

异质信息网络表征学习综述

周丽华 王家龙 王丽珍 陈红梅 孔 兵

(云南大学信息学院计算机科学与工程系 昆明 650500)

摘 要 随着信息技术的快速发展,信息网络无处不在,例如社交网络、学术网络、万维网等. 由于网络规模不断扩大以及数据的稀疏性,信息网络的分析方法面临巨大挑战. 作为应对网络规模及数据稀疏挑战的有效方法,信息网络表征学习旨在利用网络的拓扑结构、节点内容等信息将节点嵌入到低维的向量空间中,同时保留原始网络固有的结构特征和内容特征,从而使节点的分类、聚类、链路预测等网络分析任务能够基于低维、稠密的向量完成. 由多种类型的节点和连边构成的异质信息网络包含更加全面、丰富的结构和语义信息,因此异质信息网络的表征学习不仅能够有效缓解网络数据高维、稀疏性问题,还能融合网络中不同类型的异质信息,使学习到的表征更有意义和价值. 近年来,异质信息网络的表征学习受到学术界和工业界的广泛关注,成为网络分析的一个重要研究主题,研究成果不断涌现. 然而,目前还缺乏对现有成果进行全面梳理的工作,相关研究人员难以系统地了解最新研究进展,在实际应用中也难以选择合适的嵌入模型. 为此,本文对异质信息网络表征学习的方法进行了全面综述,包括相关概念、网络分类、学习方法、数据集与测评指标、典型应用,同时对未来的研究方向进行了展望. 本文工作有助于研究人员全面系统地了解异质信息网络表征学习的研究进展,也有助于从业人员更有效地解决实际问题.

关键词 异质信息网络; 表征学习; 随机游走; 负采样; 深度神经网络

中图法分类号 TP18 **DOI号** DOI号 10.11897/SP.J.1016.2022.00160

Heterogeneous Information Network Representation Learning: A Survey

ZHOU Li-Hua WANG Jia-Long WANG Li-Zhen CHEN Hong-Mei KONG Bing

(School of Information Science and Engineering, Yunnan University, Kunming 650500)

Abstract Information networks are ubiquitous in real world, such as social networks, academic networks and the World Wide Web, etc. With the rapid development of information technology, methods for analyzing information networks face huge challenges due to the ever-expanding network scale and the sparseness of data, which requires that the analysis methods have good scalability and can effectively solve the problems caused by data sparsity. As an effective way to deal with these challenges, information network representation learning aims to embed nodes or links into a low-dimensional vector space by using information such as network topology and node content, while preserving the intrinsic structural and content features of the original network, so that network analysis tasks, such as node classification, clustering and link prediction, etc. can be completed based on low-dimensional dense vectors. The heterogeneous information network, consisting of multiple types of nodes and links, contains more comprehensive and rich structure and semantic information. Therefore, the representation learning of heterogeneous information

收稿日期:2020-10-13;在线发布日期:2021-05-13. 本课题得到国家自然科学基金(62062066,61762090和61966036)、云南省大学创新计划(IRTSTYN)、国家社会科学基金(18XZZ005)、云南省高校物联网技术及应用重点实验室和云南大学研究生科研创新基金项目(2021Y024)的资助. 周丽华,博士,教授,博士生导师,中国计算机学会(CCF)会员(42099M),主要研究领域为数据挖掘、社会网络分析、人工智能. E-mail: lhzhou@ynu.edu.cn. 王家龙,博士研究生,主要研究领域为数据挖掘、社会网络分析. 王丽珍,博士,教授,博士生导师,主要研究方向为数据挖掘、数据仓库、计算机算法. 陈红梅(通信作者),博士,副教授,主要研究方向为数据结构、数据库技术、数据挖掘. E-mail: hmchen@ynu.edu.cn. 孔 兵,博士,副教授,主要研究方向为数据挖掘、社会网络分析、机器学习.

networks can not only effectively alleviate the high-dimensional and sparsity problems of network data, but also integrate different types of heterogeneous information into the same vector space, which makes the learned features more meaningful and valuable. Recently the research on the representation learning and application of heterogeneous information networks has drawn increasing attention in both academia and industry and has been an important research topic in network analysis area, and a lot of research results on the representation learning of heterogeneous information networks have been proposed. However, the lack of detailed and comprehensive survey of existing work makes it difficult for researchers to systematically understand the latest research progress, and for choosing appropriate embedding models and algorithms in practical applications. To this end, we conduct in this survey a thorough review of existing work on representation learning of heterogeneous information networks. We first introduce some basic concepts about heterogeneous information networks and network embedding learning, divide the heterogeneous information networks into nine categories (structural, attribute, multi-layer, multi-view, multiplex, attributed multiplex, multi-resolution, heterogeneous feature and dynamic heterogeneous networks) based on the heterogeneity sources, summarize the characteristics of various networks, and introduce the principle of the random walks and the negative sampling, two techniques usually used by many representation learning methods for heterogeneous information networks. Next, we comprehensively review the network representation learning methods for single, multiple and dynamic heterogeneous information networks, including random walks-based, decomposition-based and deep neural network-based methods. We also analyze the characteristics of different methods, such as used information and technique, embedded object, representation form, and time complexity etc. Then, we summarize data sets and evaluation metrics generally used by various embedding learning methods for heterogeneous information networks. In addition, several typical applications of embedding learning for heterogeneous information networks, i. e. author identification, recommendation and sentiment link prediction are presented in detail. At last, we reveal the research directions related to the embedding learning for heterogeneous information networks in the future. This survey does not only help researchers to have a better understanding of existing work, but also help applicants to better solve practical application problems.

Keywords heterogeneous information network; representation learning; random walks; negative sampling; deep neural network

1 引言

信息网络在现实世界中无处不在,比如 Facebook 和 Twitter 等社交网络、DBLP 和 Google Scholar 等论文引用网络、Amazon 和淘宝等电子商务网站、Foursquare 和 Yelp 等推荐站点以及生物网络和医疗系统等.随着信息技术的飞速发展,信息网络常规模庞大,涉及数百万甚至数亿个节点,每个节点还包含复杂的属性特征,例如,Facebook 在 2018 年第二季度每月有 22.3 亿个活跃用户,有的用户可能有成千上万的帖子和朋友^①.同时,在规模

庞大的网络中数据常常是稀疏的,比如在社交媒体中,很大比例的单词或属性通常仅对少数用户有用^[1].不断增长的网络规模以及数据的稀疏性对网络分析方法提出了挑战,要求分析方法应该具有良好的可扩展性,并能有效解决数据的稀疏性问题.

信息网络表征学习(也称为表示学习或嵌入学习)旨在利用网络的拓扑结构、节点内容等信息将节点嵌入到低维的向量空间中,同时保留网络固有的结构特征和内容特征.在嵌入过程中,高出/入度节

^① www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

点的结构和/或内容信息可用于辅助低出/入度节点的结构或语义特征的表示,从而有效缓解网络数据稀疏性问题.通过嵌入,网络中的任意节点均可以用一个低维的稠密向量来表示,使得网络结构/语义的相关性能快速计算出来,为网络可视化、节点分类、聚类以及链路预测^[2]、Facebook 搜索^[3]、社区检测^[4-5]、微博用户标记^[6]等网络分析任务提供有力的支持.图 1 示意了网络表征学习的概念,其中节点的不同颜色代表节点的不同属性,比如用户偏好.网络表征学习将所有节点映射到二维向量空间,使得网络中有边相连或属性相似的节点对在嵌入空间中距离较近.

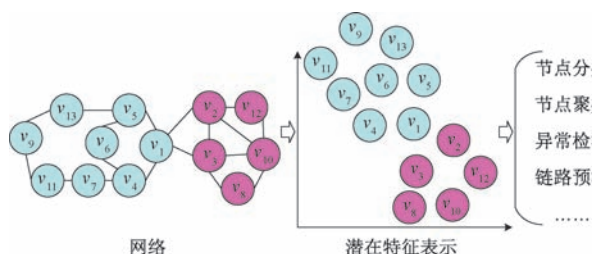


图 1 信息网络表征学习示意图

目前,大量的网络表征学习研究致力于只包含一种节点类型和一种边类型的同质信息网络,比如 DeepWalk、LINE 等模型^[7-15]基于网络的拓扑结构进行学习;TADW、GraRep 等模型^[16-30]使用网络的拓扑结构和节点的文本、图像和标签等信息完成网络的嵌入;DNE-SBP 模型^[31]面向包含正连接和负连接的符号网络学习节点表示;TWSRL 模型^[32]利用 Web 服务网络中网络结构、功能标签和服务描述信息学习 Web 服务的表示,使具有相似功能属性和网络结构的服务在低维空间中邻近;温雯等^[33]基于深度学习融合了节点的先验信息;GraphGan^[34]和 GraphWGAN^[35]利用生成对抗网络提高嵌入质量;李志宇等^[36]提出了动态特征学习方法,以适应社会网络所具备的动态性、大规模甚至超大规模等特性;崔宁宇等^[37]基于时间标签,提出了一种时间感知的知识表示学习方法;文献[2,38-43]全面总结了同质信息网络表征学习的研究进展.

同质信息网络是现实信息网络的一种简化,往往只抽取现实网络的部分信息,或者没有区分现实网络中对象及关系的差异性,从而造成信息的不完整或损失^[44],比如只考虑作者间合作关系的合著网络会丢失作者在什么主题、哪些论文上进行了合作的重要信息.为了完整自然地对现实世界的网络数据进行建模,越来越多的研究人员开始从事异质信

息网络的分析研究^[45-47].异质信息网络包含多种类型的节点和连边,不同类型的节点代表不同的对象,不同类型的连边代表对象间不同的交互关系.例如在疾病防控网络中包含医生、患者、疾病和药物等对象,医生与患者之间通过“治疗/被治疗”关系连接,患者与药物之间通过“服用/被服用”关系连接.不同类型的对象和连边具有不同的语义,区分了现实世界中数据对象及其关系的差异性;同时信息可以通过网络中不同类型的关系在不同对象之间传播,体现了对象之间的相似性或影响力.由于异质信息网络包含了更加全面、丰富的结构和语义信息,因此异质信息网络的表征学习不仅能够有效缓解网络数据高维、稀疏性问题,还能融合网络中不同类型的异质信息,使学习到的特征表示更有意义和价值.

然而,由于异质信息网络的特殊性,同质信息网络的表征学习方法不能直接应用于异质信息网络.网络中节点和边的异质性要求表征学习不仅能够全面地抽取和利用网络的多维度信息,还要对这些信息进行有效融合^[48],同时尽可能地捕获由各种属性引起的嵌入不确定性^[49].近年来,大量研究者对异质信息网络的表征学习进行了深入研究^[50-54],相关研究成果频频出现在 TKDE、TKDD、《计算机学报》、《软件学报》、《计算机研究与发展》、AAAI、IJ-CAI、SIGMOD、ICDE、VLDB、WWW、CIKM 等国内外重要的期刊和会议上.这些工作从不同的角度对现实网络建模,采用不同的方法融合异质信息,提出了不同的嵌入模型和优化方法实现嵌入.但是,目前还没有全面和详细介绍最新研究成果的综述论文,相关人员难以系统地了解最新研究进展,实际应用中也难以选择合适的嵌入模型和算法.本文对近年来关于异质信息网络表征学习的方法及研究现状进行归纳,介绍了相关知识,阐述了当前的主流方法与模型,概述了数据集与测评指标,示例了典型应用场景,同时对未来可能的研究方向进行了探讨与展望.本文工作有助于研究人员全面了解异质信息网络表征学习,也有助于从业人员更好地解决实际问题.

本文其余部分按如下方式组织.第 2 节介绍相关知识;第 3、4、5 节阐述主流方法与模型;第 6 节对各类方法的特点进行讨论;第 7 节概述相关数据集与测评指标;第 8 节示例典型应用;第 9 节对全文进行总结,并展望未来的研究方向.

2 相关知识

本文使用粗体的小写字母(例如 $\mathbf{f}$)表示向量,粗体大写字母(例如 $\mathbf{F}$)表示矩阵,大写花体字符(例如 $\mathcal{R}$)表示集合,使用普通的字符(例如 u 或 s)表示集合中的元素. 矩阵 $\mathbf{F}$ 的转置表示为 $\mathbf{F}^T$, $\mathbf{F}$ 的第 i 行表示为 $\mathbf{F}_{i,:}$ (简记为 $\mathbf{F}_i$), $\mathbf{F}$ 的第 j 列表示为 $\mathbf{F}_{:,j}$. $\mathbb{R}$ 表示实数集合.

2.1 基本概念

异质信息网络(简称异质网络, HeN)^[47]. HeN 是一个有向图 $G = (\mathcal{N}, \epsilon)$, 其中 $\mathcal{N}$ 和 ϵ 是节点集和边集, $\mathcal{N}$ 中的节点和 ϵ 中的边都有多种类型, 节点类型的集合和映射函数分别为 $\mathcal{C}$ 和 $\phi: \mathcal{N} \rightarrow \mathcal{C}$, 边类型的集合和映射函数分别为 $\mathcal{R}$ 和 $\varphi: \epsilon \rightarrow \mathcal{R}$, 其中 $|\mathcal{C}| + |\mathcal{R}| > 2$. 每个节点 $v \in \mathcal{N}$ 属于 $\mathcal{C}$ 中的一种特定类型 $C_v \in \mathcal{C}$, 每条边 $e \in \epsilon$ 属于 $\mathcal{R}$ 中的一种特定类型 $R_e \in \mathcal{R}$. 如果两条边属于同一种类型, 那么这两条边具有相同类型的开始节点和结束节点. 节点和边具体化的网络称为网络实例.

与异质信息网络密切相关的三个概念是网络模式、元路径和元图.

网络模式^[47]. 网络模式是异质信息网络的元描述, 是定义在节点类型和边类型上的一个有向图, 图中节点表示对象类型, 边表示关系类型, 记为 $S_G = (\mathcal{C}, \mathcal{R})$, 用于限定对象集合以及对象间关系的类型约束, 这些约束使得异质信息网络具有半结构化的特点, 引导网络语义的探究.

元路径^[47]. 元路径 $\mathcal{P}$ 定义为在 $S_G = (\mathcal{C}, \mathcal{R})$ 上通过边类型序列 $R_1 R_2 \dots R_L$ 连接的节点类型序列 $C_1 C_2 \dots C_{L+1}$, 表示为 $\mathcal{P} = C_1 \xrightarrow{R_1} C_2 \xrightarrow{R_2} \dots \xrightarrow{R_L} C_{L+1}$, L 表示 $\mathcal{P}$ 的长度. 当 $\mathcal{P}$ 与反向元路径 $\mathcal{P}^{-1}$ 相同时, 称 $\mathcal{P}$ 为对称元路径. 节点和边具体化的元路径 p 称为元路径的实例, 比如“作者甲 $\xrightarrow{\text{撰写}}$ 论文 1 $\xrightarrow{\text{发表}}$ KDD 会议”, 记为 $p \in \mathcal{P}$, 表示作者甲撰写的论文 1 发表在 KDD 会议上.

元图^[47]. 元图(或元结构)是定义在网络模式 $S_G = (\mathcal{C}, \mathcal{R})$ 上的有向无环图 $M_G = (\mathcal{C}_M, \mathcal{R}_M)$, 其中 $\mathcal{C}_M \subseteq \mathcal{C}$, $\mathcal{R}_M \subseteq \mathcal{R}$. 元图仅具有单个源实体类型 C_s (即入度为 0) 和单个目标实体类型 C_t (即出度为 0). $C_s = C_t$ 的元图 M_G 称为递归元图, 可以通过尾一头级联递归扩展.

图 2(a) 是一个由科技文献数据构成的典型异

质信息网络实例, 其中有四种类型的对象和四种类型的边. 图 2(b) 是该网络的网络模式, 描述了四种对象类型(会议 V 、论文 P 、作者 A 、主题 T) 和相应的关系(撰写/被撰写、出版/被出版、引用/被引用、提及/被提及). 图 2(c) 是三条元路径, 分别简记为“APA”、“APTP”和“APVPA”. APA (作者—论文—作者) 表示两个作者合著了一篇论文; APVPA (作者—论文—会议—论文—作者) 表示两个作者在同一会议上发表了论文. 图 2(d) 显示了文献网络的一个元图, 描述了三个作者之间的合著和两个作者在同一会议发表论文的关系.

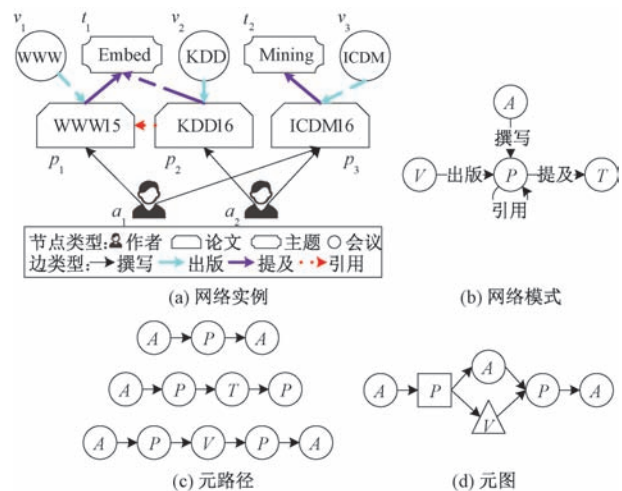


图 2 建模文献数据的异质信息网络^[48,56]

元路径刻画了对象之间的语义关系, 链接两个对象的不同元路径, 表示不同的语义关系. 一个元图可以同时建模多种关系, 因此保留了更完整的结构并捕获了更精确的节点特征, 具有更丰富的语义信息^[55].

网络表征学习常使用一阶邻近性和高阶邻近性表示节点间邻域结构的相似程度, 使用属性相似性表示节点间属性特征的相似程度.

一阶邻近性^[25]. 给定网络 $G = (\mathcal{N}, \epsilon)$, 节点 $v_i \in \mathcal{N}$ 和 $v_j \in \mathcal{N}$ 之间的一阶邻近性定义为 v_i 和 v_j 之间边上的权重, 即 $s_{ij}^{(1)} = w_{ij}$. 权重越大, v_i 和 v_j 之间的邻近程度越高.

高阶邻近性^[25]. 给定网络 $G = (\mathcal{N}, \epsilon)$, 设 $\mathbf{A}$ 为网络 G 的邻接矩阵, $\hat{\mathbf{A}}^t$ 是从 $\mathbf{A}$ 的行归一化获得的 k 步概率转移矩阵, 即 $\hat{\mathbf{A}}^k = \hat{\mathbf{A}}^1 \dots \hat{\mathbf{A}}^1$, $\mathbf{M} = \hat{\mathbf{A}}^1 + \hat{\mathbf{A}}^2 + \dots + \hat{\mathbf{A}}^k$, 则节点 $v_i \in \mathcal{N}$ 和 $v_j \in \mathcal{N}$ 之间的 k 阶邻近性定义为 $\mathbf{M}$ 的行向量 $\mathbf{M}_i$ 和 $\mathbf{M}_j$ 之间的相似性, 即 $s_{ij}^{(k)} =$

$\text{sim}(\mathbf{M}_i, \mathbf{M}_j)$, 其中相似性函数 $\text{sim}()$ 可以是欧几里得距离或余弦相似度.

属性相似性^[25]. 给定属性网络 $G = (\mathcal{N}, \epsilon, \mathbf{F})$, 其中 $\mathbf{F}$ 是网络中节点的属性特征矩阵, 节点 $v_i \in \mathcal{N}$ 和 $v_j \in \mathcal{N}$ 之间的属性相似性定义为 $\mathbf{F}$ 的行向量 $\mathbf{F}_i$ 和 $\mathbf{F}_j$ 之间的相似性, 即 $s_{ij}^{(k)} = \text{sim}(\mathbf{F}_i, \mathbf{F}_j)$.

2.2 网络类型

信息网络可以分为同质信息网络(简称同质网络, HoN)和异质信息网络两大类, 其中同质信息网络中所有节点和边都具有相同的类型, 即 $|\mathcal{C}| = 1$, $|\mathcal{R}| = 1$. 如果一个同质网络中每个节点都带有描述其性质的属性, 则称该网络为属性同质网络(AHoN). 异质信息网络的异质性可能源于网络中节点和/或边的不同类型, 也可能源于节点的多种属性, 或者源于多个网络, 因此, 异质信息网络除了用一个网络表示, 还可以用多个网络表示. 网络中的节点可能有属性, 也可能没有属性. 各种异质网络定义如下.

结构异质网络(SHeN). 只用一个网络 $G = (\mathcal{N}, \epsilon)$ 表示, 节点和/或边具有不同类型.

属性异质网络(AHeN)^[49]. 只用一个网络表示, 节点和/或边具有不同类型, 同时每个节点都有描述其属性的特征, 表示为 $G = (\mathcal{N}, \epsilon, \mathbf{F})$. 其中属性特征矩阵 $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_i, \dots, \mathbf{F}_{|\mathcal{N}|}]^T$ 将所有类型节点的属性连接在一起, $\mathbf{F}$ 的行向量 $\mathbf{F}_i$ 表示第 i 个节点 $v_i \in \mathcal{N}$ 的属性特征. 节点属性在网络结构的形成中起着重要作用.

例如, 在图 3 中 $G = (\mathcal{N}, \epsilon, \mathbf{F})$ 中 $\mathcal{N} = \{a_1, a_2, a_3, a_4, p_1, p_2, p_3, p_4, v_1, v_2\}$, $\epsilon = \{a_1 p_1, p_1 v_1, \dots\}$, $\phi: \mathcal{N} \rightarrow \{A, P, V\}$, $\varphi: \epsilon \rightarrow \{\text{撰写}, \text{发表}, \text{引用}\}$, $\mathbf{F} = \{\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_{10}\}^T$. 在节点属性中, $\{f_1, f_2\}$ 表示作者属性: 隶属关系和研究兴趣; $\{f_3, f_4\}$ 表示论文属性: 论文标题和摘要; $\{f_5, f_6\}$ 表示会议属性: 会议名称和介绍. 每个 f_i 的维度由用户定义. 每个节点由包含六种属性($f_1, \dots, f_6$)的向量 $\mathbf{F}_i$ 表示, 比如表示作者 a_1 的向量 $\mathbf{F}_1$ 中, 属性 f_1, f_2 的值由作者 a_1 给出, 而属性 f_3, f_4, f_5, f_6 设置为零.

