

基于 K 阶互信息估计的位置感知网络表征学习

储晓恺^{1,2} 范鑫鑫² 毕经平²

¹(中国科学院大学 北京 100049)

²(中国科学院计算技术研究所 北京 100190)

(chuxiaokai@ict.ac.cn)

Position-Aware Network Representation Learning via K -Step Mutual Information Estimation

Chu Xiaokai^{1,2}, Fan Xinxin², and Bi Jingping²

¹(University of Chinese Academy of Sciences, Beijing 100049)

²(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

Abstract As the network data grows rapidly and persistently, also affiliated with more sophisticated applications, the network representation learning, which aims to learn the high-quality embedding vectors, has become the popular methodology to perform various network analysis tasks. However, the existing representation learning methods have little power in capturing the positional/locational information of the node. To handle the problem, this paper proposes a novel position-aware network representation learning model by figuring out center-rings mutual information estimation to plant the node's global position into the embedding, PMI for short. The proposed PMI encourages each node to respectively perceive its K -step neighbors via the maximization of mutual information between this node and its step-specific neighbors. The extensive experiments using four real-world datasets on several representative tasks demonstrate that PMI can learn high-quality embeddings and achieve the best performance compared with other state-of-the-art models. Furthermore, a novel neighbor alignment experiment is additionally provided to verify that the learned embedding can identify its K -step neighbors and capture the positional information indeed to generate appropriate embeddings for various downstream tasks.

Key words network representation learning; mutual information estimation; node representation; information network analysis; neural networks; node classification; link prediction

摘要 随着网络结构数据持续、快速的增长,各种复杂网络数据分析与应用层出不穷.近年来,网络表征学习已经成为各类网络分析任务的主流方法.网络表征学习的主要目标是依据节点间连接关系,学习高质量的节点表征向量,从而辅助分析下游任务.然而,现有的表征学习方法未考虑节点在网络中的位置信息.为了解决这一问题,提出了一种位置感知网络表征学习模型 PMI,该模型通过最大化每个中心节点与各阶邻居之间的互信息,从而将节点的位置信息学入表征向量中.在表征训练过程中,PMI 模型激励每个中心节点记住并识别其每阶的邻居节点,从而间接记录其位置信息.在 4 个不同领域的真实数据集上进行了多标签分类、网络重构、链接预测等多个代表性网络分析任务实验,实验结果表明提出的 PMI 模型可以学到高质量的节点表征向量,与现有的表征学习模型相比,PMI 模型能够在多个下游

收稿日期:2021-04-01;修回日期:2021-06-01

基金项目:国家自然科学基金项目(62077044,61702470,62002343)

This work was supported by the National Natural Science Foundation of China (62077044,61702470, 62002343).

通信作者:毕经平(jpingbi@ict.ac.cn)

任务上有较大幅度提升.此外,还设计邻居对齐任务对 PMI 模型进行进一步的分析,结果表明 PMI 模型学到的节点表征能够有效识别不同阶的邻居节点并捕获自身的位置信息,从而为各种下游任务生成合理有效的表征.

关键词 网络表征学习;互信息估计;节点表征;信息网络分析;神经网络;节点分类;链接预测

中图法分类号 TP391

图(网络)是节点与边的集合,其中节点代表了一类特定的事物,而边则代表了节点与节点之间的关联.现实世界中的大量数据均可抽象为图结构数据,比如社交网络、生物网络、学术网络等.为了更好地支持这些基于图结构数据的应用和分析,如何合理有效地表征网络节点成为了关键所在.传统的独热码方式(one-hot),由于其空间复杂度会随着节点规模的扩增而快速增大,严重限制了大规模网络数据处理和分析的能力,并且独热码无法表达节点间的相关性,间接影响了下游任务的效果.近年来,随着深度神经网络和表征学习在图像、自然语言处理等领域取得重大的成功,不少研究者开始关注如何对图中的节点进行低维表征,使得该表征可以保留网络节点间的相关性,即网络表征学习^[1-3].现如今,网络表征学习已经在多种网络处理和分析任务上证明其有效性,包括网络节点分类^[4-5]、链接预测^[6-7]、社区发现^[8-9]等.

迄今为止,学术界已经提出许多网络表征学习方法^[1,10-11].大部分方法通常关注如何建模节点的近邻信息.然而,少有方法重视节点的位置信息在表征中的作用.对于每个节点,它相对于网络中其他节点的最短距离标定了该节点在网络中的位置.相较于近邻信息,节点的位置信息包含更宽的感受野和更全面的结构上下文信息,而这类信息在一些与结构分析相关的数据挖掘任务中是至关重要的.如图 1 所示,如果只重视近邻信息,那么用户(节点)A 和 B 因其直接相连将会获得相似的节点表征.然而,当我们考虑两者的位置信息,会发现他们在不同的距离上有着完全不同的邻居节点,这意味两者的表征应具有较大的差异性.因此,如果表征模型可以感知到 2 个节点在各阶邻居上的差异,那么学到的表征将能够支撑更多与网络结构或与节点距离相关的下游任务.然而,现有的方法通常只关注节点的近邻信息,虽然部分方法(如 GraRep^[12])通过计算节点间转移概率的方式捕捉中心节点与高阶节点间的相关性,但这种方式却模糊了各阶邻居间的界限,使得模型仍然无法捕捉节点的位置信息.

