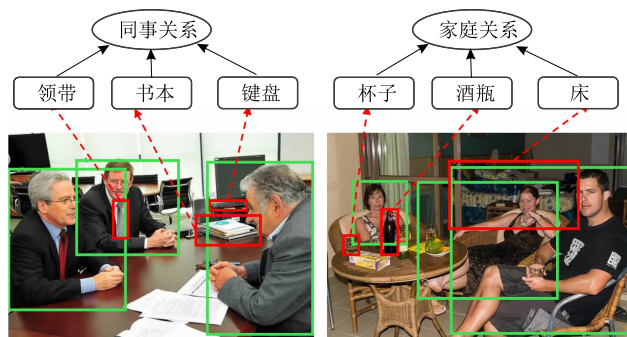


图 10 基于图像内容的社交关系理解

相关的研究工作也在近几年涌现, 综述^[162] 全面阐述了近年来利用图像和视频内容进行社交关系识别的研究进展. 工作^[163-165] 利用通用知识图谱中的背景信息辅助进行图像中的人物社交关系识别, 其中 GRM^[163] 首先将知识图谱与目标检测结合, 通过物品与社交关系在知识图谱中的联系建模背景知识, 并且设计门控图神经网络用于推理社交关系; HGAT^[164] 同时关注到物品以及图像中人物之间的交互与社交关系的潜在关联, 并设计分层图注意力神经网络进行社交关系推理; SPTS^[165] 关注到图像中群体关系的识别. 工作^[166-167] 利用视频内容进行社交关系推断, 其中 MSRT^[166] 通过将视频按帧进行分解, 并将视频中的人物交互建模成离散的图切片序列, 在此基础上通过图神经网络分析视频中人物对的社交关系; HC-GCN^[167] 聚焦于视频内容中出现的多个人物而非人物对, 并通过生成社交图的方式推理多个人物之间存在的社交关系. 社交知识图谱中源自于在线社交平台的用户交互信息包含大量的表情包(Emoji), 难以通过传统的文本处理方式建模, 考虑到这一点, 工作^[168-169] 针对于 Emoji 表情设计了嵌入式表示方法, 且成功应用于情感分类等下游任务之中.

本节从结构、时态和内容三个角度梳理了基于社交知识图谱的推理工作. 综上所述, 社交知识图谱

能够继承和丰富源自于社交网络和知识图谱的信息,在深度融合后能够实现信息互补,有效提升知识推理的性能,并且能够建模社交知识图谱中知识的动态演化。同时,随着包括知识推理在内的深度学习技术发展深入,对于算法可解释性的需求也日益提升,而社交知识图谱作为一种新的数据结构将社交网络与知识图谱进行了有机地结合,也将对社交网络与知识图谱上以知识推理为代表的典型以人为中心应用赋予新的内涵,对于推理的可解释性等需求提供了新的解决思路。具体来讲,社交知识图谱的结构信息、内容信息时序信息都能够为知识推理提供可解释的依据。从结构信息的角度举例,现实世界中两人都和某学校存在联系,则两人可能是师生或同学关系;从内容信息的角度举例,如果现实世界中两个人聊天谈及或照片出现的大量内容是家具或柴米油盐,则两人更有可能是家庭关系,如果内容是实验仪器、论文和研究,则更有可能是师生或同学关系;从时态信息的角度举例,时态信息蕴含着知识的因果联系以及演化规律,例如现实世界中两人的关系需先成为情侣才能再发展成夫妻。因此,基于社交知识图谱的可解释知识推理无疑有巨大的潜力与探索空间。然而,现有基于社交知识图谱知识推理研究工作对于社交知识图谱的利用还较为有限,缺乏对其潜力的深入挖掘,需要在后续的研究中进行深入的探索。

7 社交知识图谱应用

社交知识图谱能够为以人为中心的丰富应用提供有力支撑,本文将从社交电商、科技情报、产业投资和项目协作四个应用场景为例对于社交知识图谱的应用进行阐述。

7.1 社交电商

在互联网时代,社交电商已经成为主流购物方式之一,通过人与人之间的关系或交互行为进行商品的推荐是社交电商的核心理念。社交知识图谱在建模用户之间社交信息的同时,能够通过知识图谱信息建模用户-用户、用户-物品以及物品-物品之间深层次的联系,以及商品的背景信息,用于分析用户的兴趣演化,继而服务于商品的智能化推荐。围绕社交知识图谱的社交电商应用在学术界和工业界都受到广泛地关注。工业界中,美团的美团大脑和阿里的 AliCoCo^[170]都是社交知识图谱的典型产业落地,通

过社交信息与知识图谱的结合赋予了社交电商“智慧大脑”,服务于企业的同时,也为用户提供更为智能与便利的服务。近年来,聚焦于社交知识图谱在社交电商中应用的相关研究工作涌现^[171],其中的研究例如工作^[59,172]聚焦于利用知识图谱中的背景知识分析用户的交互信息背后所体现的用户兴趣,并用于商品推荐。

7.2 科技情报

高新科技的快速发展在带来社会进步的同时,也会增加技术的认知成本,如何从大量的信息中获取所需的科技情报,继而服务于技术攻坚、人才挖掘以及高新技术投资等任务是亟待解决的问题。社交知识图谱的引入无疑能为上述问题提供行之有效的解决方案,通过构建社交知识图谱建模学者之间的交互信息如论文、专利以及项目的合作交流,以及学者与科研团队、组织机构以及研究领域的关系,能够将大规模的科技情报以图的形式呈现,并在此基础上通过图挖掘的算法实现对于科技情报的深度挖掘,降低了解相关科技信息的认知成本。现有社交知识图谱在科技情报领域的典型应用,如清华大学的 AMiner 系统^[173]和微软的 Microsoft Academic Graph 系统^[174]都已经在科技人才挖掘和技术分析等方面有了广泛地应用,有效地降低了科技情报的认知门槛。