多层网络(MLN)^[57]. 多层网络(也称为耦合异质网络)由多个不同但相关的子网络组成, 这些子网络通过网络间的边相连. “不同”是指各个子网的节点类型不同, “相关”是指不同子网的节点之间具有特定类型的交互或关系(跨层网络依赖). 不同层上的节点来自异质数据源, 即使以不同的方式呈现, 也可以相互影响, 并非彼此独立. 跨层网络依赖在某种

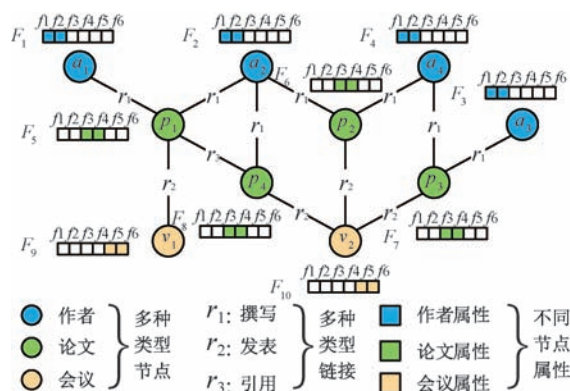


图3 属性异质信息网络^[49]

程度上蕴含了节点邻近度, 对网络内的边提供补充信息, 使隐特征更加全面和准确, 有效缓解由于网络内的边缺失带来的冷启动问题, 对于理解整个系统至关重要.

图4是表示关键基础设施系统的三层网络, 其中, 电网中的电站用于向AS网络中的路由器和运输网络中的车辆供电, AS网络提供通信机制以保持电网和运输网络的正常运转, 运输网络为发电厂供应燃料.

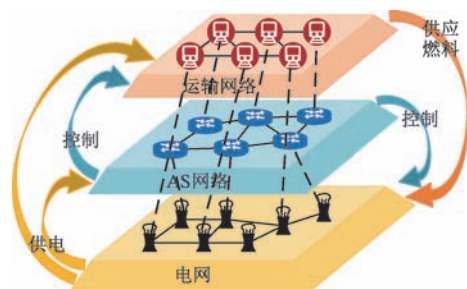


图4 多层网络^[57]

多视图网络(MVN)^[58]. 多视图网络包含多个网络或多个视图, 每个视图对应一种类型的边, 不同视图中的节点集可以相同, 也可以不同. 同一视图中的节点的类型可以相同(同质视图, homo-view), 也可以不同(异质视图, heter-view). 视图之间可以有共享节点, 不同视图的边集没有交集. 包含 $|\mathcal{V}|$ 个视图的网络记为 $G = (\mathcal{N}^{(v)}, \epsilon^{(v)})_{v \in \mathcal{V}}$, 其中 $\epsilon^{(v)}$ 包含视图 $v \in \mathcal{V}$ 中的所有边.

图5是一个三视图的例子, (a)是所有视图中的节点集都相同的同质视图, 其中每个节点代表一个人, 三个视图分别对应三种人际关系. 同事在职业视图中链接, 朋友在友谊视图中链接, 亲戚在亲属关系视图中链接; (b)是不同视图有不同的节点集, 但同一视图中节点类型相同的同质视图; (c)是不同视图

有不同的节点集,且同一视图中节点类型也不不同的异质视图。图中实线代表具体存在的边,虚线表示不同视图中节点的对应关系。

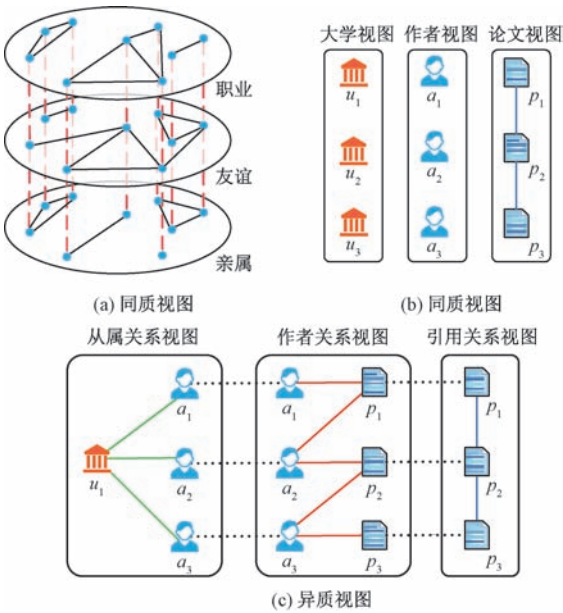


图 5 三视图网络^[58]

多重异质网络(MHeN)^[59]. 多重异质网络是一对节点之间可以有多种链接类型($|\mathcal{R}| > 1$)的异质网络,也称多关系网络,表示为 $G = (\mathcal{N}, \epsilon)$, $\epsilon = \bigcup_{r \in \mathcal{R}} \epsilon_r$, 其中 ϵ_r 是类型为 $r \in \mathcal{R}$ 的边集. 由类型为 $r \in \mathcal{R}$ 的边连接的节点构成网络 $G_r = (\mathcal{N}_r, \epsilon_r)$, $\mathcal{N}_r \subseteq \mathcal{N}$. G_r 表示节点间的一种交互,体现了网络一种独特的性质,不同的 G_r 相互补充。

属性多重异质网络(AMHeN)^[59]. 属性多重异质网络是节点具有属性的多重异质网络,表示为 $G = (\mathcal{N}, \epsilon, \mathbf{F})$. 由类型为 $r \in \mathcal{R}$ 的边连接的节点构成网络 $G_r = (\mathcal{N}_r, \epsilon_r, \mathbf{F})$, $\mathcal{N}_r \subseteq \mathcal{N}, \epsilon_r \subseteq \epsilon$.

例如,图 6 中包含 2 种节点类型(“用户”和“项”),4 种链接类型(点击、收藏、加入购物车、下单). 用户的属性包括性别、年龄、位置等,项的属性包括价格、品牌等。

多分辨率多网络($M_R M_{IN}$)^[60]. $M_R M_{IN}$ 是指从多个输入网络中抽取特定类型的子图(比如 Weisfeiler-Lehman 子树^[61]),然后根据节点—子图、子图—网络的隶属关系和子图间的相似度将分别表示对象、子图和网络的节点连接在一起,构成一个跨分辨率跨网络(CRCN)的关系网络,使来自不同网络不同分辨率的对象能够在彼此的上下文中。

图 7 示例了一个包含三个分子网络的 $M_R M_{IN}$, 图(a)中“enzyme 33”和“enzyme 54”属于 enzyme 的

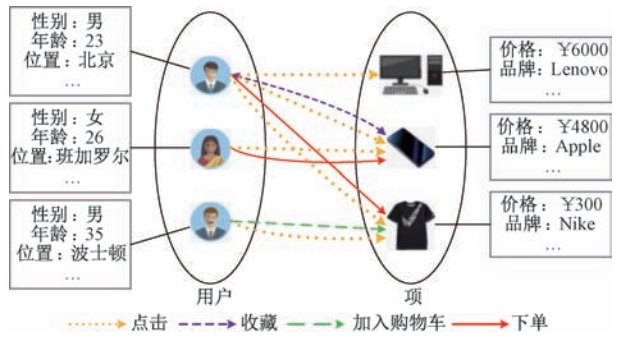


图 6 包含用户和项的属性多重异质网络^[59]

同一类别,(b1)代表 CRCN 关系网络,圆代表(a)中的节点,正方形代表子图,六边形代表网络,实线表示节点—子图、子图—网络的成员关系,虚线表示相似子图. 编号为 31 和 34 的正方形是从“enzyme 33”和“enzyme 54”中提取的两个子图(WL 子树),这两个子图结构相似(只有一个节点不同,图(b2)),因此正方形 31 和 34 由虚线连接;“enzyme 33”包含子图 31,子图 31 包含节点 16 和 17,因此正方形 31 分别与代表“enzyme 33”的六边形和圆 16、17 相连。

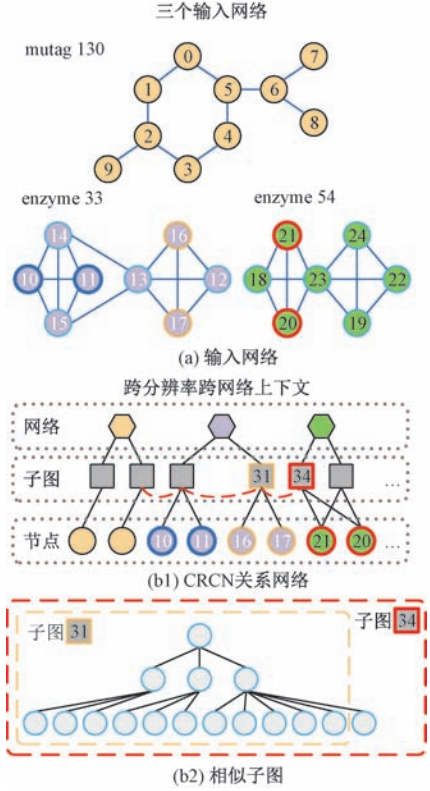


图 7 多分辨率多网络^[60]

CRCN 将多个网络中的节点、子图及网络关联在一起,从三个分辨率层次描述三者之间的关系,从而每个网络中的每个节点都可能成为其他网络节点

的上下文,彼此提供互补的信息.

异质特征网络(HeFN)^[62]. 异质特征网络是指一个同质网络中的节点具有从多个视图收集的一组特征. 令 $\{\mathbf{F}^{(i)}\}, i=1,2,\cdots,I$ 是从 I 个不同的视图为 N 个实例收集的一组相关特征矩阵,其中最后一个特征矩阵 $\mathbf{F}^{(I)}=G$ 为描述 N 个实例之间关系的加权邻接矩阵. 比如亚马逊上的产品有产品信息和客户评论等多种描述,这些描述相互补充,可用于构建 $\{\mathbf{F}^{(i)}\}$;而客户的购买记录可用于构建 G .

异质特征网络 HeFN 与属性异质网络 AHeN 和属性多重异质网络 AMHeN 的差异在于,AHeN

关注单个异质网络和单个特征矩阵,HeFN 致力于单个同质网络和多个特征矩阵,而 AMHeN 涉及多个异质网络和单个特征矩阵.

动态异质网络(DHeN)^[63]. 动态异质网络是带有时间戳的异质网络的集合,表示为 $G=(\mathcal{N},\epsilon,\mathcal{T})$,其中 $\mathcal{T}=\{1,2,\cdots,z\}$ 是时间戳的集合, $G^t=(\mathcal{N}^t,\epsilon^t)(t\in\mathcal{T})$ 表示时间戳为 t 时的静态异质网络, G^t 和 G^{t+1} 中节点的数目和边的数目可能不相等,即 $|\mathcal{N}^t|\neq|\mathcal{N}^{t+1}|,|\epsilon^t|\neq|\epsilon^{t+1}|$. 在所有时间戳 t 当中,节点的类型和边的类型保持不变.

表 1 总结了目前网络表征学习的网络类型.

表 1 网络表征学习的网络类型

网络类型	节点类型	边类型	输入网络	输入特征矩阵	分辨率	静态/动态
同质网络(HoN)	单种	单种	单个	无	单分	静态
属性 HoN(AHoN)	单种	单种	单个	单个	单分	静态
结构 HeN(SHeN)	多种	一种或多种,但每对节点之间只有一种类型的边	单个	无	单分	静态
属性 HeN(AHeN) ^[49]	多种	一种或多种,但每对节点之间只有一种类型的边	单个	单个	单分	静态
多层网络(MLN) ^[57]	多种	多种	多个	无	单分	静态
多视图网络(MVN) ^[58]	单种或多种	多种	多个	无	单分	静态
多重异质网络(MHeN) ^[59]	多种	多种,且每对节点之间有多种类型的边	多个	无	单分	静态
属性 MHeN(AMHeN) ^[59]	多种	多种,且每对节点之间有多种类型的边	多个	单个	单分	静态
多分辨率多网络($M_R M_{IN}$) ^[60]	单种	单种	多个	无	多分	静态
异质特征网络(HeFN) ^[62]	单种	单种	单个	多个	单分	静态
动态异质网络(DHeN) ^[63]	多种	多种,但每对节点之间只有一种类型的边	多个	无	单分	动态

2.3 信息网络表征学习

给定信息网络 $G=(\mathcal{N},\epsilon)$, 表征学习的目标是学习网络中节点的低维向量 $\mathbf{H}\in\mathbb{R}^{|\mathcal{N}|\times d}$,同时保留原始网络中节点之间的结构和语义相关性(比如一阶邻近性和高阶邻近性),其中 d 是嵌入维度, $|\mathcal{N}|$ 表示节点数目, $d\ll|\mathcal{N}|$. 对于属性网络, $\mathbf{H}$ 还需要保留节点属性间的相似性.

表征学习也可以学习节点、子图或网络的低维向量,比如多分辨率多网络($M_R M_{IN}$ ^[60])同时从节点、子图和网络三个互补的分辨率学习来自不同网络的节点、子图和网络. 图 8 是图 7 所示多分辨率多网络的嵌入结果. 可以看出在嵌入空间中,属于同一类别的深底黑字网络(“enzyme 54”)和深底白字网络(“enzyme 33”)彼此接近,两者都远离浅底黑字网络(“mutag 130”);源于不同网络但结构相似的子图 31 和 34 接近;属于同一子图的节点 10 和 11 靠近;来自不同网络但连接相似子图的节点 16 和 21 也靠近. 嵌入结果体现了节点、子图和网络三种不同分辨率对象间的相关性.

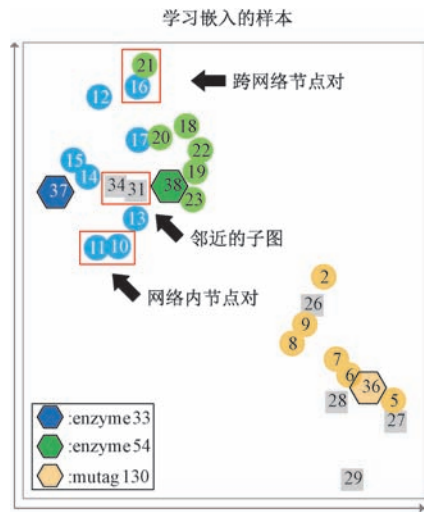


图 8 图 7 所示 $M_R M_{IN}$ 的嵌入结果^[60]

石川与孙怡舟将异质信息网络表征学习的方法分为基于随机游走、基于分解和基于深度神经网络的方法^[48],但是许多现有的模型交叉使用了这三种方法,因此本文首先根据时变特性将网络分为静态网络和动态网络,再根据网络数目将静态网络分为单网络和多网络,然后根据学习特点将单网络的表征学习分为基于概率、距离、交互路径和

图卷积的学习,根据网络特性将多网络的表征学习分为多层、多视角、多重、新兴、多分辨率网络的学习.本文第3、4和5节将按照上述的分类介绍典型的表征学习方法.

2.4 常用技术

2.4.1 随机游走

随机游走是一种经典的图分析模型,常用于刻画网络中节点间的可达性,因此也被广泛应用于网络表征学习中,用于表达采样节点间的邻域关系.在同质信息网络中,节点类型单一,游走可以沿任意的路径进行;而在异质信息网络中,由于节点/边的类型不同,游走可以对节点序列施加类型约束,确保将不同类型节点之间的语义关系融合到嵌入模型中,更好地抽取网络中的结构信息.针对描述节点属性的特征矩阵,还可以根据特征间的相似性进行游走.因此,异质信息网络中的随机游走可以分为结构游走和特征游走,前者捕捉节点间的结构邻近性,后者捕捉节点间的属性邻近性.

结构游走.结构游走基于网络的拓扑结构获取节点序列,包括基于元路径、元图或网络模式的随机游走.基于元路径的随机游走以预定义的元路径 $\mathcal{P} = C_1 \xrightarrow{R_1} C_2 \xrightarrow{R_2} \dots \xrightarrow{R_L} C_{L+1}$ 为游走约束条件,第 i 步的转移概率定义为

$$P_{C_i, C_{i+1}}(u, v | \mathcal{P}) = \begin{cases} \frac{1}{|\mathcal{N}_{C_{i+1}}(u)|}, & (u, v) \in \epsilon, \phi(u) = C_i, \phi(v) = C_{i+1} \\ 0, & \text{其他} \end{cases} \quad (1)$$

其中 C_i 是节点 $u \in \mathcal{N}$ 的类型, $\mathcal{N}_{C_{i+1}}(u)$ 表示类型为 C_{i+1} 的节点 u 的邻域. $P_{C_i, C_{i+1}}(u, v | \mathcal{P})$ 表示从类型为 C_i 的节点 u 执行随机游走时,选择下一个类型为 C_{i+1} 的节点 $v \in \mathcal{N}$ 的概率.若有多个节点符合条件,则随机选择其中一个.

如果节点 u 和 v 之间有多种类型的边,则元路径的选择及节点 u 的邻域均与边的类型 $r \in \mathcal{R}$ 有关. GATNE^[59] 定义了多重异质网络中随机游走的转移概率.

为了覆盖网络中的所有节点,随机游走应确保足够的游走次数和游走长度. HeteSpaceyWalk^[56] 提出了异质个性化空间随机游走,允许在游走过程中跳过冗余节点以提高游走效率.比如在 DBLP 网络中,元路径“APVPA”的实例 $a_1 \rightarrow p_1 \rightarrow v_1 \rightarrow p_1 \rightarrow a_2$ 表达的意思与“APA”的实例 $a_1 \rightarrow p_1 \rightarrow a_2$

表达的意思相同,所以 $p_1 \rightarrow v_1 \rightarrow p_1$ 中的后两个节点 v_1 和 p_1 冗余,可以跳过.

元图由于集成了多条元路径,因此可以捕获更复杂的异质结构和更丰富的语义.在基于元图的异质个性化空间随机游走^[56]的过程中,遇到分支节点时不是随机选择下一个节点类型,而是先以概率 α 忘记游走序列的倒数第二个节点,然后分别以概率 β 和 γ 选择下一个节点类型 C_{n+1} 和节点.概率 β 和 γ 可以通过当前节点的后继节点类型集合以及元图的集成转移超矩阵计算.

尽管元路径和元图能够有效捕获 HeN 中的异质结构和语义信息,但是设计合适的元路径或元图非常困难.基于网络模式的异质个性化空间随机游走^[56]降低了用户设计合适元路径或元图的困难,但也存在分支选择问题,需要确定选择下一个节点类型 C_{n+1} 和节点的概率 β 和 γ .

特征游走.特征游走^[62]基于描述节点属性的特征矩阵获得节点序列.令 $\{\mathbf{X}^{(i)}, i = 1, 2, \dots, I\}$ 是从 I 个不同的视图为 $\mathcal{N}$ 个实例收集的一组特征矩阵,矩阵 $\mathbf{X}^{(i)}$ 中的第 j 行 $\mathbf{x}_j^{(i)} = (x_{j1}^{(i)}, x_{j2}^{(i)}, \dots, x_{jm}^{(i)})$ 表示实例 j 的 m 个特征值.首先使用 L_2 范数对 $\{\mathbf{X}^{(i)}\}$ 的每一行进行归一化得到 $\{\bar{\mathbf{X}}^{(i)}\}$,删除其中值小于 $\beta \text{Mean}(\bar{\mathbf{X}}^{(i)})$ 的元素(因为随机游走很难模拟小值概率)得到 $\{\hat{\mathbf{X}}^{(i)}\}$,其中 β 是阈值, $\text{Mean}(\bar{\mathbf{X}}^{(i)})$ 表示 $\bar{\mathbf{X}}^{(i)}$ 中所有元素的平均值;接着使用 L_1 范数规范化 $\hat{\mathbf{X}}^{(i)}$ 的每一行,得到矩阵 $\mathbf{Y}^{(i)}$,即 $y_{st}^{(i)} = \frac{\hat{x}_{st}^{(i)}}{\sum_{r=1}^m \hat{x}_{sr}^{(i)}}$;然后从实例 j 开始游走,以概率 $P(j \rightarrow f_t^{(i)}) = y_{jt}^{(i)} (t=1, 2, \dots, m)$ 从 m 个特征中选择特征 $f_t^{(i)}$,并以概率 $P(f_t^{(i)} \rightarrow k) = \frac{y_{kt}^{(i)}}{\sum_{s=1}^N y_{st}^{(i)}}$ 选择实例 k ,完成从实例 j 到实例 k 的游走.游走过程示于图 9.

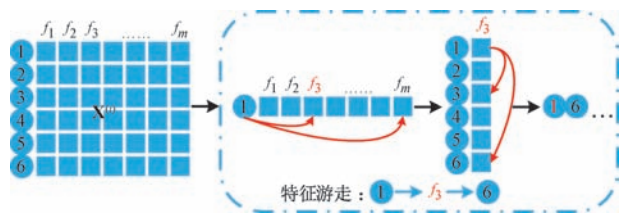


图9 特征游走^[62]

特征游走不直接计算任何实例对之间的相似度,有效缓解了大规模网络中相似度计算时间、空间复杂度高的问题,并且各个特征矩阵上的游走可以

按分布式的方式进行,使游走具有可扩展性.尽管各个特征矩阵异质,但是在各个特征矩阵上行走获得的均是同质的节点序列,可以与结构游走获得的节点序列一同处理,自然融合了网络的结构和节点的多重属性信息.