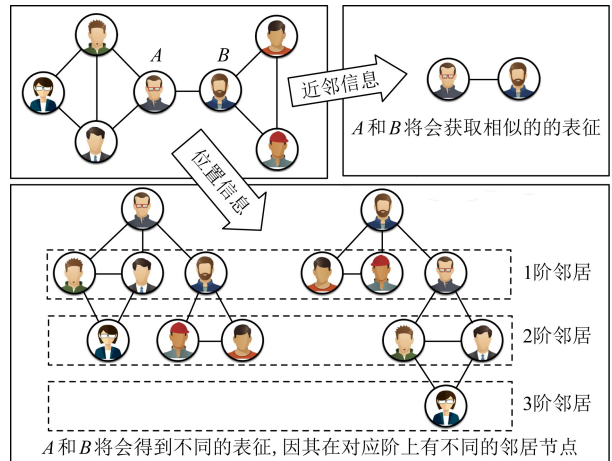


Fig. 1 The importance of positional information in network embedding

图 1 位置信息在节点表征中的作用

为了将节点的位置信息嵌入到节点表征中,本文提出了一种基于 k 阶互信息估计^[13-14]的位置感知网络表征学习模型(position-aware network representation learning via k -step mutual information estimation, PMI),该模型通过最大化节点与其各阶邻居之间的互信息,从而捕捉到节点的全局位置信息.其直观解释在于:处于不同网络位置的节点,在不同阶结构上通常拥有不同的邻居,因此在模型训练的过程中,激励每个节点去分别记住并区别自身在各阶的邻居节点.通过这种方式,模型可以从节点在各阶邻居获取不同的上下文信息,从而学得更具区分性的节点表征.本文的主要贡献涵盖 3 个方面:

1) 提出了一种基于 K 阶互信息最大化的表征学习模型 PMI,该模型不仅可以保留节点的局部结构信息,还可以捕捉节点的全局位置信息;

2) 为了捕捉节点的位置信息,PMI 最大化每个节点与其 K 阶邻居间的互信息,从而让每个节点能够识别并记住其各阶邻居;

3) 在 4 种不同类型的真实网络数据集上进行多种代表性网络分析实验,包括链接预测、多标签分类、网络重构等.实验结果表明 PMI 可以有效地获取高质量的节点表征.此外,在另一项邻居匹配任务

上的实验结果表明:相较于现有的工作,本文提出的 PMI 模型可以有效地捕捉节点的位置信息.

1 相关工作

1.1 网络表征学习

随着网络数据的大规模增加,近年来大量研究者对网络节点表征开展了研究^[1,6,10].本文主要针对同质网络下的网络表征学习问题,即节点与边的类型有且仅有一种.大部分研究类比自然语言和网络节点在统计特征上的相似性^[1],提出基于语言模型的网络表征学习.文献[1]提出了 DeepWalk 模型,首先通过随机游走(random walks)获取节点序列,再利用语言模型 SkipGram^[15]获取节点的表征.在 DeepWalk 的基础上,文献[10]修改了节点的采样策略,提出了一种带偏好的随机游走策略(biased random walks),该策略权衡了广度优先搜索(breadth first search, BFS)和深度优先搜索(depth first search, DFS).除此之外,一些文献也通过直接建模邻居间的相关性来学习节点表征.基于近邻相似的假设,文献[13]着重捕捉一阶或二阶邻居间的关系,并采用负采样策略对大型网络进行训练.文献[12]通过计算节点的 K 阶转移矩阵来建模高阶节点间的相关性,并通过矩阵分解获取节点表征.一些工作^[11]也利用自动编码器来学习节点表征,通过最小化自动编码器的输入和输出来学习节点的表征.文献[11]证明这种方式等同于学习节点间的二阶关系.不同以上基于近邻假设的方法,文献[16]认为空间结构相似的节点在表征空间中也应当相近,基于结构相似性假设,作者建立一个多层次的带权重网络来衡量节点间的结构相似性.然而,上述工作均无法捕捉节点的位置信息.针对该问题,本文提出一种基于 K 阶互信息估计的网络表征学习方法,最大化节点与其各阶邻居间的互信息,让每个节点能够记住并识别来自不同阶的邻居,从而将位置信息融入节点表征中.

除了上述针对同质网络的表征学习方法,一些研究者提出在其他不同类型网络场景下的表征学习方法,包括异构信息网络^[17-18]、多层网络^[19-20]、符号网络^[21-23]、属性网络^[24-25]等.此外,一些文献着手于为不同的应用设计对应的表征学习方法,如社区发现^[26-27]、结构识别^[28-29]等.近几年,图神经网络(graph neural networks, GNNs)^[30-32]成为了属性网络表征

学习的热门,其在半监督学习下的图分类、节点分类等问题上超越了传统方法,取得了优异的效果.

1.2 互信息估计

互信息(mutual information)是一种基于香农熵的度量方法,它捕捉变量之间的非线性依赖关系,通常可以用于度量 2 个变量间的真实依赖性.对 2 个变量 X 和 Y 的互信息估计等价于计算两者联合分布 \mathcal{J} 和边缘分布 \mathcal{M} 的 K-L 散度(Kullback-Leibler divergence):

$$I(X;Y) = KL(\mathcal{J} | \mathcal{M}). \quad (1)$$

尽管式(1)在形式上十分简洁,但它的精确计算仅适用于离散变量或是已明确先验分布的连续变量,因此,它的适用场景较为有限.为解决该问题,文献[33]提出一种基于神经网络的互信息估计模型 MINE,利用 DV(Donsker-Varadhan)表征推导式(1)的下界:

$$I(X;Y) \geq E_{\mathcal{J}}[D_{\mathbf{M}}(x,y)] - \log E_{\mathcal{M}}[e^{D_{\mathbf{M}}(x,y)}], \quad (2)$$

其中, $D_{\mathbf{M}}(\cdot, \cdot)$ 是参数为 \mathbf{M} 的双线性函数,该方法在学习更好的生成模型方面表现出了很强的一致性.