7.3 产业投资

在产业投资领域,信息的价值不言而喻,而传统产业投资的调研工作需要耗费大量的人力物力进行有价值信息的挖掘,且企业价值的变化受到多种因素影响,是典型的复杂系统^[175]。社交知识图谱概念的提出能够为产业投资提供新的范式,其能够通过图的形式建模人员、公司以及投资机构之间动态的交互和关系,通过知识抽取的方式将诸如研报和政策等文本信息转化为结构化知识用于更新社交知识图谱,并且通过知识推理的手段挖掘人、公司和投资机构之间潜在的联系,继而服务于投资推荐、风险预测、欺诈检测等典型的产业投资应用。该领域典型的应用如天眼查在企业智能分析、潜在风险挖掘等已经有了成功实践。近几年涌现了多个相关的明星初创公司,例如国外的 Kensho、国内的通联数据,这些公司都在通过诸如社交知识图谱等人工智能技术深刻地推动着产业投资领域的变革。

7.4 项目协作

如何实现高效的人员协作是大型企业内部项目统筹时需要重点考虑的问题,其中包括如何根据团

队的特点以及成员的能力进行工作的安排等具体需求. 社交知识图谱在该领域也有着广泛的应用前景, 并且许多大型企业已经在实际项目协作中对此进行了探索. 微软的 Microsoft Graph 与领英的 LinkedIn Knowledge Graph^[22] 结合是社交知识图谱在项目协作中的典型案例. 其中微软的 Microsoft Graph 对基于 Office 的企业内部人员交互信息进行了建模, 领英的 LinkedIn Knowledge Graph 则对于人员的背景信息包括掌握技能、职业生涯和毕业院校等进行了建模, 两者的结合能够实现在综合考虑人员的社交关系(例如和哪些人员合作交流较多)以及个人背景信息(例如擅长哪些技能)的项目协作安排, 避免了低效的人力资源调度, 继而显著提升企业的生产力.

综上所述, 社交知识图谱具有丰富的应用领域以及巨大的应用价值, 除此之外, 社交知识图谱在国防安全、疫情防控等领域也有着应用. 但同时也可以观察到, 现有对于社交知识图谱的应用还处于探索阶段, 大部分应用并未显式地提出社交知识图谱, 缺乏统一的概念体系指导, 因此对于社交知识图谱的利用还较为简单. 本文致力于构建统一的社交知识图谱概念体系, 并且对其性质进行深入探索, 继而挖掘社交知识图谱在应用层面的潜力.

8 研究展望

考虑到社交知识图谱的研究还处于早期, 许多问题亟待解决, 因此本节在前文的基础上, 对于未来社交知识图谱可能有价值的研究方向进行了展望.

(1) 融合物联网数据的社交知识图谱

从数据来源的角度来讲, 通过前文社交知识图谱的构建的分析, 可以发现当前研究聚焦于将在线社交网络和通用知识图谱作为数据源进行社交知识图谱的构建. 事实上, 大量的物联网数据(如传感器数据和 GPS 数据)能够反映人的现实世界行为, 对于人物画像、关系推理等任务同样有着重要的意义. 如何将物联网数据与在线社交网络和知识图谱的数据进行虚实对映, 实现多源异质数据的融合, 以及如何有效利用融合后的物联网数据服务于知识推理等任务是这一方面研究的重点与难点.

(2) 社交知识图谱时态信息建模

从社交知识图谱的动态性角度来讲, 社交知识图谱中含有丰富的时态信息, 而时态信息具有多粒

度的、多尺度的特点, 并且存在残缺、模糊的问题.

多粒度是指知识中时态信息的表述不统一. 社交知识图谱数据来自于多样的数据源, 存在多粒度时态信息并存的现象, 如“1999 年”, “1999 年 10 月 27 日”代表了两种粒度的时态信息. 多尺度是指知识演化的时态尺度即知识更新改变的频率不统一. 例如, 政要之间的共同参会、在线社交平台中互动等事件是相对频繁的, 如一位政要可能会在一周内多次与其他国家政要互动; 而政要与机构、组织之间的任职关系一般不会轻易变化. 残缺是指知识中时态信息的不完整性. 由于数据源中时态信息缺失的问题, 以及时间抽取工具的误差, 知识中的时态信息会存在全部残缺或部分残缺的问题. 全部残缺指时态信息全部缺失, 部分残缺指时态信息缺失了一部分, 如“8 月 6 日”并不清楚是哪一年的 8 月 6 日. 模糊是指知识中时态信息的表述未采取标准的日历或计时方式表述的问题. 这是由于许多时态知识抽取自社交短文本与新闻文本, 而这些文本中的时态信息描述并不保证是标准表述所导致的, 如“昨天下午”, “几天前”.

现有研究工作未充分考虑社交知识图谱中时态信息存在的上述问题, 而将知识图谱中时态知识建模为统一粒度, 忽视了多尺度时态知识并存于知识图谱的现象; 更无法处理时态知识残缺、模糊的问题, 这些问题值得在后续研究中进行深入探索.

(3) 社交知识图谱的信息互演化性

结合前文社交知识图谱性质的论述, 可以发现互演化性是社交知识图谱的重要特性之一, 且其与异质性、动态性和情感性的研究紧密关联. 然而, 现有的相关研究工作较少, 因此对于互演化性的挖掘与利用较为有限. 未来可以分别从异质性入手, 探究人与其他类型节点的联系和人与人之间关系或交互之间的影响; 从动态性入手, 探究长期稳定的关系与瞬时的交互之间的相互作用; 从情感性入手, 探究人与人之间关系的情感极性与人之间交互的情感极性之间相互的影响.