2.4.2 负采样

表征学习过程中,应尽量使每个中心节点与其邻居彼此靠近(嵌入向量相似)并远离所有其他节点.其他节点很多,为了减少计算成本,负采样(NS)^[64]随机采样少量非邻居节点(负样本),中心节点只需要远离负样本即可.在 HeN 中,负样本与中心节点的邻居具有相同的类型,比如,在图 10 中,当 v_1 为中心节点时, v_1 的 P 型邻居为 $\mathcal{N}_P(v_1) = \{p_1, p_4\}$, A 型邻居为 $\mathcal{N}_A(v_1) = \{a_1, a_2\}$,相应的负采样为 $\{p_2, p_3\}$ 和 $\{a_3, a_4\}$.

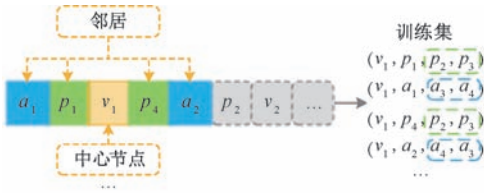


图 10 负采样^[49]

许多算法在使用 NS 时,负样本是从所有样本 $\mathcal{N}$ 中随机选择,因此节点的邻居也可能被选择为负样本,这会带来“流行邻居问题”,即度高的节点比度低的节点被选择的可能性大,导致度高节点的嵌入过度收缩,效果不佳.对于邻居节点的度也很高的节点,情况更糟.为此,鲁棒的负采样(R-NS)^[65]从集合 $\mathcal{N} - \{\mathcal{N}(u) \cup \{u\}\}$ 中为每个节点 u 选择负样本,并且对度较高的节点施加更大的惩罚,从而在不改变算法复杂性的情况下,有效避免“流行邻居问题”. HeGAN^[66] 和 AND-HeA^[67] 在博弈中同时训练具有关系感知能力的判别器和生成器,并且生成器通过学习节点分布直接从连续分布中采样“潜在”节点以生成更好的负样本.

3 单网络的表征学习

本节介绍只涉及一个网络的异质信息网络的表征学习.

3.1 基于概率的表征

3.1.1 “节点共现”概率

基于“节点共现”概率的嵌入(Skip-Gram 模型)的思想是使用中心节点来预测由随机游走路径给出

的上下文节点,目标是基于中心节点的嵌入最大化观察到的邻居节点(上下文)的概率.

Metapath2vec^[68]: Metapath2vec 基于网络结构完成嵌入.设基于元路径 $\mathcal{P}$ 的随机游走生成节点序列集 $\mathcal{V}_{\mathcal{P}}$, 上下文窗口大小为 w , $\mathcal{N}_i(u) \subseteq \mathcal{N}$ 为中心节点 $u \in \mathcal{V}_{\mathcal{P}}$ 的 w 窗口内类型为 $C_i \in \mathcal{C}$ 的节点集(邻域), $\mathbf{H}$ 是节点的嵌入矩阵, $\mathbf{h}_u$ 和 $\mathbf{h}_v$ 分别是节点 u 和 v 的嵌入向量,则 v 是 u 的邻居的条件概率定义为

$$P(v | u; \mathbf{H}) = \frac{\exp\{\mathbf{h}_v^T \mathbf{h}_u\}}{\sum_{v' \in \mathcal{N}_i(u)} \exp\{\mathbf{h}_{v'}^T \mathbf{h}_u\}} \quad (2)$$

Metapath2vec 的目标函数定义为

$$\arg \min_{\mathbf{H}} \left(- \sum_u \sum_{C_i \in \mathcal{C}} \sum_{v \in \mathcal{N}_i(u)} \log P(v | u; \mathbf{H}) \right) \quad (3)$$

使用负采样技术,式(3)可以修改为

$$\mathcal{L} = \log \sigma(\mathbf{h}_v^T \mathbf{h}_u) + \sum_{m=1}^M \mathbb{E}_{n \sim P_{neg}(u)} [\log \sigma(-\mathbf{h}_n^T \mathbf{h}_u)] \quad (4)$$

其中 $\sigma(x) = \frac{1}{1 + \exp(-x)}$, M 是负样本的采样次数, $P_{neg}(u)$ 是负样本 $n \notin \mathcal{N}_i(u)$ 的采样分布, n 的类型与 v 的类型相同. $P_{neg}(u)$ 通常根据节点在游走过程中出现的频率进行计算,即 $P_{neg}(u) = \frac{counter(u)^{\frac{3}{4}}}{\sum_{v' \in \mathcal{N}} counter(v')^{\frac{3}{4}}}$, 其中分子表示节点 u 在所有随机游走中出现的频率,而分母表示所有节点出现频率的总和.

Metapath2Vec++ 考虑了节点的类型,按照上下文 v 的节点类型 $\phi(v) = C_i$ 对式(2)进行改进,即 $P(v | u; \mathbf{H})$ 的计算要针对特定的节点类型 $\phi(v) = C_i$, 所以

$$P(v | u; \mathbf{H}) = \frac{\exp\{\mathbf{h}_v^T \mathbf{h}_u\}}{\sum_{v' \in \mathcal{N}_i(u)} \exp\{\mathbf{h}_{v'}^T \mathbf{h}_u\}} \quad (5)$$

其中 $\mathcal{N}_i$ 是网络中类型为 C_i 的节点集.由此,采样分布也要根据预测邻居 v 的类型 C_i 指定.因此,目标函数为

$$\mathcal{L} = \log \sigma(\mathbf{h}_v^T \mathbf{h}_u) + \sum_{m=1}^M \mathbb{E}_{n \sim P_{neg}(v')} [\log \sigma(-\mathbf{h}_n^T \mathbf{h}_u)] \quad (6)$$

其中 v' 表示类型为 C_i 的样本. $\mathcal{L}$ 使用随机梯度下降算法进行优化.

metapath2vec 和 metapath2vec++ 的框架如图 11 所示,在图 11(b)中,输出层为 a_4 邻居节点的所有类型分配一个多项式分布集合,而在图 11(c)

中,输出层为 a_i 邻居中每种类型的节点指定一个多项式分布集合. $\mathcal{N}_i$ 表示类型 C_i 的节点集, $\mathcal{N} = \mathcal{N}_V \cup \mathcal{N}_A \cup \mathcal{N}_O \cup \mathcal{N}_P$, k_i 指定了 C_i 类型邻居的数目, $k = k_V + k_A + k_O + k_P$.

AHNG^[49]:在属性异质网络中,针对节点复杂和属性多样会使节点的表示形式不确定的问题,

AHNG^[49]通过两层全连接神经网络对各种属性进行编码,并通过基于元路径的随机游走来捕捉 AHeN 的结构和语义信息,节点 $u \in \mathcal{N}$ 使用高斯分布 $\mathbf{h}_u = \mathcal{N}(\boldsymbol{\mu}_u, \Sigma_u)$ 表示,用于捕获节点表示的不确定性,目标是最大化预测中心节点邻域的概率. AHNG 的框架如图 12 所示.

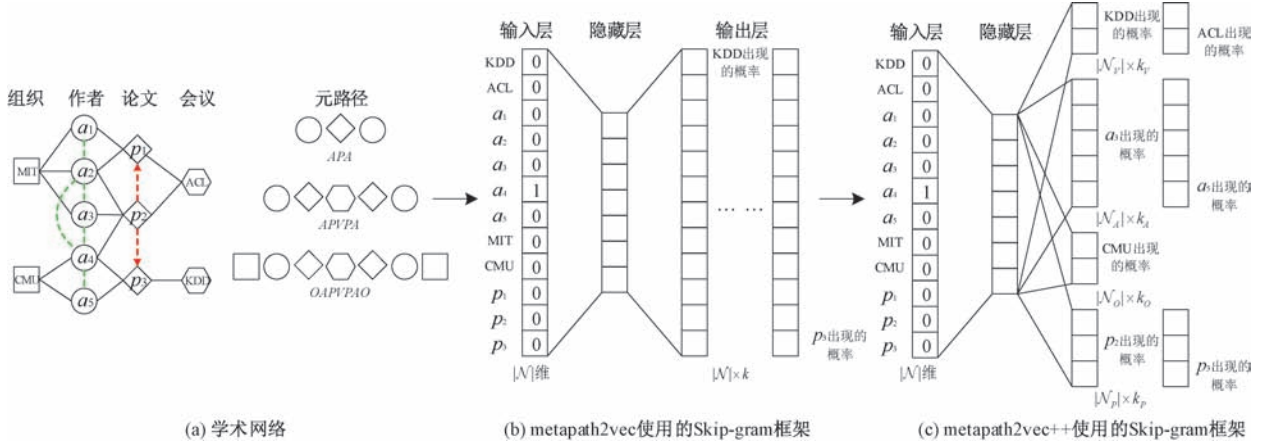


图 11 metapath2vec 和 metapath2vec++ 中的 Skip-gram 框架^[68]

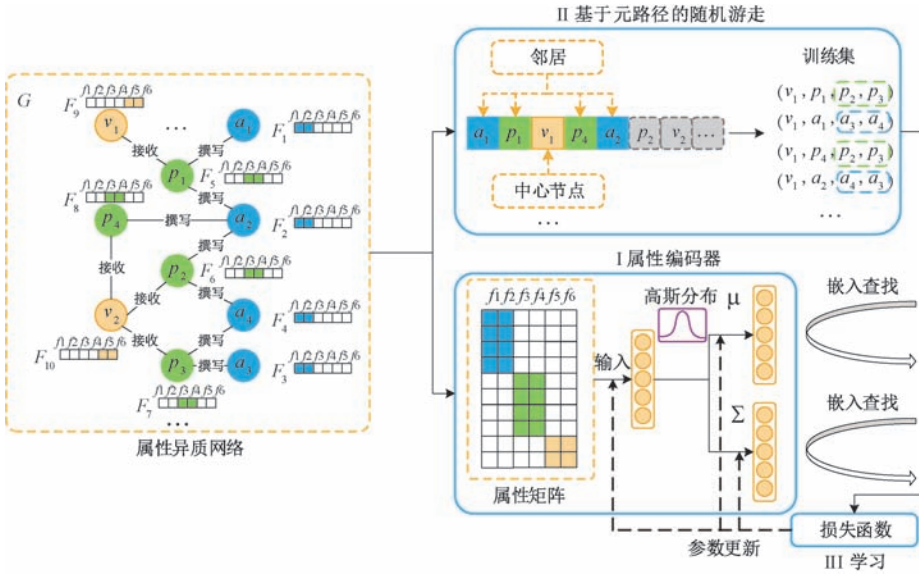


图 12 AHNG 的框架^[49]

给定网络 $G = (\mathcal{N}, \epsilon, \mathbf{F})$, h 是一个由两层神经网络构成的属性编码器,其输入是节点 u 的特征向量 $\mathbf{f}_u \in \mathbf{F}$,输出是 u 的均值 $\boldsymbol{\mu}_u$ 和协方差 Σ_u . 设 $\mathbf{h}_u$ 和 $\mathbf{h}_v$ 分别表示中心节点 u 和它的邻域 $v \in \mathcal{N}_i(u)$ 的嵌入,即 $\mathbf{h}_u = h(u)$, $\mathbf{h}_v = h(v)$, $\mathcal{N}_i(u)$ 是出现在 u 的 w 窗口内类型为 C_i ($C_i \in \mathcal{C}$) 的邻域,定义 $P(v | h(u)) = \frac{\exp[\kappa(\mathbf{h}_u, \mathbf{h}_v)]}{\sum_{v' \in \mathcal{A}} \exp[\kappa(\mathbf{h}_u, \mathbf{h}_{v'})]}$, 其中

$$\kappa(\mathbf{h}_u, \mathbf{h}_v) = \frac{1}{2} \left(\log \frac{|\Sigma_v|}{|\Sigma_u|} - l + \text{Tr}(\Sigma_u^{-1} \Sigma_v) + (\boldsymbol{\mu}_u - \boldsymbol{\mu}_v)^T \Sigma_u^{-1} (\boldsymbol{\mu}_u - \boldsymbol{\mu}_v) \right)$$

l 是路径长度,则 AHNG 的目标函数定义为

$$\mathcal{L} = \log \sigma[\kappa(\mathbf{h}_u, \mathbf{h}_v)] + \sum_{k=1}^K \mathbb{E}_{n \sim P_{\text{neg}}(v')} (\log \sigma[-\kappa(\mathbf{h}_u, \mathbf{h}_n)])$$

(7)

$\mathcal{L}$ 使用梯度下降算法进行优化.

3.1.2 “关系成立”概率

基于“关系成立”概率完成嵌入的思想是使用节点来预测某种关系是否成立,目标是基于节点、关系的嵌入最大化正样本的概率。

HIN2Vec^[69]: HIN2Vec 使用图 13 所示的全连接神经网络模型通过预测节点之间的关系(二元分类器)来学习 HeN 中节点和元路径的潜在向量. 给定网络 $G = (\mathcal{N}, \epsilon)$ 和以元路径形式指定的关系集 $\mathcal{R}$, 模型采用一对节点 $u \in \mathcal{N}$ 和 $v \in \mathcal{N}$, 以及关系 $r \in \mathcal{R}$ 作为输入, 以预测 u 和 v 之间是否存在关系 r . 输入层接收的三个 one-hot 向量 u, v 和 r 被转换为隐层中的隐向量 $W_u^T u, W_v^T v$ 和 $f_{01}(W_R^T r)$. 由于节点和关系的语义和含义不同, 因此对关系 r 的处理不同于对节点 u 和 v 的处理. $f_{01}(\cdot)$ 是正则函数(比如 sigmoid 函数), 用于限制 r 的隐向量 $W_R^T r$ 的值在 0 到 1 之间. 向量 u, v 和 r 的聚合采用 Hadamard 函数完成, 输出层通过 $P(r | u, v) = \text{sigmoid}(\sum W_u^T u \odot W_v^T v \odot f_{01}(W_R^T r))$ 实现逻辑分类. 如果 W_u 和 W_v 相同(不强调关系中输入节点的个体角色), 则 W_u 和 W_R 分别包含了学到的节点向量和元路径向量。

HIN2Vec 采用随机游走的方式生成节点序列, 进而产生训练样本集 $\mathcal{Q}$. $\mathcal{Q}$ 中的样本形式为 $\langle u, v, r, q_{uv} \rangle$, 若 u 和 v 之间存在关系 r , 则 $q_{uv} = 1$, 否则 $q_{uv} = 0$. HIN2Vec 的目标函数为

$$\mathcal{L} = \sum_{u, v, r \in \mathcal{Q}} \{q_{uv} \log P(r | u, v) + [1 - q_{uv}] \log [1 - P(r | u, v)]\} \quad (8)$$

$\mathcal{L}$ 使用随机梯度下降算法最大化, W_u, W_v 和 W_R 通过反向传播算法调整。

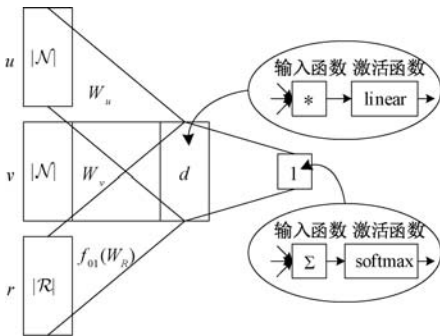


图 13 HIN2Vec 表征学习模型^[69]

M-HIN^[70]: M-HIN 基于节点和元图之间的关系学习节点的复嵌入, 对称元图的嵌入是实数, 非对称元图的嵌入是纯虚数。

设 u 和 v 分别是元图 s 的首节点和尾节点, $\mathbf{h} \in \mathbb{C}^d, \mathbf{h} = \text{Re}(\mathbf{h}) + i\text{Im}(\mathbf{h})$, 其中 $\text{Re}(\mathbf{h}) \in \mathbb{R}^d$ 和 $\text{Im}(\mathbf{h}) \in \mathbb{R}^d$ 分别表示向量 $\mathbf{h} \in \mathbb{C}^d$ 的实部和虚部, $\bar{\mathbf{h}} \in \mathbb{C}^d$ 是 $\mathbf{h}$ 的复共轭, u 和 v 关于 s 的相关性定义为

$$\begin{aligned} f(u, v | s) &= \text{Re}(\langle \mathbf{h}_u, \mathbf{h}_s, \bar{\mathbf{h}}_v \rangle) \\ &= \langle \text{Re}(\mathbf{h}_u), \text{Re}(\mathbf{h}_s), \text{Re}(\mathbf{h}_v) \rangle + \langle \text{Im}(\mathbf{h}_u), \text{Re}(\mathbf{h}_s), \\ &\quad \text{Im}(\mathbf{h}_v) \rangle + \langle \text{Re}(\mathbf{h}_u), \text{Im}(\mathbf{h}_s), \text{Im}(\mathbf{h}_v) \rangle - \langle \text{Im}(\mathbf{h}_u), \\ &\quad \text{Im}(\mathbf{h}_s), \text{Re}(\mathbf{h}_v) \rangle \end{aligned} \quad (9)$$

$\langle \cdot \rangle$ 是逐元素多线性点积, 即 $\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle = \sum_k a_k b_k c_k$.

三元组 (u, s, v) 是否成立的概率记为 $P(s | u, v) = \sigma(f(u, v | s))$, 设 $q_{uv} \in \{0, 1\}$ 表示 (u, s, v) 是否是正样本. 若 $q_{uv} = 1, P(s | u, v)$ 最大化, 否则 $P(s | u, v)$ 最小化. 目标函数与 HIN2Vec 相似。

为了区分不同元图的重要性, Yang 等^[71] 设计了评估元图和选择有用嵌入维度的实用方法, 提出了具有 $L_{2,1}$ 损失的自编码^[72] 嵌入方法, 以端到端的无监督方式寻找期望的元图和嵌入。

TaPEm^[73]: TaPEm 不通过节点嵌入之间的相似性来计算两个节点之间成对关系的可能性, 而是基于上下文路径(节点对之间的节点序列)直接学习节点对(例如“论文—作者”)的嵌入, 预测 $\mathcal{N}$ 中任意两个节点之间成对关系的可能性。

TaPEm 首先在网络 $G = (\mathcal{N}, \epsilon)$ 上执行基于元路径的随机游走, 从中提取每对目标节点对 (v, u) 关联的上下文路径集 $\mathcal{C}_{v \rightarrow u}^p$, 然后完成节点对和上下文路径的嵌入. TaPEm 的框架如图 14 所示。

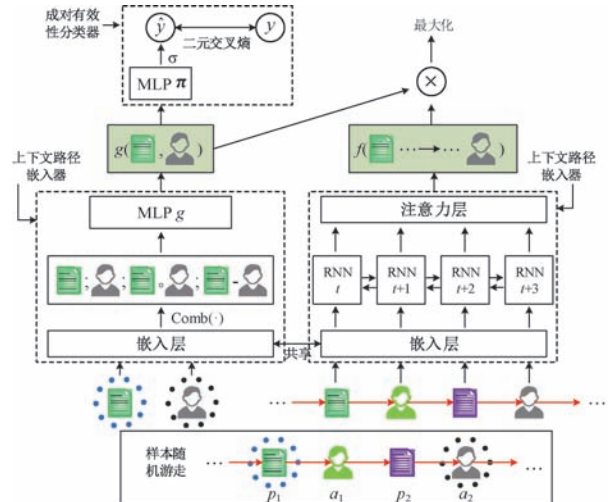


图 14 TaPEm 的框架^[73]

节点对 (v, u) 的嵌入通过一个 n 层的感知器 $g: \mathbb{R}^{4K} \rightarrow \mathbb{R}^d$ 完成,感知器的输入是 v 和 u 的嵌入向量 $p_v \in \mathbb{R}^K$ 和 $q_u \in \mathbb{R}^K$ 的组合,即 $h^{(0)} = \text{Comb}(p_v, q_u) = [p_v; q_u; p_v \circ q_u; p_v - q_u] \in \mathbb{R}^{4K}$,其中 $\circ$ 表示向量的对应元素相乘,感知器的输出为 $g(u, v) = h^{(n)} \in \mathbb{R}^d$, p_v 和 q_u 可以使用 v 和 u 的相关信息表示,比如使用论文摘要中的关键词表示论文。

上下文路径 $c \in \mathcal{C}_{v \rightarrow u}^{\mathcal{P}}$ 的嵌入通过双向门控单元^[74]和注意力组件完成.在时刻 t ,双向门控单元BiGRU根据前一时刻的隐状态 $\vec{h}_{t-1} \in \mathbb{R}^d, \overleftarrow{h}_{t-1} \in \mathbb{R}^d$ 和当前输入 $x_t \in \mathbb{R}^K$ 计算 $h_t = \text{GRU}(\vec{h}_{t-1}, \overleftarrow{h}_{t-1}, x_t)$,然后注意力组件度量上下文路径中各个节点的贡献 w_t ,其中路径嵌入定义为 $f(c) = \sum_t w_t W_{attn} h_t$.上下文路径嵌入的目标是最小化给定上下文路径 $c \in \mathcal{C}_{v \rightarrow u}^{\mathcal{P}}$ 时节点对 (v, u) 的负对数似然 $\mathcal{L}_{\text{ctx}}(v, u) = \sum_{c \in \mathcal{C}_{v \rightarrow u}^{\mathcal{P}}} -\log P((v, u) | c, \mathcal{P})$,似然概率定义为

$$P((v, u) | c, \mathcal{P}) = \frac{\exp[g(v, u) \cdot f(c)]}{\sum_{c' \in \mathcal{C}_*^{\mathcal{P}}} \exp[g(v, u) \cdot f(c')]} \quad (10)$$

其中 $\mathcal{C}_*^{\mathcal{P}}$ 是基于 $\mathcal{P}$ 的随机游走产生的上下文路径集.使用负采样,似然概率可以近似为

$$\log P((v, u) | c, \mathcal{P}) \approx \log \sigma[g(v, u) \cdot f(c)] + \sum_{m=1}^M \mathbb{E}_{n \sim P_{\text{neg}}(c)} \log \sigma[-g(v, u) \cdot f(n)] \quad (11)$$

图14中节点对验证分类器 $\pi: \mathbb{R}^d \rightarrow \mathbb{R}$ 由2层的感知器构成,用于区分节点对 (v, u) 是否是有效对,其二元交叉熵损失表示为

$$\mathcal{L}_{\text{pv}}(v, u) = q_{v, u} \sigma(\pi(g(v, u))) + (1 - q_{v, u})(1 - \sigma(\pi(g(v, u)))) \quad (12)$$

如果 (v, u) 是有效对,比如论文 v 是由作者 u 撰写,则 $q_{v, u} = 1$,否则 $q_{v, u} = 0$.