受到 MINE 的启发,文献[14]首次将互信息的神经估计应用到无监督图像表征学习中,并提出 DIM 模型.作者认为在很多场景下,并不需要关心具体的互信息值,而更关心变量间互信息的相对大小.因此, DIM 模型用 J-S 散度(Jensen-Shannon divergence)替代 K-L 散度,并采用一种对抗训练的方式最大化正样本间的互信息:

$$E_{\mathcal{P}}[-s\mathcal{P}(-D_{\mathbf{M}}(x,y))] - E_{\mathcal{P} \times \bar{\mathcal{P}}}[s\mathcal{P}(D_{\mathbf{M}}(z,y))], \quad (3)$$

其中, $s\mathcal{P}(x) = \log(1 + e^x)$ 为 softplus 激活函数,变量 z 为采样出的负样本.通过式(3), DIM 最终最大化每张图片的全局特征与其每个像素表征间的互信息,并取得了优异的效果.

受到 MINE 的启发,文献[34]提出模型 DGI 并首次将互信息估计应用到属性网络节点表征学习中,最大化网络全局表征与节点局部表征间的互信息.而文献[35]提出了模型 GMI,该模型最大化中心节点表征与其一阶邻居属性间的互信息.本文所提出的 PMI 模型与上述 DGI 和 GMI 模型存在较大的不同,如图 2 所示:1)本文针对的是网络结构表征,并不包含属性信息;2)本文方法通过最大化中心节点与其各阶邻居间的互信息,从而获取每个节点的位置信息.

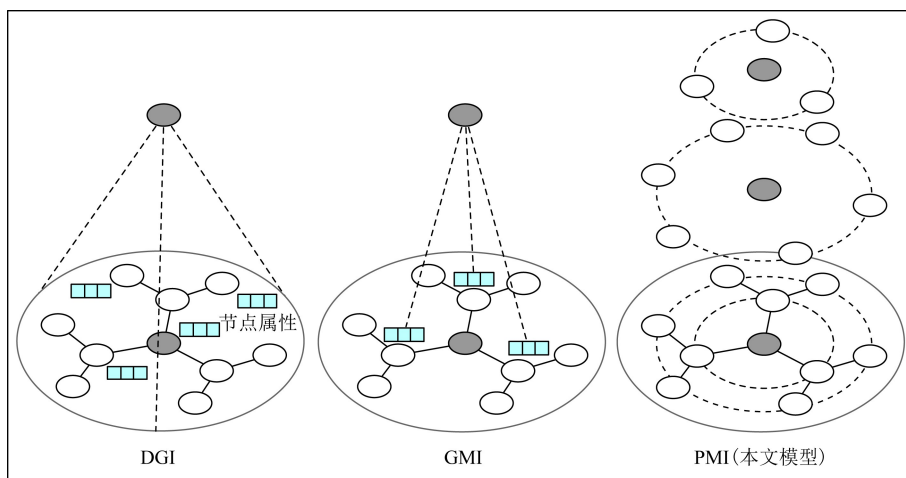


Fig. 2 Difference from current mutual information-based embedding models

图 2 本文工作与现有基于互信息的图表征模型的不同

2 问题定义

首先给出本文方法涉及到的相关定义,表 1 列出了本文的主要符号以及对应的含义.

Table 1 Primary Notations and Explanations

表 1 主要符号和对应含义

符号名	含义
G	同质网络
A	同质网络对应的邻接矩阵
v_i	节点 v_i
N_i^k	节点 v_i 的 k 阶邻居集合
X	网络的表征矩阵
$X_{i,\cdot}, x_i$	节点 v_i 对应的表征向量
$M^k, W^k, b^{(i)}$	待学习的参数矩阵
$D_M(\cdot, \cdot)$	以 M 为参数的双线性函数
$R(\cdot)$	readout 函数
$\Phi(\cdot)$	衰减函数
d	表征维度
a, b	损失函数的超参数
K	最大阶数
n_z	负采样数目

定义 1. 同质网络. 给定一个无向图 $G = (V, E; A)$, 其中 $V = \{v_1, v_2, \dots, v_n\}$ 代表节点集, E 代表边集合, 矩阵 A 是图 G 的邻接矩阵, 其中 $A_{i,j} = 1$ 当且仅当节点 v_1 和 v_2 间存在一条边, 否则 $A_{i,j} = 0$.

本文方法通过最大化每个节点与其各阶邻居间的互信息学习节点的表征, 其中第 k 阶邻居和网络表征学习的定义分别如下:

定义 2. k 阶邻居. 给定节点 v_i , 它的 k 阶邻居定义为集合 $N_i^k = \{v_{i_1}^k, v_{i_2}^k, \dots\}$ 其中任一节点到节点 v_i 的图上最短距离为 k , 其中 $k \in \{1, 2, \dots, K\}$, 其中 K 为文章中考虑到的最远距离. 同时, 为了简洁起见, 当后续谈及 k 阶邻居时, 我们简称节点 v_i 为中心节点.

定义 3. 网络表征学习. 给定图 G , 网络表征学习旨在学习图 G 中节点的表征矩阵 $X \in \mathbb{R}^{n \times d}$, 其中 d 是表征维度, 其中矩阵的每一行向量 $X_{i,\cdot}$ 表示节点 v_i 的表征. 在后续的论述中为了简明起见, 也用 x_i 代表 v_i 的表征.