(4) 社交知识图谱的知识推理可解释性

前文知识推理部分已经对社交知识图谱在可解释性上的潜力有所论述, 可解释性关乎算法的透明性、公平性以及个人隐私安全, 因此, 随着人工智能的研究进入深水区, 无论是产业界还是学术界对于算法可解释的需求都日益增加. 社交知识图谱概念的提出无疑为可解释性的研究提供了新的可行方

案,具体可以围绕算法的可解释性,利用社交知识图谱分析知识的因果演化规律,从而对人物的属性以及人与人之间的关系变化做出解释,该方向的研究仍处于早期,未来还有深入探索的空间。

(5) 基于社交知识图谱的社交挖掘算法

社交知识图谱的核心是人,并立足于服务以人为中心的应用将社交网络与知识图谱数据进行了深度的融合。传统基于社交网络的社会计算研究经历了漫长的阶段,其无论是理论还是应用体系都较为成熟,而社交知识图谱的提出为传统社交网络的研究提供了新的思路,也将使社会计算等研究领域焕发新的活力,例如在社交知识图谱上结合知识信息的团伙挖掘和意见领袖挖掘。从另一角度来想,社会计算中包含着丰富的社会学理论,对于挖掘特定个体或群体的行为模式都有着重要的意义,未来也应当立足于此探索将社交知识图谱与社会学理论相结合,丰富社交知识图谱的内涵。

(6) 社交知识图谱面向特定领域的应用

第7节对于社交知识图谱的应用场景进行了举例说明,论证了其在应用层面具有巨大的潜力,未来值得探索社交知识图谱在更广阔领域应用的可能。然而,考虑到不同领域特点与需求差异,需要在未来的研究落地中思考针对于特定领域的社交知识图谱是否会有新的特性,如何有效地利用,并且考虑如何面向于特定领域进行优化改良。

9 总 结

社交知识图谱将社交网络信息与通用知识图谱信息进行结合,以人为中心实现两个网络的深度融合与信息互补。其在继承通用知识图谱丰富的知识信息同时,通过社交网络中丰富的时态信息建模图的动态演化规律以及与人相关的丰富情感信息,是对知识图谱概念的扩展与延伸。社交知识图谱的提出也为以人为中心的应用中高性能和可解释等实际需求提供了新的解决范式。本文对社交知识图谱的研究现状进行了总结,首先在梳理相关概念的基础上给出了社交知识图谱的形式化定义,其次对于社交知识图谱的性质进行了深入分析,再次从社交知识图谱的构建、表示学习和推理三个层面对相关研究工作进行了系统性地梳理,然后对于社交知识图谱相关的应用场景进行了举例说明,证明了其广袤的应用前景,最后对于社交知识图谱未来的研究方向进行了展望。本文希望能够通过以上的分析与讨

论,建立起社交知识图谱的统一概念体系,帮助读者快速了解社交知识图谱,并且吸引更多的研究者加入,推动该研究领域的蓬勃发展。

致 谢 在此,我们向对本文的工作给予支持和建议的同行表示衷心的感谢!

参 考 文 献

- [1] Cheng Xue-Qi, Mei Hong, Zhao Wei, et al. Data science and computing intelligence: Concept, paradigm, and opportunities. *Bulletin of Chinese Academy of Sciences*, 2020, 35(12): 1470-1481(in Chinese)
(程学旗, 梅宏, 赵伟等. 数据科学与计算智能: 内涵、范式与机遇. 中国科学院院刊, 2020, 35(12): 1470-1481)
- [2] Lehmann F. Semantic networks. *Computers & Mathematics with Applications* 1992, 23(2-5): 1-50
- [3] AMIT S. Introducing the knowledge graph. Official Blog of Google, <https://www.blog.google/products/search/introducing-knowledge-graph-things-not/2012,5,16>
- [4] Bollacker K, Evans C, Paritosh P, et al. Freebase: A collaboratively created graph database for structuring human knowledge // *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. Vancouver, Canada, 2008: 1247-1250
- [5] Auer S, Bizer C, Kobilarov G, et al. DBpedia: A nucleus for a web of open data // *Proceedings of the 6th International the Semantic Web and 2nd Asian Conference on Asian Semantic Web Conference*. Busan, Korea, 2007: 722-735
- [6] Suchanek F M, Kasneci G, et al. YAGO: A large ontology from Wikipedia and WordNet. *Journal of Web Semantics*, 2008, 6(3): 203-217
- [7] Liu H, Singh P. ConceptNet—A practical commonsense reasoning tool-kit. *BT Technology Journal*, 2004, 22(4): 211-226
- [8] Chen J, Wang A, Chen J, et al. CN-Probase: A data-driven approach for large-scale Chinese taxonomy construction // *Proceedings of the 2019 IEEE 35th International Conference on Data Engineering*. Macao, China, 2019: 1706-1709
- [9] Xu B, Xu Y, Liang J, et al. CN-DBpedia: A never-ending Chinese knowledge extraction system // *Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Arras, France, 2017: 428-438
- [10] Chen H, Hu N, Qi G, et al. OpenKG chain: A blockchain infrastructure for open knowledge graphs. *Data Intelligence*, 2021, 3(2): 205-227
- [11] Wang Yuan-Zhuo, Jia Yan-Tao, Zhao Ze-Ya, et al. OpenKN—Knowledge computing engine in network big data era. *Communications of the China Computer Federation*, 2014, 10(11): 30-35(in Chinese)