TaPEm的目标函数定义为

$$\mathcal{L} = \sum_{\mathcal{P}} \sum_{w \in \mathcal{W}^{\mathcal{P}}} \sum_{v \in w} \sum_{u \in w} [\mathcal{L}_{\text{ctx}}(v, u) + \mathcal{L}_{\text{pv}}(v, u)] \quad (13)$$

其中 $\mathcal{W}^{\mathcal{P}}$ 是基于 $\mathcal{P}$ 的随机游走集, τ 是节点 v 的上下文窗口, Z_v 表示 v 在游走 w 中的位置. $\mathcal{L}$ 通过小批量Adam优化器^[75]最小化.

3.1.3 “元路径实例”概率

基于“元路径实例”概率的嵌入模型在异质网络上基于元路径进行随机游走,通过最大化元路径实例的概率来学习出现在该实例中节点的向量表示.

ESim^[76]:给定网络 $G = (\mathcal{N}, \epsilon)$ 和元路径 $\mathcal{P}$,元组 $e = \langle u, v, r \rangle$ 表示连接节点 u 和 v 的类型为 $r \in \mathcal{R}$ 的边, $p_{e_1 \sim e_L} = \langle e_1 = \langle u_1, v_1, r_1 \rangle, \dots, e_L = \langle u_L, v_L, r_L \rangle \rangle$ 是 $\mathcal{P}$ 的一条路径实例, $p_{e_1 \sim e_L}$ 的概率定义为

$$P(p_{e_1 \sim e_L} | \mathcal{P}) = P(u_1 | \mathcal{P}) \times P(p_{e_1 \sim e_L} | u_1, \mathcal{P}) \propto B(u_1, 1 | \mathcal{P})^\gamma \times P(p_{e_1 \sim e_L} | u_1, \mathcal{P}) \quad (14)$$

其中 $B(u, i | \mathcal{P})$ 表示 $\mathcal{P}$ 上第 i 个节点是 u 的路径实例数目,参数 γ 通常取3/4.假设一条路径实例上的所有节点彼此高度相关,则

$$P(p_{e_1 \sim e_L} | u_1, \mathcal{P}) = \prod_{i=1}^L \prod_{j=i}^L P(v_j | u_i, \mathcal{P}_{i,j}) \quad (15)$$

嵌入的目标是 $\arg \max_{\Pi} \sum_{u \in \mathcal{N}} \sum_{v \in \mathcal{N}_i(u)} P(p_{e_1 \sim e_L} | \mathcal{P})$.

使用负采样技术,目标函数为

$$\mathcal{L} \approx \sum_{p_{e_1 \sim e_L}} \log \sigma \left(\sum_{i=1}^L \sum_{j=i}^L f(u_i, v_j | \mathcal{P}_{i,j}) \right) + \sum_{m=1}^M \mathbb{E}_{p_{e_1 \sim e_L}^n \sim P_{\text{neg}}(u_1, \mathcal{P})} \left[\log(1 - \sigma \left(\sum_{i=1}^L \sum_{j=i}^L f(u_i^n, v_j^n | \mathcal{P}_{i,j}) \right)) \right] \quad (16)$$

其中 $p_{e_1 \sim e_L}^n$ 是路径负样本, $\mathcal{P}_{i,j}$ ($1 \leq i \leq j \leq L$)是 $\mathcal{P}$ 的子元路径, u_i^n, v_j^n 是路径负样本上的节点, $f(u_i, v_j | \mathcal{P}_{i,j})$ 是节点 u_i 和 v_j 之间关于 $\mathcal{P}_{i,j}$ 的相关性,相关性越高,节点对在多条路径实例中同时出现的可能性越大. $f(u, v | \mathcal{P})$ 定义为

$$f(u, v | \mathcal{P}) = \mu_{\mathcal{P}} + \mathbf{g}_{\mathcal{P}}^T \mathbf{h}_u + \mathbf{y}_{\mathcal{P}}^T \mathbf{h}_v + \mathbf{h}_u^T \mathbf{h}_v \quad (17)$$

其中 $\mu_{\mathcal{P}} \in \mathbb{R}$ 是元路径 $\mathcal{P}$ 的全局偏差, $\mathbf{g}_{\mathcal{P}}, \mathbf{y}_{\mathcal{P}} \in \mathbb{R}^d$ 是 d 维的局部偏差向量, $\mathbf{h}_u, \mathbf{h}_v \in \mathbb{R}^d$ 是 u 和 v 的 d 维嵌入向量. $\mathcal{L}$ 使用随机梯度下降算法进行优化.

3.2 基于距离的表征

3.2.1 分布距离

基于分布距离的嵌入模型HINE^[77]首先基于元路径随机游走来计算节点间的邻近度,并定义了节点间基于元路径邻近度的概率分布和基于嵌入向量邻近度的概率分布,通过最小化两个分布之间的距离学习节点的向量表示.

给定网络 $G = (\mathcal{N}, \epsilon)$ 和元路径 $\mathcal{P}$,设 $s(u, v | \mathcal{P})$

是节点 u 和 v 之间关于 $\mathcal{P}$ 的邻近程度(比如连接 u 和 v 的 $\mathcal{P}$ 的实例数目^[78],或随机游走时沿 $\mathcal{P}$ 从 u 移动到 v 的概率^[79]), u 和 v 在 G 中的邻近度定义为 $s(u, v) = \sum_{\mathcal{P}} s(u, v | \mathcal{P})$, 则 u 和 v 在 G 中基于元路

径邻近度的分布定义为 $\hat{P}(u, v) = \frac{s(u, v)}{\sum_{v' \in \mathcal{A}} s(u, v')}$;

u 和 v 在嵌入向量空间中邻近度的分布定义为 $P(u, v) = \frac{1}{1 + e^{-\mathbf{h}_u \cdot \mathbf{h}_v}}$. 若使用 KL 散度度量两个概

率分布之间的距离 $O = - \sum_{u, v \in \mathcal{A}} s(u, v) \log P(u, v)$, 则目标函数为

$$\mathcal{L} = -\log(1 + e^{-\mathbf{h}_u \cdot \mathbf{h}_v}) - \sum_{m=1}^M \mathbb{E}_{n \sim P_{neg}(u)} \{\log(1 + e^{\mathbf{h}_u \cdot \mathbf{h}_n})\} \quad (18)$$

$\mathcal{L}(u, v)$ 采用异步随机梯度下降算法 ASGD^[80] 进行优化.

3.2.2 双曲距离

基于双曲距离的嵌入模型 HHNE^[81] 将节点嵌入双曲空间^[82], 使中心节点与其邻居之间的双曲距离较小, 与负样本之间的双曲距离较大. 设 $\mathbb{D}^d = \{\mathbf{x} \in \mathbb{R}^d \mid ||\mathbf{x}|| < 1\}$ 是开放的 d 维单位球, 节点嵌入矩阵 $\mathbf{H} = \{\mathbf{h}_u\}_{u \in \mathcal{A}}$, $\mathbf{h}_u \in \mathbb{D}^d$, 节点 u 和 v 之间的 Poincaré ball 距离定义为

$$d_{\mathbb{D}}(\mathbf{h}_u, \mathbf{h}_v) = \cosh^{-1} \left(1 + 2 \frac{||\mathbf{h}_u - \mathbf{h}_v||^2}{(1 - ||\mathbf{h}_u||^2)(1 - ||\mathbf{h}_v||^2)} \right) \quad (19)$$

则 v 是 u 邻居的条件概率定义为

$$P(v \mid u; \mathbf{H}) = \sigma[-d_{\mathbb{D}}(\mathbf{h}_u, \mathbf{h}_v)] \quad (20)$$

嵌入的目标是 $\arg \max_{\mathbf{H}} \sum_{u \in \mathcal{A}} \sum_{v \in \mathcal{A}_r(u)} P(v \mid u; \mathbf{H})$. 使用负采样技术, 目标函数为

$$\mathcal{L} = \log \sigma[-d_{\mathbb{D}}(\mathbf{h}_v, \mathbf{h}_u)] + \sum_{m=1}^M \mathbb{E}_{n \sim P_{neg}(v^t)} \{\log \sigma[d_{\mathbb{D}}(\mathbf{h}_n, \mathbf{h}_u)]\} \quad (21)$$

$\mathcal{L}$ 基于黎曼随机梯度下降算法 RSGD^[83] 进行优化.

欧几里得和双曲空间中的嵌入在不同的空间中保留固有的网络结构和语义相关性. 由于双曲几何能够自然反映复杂网络的层次和幂律结构^[84] 等属性, 且双曲空间比欧几里得空间扩展快^[85], 因此, 复杂网络适合在双曲空间中建模^[86], 且在双曲空间的嵌入取得了较好的效果.

3.2.3 欧式距离+转化距离

关系结构感知的异质网络嵌入 RHINE^[87] 将异

质关系划分为隶属关系 AR 和交互关系 IR, 使用不同的模型分别处理 AR 和 IR, 以保留其独特的结构特征. 在 AR 中, 关系 r 连接的两种类型的节点的平均度有很大不同, 表现出一种类型的节点以另一种类型的节点为中心的结构, 暗示节点之间的从属关系, 比如“一次会议上发表了多篇论文”; 在 IR 中, 关系 r 连接的两种类型的节点的平均度近似相等, 表现出对等结构, 暗示节点之间的交互关系, 比如“作者撰写论文”.

对于 AR, 节点 u 和 v 在低维空间中的邻近度用欧氏距离度量, 即 $f(u, v) = w_{uv} ||\mathbf{h}_u - \mathbf{h}_v||_2^2$, 其中 w_{uv} 是 u 和 v 之间的权重, 损失函数定义为

$$\mathcal{L}_{\text{EuAR}} = \sum_{r \in \mathcal{R}_{\text{AR}}} \sum_{\langle u, r, v \rangle \in \mathcal{I}_{\text{AR}}} \sum_{\langle u', r, v' \rangle \in \mathcal{I}'_{\text{AR}}} \max [0, \gamma + f(u, v) - f(u', v')] \quad (22)$$

其中 $\langle u, r, v \rangle$ 表示节点 u 和 v 通过关系 $r \in \mathcal{R}_{\text{AR}}$ 连接, $\mathcal{I}_{\text{AR}}$ 是正样本集, $\mathcal{I}'_{\text{AR}}$ 是负样本集, γ 是超参数.

对于 IR, u 和 v 在低维空间中的邻近度使用转化距离度量, 定义为 $g(u, v) = w_{uv} ||\mathbf{x}_u + \mathbf{y}_r - \mathbf{x}_v||$, 损失函数定义为

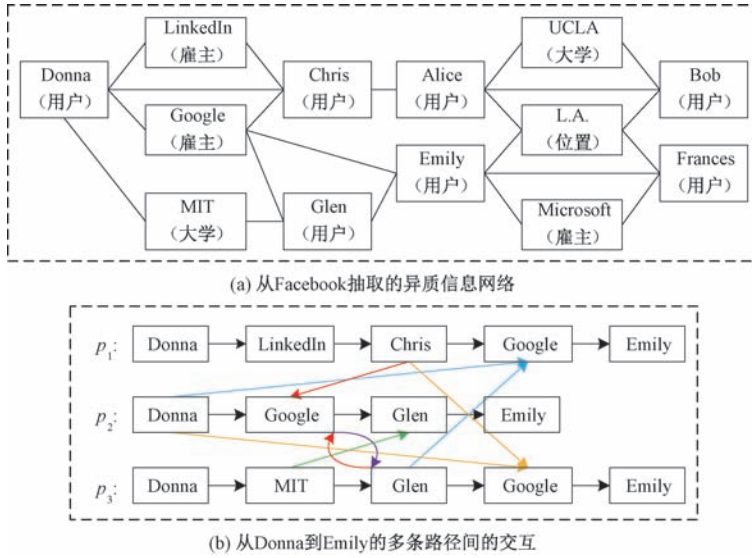
$$\mathcal{L}_{\text{TrIR}} = \sum_{r \in \mathcal{R}_{\text{IR}}} \sum_{\langle u, r, v \rangle \in \mathcal{I}_{\text{IR}}} \sum_{\langle u', r, v' \rangle \in \mathcal{I}'_{\text{IR}}} \max [0, \gamma + g(u, v) - g(u', v')] \quad (23)$$

其中 $\mathcal{I}_{\text{IR}}$ 是正样本集, $\mathcal{I}'_{\text{IR}}$ 是负样本集. RHINE 的总损失函数 $\mathcal{L} = \mathcal{L}_{\text{EuAR}} + \mathcal{L}_{\text{TrIR}}$.

3.3 交互式路径表征

在 HeN 中, 每条路径只能捕获图的部分信息, 难以全面捕捉节点之间的邻近关系, 比如图 15 中, 路径 p_1 反映不出 Donna 在 Google 工作, p_2 也反映不出 Emily 也在 Google 工作. IPE^[88] 引入交互路径来建模多条路径之间的依赖性, 即: 如果一条路径中的某个节点在其他路径中有前驱, 则增加从该节点在其他路径中的前驱到该节点的有向边. 比如, 在图 15(b) 中添加从 $(p_2; \text{Donna})$ 和 $(p_3; \text{Glen})$ 到 $(p_1; \text{Google})$ 的有向边. 通过路径之间的耦合, 每条路径都可以从与其他路径的交互作用中获得可以从原图结构中提取出来但在单条路径中丢失的信息, 比如 Donna 和 Emily 实际上是同事.

IPE 设计了图 16(b) 所示的交互式 GRU 机制学习节点的低维向量表示, 图中 $\mathbf{g}_2$ 沿图 16(a) 的路径 p_2 使用每个节点自身的特征 $\mathbf{x}_v \in \mathbb{R}^{d'}$ 及其前驱实现每个节点的嵌入. 路径的嵌入通过节点嵌入的最大池化进行计算.

图 15 异质信息网络中的路径交互^[88]

3.4 基于图卷积神经网络的表征

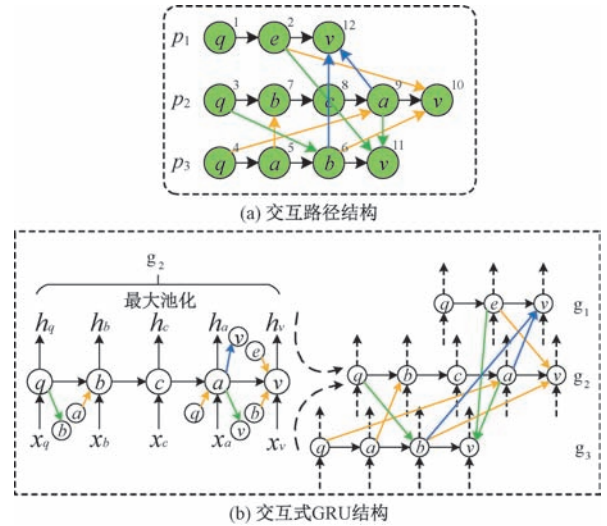
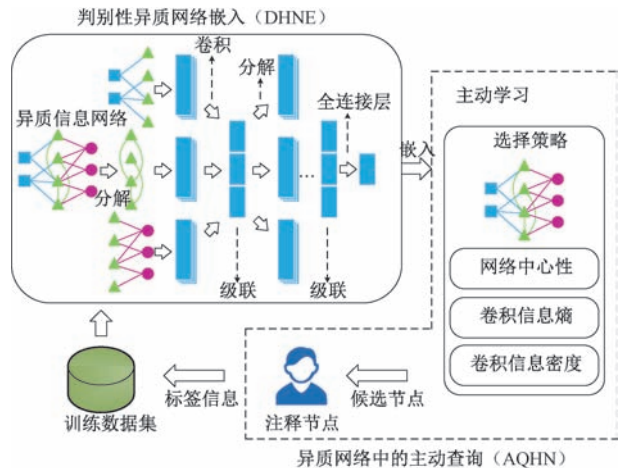
ActiveHNE^[89]: ActiveHNE 是一种半监督的网络嵌入方法,体系结构如图 17 所示. ActiveHNE 包含两个组件:判别性异质网络嵌入(DHNE)和异质网络中的主动查询(AQHN),其中 DHNE 基于图卷积神经网络实现半监督异质网络嵌入,AQHN 通过主动查询获取最有价值的监督信息并提供给 DHNE 以提高嵌入质量. 在每次迭代中,一旦 DHNE 获得了网络嵌入,AQHN 就会选择要查询的最有价值的节点,然后使用新标签更新 DHNE. 为了区别对待不同类型的节点和关系,DHNE 将原始的 HeN 分解为同质网络和二分网络(包含两种类型的节点). 对于每个卷积层,DHNE 分别学习每个分解网络中节点的深层语义,然后将所有网络中每个节点的输出向量进行级联.

设 $\{\mathcal{G}_t \mid t = 1, 2, \dots, T\}$ 是划分后的同质和二分网络集合, $\{\mathbf{A}_t \mid t = 1, 2, \dots, T\}$ 表示与 $\{\mathcal{G}_t\}$ 对应的邻接矩阵, $\mathbf{D}_t = \text{diag}(\sum_i \mathbf{A}_t(i, j))$ 表示度矩阵, $\mathbf{P}_t = \mathbf{D}_t^{-1} \mathbf{A}_t$ 表示转移概率矩阵, $\mathbf{X}_t \in \mathbb{R}^{N_t \times D}$ 是网络 $\mathcal{G}_t$ 的输入信号, N_t 和 D 分别是 $\mathcal{G}_t$ 中的节点数和每个节点的特征数. 网络 $\mathcal{G}_t$ 的卷积操作定义为

$$\mathbf{H}_t = \sigma\left(\sum_{k=1}^K \mathbf{P}_t^k \mathbf{X}_t \Theta_t\right) \quad (24)$$

其中 $\Theta_t \in \mathbb{R}^{D \times d}$ 和 $\mathbf{H}_t \in \mathbb{R}^{N_t \times d}$ 分别表示可训练的权矩阵和输出信号,激活函数采用 $\text{ReLU}(\cdot)$ 函数.

在每个网络上分别执行卷积后,根据节点所属的网络将卷积信号的向量按顺序进行连接,以获得每个节点的最终输出信号. 对于不是网络元素的节

图 16 交互式 GRU 结构^[88]图 17 ActiveHNE 的体系结构^[89]

点,使用零向量表示相应的输出信号.令 $\mathbf{Z}_t$ 表示 $\mathcal{G}_t$ 中节点的级联卷积信号,则 $\mathcal{G}_t$ 上第 l 层的卷积定义为

$$\mathbf{H}_t^{(l)} = \sigma\left(\sum_{k=1}^K \mathbf{P}_t^k \mathbf{Z}_t^{(l)} \Theta_t^{(l)}\right), \quad l=0,1,2,\dots \quad (25)$$

其中 $\mathbf{Z}_t^{(l)} \in \mathbb{R}^{N_t \times Td^{(l-1)}}$, $\Theta_t^{(l)} \in \mathbb{R}^{Td^{(l-1)} \times d^{(l)}}$, $\mathbf{H}_t^{(l)} \in \mathbb{R}^{N_t \times d^{(l)}}$ 分别表示第 l 层的输入信号、可训练的权重矩阵和输出信号, $\mathbf{Z}_t^{(0)} = \mathbf{X}_t$, $\Theta_t^{(0)} \in \mathbb{R}^{D \times d^{(1)}}$.

经过 β 层的卷积和级联后,可以得到所有节点的最终输出向量 $\mathbf{E} = \mathbf{Z}^\beta \in \mathbb{R}^{N \times Td^{(\beta)}}$. 最后用一个全连接层来预测节点的标签信息 $\mathbf{F} = \sigma(\mathbf{E}\Theta^{\text{pre}}) \in \mathbb{R}^{N \times C}$, 其中 $\Theta^{\text{pre}} \in \mathbb{R}^{Td^{(\beta)} \times C}$ 是隐层到输出层的权重矩阵,

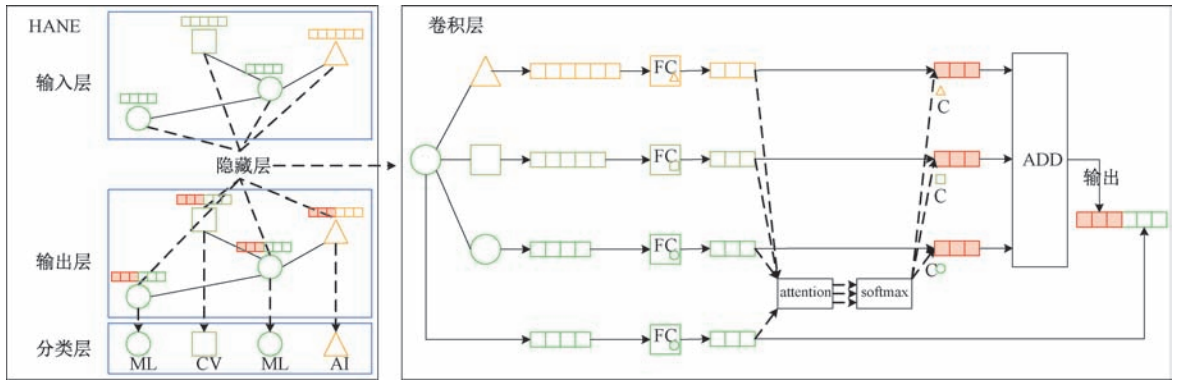


图 18 HANE 的框架^[90]

4 多网络的表征学习

4.1 多层网络表征

多层网络表征学习的目标是学习 b 层网络中各层节点的低维向量 $\mathbf{H}_k \in \mathbb{R}^{n_k \times d_k}$ ($k=1, \dots, b$), 同时保留原始网络中节点之间层内和层间的结构和语义相关性. n_k 是第 k 层内的节点数, d_k 是第 k 层的嵌入维度.