现有的网络表征学习模型无法学习到节点的位置信息. 为了解决该问题, 我们认为关键在于训练的过程中, 适当地激励节点记住并区别其在各阶上的邻居节点. 为此, 我们尝试将互信息神经估计融入网络表征学习中, 为各种下游分析任务生成位置感知的节点表征.

3 基于互信息估计的位置感知表征学习方法

3.1 总体框架

如图 3 所示, 本文提出的 PMI 模型主要分为 2 个模块: 1) K 阶互信息估计模块用于捕捉每个节点的位置信息; 2) 网络重构模块用于学习网络的基本结构信息. 因此, 最终优化函数也分为 2 部分:

$$J = \alpha J_{MI} + \beta J_{AE}, \quad (4)$$

其中 J_{MI} 代表 K 阶互信息估计模块的优化函数, 而 J_{AE} 则对应了网络重构模块的优化目标, 超参数 α

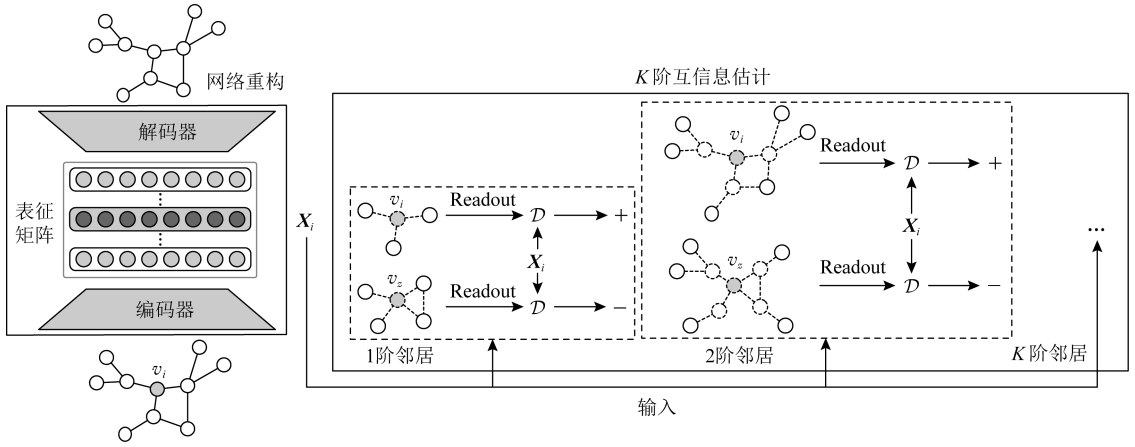


Fig. 3 An illustration of the proposed model PMI

图3 本文提出的PMI模型

和 β 用来权衡两者.接下来,我们逐一介绍面向 K 阶邻居的互信息最大化和基于自动编码机的结构学习模块,最后对本文方法的复杂度进行分析并给出优化方案.

3.2 面向 K 阶邻居的互信息估计

为了捕捉节点的位置信息,PMI模型在训练过程中让每个节点感知并记录其在第 k 阶的邻居节点.具体来说,我们采取一种环形的互信息估计策略,依次最大化每个节点与其各阶邻居间的互信息.另一方面,由于节点的高阶邻居数目会随着阶数增长而快速增长,直接估算节点与每个邻居的互信息是极其耗时的,无法扩展到大型网络.因此,我们可以从一个宏观的角度,将节点的第 k 阶邻居视作一个整体,最大化中心节点与第 k 阶邻居整体的互信息.具体来说,给定一个节点 v_i 和第 k 阶邻居集合 $N_i^k = \{v_{i_1}^k, v_{i_2}^k, \dots\}$,首先使用一个Readout函数 $R(\cdot)$ 规约出所有 k 阶邻居的特征:

$$y_i^k = R(\{x_{i_1}^k, x_{i_2}^k, \dots\}), \quad (5)$$

其中,每个 $x_{i_j}^k$ 对应某个第 k 阶邻居节点 $v_{i_j}^k$,取各维度的均值作为Readout函数.关于 $x_{i_j}^k$ 的获取方式,将会在3.4节详细描述.

接下来,介绍最大化中心节点 v_i 与第 k 阶邻居节点集合的互信息.在式(3)的基础上,我们定义互信息估算公式为

$$I(N_i^k; v_i) = E_p[-sp(-D_{M^k}(y_i^k, x_i))] - \sum_{v_z \sim \tilde{\mathcal{P}}} E_{\tilde{\mathcal{P}}}[sp(D_{M^k}(y_z^k, x_i))], \quad (6)$$

其中,节点 v_z 是按照分布 $\tilde{\mathcal{P}}$ 随机采集的负样本节点,该分布服从均匀分布且有 $\mathcal{P} = \tilde{\mathcal{P}}$,而 n_z 则表示对每个中心节点 v_i 所采集的负样本的数目, M^k 表

示对应于第 k 阶邻居的双线性函数的参数矩阵.

式(6)鼓励每个中心节点 v_i 通过区分其真实的第 k 阶邻居和虚假的第 k 阶邻居(即某个其他节点 v_z 的第 k 阶邻居),从而达到让节点 v_i 的表征携带其 k 阶邻居的信息.但值得注意的是式(6)是非对称的, $I(N_i^k; v_i)$ 和 $I(v_i; N_i^k)$ 有着不同的物理含义. $I(N_i^k; v_i)$ 可以看作是模型要求第 k 阶邻居去记住中心节点的信息,因此,从直观理解式(6)更符合本文方法的目的.同时,在实际应用中,我们也发现 $I(N_i^k; v_i)$ 优于 $I(v_i; N_i^k)$.