- (王元卓, 贾岩涛, 赵泽亚等. OpenKN——网络大数据时代的知识计算引擎. 中国计算机学会通讯, 2014, 10(11): 30-35)
- [12] Ji S, Pan S, Cambria E, Yu P S, et al. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 33(2): 494-514
 - [13] Barnes J A. Class and committees in a Norwegian island parish. *Human Relations*, 1954, 7(1): 39-58
 - [14] Wasserman S, Faust K, et al. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press, 1994
 - [15] Yang Z, Tang J, Cohen W. Multi-modal Bayesian embeddings for learning social knowledge graphs//*Proceedings of the 25th International Joint Conference on Artificial Intelligence*. New York, USA, 2016: 2287-2293
 - [16] Harandizadeh B, Singh S. Tweeki: Linking named entities on Twitter to a knowledge graph//*Proceedings of the 6th Workshop on Noisy User-Generated Text*. Online, 2020: 222-231
 - [17] Nechaev Y, Corcoglioniti F, Giuliano C. SocialLink: Exploiting graph embeddings to link DPpedia entities to Twitter profiles. *Progress in Artificial Intelligence*, 2018, 7(4): 251-272
 - [18] Nguyen H L, Jung J E. SocioScope: A framework for understanding internet of social knowledge. *Future Generation Computer Systems*, 2018, 83: 358-365
 - [19] Nguyen H L, Jung J J. Extending SocioScope framework for generating knowledge graph from social data//*Proceedings of the 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems*. Las Palmas de Gran Canaria, Spain, 2018: 621-624
 - [20] Nguyen H L, Jung J J. Social event decomposition for constructing knowledge graph. *Future Generation Computer Systems*, 2019, 100: 10-18
 - [21] Alonso O, Kandylas V, Tremblay S E. Social knowledge graph explorer//*Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Paris, France, 2019: 1317-1320
 - [22] He Q, Yang J, Shi B. Constructing knowledge graph for social networks in a deep and holistic way//*Proceedings of the Web Conference 2020*. Taipei, China, 2020: 307-308
 - [23] Balog K, Kenter T. Personal knowledge graphs: A research agenda//*Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*. Santa Clara, USA, 2019: 217-220
 - [24] Shi B, Yang J, Weninger T, et al. Representation learning in heterogeneous professional social networks with ambiguous social connections//*Proceedings of the 2019 IEEE International Conference on Big Data*. Los Angeles, USA, 2019: 1928-1937
 - [25] Kazemi S M, Goel R, Jain K, et al. Representation learning for dynamic graphs: A survey. *Journal of Machine Learning Research*, 2020, 21(70): 1-73
 - [26] Skardinga J, Gabrys B, Musial K. Foundations and modelling of dynamic networks using dynamic graph neural networks: A survey. *IEEE Access*, 2021, 9: 79143-79168
 - [27] Hamblin C L. Instants and intervals. *The Study of Time*, 1971, 24(1): 127-134
 - [28] Cheng Su-Qi, Shen Hua-Wei, Zhang Guo-Qing, et al. Survey of signed network research. *Journal of Software*, 2014, 25(1): 1-15(in Chinese)
(程苏琦, 沈华伟, 张国清等. 符号网络研究综述. 软件学报, 2014, 25(1): 1-15)
 - [29] Trivedi R, Farajtabar M, Biswal P, et al. DyREP: Learning representations over dynamic graphs//*Proceedings of the International Conference on Learning Representations*. New Orleans, USA, 2019
 - [30] Lin B Y, Xu F F, Luo Z, et al. Multi-channel BiLSTM-CRF model for emerging named entity recognition in social media//*Proceedings of the 3rd Workshop on Noisy User-Generated Text*. Copenhagen, Denmark, 2017: 160-165
 - [31] Moon S, Neves L, Carvalho V. Multimodal named entity recognition for short social media posts. *arXiv preprint arXiv: 1802.07862*, 2018
 - [32] Aguilar G, Maharjan S, López-Monroy A P, et al. A multi-task approach for named entity recognition in social media data. *arXiv preprint arXiv:1906.04135*, 2019
 - [33] Peng N, Dredze M. Named entity recognition for Chinese social media with jointly trained embeddings//*Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal, 2015: 548-554
 - [34] Tu C, Zhang Z, Liu Z, et al. TransNet: Translation-based network representation learning for social relation extraction//*Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Melbourne, Australia, 2017: 2864-2870
 - [35] Shen Y, Jiang X, Wang Y, et al. Dynamic relation extraction with a learnable temporal encoding method//*Proceedings of the 2020 IEEE International Conference on Knowledge Graph*. Nanjing, China, 2020: 235-242
 - [36] Wan H, Zhang M, Du J, et al. FL-MSRE: A few-shot learning based approach to multimodal social relation extraction//*Proceedings of the AAAI Conference on Artificial Intelligence*. Online, 2021: 13916-13923
 - [37] Yu D, Sun K, Cardie C, et al. Dialogue-based relation extraction. *arXiv preprint arXiv:2004.08056*, 2020
 - [38] Zhao L, Xu W, Gao S, et al. Utilizing graph neural networks to improving dialogue-based relation extraction. *Neurocomputing*, 2021, 456:299-311
 - [39] Chen H, Hong P, Han W, et al. Dialogue relation extraction with document-level heterogeneous graph attention networks. *arXiv preprint arXiv:2009.05092*, 2020
 - [40] Ritter A, Etzioni O, Clark S. Open domain event extraction from Twitter//*Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA, 2012: 1104-1112