MANE^[57]: MANE 同时对层内、层间的连接进行建模. 设 $\mathcal{A} = \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_b\}$ 表示每一层内节点之间的邻接矩阵集, $\mathcal{D} = \{\mathbf{D}_{i,j}, (i,j=1, \dots, b) (i \neq j)\}$ 表示层间的邻接矩阵集, $\mathbf{L}_k = \mathbf{Z}_k^{-1/2} \mathbf{A}_k \mathbf{Z}_k^{-1/2}$ 是第 k 层网络结构的规范化拉普拉斯矩阵, $\mathbf{Z}_k$ 是对角矩阵, 元素为 $\mathbf{Z}_k(i,i) = \sum_{j=1}^{n_k} \mathbf{A}_k(i,j)$. 约束 $\mathbf{H}_k^T \mathbf{H}_k = \mathbf{I}$ 用于避免 $\mathbf{H}_k$ 的任意缩放, $\mathbf{I}$ 是单位矩阵, $\mathbf{K}_{ij} \in \mathbb{R}^{d_i \times d_j}$ 表示第 i,j 两层嵌入的相互作用矩阵. MANE 的目标函数定义为

$\mathbf{F}_{ic}$ 表示第 i 个节点属于第 c 类的概率, 第 i 个节点的标签可以预测为 $y_i = \arg \max_c \mathbf{F}_{ic}$. 损失函数定义为所有标签节点上的交叉熵误差:

$$\mathcal{L} = - \sum_{i=1}^L \sum_{c=1}^C \mathbf{Y}_{ic} \ln \mathbf{F}_{ic} \quad (26)$$

其中 $\mathbf{Y} \in \{0,1\}^{N \times C}$ 存储节点的真实标签. 如果第 i 个节点属于第 c 类, 则 $\mathbf{Y}_{ic}=1$, 否则 $\mathbf{Y}_{ic}=0$. 使用梯度下降法更新参数 $\Theta_t^{(l)}$ 和 Θ^{pre} , 以最小化 $\mathcal{L}$.

HANE^[90]: HANE 首先将异质网络的输入属性转换到一个公共空间, 然后利用图卷积网络学习节点的嵌入, 并引入注意力机制给邻域中不同类型的节点分配不同的重要性. HANE 的框架如图 18 所示.

$$\begin{aligned} \max_{\mathbf{H}_i, \mathbf{K}_{ij}} & \left(\sum_{i=1}^b \text{tr}(\mathbf{H}_i^T \mathbf{L}_i \mathbf{H}_i) - \alpha \sum_{i,j=1}^b ||\mathbf{D}_{ij} - \mathbf{H}_i \mathbf{K}_{ij} \mathbf{H}_j^T||_F^2 \right) \\ \text{s.t.} \quad & \mathbf{H}_k^T \mathbf{H}_k = \mathbf{I} \quad (\forall k = 1, 2, \dots, b) \end{aligned} \quad (27)$$

其中 $\text{tr}(\cdot)$ 表示矩阵的迹, 第一项和第二项分别用于优化层内和层间连接, 参数 $\alpha > 0$ 用于平衡层内连接和层间连接的贡献. $\mathbf{H}_i$ 和 $\mathbf{K}_{ij}$ 通过交替优化算法求解.

EOE^[91]: EOE 同时对各个子网内的节点对和子网间的节点对进行表征学习, 并引入和谐嵌入矩阵将仅对网络内部边进行编码的嵌入进一步嵌入到另一个潜在空间, 使得不同层网络的表示向量具有可比性.

PTE^[92]: PTE 通过半监督方法学习异质文本网络中词的低维嵌入.

PMNE^[93]: 提出了“网络聚合”、“结果聚合”和“层协同分析”三种方法将多层网络投影到连续的向量空间中, 不仅使用一阶和二阶随机游走在一个层上遍历, 而且还具有利用层间交互作用在层与层之间遍历的能力.

4.2 多视图网络表征

4.2.1 同质多视图网络表征

在多视图网络中,同一对节点之间的边可能会在不同的视图中观察到,因此不同的视图可能会相互补充.同时,不同的视图可能具有不同的语义,部分节点对可能在不同的视图中具有完全不同的边.

M-DeepWalk^[94]:将每个视图的网络看作一个层,基于局部邻域结构之间的相似性引入层间边建立多层超图 $\mathcal{G}^{\text{sup}}$,以显式建模层间依赖关系,并利用 DeepWalk 学习 $\mathcal{G}^{\text{sup}}$ 中 $|\mathcal{N}|$ 个节点的嵌入,然后再使用深度网络嵌入方法优化每一层中每个节点的嵌入. M-DeepWalk 模型如图 19 所示.

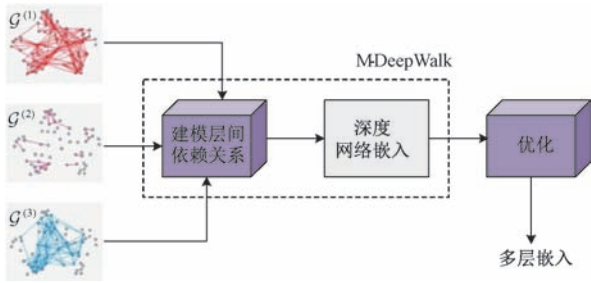


图 19 M-DeepWalk 模型^[94]

MVE^[95]:MVE 首先使用视图侧的节点表达来保留节点在不同视图中的邻近度,然后组合视图侧的节点向量通过投票获得鲁棒的节点表达.由于不同视图的信息质量可能不同,投票期间不同视图的权重通过注意力机制利用少量的标签数据进行学习. MVE 的框架如图 20 所示.

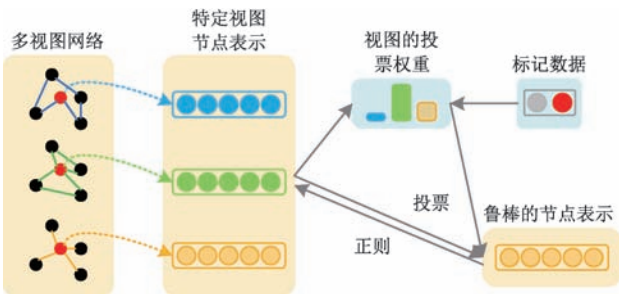


图 20 MVE 的框架^[95]

MNE^[96]:MNE 将多类型关系的信息投影到一个统一的嵌入空间中.为了结合不同类型关系的信息,同时保持其独特的属性,对于每个节点 $v \in \mathcal{N}$, MNE 使用在所有关系中共享的公共嵌入向量 $\mathbf{b}_v \in \mathbb{R}^d$ 来联系网络的各种关系,针对关系类型 $r \in \mathcal{R}$ 使用附加向量 $\mathbf{u}_v^r \in \mathbb{R}^s$ (s 是关系类型向量的维度)学习节点在关系 r 中的独特性质,从而节点 v 关于关系 r

的嵌入向量 $\mathbf{h}_v^r = \mathbf{b}_v + \omega^r \cdot (\mathbf{X}^r)^T \mathbf{u}_v^r$,其中 ω^r 是第 r 种关系的权重, $\mathbf{X}^r \in \mathbb{R}^{s \times d}$ 是关系 r 的变换矩阵.

Mvn2vec^[97]:Mvn2vec 提出了 MVN2VEC-CON 和 MVN2VEC-REG 两种嵌入方法,其中 CON 和 REG 分别代表约束和正则化. MVN2VEC-CON 模型不对随机游走的中心节点 $u \in \mathcal{N}$ 在视图 $v \in \mathcal{V}$ 中的嵌入 $\{\mathbf{h}_u^v\}_{u \in \mathcal{N}}$ 进行约束,以保留每个个体视图的语义,但是对 u 的上下文节点的嵌入 $\{\mathbf{h}_c^v\}_{c \in \mathcal{N}(u)}$ 强制施加约束,以便在不同视图之间共享参数. MVN2VEC-REG 模型没有对参数如何在不同视图之间共享硬性设置约束,但是对不同视图上的嵌入进行规范化处理.模型学习之后,节点 u 的最终嵌入为各个视图中嵌入的级联,即

$$\mathbf{h}_u = \bigoplus_{v \in \mathcal{V}} \mathbf{h}_u^v.$$

4.2.2 异质多视图网络表征

TransN^[58] 使用了 Single-View 和 Cross-View 两个算法分别捕捉视图内的邻近性和视图间的相关性,模型框架如图 21 所示,图中圆形、正方形和六边形表示不同类型的节点,连接不同类型的实线表示不同类型的边,虚线表示视图对共享的公共节点.共享公共节点的一对视图 $G_i = (\mathcal{N}_i, \epsilon_i)$ 和 $G_j = (\mathcal{N}_j, \epsilon_j)$ 称为视图对,记为 $G_{i,j}$. 设 $\mathcal{M}_{i,j}$ 是视图 G_i 和 G_j 中公共节点的集合, $\mathcal{A}_{i,j}$ 是与 $\mathcal{M}_{i,j}$ 中的任何节点相邻的节点集合,由 $\mathcal{M}_{i,j} \cap \mathcal{A}_{i,j}$ 中的节点以及它们之间的边组成的视图 $G'_{i,j}$ (或 $G'_{j,i}$) 是 G_i (或 G_j) 的子网,称为成对子视图.

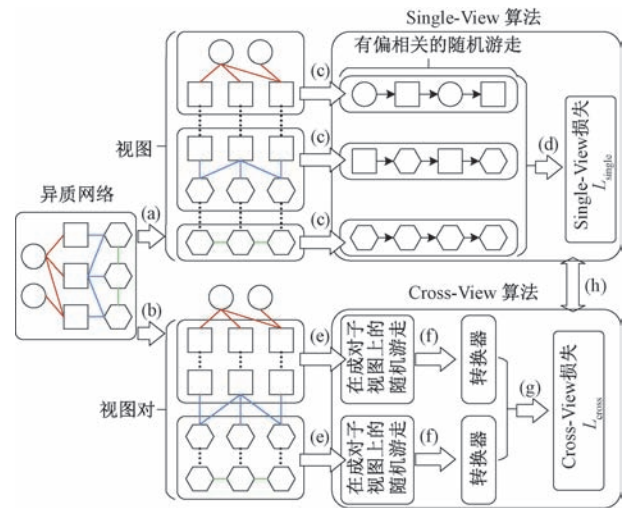


图 21 TransN 的框架^[58]

Single-View 算法用于保存每个视图内的邻近信息,算法的输入是从各个视图中进行随机游走采样的节点序列,目标函数 $\mathcal{L}_{\text{single}}$ 定义为

$$\mathcal{L}_{\text{single}} = -\log\left(\prod_{u \in \mathcal{A}_i} \prod_{v \in \mathcal{A}(u)} P(v | \mathbf{h}_u^i)\right) \quad (28)$$

其中 $\mathcal{A}(u)$ 是节点 u 的上下文节点集合, $P(v | \mathbf{h}_u^i)$ 是在视图 G_i 中给定 u 的嵌入 $\mathbf{h}_u^i$ 的条件下节点 v 是节点 u 的上下文的条件概率, 可以用 softmax 函数定义.

Cross-View 算法用于在不同视图之间投影节点的嵌入, 实现在视图之间传递学习到的信息. 算法的输入是从成对子视图(非原始视图)中进行随机游走采样的节点序列. 由于公共节点是视图之间传递信息的桥梁, 因此 Cross-View 算法重点关注公共节点及其相邻节点. 虽然公共节点在不同视图中的嵌入可能不同, 但由于这些嵌入表示的是同一实体, 因此, 从视图 G_i 投影到视图 G_j 的嵌入矩阵在反向投影时应投影到自身, 即 $\mathbf{H}_i \approx T_{j \rightarrow i}(T_{i \rightarrow j}(\mathbf{H}_i))$. 设 p_i 是 G_i 中元路径 $\mathcal{P}_i$ 的实例, $\mathcal{L}_{i \rightarrow j}(p_i)$ 是从 G_i 变换到 G_j 的损失, 则 Cross-View 的损失 $\mathcal{L}_{\text{cross}}$ 定义为

$$\begin{aligned} \mathcal{L}_{\text{cross}} = & \frac{1}{|\mathcal{P}_i|} \sum_{p_i \in \mathcal{P}_i} (\mathcal{L}_{i \rightarrow j}(p_i) + \mathcal{L}_{j \rightarrow i}(p_i)) \\ & + \frac{1}{|\mathcal{P}_j|} \sum_{p_j \in \mathcal{P}_j} (\mathcal{L}_{j \rightarrow i}(p_j) + \mathcal{L}_{i \rightarrow j}(p_j)) \end{aligned} \quad (29)$$

TransN 的整体损失函数定义为

$$\mathcal{L}_{\text{overall}} = \mathcal{L}_{\text{single}} + \mathcal{L}_{\text{cross}} \quad (30)$$

$\mathcal{L}_{\text{overall}}$ 通过随机优化算法 Adam^[75] 和反向传播算法 BP^[98] 最小化.

4.3 多重异质网络表征

多重异质网络嵌入 GATNE^[59] 提出了 GATNE-T 和 GATNE-I 两种模型处理节点间的多种关系. GATNE-T 只使用网络的结构信息, 而 GATNE-I 既使用结构信息, 也使用节点属性. 两种模型都将节点 v 关于边类型 r 的整体嵌入分为基嵌入 $\mathbf{b}_v$ 和边嵌入 $\mathbf{U}_v$ 两部分, 基嵌入 $\mathbf{b}_v$ 在不同的边类型之间共享.

在 GATNE-T 中, 节点 v 关于边类型 r 的第 k 层边嵌入 $\mathbf{u}_{v,r}^{(k)} \in \mathbb{R}^d$ ($1 \leq k \leq K$) 通过聚集邻居的边嵌入获得, 即 $\mathbf{u}_{v,r}^{(k)} = \text{aggregator}(\{\mathbf{u}_{v',r}^{(k-1)}, \forall v' \in \mathcal{N}_r(v)\})$, 其中 $\mathcal{N}_r(v)$ 表示节点 v 关于边类型 r 的邻居集. 函数 $\text{aggregator}(\cdot)$ 可以是平均或最大池化, 初始边嵌入 $\mathbf{u}_{v,r}^{(0)}$ 随机初始化. 第 K 层的边嵌入 $\mathbf{u}_{v,r}^{(K)}$ 记为 $\mathbf{u}_{v,r}$, 节点 v 关于 m 种类型的边的嵌入级联为 $d \times m$ 维的 $\mathbf{U}_v = (\mathbf{u}_{v,1}, \mathbf{u}_{v,2}, \dots, \mathbf{u}_{v,m})$. 节点 v 关于边类型 r 的整体嵌入 $\mathbf{h}_{v,r} = \mathbf{b}_v + \alpha_r \mathbf{M}_r^T \mathbf{U}_v \mathbf{a}_{v,r}$, 其中 α_r 是超参数, $\mathbf{M}_r \in \mathbb{R}^{d \times d}$ 是可训练变换矩阵, $\mathbf{a}_{v,r} = \text{softmax}(\mathbf{w}_r^T \tanh(\mathbf{W}_r \mathbf{U}_v))^T \in \mathbb{R}^m$ 是 $\mathbf{U}_i$ 中向量

的线性组合系数, $\mathbf{w}_r$ 和 $\mathbf{W}_r$ 是边类型 r 的可训练参数.

在 GATNE-I 中, 基嵌入 $\mathbf{b}_v$ 和节点 v 关于边类型 r 的初始边嵌入 $\mathbf{u}_{v,r}^{(0)}$ 定义为 v 的属性 $\mathbf{x}_v$ 的函数, $\mathbf{b}_v = f_z(\mathbf{x}_v)$, $\mathbf{u}_{v,r}^{(0)} = g_{z,r}(\mathbf{x}_v)$, 其中 f_z 和 $g_{z,r}$ 是变换函数, $z = \phi(v)$ 是节点 v 的类型. 节点 v 关于边类型 r 的整体嵌入 $\mathbf{h}_{v,r} = f_z(\mathbf{x}_v) + \alpha_r \mathbf{M}_r^T \mathbf{U}_v \mathbf{a}_{v,r} + \beta_r \mathbf{D}_z^T \mathbf{x}_v$, 其中 β_r 是系数, $\mathbf{D}_z$ 是关于 v 的类型 z 的特征变换矩阵, $\beta_r \mathbf{D}_z^T \mathbf{x}_v$ 是 v 关于类型 z 的属性项.

GATNE-T 和 GATNE-I 的差异在于基嵌入 $\mathbf{b}_v$ 和初始边嵌入 $\mathbf{u}_{v,r}^{(0)}$ 的生成. GATNE-T 基于网络结构直接对每个节点训练基嵌入 $\mathbf{b}_v$ 和初始边嵌入 $\mathbf{u}_{v,r}^{(0)}$, 无法处理训练期间未观察到的节点. GATNE-I 不直接训练 $\mathbf{b}_v$ 和 $\mathbf{u}_{v,r}^{(0)}$, 而是训练将原始特征变换为 $\mathbf{b}_v$ 和 $\mathbf{u}_{v,r}^{(0)}$ 的变换函数 f_z 和 $g_{z,r}$, 因此能够处理训练期间没有出现的节点, 只要它们具有对应的原始特征.

GATNE 使用基于元路径的随机游走产生节点序列, 然后在节点序列上通过 Skip-Gram 学习嵌入. 设中心节点为 c , 则目标函数定义为

$$\begin{aligned} \mathcal{L} = & -\log \sigma(\mathbf{h}_c^T \cdot \mathbf{h}_{v,r}) - \sum_{m=1}^M \mathbb{E}_{n \sim P_{neg}^{(v)}} \\ & [\log \sigma(-\mathbf{h}_n^T \cdot \mathbf{h}_{v,r})] \end{aligned} \quad (31)$$

4.4 新兴异质社交网络表征

BL-MNE^[99] 基于深度自编码实现新兴异质社交网络的嵌入. 新兴网络由于开始提供服务时间不长, 节点间的连接较少, 网络结构稀疏, 因此嵌入性能欠佳, 但是新兴网络的用户可能也在其他成熟的网络(例如 Facebook, Twitter). 成熟网络信息丰富且数量充足, 因此成熟网络与共享用户的新兴网络之间的信息交换可以有效克服信息稀疏性问题.

设网络 $\mathcal{G} = ((G^{(1)}, G^{(2)}), (\mathcal{A}^{(1,2)}))$ 中, $G^{(1)}$ 是新兴网络, $G^{(2)}$ 是成熟网络. 在 $G^{(1)}$ 中指定元路径集 $\{\Phi_0, \Phi_1, \dots, \Phi_R\}$, 并从网络中抽取一组连接网络中用户的元路径实例. 设 $\mathcal{P}_{\Phi_k}^{(1)}(u, v)$ 表示连接 $u^{(1)}$ 和 $v^{(1)}$ 的元路径 Φ_k 的实例集合, $\mathcal{P}_{\Phi_k}^{(1)}(u^{(1)}, \cdot)$ (或 $\mathcal{P}_{\Phi_k}^{(1)}(\cdot, v^{(1)})$) 表示从 $u^{(1)}$ 出(或进入 $v^{(1)}$)的元路径实例集, 则 $u^{(1)}$ 和 $v^{(1)}$ 基于元路径 Φ_k 的元邻近度定义为

$$p_{\Phi_k}^{(1)}(u^{(1)}, v^{(1)}) = \frac{2 | \mathcal{P}_{\Phi_k}^{(1)}(u^{(1)}, v^{(1)}) |}{| \mathcal{P}_{\Phi_k}^{(1)}(u^{(1)}, \cdot) | + | \mathcal{P}_{\Phi_k}^{(1)}(\cdot, v^{(1)}) |} \quad (32)$$

网络 $G^{(1)}$ 中所有用户之间基于元路径 Φ_k 的元邻近度表示为矩阵 $\mathbf{P}_{\Phi_k}^{(1)} \in \mathbb{R}^{|\mathcal{U}^{(1)}| \times |\mathcal{U}^{(1)}|}$, 其元素

$p_{\Phi_k}^{(1)}(u, v) = p_{\Phi_k}^{(1)}(u^{(1)}, v^{(1)})$. 根据网络 $G^{(1)}$ 定义的所有元邻近矩阵记为 $\{\mathbf{P}_{\Phi_k}^{(1)}\}_{\Phi_k}$. 同理, 根据网络 $G^{(2)}$ 定义的所有元邻近矩阵记为 $\{\mathbf{P}_{\Phi_k}^{(2)}\}_{\Phi_k}$. $\{\mathbf{P}_{\Phi_k}^{(1)}\}_{\Phi_k}$ 和 $\{\mathbf{P}_{\Phi_k}^{(2)}\}_{\Phi_k}$ 将被馈入图 22 所示的深度学习框架 DIME, 学习用户的嵌入向量.