因此,面向第 k 阶邻居的互信息最大化的优化目标定义为

$$J_{MI}^k = \arg \max_{M^k, X} \sum_{v_i \in V} I(N_i^k; v_i), \quad (7)$$

其中 X 是待学习的表征矩阵.最终,总优化目标定义为与阶数 k 相关的加权和:

$$J_{MI} = \sum_{k=1}^K \Phi(k) J_{MI}^k. \quad (8)$$

考虑到位于更高阶的邻居节点对于中心节点的影响力往往越弱,在式(8)中,我们添加了一个与阶数相关的衰减函数 $\Phi(k)$ 来控制不同阶的邻居对于中心节点影响力的强弱,该衰减函数满足 $\Phi(k+1) \leq \Phi(k)$,并且当 $k > 0$ 时有 $\Phi(k) > 0$.

3.3 基于自动编码机的结构学习

面向 K 阶邻居的互信息最大化模块尽管可以让节点表征获取节点对应的位置信息,但无法学习到网络中最基本的结构信息,即边信息,而边信息在多种下游任务中是极其重要和基础的,丢失该信息将会极大地影响下游任务的效果.因此,PMI模型需要另一个模块来学习网络中的基本结构信息.尽管

当前对于结构信息的表征学习模型^[2,11-12]层出不穷,但我们发现自动编码器模型是最契合本文方法的一类模型,因为它可以直接建模近邻间的关系而不对互信息估计产生影响.同时,这也有助于我们设计一种端到端的学习模式.

尽管已有大量工作基于不同假设设计不同的自编码器模型^[11,36-37],但 PMI 模型仅需要一个能够捕捉基本结构信息的模块,因此我们采用一种简单的自编码器模型,通过直接重建网络中的边来学习网络结构.

自编码器的编码器部分采用 2 层的 DNN(deep neural network),其中第 i 层的 DNN 定义为

$$\mathbf{X}^{(i)} = \sigma(\mathbf{W}^{(i)} \mathbf{X}^{(i-1)} + \mathbf{b}^{(i)}), \quad (9)$$

其中, $\mathbf{X}^{(i)}$ 是第 i 层 DNN 的输入特征,且有 $\mathbf{X}^{(0)} = \mathbf{A}$ 为网络的邻接矩阵, $\mathbf{W}^{(i)}$ 和 $\mathbf{b}^{(i)}$ 为待学习的参数, $\sigma(\cdot)$ 为激活函数,这里我们使用修正线性单元 $ReLU(\cdot)$ 作为激活函数.第 2 层 DNN 的输出就是节点的表征 $\mathbf{X} = \mathbf{X}^{(2)}$.

对于解码器模块,我们采用一个简单的点乘形式重构整个网络:

$$\mathbf{Z} = \sigma(\mathbf{X}^T \mathbf{X}), \quad (10)$$

其中 $\sigma(\cdot) = 1/(1 + e^{-x})$ 为 sigmoid 函数.这种简单的点乘形式不仅有效减少本模型的参数量,而且可以避免繁重的编码器预训练过程^[11,38].

最后,自编码器的优化目标为重构出的网络与真实网络的距离:

$$AE = \arg \min_{\mathbf{W}^{(i)}, \mathbf{b}^{(i)}} \|\mathbf{A} - \mathbf{Z}\|_2. \quad (11)$$

3.4 复杂度分析

模型的复杂度主要分为 2 个部分: K 阶互信息估计模块和自动编码器模块.对于前者,每一阶的互信息估算的复杂度为 $O(d^2 |V|)$,对于自动编码器模块, DNN 层的计算复杂度为 $O(d^{(i-1)} d^{(i)} |V|)$,其中 $d^{(i-1)}$ 和 $d^{(i)}$ 代表第 i 层 DNN 的输入和输出维度.因此,整个模型的复杂度为

$$O(Kd^2 |V| + \sum_{i=1}^2 d^{(i-1)} d^{(i)} |V| + d |V|^2).$$

4 实验评估

为了评估本文提出的 PMI 模型的有效性,我们将模型训练得到的节点表征作为下游网络分析任务

的输入,测试其在多标签分类、链接预测和网络重构任务上的效果.此外,为了验证学到的表征是否能够捕捉其全局位置信息,我们在 4.5 节进行对于表征中位置信息的进一步分析,测试模型在邻居对齐任务上的效果.最后,我们对 PMI 模型中的参数敏感度进行实验分析.

4.1 实验设置

我们使用的真实网络数据来自多个不同领域,统计数据如表 2 所示,各个数据集的具体描述如下:

- 1) PPI^[10].该网络是蛋白质相互作用网络(protein-protein interactions)的子图,数据来源于文献[3].
- 2) IMDB^[39].该网络数据抽取自 MovieLens^①,其中每个节点代表一部电影,节点间的连边表示 2 部电影出自同一位导演.
- 3) WordNet[®].该网络为维基百科中的单词共现网络,每个节点表示一个单词,而单词间的连接表示 2 个单词在同一个文档中共同出现.
- 4) VoteNet^[1].该数据来自于斯坦福的公开数据集 who-votes-on-whom,是一个投票网络.