- [41] Gottschalk S, Demidova E. EventKG: A multilingual event-centric temporal knowledge graph//Proceedings of the European Semantic Web Conference. Crete, Greece, 2018: 272-287
- [42] Hu Zhi-Lei, Jin Xiao-Long, Chen Jian-Yun, et al. Construction, reasoning and applications of event graphs. *Big Data Research*, 2021, 7(3): 80-96(in Chinese)
(胡志磊, 靳小龙, 陈剑赞等. 事件图谱的构建、推理与应用. *大数据*, 2021, 7(3): 80-96)
- [43] Shu K, Wang S, Tang J, et al. User identity linkage across online social networks: A review. *ACM SIGKDD Explorations Newsletter*, 2017, 18(2): 5-17
- [44] Sun L, Zhang Z, Ji P, et al. DNA: Dynamic social network alignment//Proceeding of the 2019 IEEE International Conference on Big Data (Big Data). Los Angeles, USA, 2019: 1224-1231
- [45] Sun Z, Zhang Q, Hu W, et al. A benchmarking study of embedding based entity alignment for knowledge graphs. *arXiv preprint arXiv:2003.07743*, 2020
- [46] Zhao X, Zeng W, Tang J, et al. An experimental study of state-of-the-art entity alignment approaches. *IEEE Transactions on Knowledge & Data Engineering*, 2020, 34(6): 2610-2625
- [47] Sun Z, Hu W, Zhang Q, et al. Bootstrapping entity alignment with knowledge graph embedding//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018: 4396-4402
- [48] Zhang Q, Sun Z, Hu W, et al. Multi-view knowledge graph embedding for entity alignment. *arXiv preprint arXiv:1906.02390*, 2019
- [49] Tang X, Zhang J, Chen B, et al. Bert-INT: A BERT-based interaction model for knowledge graph alignment//Proceedings of the 29th International Joint Conference on Artificial Intelligence. Yokohama, Japan, 2020: 3174-3180
- [50] Wang Z, Lv Q, Lan X, et al. Cross-lingual knowledge graph alignment via graph convolutional networks//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 349-357
- [51] Wu Y, Liu X, Feng Y, et al. Relation-aware entity alignment for heterogeneous knowledge graphs. *arXiv preprint arXiv:1908.08210*, 2019
- [52] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks//Proceedings of the European Semantic Web Conference. Crete, Greece, 2018: 593-607
- [53] Ye R, Li X, Fang Y, et al. A vectorized relational graph convolutional network for multi-relational network alignment //Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China, 2019: 4135-4141
- [54] Wu Y, Liu X, Feng Y, et al. Jointly learning entity and relation representations for entity alignment. *arXiv preprint arXiv:1909.09317*, 2019
- [55] Cao Y, Liu Z, Li C, et al. Multi-channel graph neural network for entity alignment. *arXiv preprint arXiv:1908.09898*, 2019
- [56] Sun Z, Wang C, Hu W, et al. Knowledge graph alignment network with gated multi-hop neighborhood aggregation//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020: 222-229
- [57] Tang J, Fang Z, Sun J. Incorporating social context and domain knowledge for entity recognition//Proceedings of the 24th International Conference on World Wide Web. New York, USA, 2015: 517-526
- [58] He G, Li J, Zhao W X, et al. Mining implicit entity preference from user-item interaction data for knowledge graph completion via adversarial learning//Proceedings of the Web Conference 2020. Tapei, China, 2020: 740-751
- [59] Zhou K, Zhao W X, Bian S, et al. Improving conversational recommender systems via knowledge graph based semantic fusion//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Online, 2020: 1006-1014
- [60] Huang C, Xu H, Xu Y, et al. Knowledge-aware coupled graph neural network for social recommendation//Proceedings of the AAAI Conference on Artificial Intelligence. Online, 2021: 4115-4122
- [61] Xiao Z, Song W, Xu H, et al. TIMME: Twitter ideology-detection via multi-task multi-relational embedding//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Online, 2020: 2258-2268
- [62] Wang Y, Ma F, Gao J. Efficient knowledge graph validation via cross-graph representation learning//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. Online, 2020: 1595-1604
- [63] Lin Hai-Lun, Wang Yuan-Zhuo, Jia Yan-Tao, et al. Network big data oriented knowledge fusion methods: A survey. *Chinese Journal of Computers*, 2017, 40(1): 1-27(in Chinese)
(林海伦, 王元卓, 贾岩涛等. 面向网络大数据的知识融合方法综述. *计算机学报*, 2017, 40(1): 1-27)
- [64] Sun Y, Han J. Mining heterogeneous information networks: Principles and methodologies. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 2012, 3(2): 1-159
- [65] Sun Y, Han J. Mining heterogeneous information networks: A structural analysis approach. *ACM SIGKDD Explorations Newsletter*, 2013, 14(2): 20-28
- [66] Shi Chuan, Wang Rui-Jia, Wang Xiao. Survey on heterogeneous information networks analysis and applications. *Journal of Software*, 2022, 33(2): 598-621(in Chinese)
(石川, 王睿嘉, 王啸. 异质信息网络分析与应用综述. *软件学报*, 2022, 33(2): 598-621)
- [67] Yang C, Xiao Y, Zhang Y, et al. Heterogeneous network representation learning: A unified framework with survey and benchmark. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(10): 4854-4873
- [68] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality//Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2013: 3111-3119

- [69] 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. Nova Scotia, Canada, 2017: 135-144
- [70] Fu T Y, Lee W C, Lei Z. HIN2Vec: Explore meta-paths in heterogeneous information networks for representation learning //Proceedings of the 2017 ACM Conference on Information and Knowledge Management. Singapore, 2017: 1797-1806
- [71] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data. *Advances in Neural Information Processing Systems*, 2013, 26: 2787-2795
- [72] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes//Proceedings of the AAAI Conference on Artificial Intelligence. Québec, Canada, 2014: 1112-1119
- [73] Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA, 2015: 2181-2187
- [74] Jia Y, Wang Y, Lin H, et al. Locally adaptive translation for knowledge graph embedding//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA, 2016: 992-998
- [75] Shi C, Lu Y, Hu L, et al. RHINE: Relation structure-aware heterogeneous information network embedding. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34 (1): 433-447
- [76] Chen D, Li Y, Ding B, et al. An adaptive embedding framework for heterogeneous information networks//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. Online, 2020: 165-174
- [77] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324
- [78] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735-1780
- [79] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016
- [80] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017
- [81] Lamb L C, d'Avila Garcez A, Gori M, et al. Graph neural networks meet neural-symbolic computing: A survey and perspective//Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence. Yokohama, Japan, 2021: 4877-4884
- [82] Wang X, Ji H, Shi C, et al. Heterogeneous graph attention network//Proceedings of the World Wide Web Conference. New York, USA, 2019: 2022-2032
- [83] Hu Z, Dong Y, Wang K, et al. Heterogeneous graph transformer//Proceedings of the Web Conference 2020. Taipei, China, 2020: 2704-2710
- [84] Zhang C, Song D, Huang C, et al. Heterogeneous graph neural network//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, USA, 2019: 793-803
- [85] Huang T, Zhuang Z, Zhang S, et al. Homogenization with explicit semantics preservation for heterogeneous information network//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. Online, 2020: 2065-2068
- [86] Yu D, Yang Y, Zhang R, et al. Knowledge embedding based graph convolutional network//Proceedings of the Web Conference 2021. Ljubljana, Slovenia, 2021: 1619-1628
- [87] Zhang J, Chen B, Zhang L, et al. Neural, symbolic and neural-symbolic reasoning on knowledge graphs. *AI Open*, 2021, 2: 14-35
- [88] Guo S, Wang Q, Wang L, et al. Jointly embedding knowledge graphs and logical rules//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, USA, 2016: 192-202
- [89] Guo S, Wang Q, Wang L, et al. Knowledge graph embedding with iterative guidance from soft rules//Proceedings of the AAAI Conference on Artificial Intelligence. Orleans, USA, 2018: 4816-4823
- [90] Zhang W, Paudel B, Wang L, et al. Iteratively learning embeddings and rules for knowledge graph reasoning//Proceedings of the World Wide Web Conference. San Francisco, USA, 2019: 2366-2377
- [91] Shen Y, Jiang X, Li Z, et al. UniSKGRep: A unified representation learning framework of social network and knowledge graph. *Neural Networks*, 2022, 158: 142-153
- [92] Heidari F, Papagelis M. Evolving network representation learning based on random walks. *Applied Network Science*, 2020, 5(1): 1-38
- [93] Goyal P, Kamra N, He X, et al. DynGEM: Deep embedding method for dynamic graphs. *arXiv preprint arXiv:1805.11273*, 2018
- [94] Goyal P, Chhetri S R, Canedo A. dyngraph2vec: Capturing network dynamics using dynamic graph representation learning. *Knowledge-Based Systems*, 2020, 187: 104816
- [95] Zhou L, Yang Y, Ren X, et al. Dynamic network embedding by modeling triadic closure process//Proceedings of the AAAI Conference on Artificial Intelligence. Orleans, USA, 2018: 571-578
- [96] Xu Z, Ou Z, Su Q, et al. Embedding dynamic attributed networks by modeling the evolution processes. *arXiv preprint arXiv:2010.14047*, 2020
- [97] Sankar A, Wu Y, Gou L, et al. DySAT: Deep neural representation learning on dynamic graphs via self-attention networks //Proceedings of the 13th International Conference on Web Search and Data Mining. Houston, USA, 2020: 519-527