设用户 $u_i^{(1)} \in \mathcal{N}^{(1)}$ 基于元路径 Φ_k 的输入特征向量表示为 $\mathbf{x}_{u, \Phi_k}^{(1)}$ (矩阵 $\mathbf{P}_{\Phi_k}^{(1)}$ 中对应 $u^{(1)}$ 的行), $\hat{\mathbf{x}}_{u, \Phi_k}^{(1)}$ 是 $\mathbf{x}_{u, \Phi_k}^{(1)}$ 的重构向量. DIME 的损失函数定义为

$$\mathcal{L}^{(1)} = \sum_{\Phi_k \in \{\Phi_0, \dots, \Phi_R\}} \sum_{u \in \mathcal{N}} \|(\mathbf{x}_{u, \Phi_k}^{(1)} - \hat{\mathbf{x}}_{u, \Phi_k}^{(1)}) \odot \mathbf{b}_{u, \Phi_k}^{(1)}\|_2^2 \quad (33)$$

其中向量 $\mathbf{b}_{u, \Phi_k}^{(1)}$ 是对应于特征向量 $\mathbf{x}_{u, \Phi_k}^{(1)}$ 的权重向

量. 同理, 可以定义网络 $G^{(2)}$ 的损失函数 $\mathcal{L}^{(2)}$.

设 $G^{(1)}$ 和 $G^{(2)}$ 中所有用户节点的嵌入分别表示为矩阵 $\mathbf{H}^{(1)} \in \mathbb{R}^{|\mathcal{N}^{(1)}| \times d^{(1)}}$, $\mathbf{H}^{(2)} \in \mathbb{R}^{|\mathcal{N}^{(2)}| \times d^{(2)}}$. 如果 $u^{(1)}$ 和 $v^{(2)}$ 是相同用户, 则 $\mathbf{T}^{(1,2)}(u, v) = 1$, 向量 $\mathbf{H}^{(1)}(u, :)$ 和 $\mathbf{H}^{(2)}(v, :)$ 将被投影到嵌入空间的邻近区域. 网络 $G^{(1)}$ 和 $G^{(2)}$ 之间引入的信息融合损失表示为

$$\mathcal{L}^{(1,2)} = \|(\mathbf{T}^{(1,2)})^T \mathbf{H}^{(1)} \mathbf{W}^{(1,2)} - \mathbf{H}^{(2)}\|_F^2 \quad (34)$$

其中 $\mathbf{W}^{(1,2)} \in \mathbb{R}^{d^{(1)} \times d^{(2)}}$ 是变换矩阵, 用于使 $\mathbf{H}^{(1)}$ 和 $\mathbf{H}^{(2)}$ 具有相同的维度. 通过最小化 $\mathcal{L}^{(1,2)}$, 来自成熟网络 $G^{(2)}$ 的锚点用户的嵌入向量可以用来调整其在新兴网络 $G^{(1)}$ 中的嵌入向量.

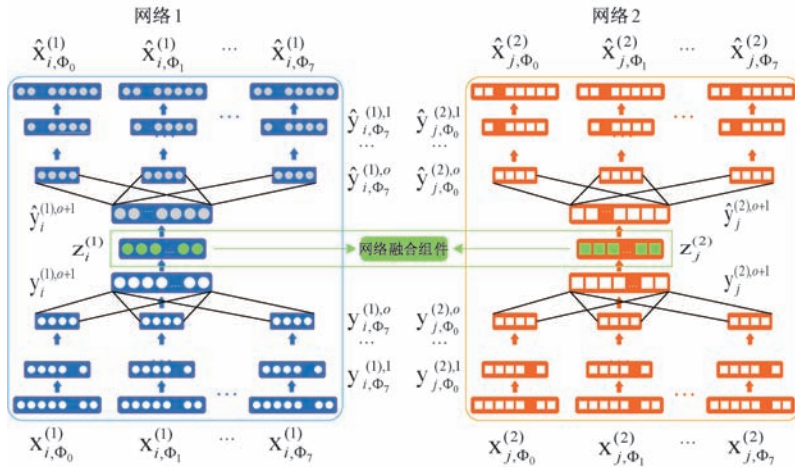


图 22 DIME 的框架^[99]

DIME 的完整目标函数包括网络 $G^{(1)}$ 和 $G^{(2)}$ 的损失项 $\mathcal{L}^{(1)}$ 、 $\mathcal{L}^{(2)}$, 以及信息融合损失 $\mathcal{L}^{(1,2)}$, 表示为

$$\mathcal{L}(G^{(1)}, G^{(2)}) = \mathcal{L}^{(1)} + \mathcal{L}^{(2)} + \alpha \mathcal{L}^{(1,2)} + \beta \mathcal{L}_{\text{reg}} \quad (35)$$

参数 α 和 β 是信息融合损失项和正则项的权重, 加入正则项 $\mathcal{L}_{\text{reg}}$ 的目的是避免过度拟合. $\mathcal{L}(G^{(1)}, G^{(2)})$ 使用随机梯度下降算法进行优化.

5 动态网络的表征学习

Change2vec^[63]: 为了有效应对网络的动态变化, Change2vec 将动态异质网络分解为具有不同时间戳的网络快照, 动态学习能够在各个时间戳捕获对象之间结构关系的低维向量表示.

Change2vec 使用三合会^[100]来表示网络的核心结构, 三合会封闭或开放的过程反映了网络结构的变化^[101]. 图 23 展示了一个动态的异质社交网络, 其中圆代表用户(U), 正方形代表主题(T), 该网络包

含两种类型的关系: 用户之间(U-U)的友谊关系(实线)和用户与主题之间(U-T)的订阅关系(虚线). 从时间戳 t 到 $t+1$, 网络引入了两条新边 (e_{BC} 和 e_{BT}), 产生了两个三合会的封闭; 从时间戳 $t+1$ 到 $t+2$, 边 e_{AC} 消失, 导致一个三合会的开放.

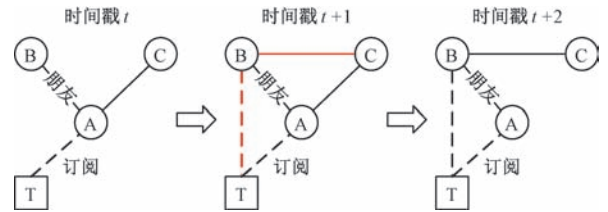


图 23 动态异质社交网络^[63]

一般而言, 网络随着时间的推移平稳演化, 而不是每两个时间戳之间都存在不稳定的重组^[102]. 因此, Change2vec 通过仅更新两个连续时间戳中涉及三合会关闭或开放过程的节点的向量表示来捕获网络结构的变化. 初始嵌入基于 $t=1$ 时的快照 G^1 运行 metapath2vec⁺⁺ 获得, G^t 和 G^{t+1} 之间的差异

通过一组变化的节点 $\mathcal{N}_{\text{change}}$ 来建模. $\mathcal{N}_{\text{change}}$ 包括新增或删除的节点和它们的一跳邻居, 以及导致三合会封闭或开放的新增边或删除边的对应节点. 利用 `metapath2vec++` 算法生成 $\mathcal{N}_{\text{change}}$ 中节点的嵌入向量, 然后在 t 时刻的结果 $\mathbf{H}^t$ 中添加新增节点的嵌入向量、删除已删除节点的向量、替换已存在但已变化的节点的嵌入向量来生成 $\mathbf{H}^{t+1}$.

Change2vec 的目标是在每个时间戳 t 最大化节点与其邻居节点间的条件概率, 即

$$\operatorname{argmax}_{\mathbf{H}} \sum_{u \in \mathcal{N}^t} \sum_{c \in \mathcal{C}} \sum_{v \in \mathcal{N}_c^t(u)} \log P(v \mid u; \mathbf{H}) \quad (36)$$

其中 $\mathcal{N}_c^t(u)$ 表示在时间戳 t 节点 u 的 C 类型的一

跳邻居集, $P(v \mid u; \mathbf{H})$ 定义为

$$P(v \mid u; \mathbf{H}) = \frac{\exp(\mathbf{h}_v^t \cdot \mathbf{h}_u^t)}{\sum_{v' \in \mathcal{N}_c^t} \exp(\mathbf{h}_{v'}^t \cdot \mathbf{h}_u^t)} \quad (37)$$

其中 $\mathcal{N}_c^t$ 表示在时间戳 t 类型为 C 的节点集. 式(37)中的目标使用异质负采样和随机梯度下降算法进行优化.

6 讨 论

第 3、4 和 5 节介绍了异质信息网络的典型表征学习方法, 各种方法的特点如表 2 所示.

表 2 典型表征学习方法的特点(注:表中“/”表示文献原文没有分析时间复杂度)								
分类		方法名称	所用信息	所用技术	嵌入对象	表示形式	方法特点	学习方式 时间复杂度
单网络嵌入	节点共现	Metapath2vec ^[68]	结构	随机游走 + Skip-Gram	节点	实向量	概率计算忽略了节点类型信息	监督 /
		Metapath2vec++ ^[68]	结构	随机游走 + Skip-Gram	节点	实向量	概率计算、采样分布针对特定的节点类型	监督 /
		AHNG ^[49]	结构 + 属性	随机游走 + 全连接神经网络(FCN)	节点	高斯分布	捕捉了节点表示的不确定性	监督 /
	关系成立	HIN2Vec ^[69]	结构	随机游走 + Skip-Gram	节点, 关系	实向量	保留更多的上下文信息, 区分节点间的不同关系	监督 /
		M-HIN ^[70]	结构	随机游走 + Skip-Gram	节点, 元路径	复向量	利用元图减少结构或语义信息的丢失, 利用复嵌入区分对称和非对称元图	监督 /
		TaPEm ^[73]	结构	随机游走, 感知器, BiGRU, 注意力机制	节点对, 上下文路径	实向量	捕捉节点间细粒度的成对关系	监督 /
		HeGAN ^[66] , AND-HeA ^[67]	结构	博弈, 生成对抗网络	节点	实向量	关系感知, 直接从连续分布中采样“潜在”节点	监督 HeGAN: $O(\mathcal{N})$
	路径实例	ESim ^[76]	结构	随机游走	节点	实向量	捕捉在多条路径实例中共现节点的相似性	监督 $O(n(\mathcal{N} + \epsilon)L)$ (预处理) + $O(TtML^2d)$
	基于距离	分布 HINE ^[77]	结构	随机游走	节点	实向量	捕捉所有节点在游走和嵌入空间中的邻近关系	监督 $O(\mathcal{N} \cdot L^D)$
		双曲 HHNE ^[81]	结构	随机游走 + Skip-Gram	节点	实向量	节点嵌入到双曲空间, 利用双曲距离度量节点间的邻近度	监督 $O(WLknd \mathcal{N})$
		欧式 + 转化 RHINE ^[87]	结构	欧式距离 + 转化距离	节点, 关系	实向量	通过区分与分类处理异质关系保留不同关系独特的结构特征	监督 /
图卷积	交互式路径	IPE ^[88]	属性 + 结构	随机游走 + 交互式 GRU	节点, 交互路径	实向量	通过交互路径建模多条路径之间的相互依赖, 全面捕捉对象间的语义关系	监督 $O(\mathcal{N} \zeta + m \mathcal{P}(q,w) ^2ls)$
		Active-HNE ^[89]	结构 + 属性	图卷积神经网络(GCN), FCN	节点	实向量	分别学习每个分解网络中节点的嵌入, 然后将结果进行级联	半监督 /
		HANE ^[90]	结构 + 属性	GCN, 注意力机制	节点	实向量	输入属性变换到一个公共空间后统一学习节点的嵌入	监督 /

续表										
分类	方法名称	所用信息	所用技术	嵌入对象	表示形式	方法特点	学习方式	时间复杂度		
多层网络	MANE ^[57] , EOE ^[91] , PTE ^[92] , PMNE ^[93]	结构	交替优化,随机游走	节点	实向量	多个网络同时学习,任一网络都可以通过网络间的边向另一网络提供补充信息	半监督	MANE: $\#iter \times O(\sum_{i=1}^b \langle m_i, n_i + \sum_{j=1}^b (n_i^2 d_j + m_{ij} d_j) + n_i^2 d_i \rangle)$		
								EOE: $O(nd_u \times d_v \#iter)$		
多网络嵌入	多视图网络	同质	M-Deep-Walk ^[94] , MVE ^[95] , MNE ^[96] , Mvn2vec ^[97]	结构	DeepWalk + FCN, 注意力机制	节点	实向量	保护视图与视图之间的共享信息及每个视图中 的独特信息	半监督	MVE: $O(\epsilon \cdot dN)$ MNE: $O(M \mathcal{N})$ Mvn2vec: $O(\frac{ \tau \cdot L}{B(L-B)}) + O(\tau ^2)$
		异质	TransN ^[58]	结构	双重学习机制	节点	实向量	不同视图之间节点嵌入的投影,实现视图之间传递学习到的信息	无监督	$O(\delta Pp(z+z') + dPp(\varepsilon \log_2(\mu) + z'H_p))$
	多重网络	GATNE-T ^[59]	结构	随机游走 + Skip-Gram, 注意力机制	节点	实向量	不能处理未观察到的节点	监督	$O(nmdL)$	
		GATNE-I ^[59]	结构 + 属性	元路径随机游走 + Skip-Gram, 注意力机制	节点	实向量	能够处理训练期间未出现但具有原始特征的节点	监督	$O(nmdL)$	
	新兴网络	BL-MNE ^[99]	结构	随机游走 + 自编码	节点	实向量	通过成熟网络与新兴网络之间的信息交换,克服信息稀疏性问题	监督	/	
多分辨率	MrMine ^[60]	结构	随机游走 + Skip-Gram	节点, 子图, 网络	实向量	考虑不同网络 and 不同分辨率的内在关系	监督	$O(H \mathcal{N} \log(\mathcal{N}))$		
动态网络嵌入	Change2vec ^[63]	结构	增量学习	节点	实向量	适用于大规模动态网络	监督	/		

AHNG^[49]、IPE^[88]、ActiveHNE^[89]、HANE^[90]、GATNE-I^[59]适用于考虑属性的异质网络,其他方法适用于仅考虑结构的异质网络;多网络的嵌入方法利用多个网络提供的互补信息保存网络内的独特信息和网络间的共享信息;许多方法交叉使用了随机游走、分解和深度神经网络(比如 ActiveHNE^[89]将原始的 HeN 分解为同质网络和二分网络,然后应用图卷积神经网络分别学习每个分解网络中节点的深层语义,最后将所有网络中每个节点的输出向量进行级联),使用的深度模型包括全连接神经网络、感知器、门控单元、图卷积神经网络、深度自编码、生成对抗网络等,有的模型还引入了注意力机制;除了 TransN^[58]不需要任何标签信息外,其他模型都需要或部分需要标签信息;嵌入对象多种多样,包括节点、边、关系、元路径、节点对、子图和网络,比如 TaPEm^[73]将 Skip-Gram 模型的中心节点扩展为一对节点,上下文节点

扩展为上下文路径;多数模型都将对象嵌入欧几里得空间,但 M-HIN^[70]基于元图将对象嵌入复空间,HHNE^[81]将对象嵌入双曲空间,AHNG^[49]利用高斯分布捕捉由于异质属性引起的节点表示的不确定性;许多方法都基于随机游走和负采样完成嵌入,随机游走可以基于元路径、元图、网络模式或特征矩阵进行,其中基于网络模式的随机游走降低了用户设计合适元路径或元图的困难,特征游走将节点间的属性相似度转换为节点序列描述,从而自然融合网络的结构和节点的属性信息;R-NS^[65]使用可扩展的自适应负采样,有效避免传统负采样技术中的“流行邻居问题”;HeGAN^[66]和 AND-HeA^[67]通过生成对抗网络中的生成器选择高质量的负样本;IPE^[88]在不同路径的对象之间添加依赖关系实现路径之间的强耦合,全面捕捉对象间的语义关系;Change2vec^[63]将动态网络分解为具有不同时间戳的网络快照,通过学习各个快照中

节点的嵌入捕捉网络的动态变化.

7 数据集和测评指标

7.1 数据集

表 3 总结了本文所综述的各种异质网络表征学习模型使用的数据集,其中括号外的符号表示模型名称,括号内的符号表示模型使用的数据集.从表 3 可以看到,数据集 DBLP、AMiner、Blogcatalog、ACM、Douban、Yelp、Twitter、YouTube、Flickr 被

多个模型使用,不同的模型根据需要使用了数据集的不同子集,比如 PathSim^[78] 使用了 DBLP 的 DBLP-4-Area 子集,包含 20 个出版源、top-5000 名作者、5237 篇论文、4479 个主题,其中 1713 名作者有簇标签;HeteSpaceyWalk^[56] 使用的子集包含 14376 篇论文、20 个出版源、14475 名作者、8920 个术语和 171000 条边;EOE^[91] 采用的作者网络由 4941 位作者组成,共有 17372 位合著者,词网络的节点和边的数目分别为 6615 和 78217,作者和词之间的链接数为 92899.

表 3 异质网络表征学习模型使用的数据集

网络类型	模型(数据集)
结构 HeN(SHeN)	HeteSpaceyWalk (ACM、DBLP、Douban、Yelp) ^[56] , Metapath2vec (AMiner) ^[68] , HIN2Vec (Blogcatalog、Yelp、DBLP、U. S. Patent) ^[69] , M-HIN (DBLP、Yelp、YAGO、Freebase) ^[70] , TaPEm (AMiner-Top) ^[73] , ES-im (Yelp) ^[76] , HINE (MOVIE、Yelp、GAME) ^[77] , HHNE (DBLP-4-Area、MovieLens) ^[81] , RHINE (DBLP、Yelp、AMiner) ^[87] , TGPA (AMiner) ^[103] , HERec (Douban Movie、Douban Book、Yelp) ^[104] , HueRec (Yelp、MovieLens、Douban Movie) ^[105]
	AHNG (AMiner) ^[49] , IPE (LinkedIn、Facebook、Taobao) ^[88] , ActiveHNE (DBLP、Cora、MovieLens) ^[89] , HANE (AMiner) ^[90] , HNE (BlogCatalog、NUS-WIDE) ^[106]
属性 HeN(AHeN)	
多层网络 (MLN)	MANE (AMiner、Infra) ^[57] , EOE (DBLP、SLAP、BlogCatalog) ^[91] , PTE (20NG、Wiki、IMDB、Corporate、Economics、Government、Market、DBLP、MR、Twitter) ^[92] , PMNE (AUCS、Terrorist、Student、VC、LN) ^[93]
多视图网络 (MVN)	TransN (AMiner、BLOG、App-Daily、App-Weekly) ^[58] , M-DeepWalk (LN、Reinnovation、Congress Votes、UCI Mammography、UCI Balance Scale、Vickers、LAZEGA、C. ELEGANS、EU Airlines) ^[94] , MVE (DBLP、Flickr、PPI、YouTube、Twitter) ^[95] , MNE (CKM、Twitter、Private) ^[96] , Mvn2vec (YouTube、Snapchat) ^[97]
多重异质网络 (MHeN)	GATNE-T (Amazon、YouTube、Twitter、Alibaba) ^[59] , BL-MNE (Twitter、Foursquare) ^[99]
属性 MHeN (AMHeN)	GATNE-T (Amazon、YouTube、Twitter、Alibaba) ^[59] , SHINE (Weibo-STC、Wiki-RfA) ^[107]
多分辨率多网络 (M _R M _{IN})	MrMine (DBLP、Flickr、LastFm、Douban、MySpace、AMiner、Bioinformatics) ^[60]
异质特征网络 (HeFN)	FeatWalk (Reuters、Flickr、ACM、Yelp) ^[62]
动态异质网络 (DHeN)	Change2vec (DBLP、AMiner) ^[63]

7.2 测评指标

针对网络表征学习的不同应用,需使用不同的测评指标.有些应用还会使用多个测评指标共同评价网络表征学习方法的性能,表 4 概括了针对不同应用经常采用的测评指标,其中 AUC 度量一个正实例排名比随机选择的负实例排名靠前的概率,F1 度量精度和召回率的谐波均值(仅当精度和召回率都很高时,F1 才高),Recall@k 度量预测排名列表中排名靠前的 k 个真实实例在总真实实例数中所占的比例,Precision@k 度量预测排名列表中排名靠前的 k 个真实实例在总实例数中所占的比例,MAP@k 度量预测排名列表中排名靠前的 k 个真实实例的平均精度的平均,NDCG@k 是考虑正确推荐项目位置的测评指标. Precision@k 和 Recall@k 用于测试模型区分用户兴趣项的能力,而 MAP@k 和 NDCG@k 则侧重于优先项在排名列表中的位置. HT(命中率)表示推荐列表中至少具有一个正确推

荐项目的用户所占的百分比. MAE 表示平均绝对误差, RMSE 表示根均方误差, ARI 表示调整后的兰特指数, NMI 表示规范化互信息.