Table 2 Network Statistics

表 2 各数据集的网络统计数据

数据集	节点数	边数	直径	类别数
PPI	3 890	38 739	8	50
IMDB	18 352	1 104 162	8	9
WordNet	4 777	184 812	3	40
VoteNet	7 115	103 689	7	

为了验证 PMI 模型的效果,我们与当前 7 种具有代表性的网络表征学习方法进行比较,包括基于随机游走的方法(DeepWalk^[1]和 Node2Vec^[10])、基于近邻估计的方法(LINE^[2]和 GraRep^[12])、基于自动编码器的方法(SDNE^[11])和基于互信息的方法(DGI^[34]和 GMI^[35]).为了公平比较,包括本文方法在内的所有方法的节点表征维度 d 均设置为 128.对于模型 Node2Vec,我们采用 grid search 方法在参数空间 $p, q \in \{0.25, 0.5, 1, 2, 4\}$ 中搜索其最佳参数;对于 LINE,我们分别优化 1 阶近邻和 2 阶近邻,并将两者学到的表征进行拼接获取最终表征;对于模型 GraRep,我们同样采用 grid search 在参数空间 $K \in \{1, 2, 3, 4, 5\}$ 中搜索其最佳超参.由于本文 PMI 模型仅考虑网络结构信息,我们将归一化后的

① <https://grouplens.org/datasets/movielens/>

② <http://www.matmahoney.net/dc/textdata/>

邻接矩阵作为模型 SDNE, DGI, GMI 和 PMI 的输入.此外,遵循文献[11]中的策略,我们使用深度置信网络(deep belief network, DBN)^[38]对 SDNE 模型进行参数初始化.本文方法的其他超参设置为:最大阶数 $K=3$,负采样数目 $n_z=5$,自动编码器的各层输出维度分别为 256 和 128,超参数 $\alpha=\beta=1$,并使用 $\Phi(k)=k^{-1}$ 作为衰减函数.

4.2 多标签分类

我们首先测试各模型在多标签分类任务上的效果.采用 multi-class SVM^[40]作为模型分类器,并将节点表征作为 SVM 的输入特征来训练分类器.选择 20% 的标签节点作为训练样本,剩下的 80% 用作测试和效果评估.对于每个模型,重复上述实验 5 次并计算每个模型的 *Micro-F₁*,实验结果如表 3 所示.结果表明,本文方法在所有数据集上均取得最好的效果,在各数据集上分别有 1.63%, 1.36% 和 1.59% 的提升.进一步分析可以发现:基于随机游走的方法

(DeepWalk 和 Node2Vec)在各数据集上的效果优于其他对比方法;而 LINE 和 SDNE 的表现也较为稳定,但由于这 2 种方法过于强调近邻相似,导致整体效果不佳;相比之下, GraRep 的表现各个数据集上非常不稳定,比如其在 PPI 上的分类效果较差,这主要因为 GraRep 引入的高阶信息中包含了大量的噪音,这些信息影响了节点分类的效果,也影响了节点表征的区分性;另外 2 种基于互信息的方法 DGI 和 PMI 在多标签分类任务上的表现也欠佳,这 2 种方法的表征依赖于高质量的节点属性信息,当输入的网络结构信息并不包含具有鲜明特性的局部结构信号时,它们无法学到具有区分性的节点表征.相比之下,由于本文方法采用了一种 K 阶互信息最大化策略,其目的是让每个节点记住它的各阶邻居,PMI 模型可以有效地从上下文信息中学到结构相似性,从而最终在多标签分类任务上取得较好的效果.

Table 3 Results of Multi-label Classification

数据集	DeepWalk	Node2Vec	LINE	SDNE	GraRep	DGI	GMI	PMI	%
IMDB	52.40	53.89	50.73	50.33	52.74	45.70	46.34	55.52	
PPI	16.49	18.26	15.13	10.92	10.59	11.56	13.82	19.62	
WordNet	43.45	46.35	45.72	38.37	44.20	40.45	40.89	47.94	

注:黑体数字表示最优的结果.

4.3 网络重构

网络重构任务测试学到的节点表征是否可以保留原网络的结构信息.在本文中,实验设置与文献[11]类似:每个模型依据学到的表征相似性推理出 k 条可能在网络中存在的边,计算精确率 *precision*

@ k ,该指标表示在 k 次预测的边中有多少在原网络中是真实存在的.我们以 IMDB 和 VoteNet 数据集为例,比较每个模型的推理精确率.