- [98] Fathy A, Li K. TemporalGAT: Attention-based dynamic graph representation learning//Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining. Singapore, 2020: 413-423
- [99] Pareja A, Domeniconi G, Chen J, et al. EvolveGCN: Evolving graph convolutional networks for dynamic graphs//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020: 5363-5370
- [100] Nguyen G H, Lee J B, Rossi R A, et al. Continuous-time dynamic network embeddings//Proceedings of the Web Conference 2018. Lyon, France, 2018: 969-976
- [101] Chang X, Liu X, Wen J, et al. Continuous-time dynamic graph learning via neural interaction processes//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. Online, 2020: 145-154
- [102] Lu Y, Wang X, Shi C, et al. Temporal network embedding with micro-and macro-dynamics//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China, 2019: 469-478
- [103] Zhou D, Zheng L, Han J, et al. A data-driven graph generative model for temporal interaction networks//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Online, 2020: 401-411
- [104] Kumar S, Zhang X, Leskovec J. Predicting dynamic embedding trajectory in temporal interaction networks//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, USA, 2019: 1269-1278
- [105] Rossi E, Chamberlain B, Frasca F, et al. Temporal graph networks for deep learning on dynamic graphs. arXiv preprint arXiv:2006.10637, 2020
- [106] Xu D, Ruan C, Korpoglu E, et al. Inductive representation learning on temporal graphs. arXiv preprint arXiv:2002.07962, 2020
- [107] Paranjape A, Benson A R, Leskovec J. Motifs in temporal networks//Proceedings of the 10th ACM International Conference on Web Search and Data Mining. Cambridge, UK, 2017: 601-610
- [108] Wang Y, Chang Y Y, Liu Y, et al. Inductive representation learning in temporal networks via causal anonymous walks. arXiv preprint arXiv:2101.05974, 2021
- [109] Huang H, Fang Z, Wang X, et al. Motif-preserving temporal network embedding//Proceedings of the 29th International Joint Conference on Artificial Intelligence. Yokohama, Japan, 2020: 1237-1243
- [110] Fu D, Zhou D, He J. Local motif clustering on time-evolving graphs//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Online, 2020: 390-400
- [111] Ivanov S, Burnaev E. Anonymous walk embeddings//Proceedings of the International Conference on Machine Learning. Stockholmssåsan, Sweden, 2018: 2186-2195
- [112] Zhao Ze-Ya, Jia Yan-Tao, Wang Yuan-Zhuo, et al. Temporal link prediction based on dynamic heterogeneous information network. Journal of Computer Research and Development, 2015, 52(8): 1735-1741(in Chinese)
(赵泽亚, 贾岩涛, 王元卓等. 基于动态异构信息网络的时序关系预测. 计算机研究与发展, 2015, 52(8): 1735-1741)
- [113] Yin Y, Ji L X, Zhang J P, et al. DHNE: Network representation learning method for dynamic heterogeneous networks. IEEE Access, 2019, 7: 134782-134792
- [114] Xue H, Yang L, Jiang W, et al. Modeling dynamic heterogeneous network for link prediction using hierarchical attention with temporal rnn. arXiv preprint arXiv:2004.01024, 2020
- [115] Li Q, Shang Y, Qiao X, et al. Heterogeneous dynamic graph attention network//Proceedings of the 2020 IEEE International Conference on Knowledge Graph. Nanjing, China, 2020: 404-411
- [116] Trivedi R, Dai H, Wang Y, et al. Know-Evolve: Deep temporal reasoning for dynamic knowledge graphs//Proceedings of the International Conference on Machine Learning. Sydney, Australia, 2017: 3462-3471
- [117] Yan C, Yan J, Xu Y, et al. Learning to encode evolutionary knowledge for automatic commenting long novels. arXiv preprint arXiv:2004.09974, 2020
- [118] Tang J, Chang Y, Aggarwal C, et al. A survey of signed network mining in social media. ACM Computing Surveys, 2016, 49(3): 1-37
- [119] West R, Paskov H S, Leskovec J, et al. Exploiting social network structure for person-to-person sentiment analysis. Transactions of the Association for Computational Linguistics, 2014, 2: 297-310
- [120] Heider F. Attitudes and cognitive organization. The Journal of Psychology, 1946, 21(1): 107-112
- [121] Cygan M, Pilipczuk M, Pilipczuk M, et al. Sitting closer to friends than enemies, revisited//Proceedings of the International Symposium on Mathematical Foundations of Computer Science. Bratislava, Slovakia, 2012: 296-307
- [122] Leskovec J, Huttenlocher D, Kleinberg J. Signed networks in social media//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Atlanta, USA, 2010: 1361-1370
- [123] Wang S, Tang J, Aggarwal C, et al. Signed network embedding in social media//Proceedings of the 2017 SIAM International Conference on Data Mining. Houston, USA, 2017: 327-335
- [124] Yuan S, Wu X, Xiang Y. SNE: Signed network embedding //Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining. Jeju Island, Korea, 2017: 183-195
- [125] Kim J, Park H, Lee J E, et al. SIDE: Representation learning in signed directed networks//Proceedings of the 2018 World Wide Web Conference. Lyon, France, 2018: 509-518