表 4 异质网络表征学习的测评指标

应用	测评指标
节点分类	Micro-F1、Macro-F1、Accuracy、F1
节点聚类	NMI、ARI、
链路预测	AUC、F1、Precision@k、MAP@k、Recall@k
网络重构	AUC、Precision@k
推荐	MAE、RMSE、HT、Precision@k、MAP@k、Recall@k、NDCG@k
社区检测	Density、Separability、Coverage、Expansion、Modularity、NMI
查询节点检索	Accuracy
网络对齐	Accuracy
作者身份识别	AUC、F1、Precision@k、MAP@k、Recall@k
可视化	t-SNE

8 典型应用

异质信息网络的嵌入表示能够较好地刻画网络中不同类型节点之间的复杂关联,因此广泛应用于各类任务场景,比如节点分类^[92]、节点聚类^[87]、链路预测^[69,107]、推荐系统^[104-105,108-113]、异常检测^[114]、作者识别^[73,103]、网络重构^[81]、查询节点检索^[60]、网络对齐^[60]、可视化^[87]等。

8.1 作者识别

许多同行评审的会议/期刊都采用双盲评审,即论文被匿名化,审稿人看不到论文的作者,只能看到论文的内容/属性(例如标题,文本信息和参考文献)。作者识别问题旨在设计一种模型来自动识别匿名论文的潜在作者,该问题在评阅系统、通用信息检索和推荐系统中有着广泛的意义。

在文献网络中,论文 p 在给定网络 G 中的邻居表示为 $\mathcal{X}_p = \{\mathcal{X}_p^{(1)}, \mathcal{X}_p^{(2)}, \dots, \mathcal{X}_p^{(R)}\}$, 其中 $\mathcal{X}_p^{(t)}$ 是第 t 种类型的邻居节点。在作者识别问题中,节点类型包括关键词、参考文献、会议和年份。 $\mathcal{A}_p$ 表示论文 p 真实的作者集。作者识别问题定义为:给定一组表示为 $(\mathcal{X}, \mathcal{A})$ 的论文,其中 $\mathcal{X} = \{\mathcal{X}_p\}$, $\mathcal{A} = \{\mathcal{A}_p\}$, 作者识别的目标是根据 $\mathcal{X}_p$ 中的信息学习一个能够对每篇匿名论文 p 的潜在作者进行排名的模型,使得排名高的作者在 $\mathcal{A}_p$ 中。

Chen 等^[103] 基于论文和作者的嵌入完成作者识别,模型框架如图 24 所示。模型首先将每个节点映射到潜在特征空间,然后根据其在网络中观察到的邻居,逐步构建匿名论文的特征表示,最后基于论文的特征表示计算潜在作者的评分。论文 p 的特征表示的学习包含两个聚合阶段:第一阶段通过平均 $\mathcal{X}_p^{(t)}$ 中的节点嵌入为第 t 种节点类型建立特征向量,即 $\mathbf{h}_p^{(t)} = \sum_{u \in \mathcal{X}_p^{(t)}} \mathbf{h}_u / |\mathcal{X}_p^{(t)}|$;第二阶段使用不同类型节点的特征向量的加权组合构建论文 p 的特征向量,即 $\mathbf{h}_p = \sum_i \omega_i \mathbf{h}_p^{(i)}$ 。论文 p 与潜在作者 a 之间的评分可以通过式(38)计算:

$$s(p, a) = \mathbf{h}_a^T \mathbf{h}_p \quad (38)$$

节点嵌入通过路径增强的通用网络嵌入完成,该方法考虑了多跳的元路径(例如作者^{撰写}论文^{包含}关键词),并基于元路径增强的邻接矩阵 $\{\mathbf{M}^{(r)}\}$ (每个 $\mathbf{M}^{(r)}$ 表示特定元路径 r 下的网络连通

性,对于 $\forall r$, $\sum_{i,j} \mathbf{M}_{i,j}^{(r)} = 1$) 中的权重随机采样获取节点序列,通过最大化节点间基于元路径可达概率学习节点嵌入。

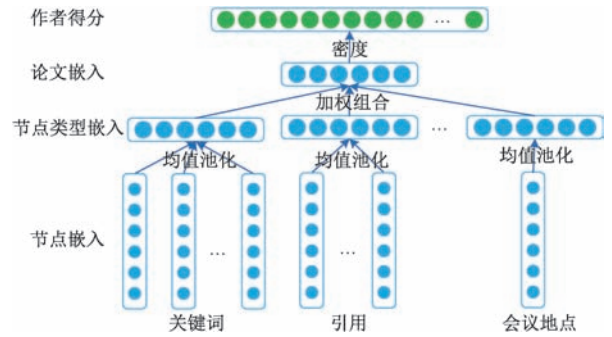


图 24 基于特定任务嵌入的作者识别框架^[103]

8.2 推荐

在面向推荐的异质信息网络 $G = (\mathcal{N}, \epsilon)$ 中,用户和项以及它们之间的评分关系需要特别关注。设 $\mathcal{U} \subset \mathcal{N}$ 和 $\mathcal{I} \subset \mathcal{N}$ 分别是用户和项的集合,元组 $\langle u, i, r_{u,i} \rangle$ 表示用户 $u \in \mathcal{U}$ 对项 $i \in \mathcal{I}$ 的评分为 $r_{u,i}$, $\mathcal{K} = \{\langle u, i, r_{u,i} \rangle\}$ 为评分记录的集合。给定 G , 推荐系统的目标是预测 u 对未评分项 $i' \in \mathcal{I}$ 的评分 $r_{u,i'}$ 。

8.2.1 基于独立嵌入的推荐

基于独立嵌入的推荐根据每条元路径独立地为用户和项学习低维向量表示,然后融合多条元路径的嵌入完成推荐。

HERec^[104] 基于元路径的随机游走生成只包含用户或项的同质节点序列,通过最大化节点与给定元路径采样的序列中相邻节点的共现概率学习节点的嵌入表示。设元路径集合为 $\mathcal{P}$, $\mathbf{h}_v^{(l)} (v = u, i)$ 表示节点 $v \in \mathcal{N}$ 关于第 l 条元路径的表示向量, $\{\mathbf{h}_v^{(l)}\}_{l=1}^{|\mathcal{P}|}$ 是节点 v 的表示集,该表示集通过变换函数 $g(\cdot)$ 变换为用户 u 和项 i 的异质嵌入 $\mathbf{h}_u^{(0)} \leftarrow g(\{\mathbf{h}_u^{(l)}\})$ 和 $\mathbf{h}_i^{(0)} \leftarrow g(\{\mathbf{h}_i^{(l)}\})$, $g(\cdot)$ 可以是简单的线性融合 $g(\{\mathbf{h}_v^{(l)}\}) = \frac{1}{|\mathcal{P}|} \sum_{l=1}^{|\mathcal{P}|} (\mathbf{M}^{(l)} \mathbf{h}_v^{(l)} + \mathbf{b}^{(l)}) (v = u, i)$, 也可以是加权线性融合,或加权非线性融合,其中 $\mathbf{M}^{(l)} \in \mathbb{R}^{D \times d}$, $\mathbf{b}^{(l)} \in \mathbb{R}^D$ 分别是关于第 l 条元路径的变换矩阵和偏置向量。

用户对项的评分 $\hat{r}_{u,i}$ 通过矩阵分解计算:

$$\hat{r}_{u,i} = \mathbf{x}_u^T \cdot \mathbf{y}_i + \alpha \cdot (\mathbf{h}_u^{(0)})^T \cdot \boldsymbol{\gamma}_i^{(0)} + \beta \cdot (\boldsymbol{\gamma}_u^{(0)})^T \cdot \mathbf{h}_i^{(0)} \quad (39)$$

其中 $\mathbf{x}_u \in \mathbb{R}^d$ 和 $\mathbf{y}_i \in \mathbb{R}^d$ 表示与用户 u 和项 i 对应的隐因子, $\boldsymbol{\gamma}_u^{(0)}$ 和 $\boldsymbol{\gamma}_i^{(0)}$ 是用户侧和项目侧的潜在因

子,分别与异质嵌入 $\mathbf{h}_u^{(u)}$ 和 $\mathbf{h}_i^{(i)}$ 对应, α 和 β 是整合三参数的调整参数.

HERec 的框架如图 25 所示. 学习模型参数的目标函数为

$$\begin{aligned} \mathcal{L} = & \sum_{(u,i,r_{u,i}) \in \mathcal{R}} (r_{u,i} - \hat{r}_{u,i})^2 + \lambda \sum_u (||\mathbf{x}_u||_2 + ||\mathbf{y}_i||_2 \\ & + ||\boldsymbol{\gamma}_u^{(u)}||_2 + ||\boldsymbol{\gamma}_i^{(i)}||_2 + ||\boldsymbol{\Theta}^{(u)}||_2 + ||\boldsymbol{\Theta}^{(i)}||_2) \end{aligned} \quad (40)$$

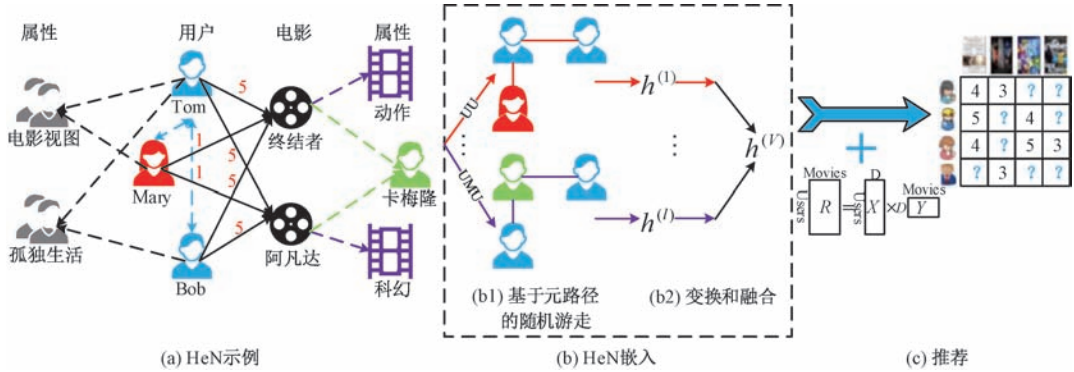


图 25 HERec 的框架^[104]

路径之间的关系缓解一条元路径上的数据稀疏性和噪声的问题. HueRec 的框架如图 26 所示, 图中 $\mathbf{C}_p$ 是基于元路径 p 的交换矩阵, $c_{i,j,p}$ 表示对象 x_i 通过元路径 p 到达对象 y_j 的路径实例数, $\hat{c}_{i,j,p}$ 表示对应的估计值. $\mathbf{R}$ 是反馈矩阵, 如果用户 u_i 对项目 i_j 有反馈, 则 $r_{i,j}=1$, 否则 $r_{i,j}=0$.

设 $\mathbf{h}_u^u \in \mathbb{R}^d, \mathbf{h}_i^f \in \mathbb{R}^d, \mathbf{h}_p^p \in \mathbb{R}^d$ 分别表示用户 u 、项 i 和路径 p 的 d 维潜在表示, 路径计数 $c_{u,i,p}$ 估计为

$$\hat{c}_{u,i,p} = f(\mathbf{h}_u^u, \mathbf{h}_i^f, \mathbf{h}_p^p) = \sum_{q=1}^d h_{u,q}^u h_{i,q}^f h_{p,q}^p \quad (41)$$

$h_{u,q}^u$ 表示用户潜在向量 $\mathbf{h}_u^u$ 的第 q 位, 其余类似. 用户 u 对项 i 的兴趣使用评分的加权总和预测:

$$\hat{r}_{u,i} = \sum_{p=1}^l (\omega_{u,p} + \omega_p) \hat{c}_{u,i,p} \quad (42)$$

其中 $\omega_{u,p}$ 是用户 u 对路径 p 的个性化权重, ω_p 是路径 p 的权重. $\omega_{u,p}$ 和 ω_p 通过学习获得, 学习的损失函数定义为

$$\begin{aligned} Loss_{Total} = & \frac{1}{|\mathcal{U}| |\mathcal{I}|} \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}} \\ & (Loss_{u,i}^{Rec} + \lambda_1 Loss_{u,i}^{HeN}) + \lambda_2 ||\boldsymbol{\Theta}||_2^2 \end{aligned} \quad (43)$$

其中, $Loss_{u,i}^{HeN}$ 是用户-项目对 (u,i) 的损失函数, $Loss_{u,i}^{Rec}$ 用于度量关于元路径的偏好学习和推荐结果的有效性. $Loss_{Total}$ 使用随机优化算法 Adam^[75] 最小化.

其中 λ 是正则参数, $\boldsymbol{\Theta}^{(u)}$ 和 $\boldsymbol{\Theta}^{(i)}$ 分别是函数 $g(\cdot)$ 关于用户和项的参数. $\mathcal{L}$ 可以通过随机梯度下降算法进行最小化.

8.2.2 基于统一嵌入的推荐

基于统一嵌入的推荐通过直接将用户和项嵌入到一个统一的潜在空间完成推荐. HueRec^[105] 基于每个用户或项在不同的元路径下都有一些共同特征的假设, 将用户和项嵌入到一个统一的空间, 利用元

8.3 情感链接预测

在线社交网络中, 人们经常表达对他人的态度, 这在用户之间形成了巨大的情感联系. 与表示朋友或跟随关系的显式社交连接不同, 情感链接是用户发布的语义内容所隐含的, 并且涉及不同的类型: 积极情感链接表示喜欢、信任或支持的态度, 消极情感链接表示不喜欢或不赞成他人. 积极或消极情感可以通过正或负两种符号表示. 在个人广告和舆论分析等许多领域, 预测情感联系的符号是一项基本任务. SHINE^[107] 通过实体级情感提取方法建立了一个由用户的情感关系、社交关系和个人资料组成的情感数据集, 利用三个深度自编码器^[72] 分别从情感网络 G_s 、社交网络 G_r 和资料网络 G_p 中提取用户 u 的非线性表示 $\hat{\mathbf{x}}_u$ 、 $\hat{\mathbf{y}}_u$ 和 $\hat{\mathbf{z}}_u$, 再通过聚合函数对这些表示进行融合得到最终的节点表示 $\mathbf{h}_u$, 比如 $\mathbf{h}_u = \hat{\mathbf{x}}_u + \hat{\mathbf{y}}_u + \hat{\mathbf{z}}_u$, 然后通过计算 $\mathbf{h}_u$ 和 $\mathbf{h}_v$ 的相似度 (例如, 内积 $\bar{s}_{uv} = \mathbf{h}_u^T \mathbf{h}_v + b$ 或欧几里得距离 $\bar{s}_{uv} = -||\mathbf{h}_u - \mathbf{h}_v||_2 + b$, 其中 b 为可训练的偏差参数) 完成 u 和 v 之间的情感预测. SHINE 的框架如图 27 所示, 整个模型基于预测情感和真实情感进行训练.

9 总结和展望

本文对近年来关于异质信息网络表征学习的网络类型进行归纳, 介绍了异质信息网络表征学习所

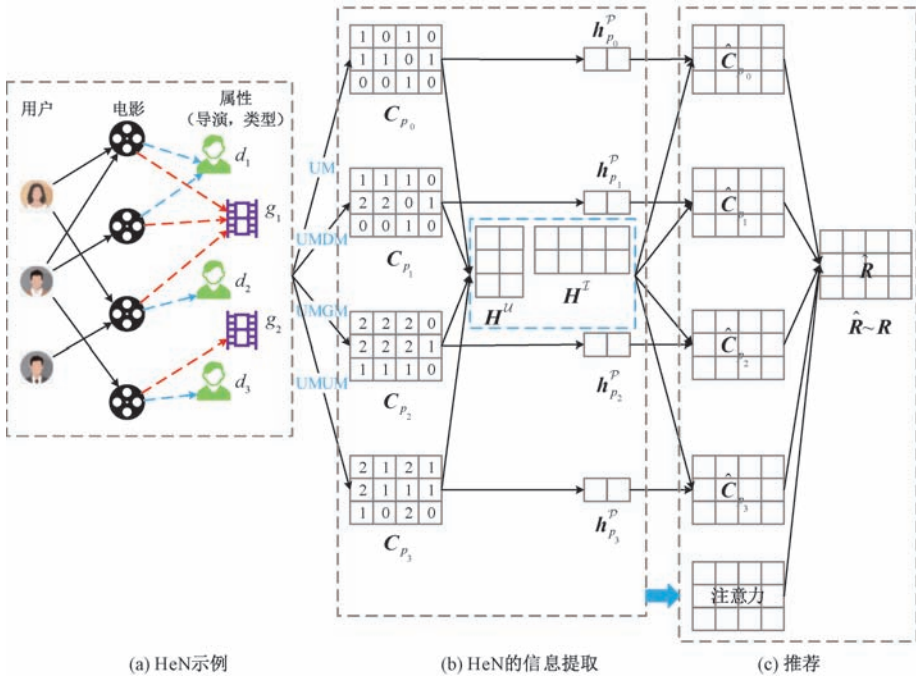


图 26 HueRec 的框架^[105]

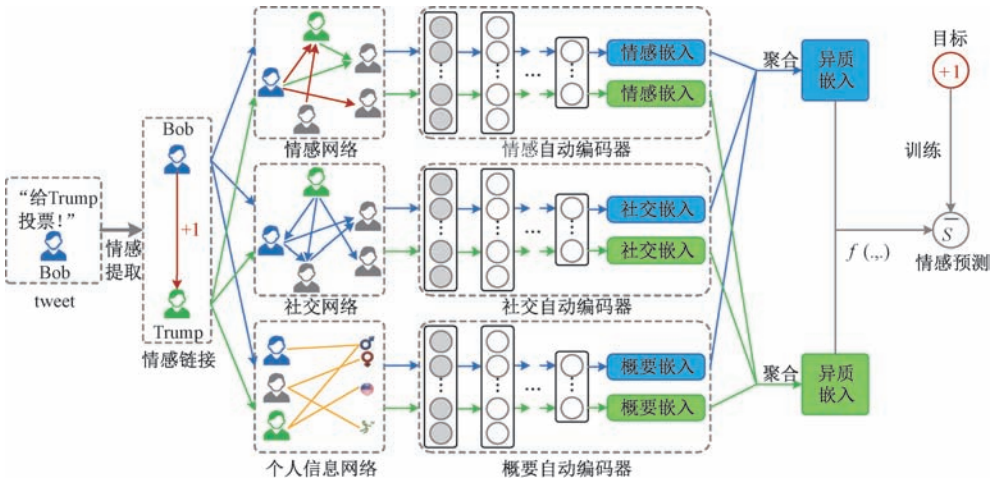


图 27 端到端的 SHINE 框架(图中仅显示三个自编码器的编码器部分,省略了解码器部分)^[107]

涉及的相关概念,阐述了当前的主流方法与模型,对各种方法的特点进行分析,概述了相关数据集与测评指标,同时示例了典型应用场景.本文工作有助于研究人员对异质信息网络表征学习有一个全面的了解,奠定进一步研究的基础.

尽管异质信息网络表征学习已有大量研究成果,但是异质信息网络的表征学习还有广阔的研究空间.未来的发展方向包括:

(1)融合多种信息的表征学习.异质信息网络中包含不同类型的节点、连边及属性信息,各类信息异质、互补,从不同的角度刻画网络的意义,体现了大数据的“多样性”.只有将多种信息考虑进来,学习到

的网络节点表示才能更加准确地对节点进行描述.多种异质信息的融合对异质信息网络分析的质量起着举足轻重的作用.

(2)大规模动态网络的表征学习.目前网络表征学习相关工作主要集中于小规模静态网络,但是现实世界中的网络往往规模较大并且动态变化.研究面向异质信息网络的增量计算和在线计算的表征学习方法将是一个重要的研究方向.

(3)节点、关系、元路径重要性度量.异质信息网络中不同类型的节点、连边、属性和元路径对于不同的分析任务具有不同的重要性.研究如何根据具体应用任务自动度量不同节点和连边在不同任务中的

重要性,捕捉不同类型关系中的一致性和互补性,以及如何有效地自动筛选和融合不同元路径下的网络节点表示是一个值得关注的研究方向。

(4)不同表示空间的耦合. 异质信息网络的表征学习中,不同类型的节点、连边有不同的特征表示空间,不同特征表示空间之间的耦合关系有助于完整了解对象之间的拓扑结构,准确度量节点之间的邻近关系,捕捉网络的复杂语义关系. 当前的研究对多种语义关系的关联及表示空间之间的耦合关系考虑不够,还需深入探究。

(5)其他信息网络的表征学习. 近年来,许多新型网络模型相继出现,比如公共—私有网^[115-116]、不确定图^[117]. 在一个公共—私有网络(例如,Facebook)中,存在一个公共图 G ,其包含网络的所有用户可见的节点集和边集;每个节点 v 与一个私有图 G_v 关联,其中 G_v 的节点是来自公共图 G 的顶点,而 G_v 只为 v 所知. 在不确定图中,由于噪声的存在,节点之间是否存在边是不确定的,因此每条边都与表示其存在概率的值相关联. 目前,还没有关于这类网络的表征学习研究。

(6)结合具体应用. 目前许多应用主要基于通用表征学习完成,在一些特定任务(比如社区发现、异常检测等)中效果往往不尽人意,因此研究如何将节点表示学习与具体应用结合、将异质网络表征学习技术真正应用于实际业务、以及增强算法结果的可解释性都是值得关注的研究方向^[48]。

(7)代码和数据集的在线资料库. 当前许多表征学习算法的代码和数据集均未公开. 因此,期望建立一个代码和数据集的在线资料库. 一方面,研究人员可以使用资料库中的代码和数据集作为研究的基础;另一方面,从业人员可以轻松地将资料库中的代码插入其应用程序直接应用,无需重新实现。