如图 4 所示,PMI 模型在 2 个数据集上都展现出更加稳定的提升.尽管 SDNE 和 LINE 在 k 较小

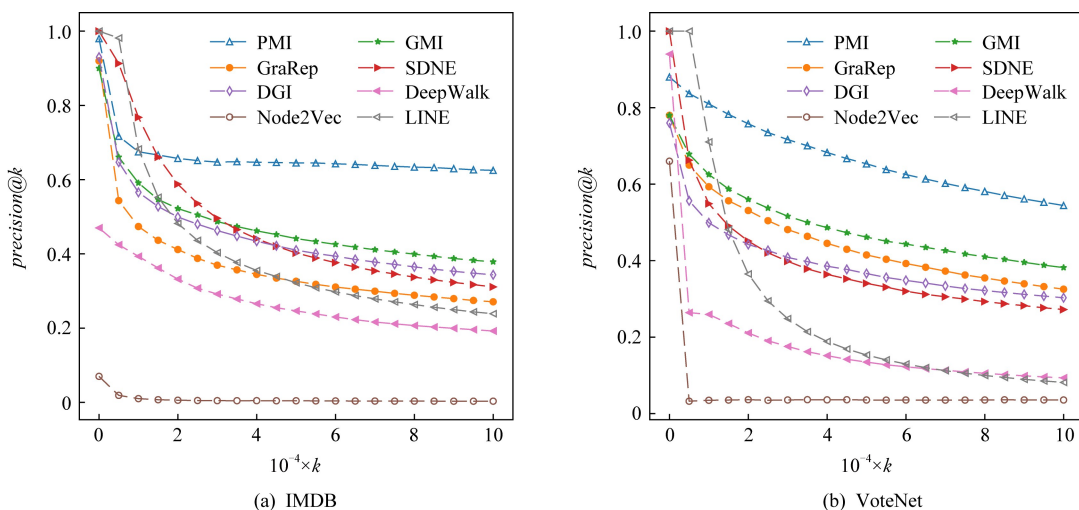


Fig. 4 Performance on network reconstruction

图 4 网络重构结果

时表现较优,但当更多的边需要被推测时(即更大的 k),2 种方法的精确率都出现大幅度下降.另一个比较有趣的发现是,2 种基于互信息的方法(DGI 和 GMI)也取得了较为不错的效果,这说明通过加强节点与局部邻居的互信息或增强节点与全局结构的互信息都有助于表征获取丰富的网络结构信息.另一方面,PMI 模型由于采用一种 K 阶互信息最大化的策略,使得每个节点的表征都携带其全局位置信息,可以在网络重构任务上取得最好的效果.

4.4 链接预测

链接预测是网络分析中的重要任务,其目的在于预测网络中缺失的或潜在的连接关系,我们通过探究表征是否可以捕捉一些潜在的关联信息来进一步比较表征质量.对每个网络数据集,我们首先随机删除 20% 的边作为测试集,剩下的部分则作为每个模型的输入用作表征学习.对测试集中的每条边,我们采样一条网络中不存在的边作为负样本,将链接预测任务转为二分类任务进行评估,并使用 AUC (area under curve) 作为评价指标.实验结果如表 4

所示.

从实验结果中可以发现,本文提出的 PMI 模型在 3 个数据集上均取得最好的效果,说明采用的 K 阶互信息最大方法可以有效地捕捉网络结构信息.此外,其他 2 种基于互信息的方法(DGI 和 GMI)均可取得非常不错的效果,结合 4.3 节在网络重构任务上的效果,我们可以得到这样一个结论:最大化节点与其近邻或全局结构的互信息有助于模型捕捉网络更深层的结构信息.其他几种方法,如 DeepWalk, Node2Vec, GraRep 也在其中几个数据上取得不错的效果,但整体效果并不稳定.我们认为出现该问题的原因在于表征无法获得节点的全局位置信息.直观上,在图上距离越近的节点越容易产生关联,因此节点的位置信息对于链接预测任务而言是非常重要的.而相反,本文提出的 PMI 方法通过最大化中心节点与各阶邻居的互信息,让中心节点能够记住其任意阶的邻居,从而使得学到的表征可以保留全局位置信息,捕捉更多网络结构中的隐藏信息,最终在链接预测任务上取得令人满意的效果.

Table 4 Results of Link Prediction

表 4 链接预测结果

数据集	DeepWalk	Node2Vec	LINE	SDNE	GraRep	DGI	GMI	PMI
IMDB	93.31	91.49	88.69	89.00	83.22	84.78	92.18	95.51
VoteNet	73.22	80.86	74.41	80.71	82.08	83.90	84.46	86.54
WordNet	72.20	73.45	72.95	71.12	74.25	76.48	79.22	84.11

注:黑体数字表示最优的结果.