- [126] Derr T, Ma Y, Tang J. Signed graph convolutional networks // Proceedings of the 2018 IEEE International Conference on Data Mining. Singapore, Singapore, 2018: 929-934
- [127] Huang J, Shen H, Hou L, et al. Signed graph attention networks // Proceedings of the International Conference on Artificial Neural Networks. Munich, Germany, 2019: 566-577
- [128] Meng Qing, Liu Bo, Zhang Heng-Yuan, et al. Multi-relational group influence modeling and analysis in online social networks. Chinese Journal of Computers, 2021, 44(6): 1064-1079 (in Chinese)
(孟青, 刘波, 张恒远等. 在线社交网络中群体影响力的建模与分析. 计算机学报, 2021, 44(6): 1064-1079)
- [129] Zhao Shu, Liu Xiao-Man, Duan Zhen, et al. A survey on social ties mining. Chinese Journal of Computers, 2017, 40(3): 535-555 (in Chinese)
(赵姝, 刘晓曼, 段震等. 社交关系挖掘研究综述. 计算机学报, 2017, 40(3): 535-555)
- [130] Guan Sai-Ping, Jin Xiao-Long, Jia Yan-Tao, et al. Knowledge reasoning over knowledge graph: A survey. Journal of Software, 2018, 29(10): 2966-2994 (in Chinese)
(官赛萍, 靳小龙, 贾岩涛等. 面向知识图谱的知识推理研究进展. 软件学报, 2018, 29(10): 2966-2994)
- [131] Sun Y, Barber R, Gupta M, et al. Co-author relationship prediction in heterogeneous bibliographic networks // Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining. NW Washington, USA, 2011: 121-128
- [132] Wan C, Li X, Kao B, et al. Classification with active learning and meta-paths in heterogeneous information networks // Proceedings of the 24th ACM International Conference on Information and Knowledge Management. Melbourne, Australia, 2015: 443-452
- [133] Zhao J, Wang X, Shi C, et al. Heterogeneous graph structure learning for graph neural networks // Proceedings of the 35th AAAI Conference on Artificial Intelligence. Online, 2021: 4697-4705
- [134] Hwang D, Park J, Kwon S, et al. Self-supervised auxiliary learning with meta-paths for heterogeneous graphs. arXiv preprint arXiv:2007.08294, 2020
- [135] Teru K, Denis E, Hamilton W. Inductive relation prediction by subgraph reasoning // Proceedings of the International Conference on Machine Learning. Online, 2020: 9448-9457
- [136] Wang Z, Zong B, Sun H. Modeling context pair interaction for pairwise tasks on graphs // Proceedings of the 14th ACM International Conference on Web Search and Data Mining. Online, 2021: 851-859
- [137] Tang J, Lou T, Kleinberg J. Inferring social ties across heterogenous networks // Proceedings of the 5th ACM International Conference on Web Search and Data Mining. Seattle, USA, 2012: 743-752
- [138] Wang H, Zhang F, Hou M, et al. SHINE: Signed heterogeneous information network embedding for sentiment link prediction // Proceedings of the 11th ACM International Conference on Web Search and Data Mining. Los Angeles, USA, 2018: 592-600
- [139] McAuley J J, Leskovec J. Learning to discover social circles in ego networks // Proceedings of the 26th Conference on Neural Information Processing Systems. Lake Tahoe, USA, 2012: 548-556
- [140] Long Q, Jin Y, Song G, et al. Graph structural-topic neural network // Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Online, 2020: 1065-1073
- [141] Yu J, Yin H, Li J, et al. Self-supervised multi-channel hypergraph convolutional network for social recommendation // Proceedings of the Web Conference 2021. Ljubljana, Slovenia, 2021: 413-424
- [142] Tylenda T, Angelova R, Bedathur S. Towards time-aware link prediction in evolving social networks // Proceedings of the 3rd Workshop on Social Network Mining and Analysis. Paris, France, 2009: 1-10
- [143] Dunlavy D M, Kolda T G, Acar E. Temporal link prediction using matrix and tensor factorizations. ACM Transactions on Knowledge Discovery from Data, 2011, 5(2): 1-27
- [144] Tresp V, Ma Y, Baier S, et al. Embedding learning for declarative memories // Proceedings of the European Semantic Web Conference. Portorož, Slovenia, 2017: 202-216
- [145] Dasgupta S S, Ray S N, Talukdar P. HyTE: Hyperplane-based temporally aware knowledge graph embedding // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 2001-2011
- [146] Leblay J, Chekol M W. Deriving validity time in knowledge graph // Proceedings of the Web Conference 2018. Lyon, France, 2018: 1771-1776
- [147] García-Durán A, Dumanacić S, Niepert M. Learning sequence encoders for temporal knowledge graph completion. arXiv preprint arXiv:1809.03202, 2018
- [148] Jin W, Qu M, Jin X, et al. Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. arXiv preprint arXiv:1904.05530, 2019
- [149] Wu J, Cao M, Cheung J C K, et al. TeMP: Temporal message passing for temporal knowledge graph completion. arXiv preprint arXiv:2010.03526, 2020
- [150] Jung J, Jung J, Kang U. T-GAP: Learning to walk across time for temporal knowledge graph completion. arXiv preprint arXiv:2012.10595, 2020
- [151] Zuo Y, Liu G, Lin H, et al. Embedding temporal network via neighborhood formation // Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London United Kingdom, 2018: 2857-2866