参 考 文 献

- [1] Zhang L, Zhang Q, Zhang L, et al. Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding. *Pattern Recognition*, 2015, 48(10): 3102-3112
- [2] Qi J, Liang X, Li Z, et al. Representation learning of large-scale complex information network: Concepts, methods and challenges. *Chinese Journal of Computers*, 2018, 41(10): 2394-2420 (in Chinese)
(齐金山, 梁循, 李志宇, 等. 大规模复杂信息网络表示学习: 概念、方法与挑战. *计算机学报*, 2018, 41(10): 2394-2420)
- [3] Huang J-T, Sharma A, Sun S, et al. Embedding-based retrieval in facebook search//*Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. CA, USA, 2020: 2553-2561
- [4] Jin D, You X, Li W, et al. Incorporating network embedding into markov random field for better community detection//*Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Hawaii, USA, 2019: 160-167
- [5] Sun H, He F, Huang J, et al. Network embedding for community detection in attributed networks. *ACM Transactions on Knowledge Discovery from Data*, 2020, 14(3): 36:1-36:25
- [6] Zhao W X, Hou Y, Chen J, et al. Learning semantic representations from directed social links to tag microblog users at scale. *ACM Transactions on Information Systems*, 2020, 38(2): 17:1-17:30
- [7] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations//*Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA, 2014: 701-710
- [8] Tang J, Qu M, Wang M, et al. Line: Large-scale information network embedding//*Proceedings of the 24th International Conference on World Wide Web*. Florence, Italy, 2015: 1067-1077
- [9] Grover A, Leskovec J. Node2vec: Scalable feature learning for networks//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, USA, 2016: 855-864
- [10] Qiu J, Dong Y, Ma H, et al. Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec//*Proceedings of the 11th ACM International Conference on Web Search and Data Mining*. Marina Del Rey, USA, 2018: 459-467
- [11] Qiu J, Dong Y, Ma H, et al. Netsmf: Large-scale network embedding as sparse matrix factorization//*Proceedings of the 28th International Conference on World Wide Web*. San Francisco, USA, 2019: 1509-1520
- [12] Shen X, Pan S, Liu W, et al. Discrete network embedding//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm, Sweden, 2018: 3549-3555
- [13] Tu C, Zeng X, Wang H, et al. A unified framework for community detection and network representation learning. *IEEE Transactions on Knowledge and Data Engineering*, 2019, 31(6): 1051-1065
- [14] Chanpuriya S, Musco C. Infinitewalk: Deep network embeddings as Laplacian embeddings with a nonlinearity//*Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Virtual Event, USA, 2020: 1325-1333
- [15] Wang C, Wang C, Wang Z, et al. Edge2vec: Edge-based social network embedding. *ACM Transactions on Knowledge Discovery from Data*, 2020, 14(4): 45:1-45:24
- [16] Yang C, Liu Z, Zhao D, et al. Network representation learning with rich text information//*Proceedings of the 24th International Joint Conference on Artificial Intelligence*. Buenos

- Aires, Argentina, 2015; 2111-2117
- [17] Cao S, Lu W, Xu Q. Grarep: Learning graph representations with global structural information//Proceedings of the 24th ACM International Conference on Information and Knowledge Management. Melbourne, Australia, 2015; 891-900
- [18] Cao S, Lu W, Xu Q. Deep neural networks for learning graph representations//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Arizona, USA, 2016; 1145-1152
- [19] Wang D, Cui P, Zhu W. Structural deep network embedding//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016; 1225-1234
- [20] Pan S, Wu J, Zhu X, et al. Tri-party deep network representation//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, USA, 2016; 1895-1901
- [21] Huang X, Li J, Hu X. Accelerated attributed network embedding//Proceedings of the 2017 SIAM International Conference on Data Mining. Houston, USA, 2017; 633-641
- [22] Huang X, Li J, Hu X. Label informed attributed network embedding//Proceedings of the 10th ACM International Conference on Web Search and Data Mining. Cambridge, UK, 2017; 731-739
- [23] Wang X, Cui P, Wang J, et al. Community preserving network embedding//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA, 2017; 203-209
- [24] Liao L, He X, Zhang H, et al. Attributed social network embedding. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(12): 2257-2270
- [25] Gao H, Huang H. Deep attributed network embedding//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018; 3364-3370
- [26] Zhang Z, Yang H, Bu J, et al. Anrl: Attributed network representation learning via deep neural networks//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018; 3155-3161
- [27] Zhou S, Yang H, Wang X, et al. Prre: Personalized relation ranking embedding for attributed networks//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino, Italy, 2018; 823-832
- [28] Li Y, Wang Y, Zhang T, et al. Learning network embedding with community structural information//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China, 2019; 2937-2943
- [29] Jin D, Li B, Jiao P, et al. Network-specific variational auto-encoder for embedding in attribute networks//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China, 2019; 2663-2669
- [30] Meng Z, Liang S, Zhang X, et al. Jointly learning representations of nodes and attributes for attributed networks. ACM Transactions on Information Systems, 2020, 38(2): 16:1-16:32
- [31] Shen X, Chung F-L. Deep network embedding for graph representation learning in signed networks. IEEE Transactions on Cybernetics, 2020, 50(4): 1556-1568
- [32] Shi M, Tang Y, Zhu X, et al. Topic-aware web service representation learning. ACM Transactions on the Web, 2020, 14(2): 9:1-9:23
- [33] Wen W, Huang J, Cai R, et al. Graph embedding by incorporating prior knowledge on vertex information. Journal of Software, 2018, 29(3): 786-798 (in Chinese)
(温雯, 黄家明, 蔡瑞初等. 一种融合节点先验信息的图表示学习方法. 软件学报, 2018, 29(3): 786-798)
- [34] Wang H, Wang J, Wang J, et al. Graphgan: Graph representation learning with generative adversarial nets//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018; 2508-2515
- [35] Yan R, Shen H, Qi C, et al. Graphwgan: Graph representation learning with wasserstein generative adversarial networks//Proceedings of the IEEE International Conference on Big Data and Smart Computing. Busan, Korea, 2020; 315-322
- [36] Li Z, Liang X, Xu Z, et al. Dnps: Damping based negative-positive sampling of large-scale dynamic social network embedding. Chinese Journal of Computers, 2017, 40(4): 805-823 (in Chinese)
(李志宇, 梁循, 徐志明等. DNPS:基于阻尼采样的大规模动态社会网络结构特征表示学习. 计算机学报, 2017, 40(4): 805-823)
- [37] Cui Y, Li J, Shen L, et al. Duration-hyte: a time-aware knowledge representation learning method based on duration modeling. Journal of Computer Research and Development, 2020, 57(6): 1239-1251 (in Chinese)
(崔员宁, 李静, 沈力等. Duration-HyTE:基于持续时间建模的时间感知知识表示学习方法. 计算机研究与发展, 2020, 57(6): 1239-1251)
- [38] Hamilton W L, Ying R, Leskovec J. Representation learning on graphs: Methods and applications. IEEE Data Engineering Bulletin, 2017, 40(3): 52-74
- [39] Tu C, Yang C, Liu Z, et al. Network representation learning: An overview. Scientia Sinica Informationis, 2017, 47(8): 980-996 (in Chinese)
(涂存超, 杨成, 刘知远等. 网络表示学习综述. 中国科学: 信息科学, 2017, 47(8): 980-996)
- [40] Goyal P, Ferrara E. Graph embedding techniques, applications, and performance: A survey. Knowledge Based Systems, 2018, 151: 78-94
- [41] Cai H, Zheng V W, Chang K C-C. A comprehensive survey of graph embedding: Problems, techniques, and applications. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(9): 1616-1637

- [42] Cui P, Wang X, Pei J, et al. A Survey on Network Embedding. *IEEE Transactions on Knowledge and Data Engineering*, 2019, 31(5): 833-852
- [43] Zhang D, Yin J, Zhu X, et al. Network Representation Learning: A Survey. *IEEE Transactions on Big Data*, 2020, 6(1): 3-28
- [44] Shi C, Sun Y, Yu P. Research status and future development of heterogeneous information network. *Communications of the CCF*, 2017, 13(11): 35-40 (in Chinese)
(石川, 孙怡舟, Yu P S. 异质信息网络的研究现状和未来发展. *中国计算机学会通讯*, 2017, 13(11): 35-40)
- [45] Shi C, Yu P S. *Heterogeneous Information Network Analysis and Applications*. Data Analytics, Switzerland: Springer International Publishing AG, 2017
- [46] Shi C, Li Y, Zhang J, et al. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(1): 17-37
- [47] Sun Y, Han J. *Mining Heterogeneous Information Networks: Principles and Methodologies*. Beijing: China Machine Press, 2017 (in Chinese)
(孙艺洲, 韩家玮. 异构信息网络挖掘: 原理和方法. 北京: 机械工业出版社, 2017)
- [48] Shi C, Sun Y. Research progress of heterogeneous information network embedding. *Communications of the CCF*, 2018, 14(3): 16-21 (in Chinese)
(石川, 孙怡舟. 异质网络表征学习的研究进展. *中国计算机学会通讯*, 2018, 14(3): 16-21)
- [49] Liu M, Liu J, Chen Y, et al. Ahng: Representation learning on attributed heterogeneous network. *Information Fusion*, 2019, 50: 221-230
- [50] Chen L, Zhu P, Qian T, et al. Edge sampling based network embedding model. *Journal of Software*, 2018, 29(3): 756-771 (in Chinese)
(陈丽, 朱裴松, 钱铁云, 等. 基于边采样的网络表示学习模型. *软件学报*, 2018, 29(3): 756-771)
- [51] Yan S, Liu C, Jiang A, et al. Discriminative cross-modal Hashing with coupled semantic correlation. *Chinese Journal of Computers*, 2019, 42(1): 164-175 (in Chinese)
(严双咏, 刘长红, 江爱文, 等. 语义耦合相关的判别式跨模态哈希学习算法. *计算机学报*, 2019, 42(1): 164-175)
- [52] Dong Y, Hu Z, Wang K, et al. Heterogeneous network representation learning//*Proceedings of the 29th International Joint Conference on Artificial Intelligence*. Yokohama, Japan, 2020: 4861-4867
- [53] Yang C, Xiao Y, Zhang Y, et al. Heterogeneous network representation learning; Survey, benchmark, evaluation, and beyond. *IEEE Transactions on Knowledge and Data Engineering*, CoRR, 2020, abs/2004.00216
- [54] Park C, Kim D, Han J, et al. Unsupervised attributed multiplex network embedding//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York, USA, 2020: 5371-5378
- [55] Fang Y, Lin W, Zheng V W, et al. Semantic proximity search on graphs with metagraph-based learning//*Proceedings of the 32nd IEEE International Conference on Data Engineering*. Helsinki, Finland, 2016: 277-288
- [56] He Y, Song Y, Li J, et al. Hetspaceywalk: A heterogeneous spacey random walk for heterogeneous information network embedding//*Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. Beijing, China, 2019: 639-648
- [57] Li J, Chen C, Tong H, et al. Multi-layered network embedding//*Proceedings of the 2018 SIAM International Conference on Data Mining*. San Diego, USA, 2018: 684-692
- [58] Li Z, Zheng W, Lin X, et al. Transn: Heterogeneous network representation learning by translating node embeddings//*Proceedings of the 36th IEEE International Conference on Data Engineering*. Dallas, USA, 2020: 589-600
- [59] Cen Y, Zou X, Zhang J, et al. Representation learning for attributed multiplex heterogeneous network//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Anchorage, USA, 2019: 1358-1368
- [60] Du B, Tong H. Mrmine: Multi-resolution multi-network embedding//*Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. Beijing, China, 2019: 479-488
- [61] Shervashidze N, Schweitzer P, Leeuwen E J v, et al. Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research*, 2011, 12: 2539-2561
- [62] Huang X, Song Q, Yang F, et al. Large-scale heterogeneous feature embedding//*Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Hawaii, USA, 2019: 3878-3885
- [63] Bian R, Koh Y S, Dobbie G, et al. Network embedding and change modeling in dynamic heterogeneous networks//*Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Paris, France, 2019: 861-864
- [64] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality//*Proceedings of the 27th Annual Conference on Neural Information Processing Systems*. Nevada, USA, 2013: 3111-3119
- [65] Armandpour M, Ding P, Huang J, et al. Robust negative sampling for network embedding//*Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Hawaii, USA, 2019: 3191-3198
- [66] Hu B, Fang Y, Shi C. Adversarial learning on heterogeneous information networks//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Anchorage, USA, 2019: 120-129
- [67] Wang H, Wang R, Wen C, et al. Author name disambiguation on heterogeneous information network with adversarial representation learning//*Proceedings of the 34th AAAI Con-*

- ference on Artificial Intelligence. New York, USA, 2020: 238-245
- [68] 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. Halifax, Canada, 2017: 135-144
- [69] Fu T-Y, Lee W-C, Lei Z. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning//Proceedings of the 26th ACM International Conference on Information and Knowledge Management. Singapore, 2017: 1797-1806
- [70] Fang Y, Zhao X, Huang P, et al. M-Hin: Complex embeddings for heterogeneous information networks via meta-graphs//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris, France, 2019: 913-916
- [71] Yang C, Feng Y, Li P, et al. Meta-graph based hin spectral embedding: Methods, analyses, and insights//Proceedings of the 18th IEEE International Conference on Data Mining. Singapore, 2018: 657-666
- [72] Salakhutdinov R, Hinton G E. Semantic hashing. International Journal of Approximate Reasoning, 2009, 50(7): 969-978
- [73] Park C, Kim D, Zhu Q, et al. Task-guided pair embedding in heterogeneous network//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China, 2019: 489-498
- [74] Cho K, Merriënboer B v, Gülçehre Ç, et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar, 2014: 1724-1734
- [75] Kingma D P, Ba J. Adam: A method for stochastic optimization//Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA, 2015
- [76] Shang J, Qu M, Liu J, et al. Meta-path guided embedding for similarity search in large-scale heterogeneous information networks. CoRR, 2016, abs/1610.09769
- [77] Huang Z, Mamoulis N. Heterogeneous information network embedding for meta path based proximity. CoRR, 2017, abs/1701.05291
- [78] Sun Y, Han J, Yan X, et al. Pathsime: Meta path-based top-k similarity search in heterogeneous information networks//Proceedings of the 37th International Conference on Very Large Data Bases. Seattle, USA, 2011: 992-1003
- [79] Lao N, Cohen W W. Relational retrieval using a combination of path-constrained random walks. Machine Learning, 2010, 81(1): 53-67
- [80] Recht B, Ré C, Wright S J, et al. Hogwild: A lock-free approach to parallelizing stochastic gradient descent//Proceedings of the 25th Annual Conference on Neural Information Processing Systems. Granada, Spain, 2011: 693-701
- [81] Wang X, Zhang Y, Shi C. Hyperbolic heterogeneous information network embedding//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Hawaii, USA, 2019: 5337-5344
- [82] Cannon J W, Floyd W J, Kenyon R, et al. Hyperbolic geometry. Flavors of geometry, 1997, 31: 59-115
- [83] Bonnabel S. Stochastic gradient descent on riemannian manifolds. IEEE Transactions on Automatic Control, 2013, 58(9): 2217-2229
- [84] Bronstein M M, Bruna J, LeCun Y, et al. Geometric deep learning: Going beyond euclidean data. IEEE Signal Processing Magazine, 2017, 34(4): 18-42
- [85] Nickel M, Kiela D. Poincaré embeddings for learning hierarchical representations//Proceedings of the 31st Annual Conference on Neural Information Processing Systems. Long Beach, USA, 2017: 6338-6347
- [86] Krioukov D V, Papadopoulos F, Kitsak M, et al. Hyperbolic geometry of complex networks. Physical Review E, 2010, 82(3): 036106
- [87] Lu Y, Shi C, Hu L, et al. Relation structure-aware heterogeneous information network embedding//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Hawaii, USA, 2019: 4456-4463
- [88] Liu Z, Zheng V W, Zhao Z, et al. Interactive paths embedding for semantic proximity search on heterogeneous graphs//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. London, UK, 2018: 1860-1869
- [89] Chen X, Yu G, Wang J, et al. Activehne: Active heterogeneous network embedding//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China, 2019: 2123-2129
- [90] Wang Y, Duan Z, Liao B, et al. Heterogeneous attributed network embedding with graph convolutional networks//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Hawaii, USA, 2019: 10061-10062
- [91] Xu L, Wei X, Cao J, et al. Embedding of embedding (Eoe): joint embedding for coupled heterogeneous networks//Proceedings of the 10th ACM International Conference on Web Search and Data Mining. Cambridge, UK, 2017: 741-749
- [92] Tang J, Qu M, Mei Q. Pte: Predictive text embedding through large-scale heterogeneous text networks//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, Australia, 2015: 1165-1174
- [93] Liu W, Chen P-Y, Yeung S, et al. Principled multilayer network embedding//Proceedings of the 17th IEEE International Conference on Data Mining Workshops. New Orleans, USA, 2017: 134-141
- [94] Song H, Thiagarajan J J. Improved deep embeddings for inferring with multi-layered graphs//Proceedings of the 2019

- IEEE International Conference on Big Data. Los Angeles, USA, 2019: 5394-5400
- [95] Qu M, Tang J, Shang J, et al. An attention-based collaboration framework for multi-view network representation learning//Proceedings of the 26th ACM International Conference on Information and Knowledge Management. Singapore, 2017: 1767-1776
- [96] Zhang H, Qiu L, Yi L, et al. Scalable multiplex network embedding//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018: 3082-3088
- [97] Shi Y, Han F, He X, et al. Mvn2vec: Preservation and collaboration in multi-view network embedding. CoRR, 2018, abs/1801.06597
- [98] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. *Nature*, 1986, 323 (6088): 533-536
- [99] Zhang J, Xia C, Zhang C, et al. Bl-Mne: Emerging heterogeneous social network embedding through broad learning with aligned autoencoder//Proceedings of the 17th IEEE International Conference on Data Mining. New Orleans, USA, 2017: 605-614
- [100] Watts D J, Strogatz S H. Collective dynamics of 'Small-World' networks. *Nature*, 1998, 393(6684): 440-442
- [101] Huang H, Tang J, Liu L, et al. Triadic closure pattern analysis and prediction in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(12): 3374-3389
- [102] Zhou L-k, Yang Y, Ren X, et al. Dynamic network embedding by modeling triadic closure process//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 571-578
- [103] Chen T, Sun Y. Task-guided and path-augmented heterogeneous network embedding for author identification//Proceedings of the 10th ACM International Conference on Web Search and Data Mining. Cambridge, UK, 2017: 295-304
- [104] Shi C, Hu B, Zhao W X, et al. Heterogeneous information network embedding for recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2019, 31(2): 357-370
- [105] Wang Z, Liu H, Du Y, et al. Unified embedding model over heterogeneous information network for personalized recommendation//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China, 2019: 3813-3819
- [106] Chang S, Han W, Tang J, et al. Heterogeneous network embedding via deep architectures//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, Australia, 2015: 119-128
- [107] 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. Marina Del Rey, USA, 2018: 592-600
- [108] Zhao H, Yao Q, Li J, et al. Meta-graph based recommendation fusion over heterogeneous information networks//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax, Canada, 2017: 635-644
- [109] Zhang Y, Ai Q, Chen X, et al. Joint representation learning for top-N recommendation with heterogeneous information sources//Proceedings of the 26th ACM International Conference on Information and Knowledge Management. Singapore, 2017: 1449-1458
- [110] Chen T, Sun Y, Shi Y, et al. On sampling strategies for neural network-based collaborative filtering//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax, Canada, 2017: 767-776
- [111] Jiang Z, Liu H, Fu B, et al. Recommendation in heterogeneous information networks based on generalized random walk model and bayesian personalized ranking//Proceedings of the 11th ACM International Conference on Web Search and Data Mining. Marina Del Rey, USA, 2018: 288-296
- [112] Li Y, Chen W, Yan H, et al. Learning graph-based embedding for personalized product recommendation. *Chinese Journal of Computers*, 2019, 42(8): 1767-1778 (in Chinese)
(李宇琦, 陈维政, 闫宏飞, 等. 基于网络表示学习的个性化商品推荐. *计算机学报*, 2019, 42(8): 1767-1778)
- [113] Qian T, Liang Y, Li Q, et al. Attribute graph neural networks for strict cold start recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2020: 1-14
- [114] Chen T, Tang L-A, Sun Y, et al. Entity embedding-based anomaly detection for heterogeneous categorical events//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, USA, 2016: 1396-1403
- [115] Archer A, Lattanzi S, Likarish P, et al. Indexing public-private graphs//Proceedings of the 26th International Conference on World Wide Web. Perth, Australia, 2017: 1461-1470
- [116] Chierichetti F, Epasto A, Kumar R, et al. Efficient algorithms for public-private social networks//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, Australia, 2015: 139-148
- [117] Hu J, Cheng R, Huang Z, et al. On embedding uncertain graphs//Proceedings of the 26th ACM International Conference on Information and Knowledge Management. Singapore, 2017: 157-166



ZHOU Li-Hua, Ph. D. , professor, Ph. D. supervisor. Her research interests include data mining, social network analysis, and artificial intelligence.

WANG Jia-Long, Ph. D. candidate. His research interests include data mining and social network analysis.

WANG Li-Zhen, Ph. D. , professor, Ph. D. supervisor. Her research interests include data mining, data warehouse,

and computer algorithms.

CHEN Hong-Mei, Ph. D. , associate professor. Her research interests include data structure, database technology, and data mining. Corresponding Author.

KONG Bing, Ph. D. , associate professor. His research interests include data mining, social network analysis, and machine learning.

Background

Information network representation learning aims to embed nodes into a low-dimensional vector space by using information such as network topology and node content, while preserving the intrinsic structural and content features of the original network, so that network analysis tasks, such as node classification, clustering and link prediction, etc. can be completed based on low-dimensional dense vectors. A large number of existing studies on the network representation learning are dedicated to homogeneous information networks that contain only one node type and one edge type. With the rapid development of information technology, various heterogeneous information networks, consisting of multiple types of nodes and links, can be seen everywhere. These heterogeneous information networks contain more comprehensive and rich structure and semantic information, which can model real-world network data more completely and naturally. Recently the research on the representation learning and application of heterogeneous information networks has received increasing attention and has become an important research topic in network analysis area. A lot of research methods on the representation learning of heterogeneous information net-

work have been proposed.

This paper deeply studies various kinds of representative models and methods for the representation learning of heterogeneous information networks in recent years, analyzes their intrinsic characteristics, reviews typical applications, data sets and evaluation metrics, and points out new research directions. This survey does not only help researchers to have a better understanding of existing work on heterogeneous information network representation learning, but also provides practitioners a better judgement on choosing the proper embedding models and algorithms.

This work is in part supported by the National Natural Science Foundation of China with No. 62062066, No. 61762090, No. 61966036, Program for Innovation Research Team (in Science and Technology) in University of Yunnan Province with No. IRTSTYN, the National Social Science Foundation of China with No. 18XZZ005, University Key Laboratory of Internet of Things Technology and Application, Yunnan Province and the Postgraduate Research and Innovation Foundation of Yunnan University with No. 2021Y024.