- [152] Han Z, Ma Y, Wang Y, et al. Graph Hawkes neural network for forecasting on temporal knowledge graphs. arXiv preprint arXiv:2003.13432, 2020
- [153] He S, Liu K, Ji G, et al. Learning to represent knowledge graphs with Gaussian embedding//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Melbourne, Australia, 2015: 623-632
- [154] Lacroix T, Obozinski G, Usunier N. Tensor decompositions for temporal knowledge base completion//Proceedings of the International Conference on Learning Representations. Online, 2019
- [155] Xu C, Chen Y Y, Nayyeri M, et al. Temporal knowledge graph completion using a linear temporal regularizer and multivector embeddings//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Online, 2021: 2569-2578
- [156] Tu C, Liu H, Liu Z, et al. CANE: Context-aware network embedding for relation modeling//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, 2017: 1722-1731
- [157] Lai Y Y, Neville J, Goldwasser D. TransConv: Relationship embedding in social networks//Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu, USA, 2019: 4130-4138
- [158] Lin L, Wang H. Graph attention networks over edge content-based channels//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Online, 2020: 1819-1827
- [159] Lai Y Y, Neville J. MERL: Multi-view edge representation learning in social networks//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. Online, 2020: 675-684
- [160] Jiang X, Shen Y, Wang Y, et al. BaKGraSTeC: A background knowledge graph based method for short text classification //Proceedings of the 2020 IEEE International Conference on Knowledge Graph. Nanjing, China, 2020: 360-366
- [161] Yang C, Zhang J, Wang H, et al. Relation learning on social networks with multi-modal graph edge variational autoencoders//Proceedings of the 13th International Conference on Web Search and Data Mining. Houston, USA, 2020: 699-707
- [162] Wang Zheng, Wu Bin, Wang Wen-Zhe, et al. A survey of social relation understanding based on image and video information. Chinese Journal of Computers, 2021, 44(6): 1168-1199(in Chinese)
(王正, 吴斌, 王文哲等. 基于图像和视频信息的社交关系理解研究综述. 计算机学报, 2021, 44(6): 1168-1199)
- [163] Wang Z, Chen T, Ren J, et al. Deep reasoning with knowledge graph for social relationship understanding. arXiv preprint arXiv:1807.00504, 2018
- [164] Mi L, Chen Z. Hierarchical graph attention network for visual relationship detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 13886-13895
- [165] Tang Y, Lu J, Wang Z, et al. Learning semantics-preserving attention and contextual interaction for group activity recognition. IEEE Transactions on Image Processing, 2019, 28(10): 4997-5012
- [166] Liu X, Liu W, Zhang M, et al. Social relation recognition from videos via multi-scale spatial-temporal reasoning//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 3566-3574
- [167] Wu S, Chen J, Xu T, et al. Linking the characters: Video-oriented social graph generation via hierarchical-cumulative gcn//Proceedings of the 29th ACM International Conference on Multimedia. Online, 2021: 4716-4724
- [168] Eisner B, Rocktäschel T, Augenstein I, et al. emoji2vec: Learning emoji representations from their description. arXiv preprint arXiv:1609.08359, 2016
- [169] Miller H J, Thebault-Spieker J, Chang S, et al. "blissfully happy" or "ready tofight": Varying interpretations of emoji//Proceedings of the Tenth International AAAI Conference on Web and Social Media. Cologne, Germany, 2016: 259-268
- [170] Luo X, Liu L, Yang Y, et al. AliCoCo: Alibaba e-commerce cognitive concept net//Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. Portland, USA, 2020: 313-327
- [171] Qin Chuan, Zhu Heng-Shu, Zhuang Fu-Zhen, et al. A survey on knowledge graph-based recommender systems. Scientia Sinica Informationis, 2020, 50(7): 937-956 (in Chinese)
(秦川, 祝恒书, 庄福振等. 基于知识图谱的推荐系统研究综述. 中国科学: 信息科学, 2020, 50(7): 937-956)
- [172] Li J, Zhao W X, Wei Z, et al. Knowledge-based review generation by coherence enhanced text planning//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. Virtual Event, Canada, 2021: 183-192
- [173] Tang J. AMiner: Toward understanding big scholar data//Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. New York, USA, 2016: 467-467
- [174] Wang K, Shen Z, Huang C, et al. Microsoft academic graph: When experts are not enough. Quantitative Science Studies, 2020, 1(1): 396-413
- [175] Cheng X, Liu S, Sun X, et al. Combating emerging financial risks in the big data era: A perspective review. Fundamental Research, 2021, 1(5): 595-606



JIANG Xu-Hui, Ph. D. candidate.

His research interests include social knowledge graph and dynamic graph representation learning.

LI Zi-Jian, M. S. candidate. His research interests include social knowledge graph and signed network representation learning.

WANG Yuan-Zhuo, Ph. D. , professor. His research interests include knowledge graph and social computing.

YIN Zhi-Yi, Ph. D. , senior engineer. Her research interests include intelligent algorithm security governance.

SHEN Hua-Wei, Ph. D. , professor. His research interests include social network analysis and social media computing.

SHEN Ying-Han, Ph. D. candidate. His research interests include social knowledge graph and dynamic graph representation learning.

Background

A large amount of data in the real world exists in the form of graphs. Social Networks and Knowledge Graphs, which are two representative graph structure data, respectively model the social interactions between people and the relationships between knowledge. The Social Network (SN) reflects social connections among people, and the Knowledge Graph (KG) describes the relational knowledge of entities in the real world. Considering that the information complementation of SN and KG can benefit human-oriented applications such as recommendation systems and social analysis. It is essential to propose a unified framework for the integration of SN and KG. However, there is currently no survey focusing on this research trend. In this paper, we introduce Social Knowledge Graph (SKG) and give a systematic analysis of it. Specifically,

we first introduce relevant concepts about SKG, and propose a formal definition of it. Next, we comprehensively analyze the characteristics of SKG from four perspectives including heterogeneous, dynamic, emotional and coevolutionary, and introduce relevant representation learning studies. Then, around the life cycle of the social knowledge graph, we introduce the relevant representative studies of the SKG construction, integration and reasoning. Finally, we introduce the related applications of SKG, discuss its valuable future direction.

This paper is funded by the National Natural Science Foundation of China (Nos. U1836206, U21B2046, and 62172393) and Zhongyuanyingcai Program-Funded to Central Plains Science and Technology Innovation Leading Talent Program (No. 204200510002).