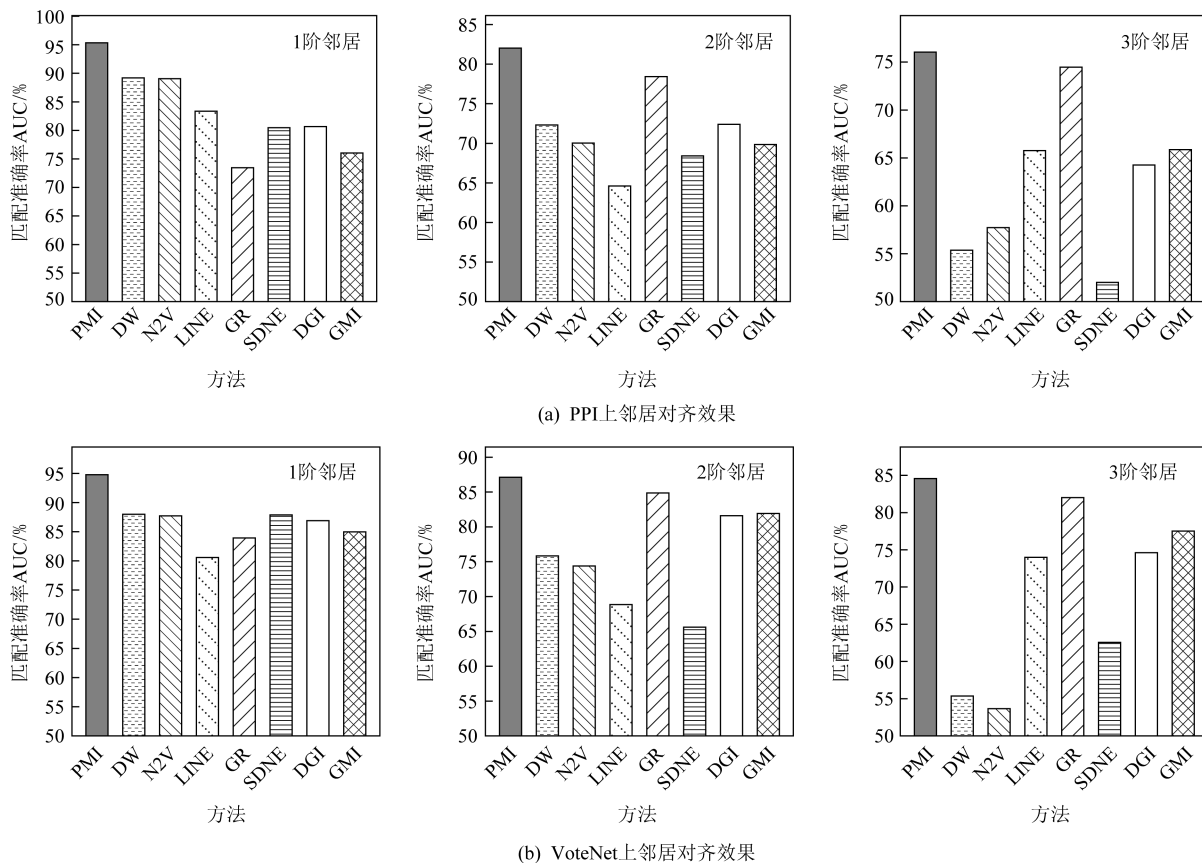
4.5 案例研究

为了研究本文提出的 PMI 模型是否可以使得表征学习到节点的全局位置信息,本节进行一个新颖的实验:邻居对齐.该实验的目标是让每个节点识别来自同一阶的节点并区分来自不同阶的节点.通过该实验,我们可以从侧面知晓学到的表征是否包含了位置信息.

由于很难直接界定不同阶表征间的边界,我们将邻居对齐问题转为一个二值分类任务,通过训练一个第 k 阶邻居分类器来评判节点表征是否可以区分来自 k 阶和非 k 阶的邻居,比如我们希望测试表征识别其 2 阶邻居的能力,那么对于节点 v_i 的所有 2 阶邻居 $v_j \in N_i^2$,我们分别随机采样一个非 2 阶节点 $v_z \notin N_i^2$,将每一对节点 (v_i, v_j) 作为正样本, (v_i, v_z) 作为负样本,问题就转为从 2 个样本中区分出正样本.我们采用一个简单的二值分类器:一个输出维度为 128 的双线性层用来学习节点对的相

关性,输出经过一层 DNN 并用交叉熵作为损失函数.我们在 PPI 和 VoteNet 上测试各个方法区分 1 阶、2 阶和 3 阶邻居的效果.对每一组实验,我们采用 5 折交叉验证并计算平均 AUC 值.最终的结果如图 5 所示,从结果中我们有如下 3 点发现:

- 1) 本文方法在 2 个数据集上均取得最优的效果.这说明通过本文方法学到的节点表征可以有效记住各阶的邻居信息,从而区分不同阶的节点.
- 2) 大部分方法都可以有效地识别 1 阶邻居,这主要是由于近邻假设导致 1 阶邻居节点通常与中心节点在向量空间中更近,因此较为容易区分.但对于其他阶的邻居节点,我们发现大部分方法(尤其是基于随机游走和自编码器方法)的效果存在明显下降,比如 SDNE 在 PPI 上区分 3 阶邻居时仅有 52% 左右的准确率,而 DeepWalk 在 2 个数据集上识别 3 阶邻居时均只有 55% 左右的准确率.这些实验结果均说明当前方法通常会混淆来自不同阶的邻居节点.



注: DW, N2V和GR分别是模型DeepWalk, Node2Vec和GraRep的简写。

Fig. 5 Results of neighbor alignment

图5 邻居对齐结果

3) GraRep 在区分 3 阶邻居的效果上与本文方法相当,这是因为 GraRep 建模节点对间的转移概率,而较高阶的转移概率会远小于低阶的转移概率,因此 GraRep 可以比较有效地识别 3 阶邻居节点.尽管如此,在 2 个数据集上 GraRep 均无法有效地区分近邻节点(如 1 阶邻居),这说明 GraRep 易混淆 1 阶邻居和其他阶的邻居节点,无法有效地捕捉到节点位置信息.在 4.2 和 4.4 节的实验中,相关任务均要求模型能够有效地捕捉近邻信息,本节的实验结果也解释了为什么 GraRep 在之前的实验中效果并不理想.

小结:本文提出的 PMI 模型可以激励节点表征记录其各阶邻居的信息,从而有效地识别来自于不同阶的邻居节点.结合 PMI 模型在 4.2~4.4 节实验中杰出的表现,我们可以发现节点的位置信息对于下游网络分析任务而言是极其重要的.

4.6 参数学习

本节探究不同的超参设置对 PMI 模型的影响.我们选择 PPI 和 VoteNet 两个数据集,并在 2 个常见

分析任务(多标签分类和链接预测)上进行参数学习.

我们首先研究衰减函数 $\Phi(x)$ 和最大阶数 K 对于模型的影响,比较 5 种具有不同衰减速度的衰减函数在最大阶数 $K \in \{1, 2, 3, 4, 5\}$ 上的效果,相关实验结果如图 6(a)所示.从结果中我们发现不同网络分析任务对于 2 类超参的偏好是不同的.在多标签分类任务中,当不考虑衰减函数时(即 $\Phi(x)=1$),整体效果会随着阶数 K 的增大而快速降低;相反, $\Phi(x)=1$ 在链接预测任务上可以取得不错的效果,其准确率随 K 的增大而升高.总体而言,衰减函数 $\Phi(x)=x^{-1}, 2^{1-x}, e^{1-x}$ 较为契合本文 PMI 模型.同时,对多标签分类任务来说,阶数 $K=2$ 较为合适,这也暗示近邻信息往往决定节点的类别.而相反,随着 K 的增大,链接预测的效果普遍上升,说明高阶信息对于链接预测任务是比较重要的,因为高阶信息可以让表征在更加宽的感受野中学习节点的位置信息.

接下来,我们研究损失函数中超参 α 和 β 对于效果的影响,两者用于平衡互信息估计和网络重构

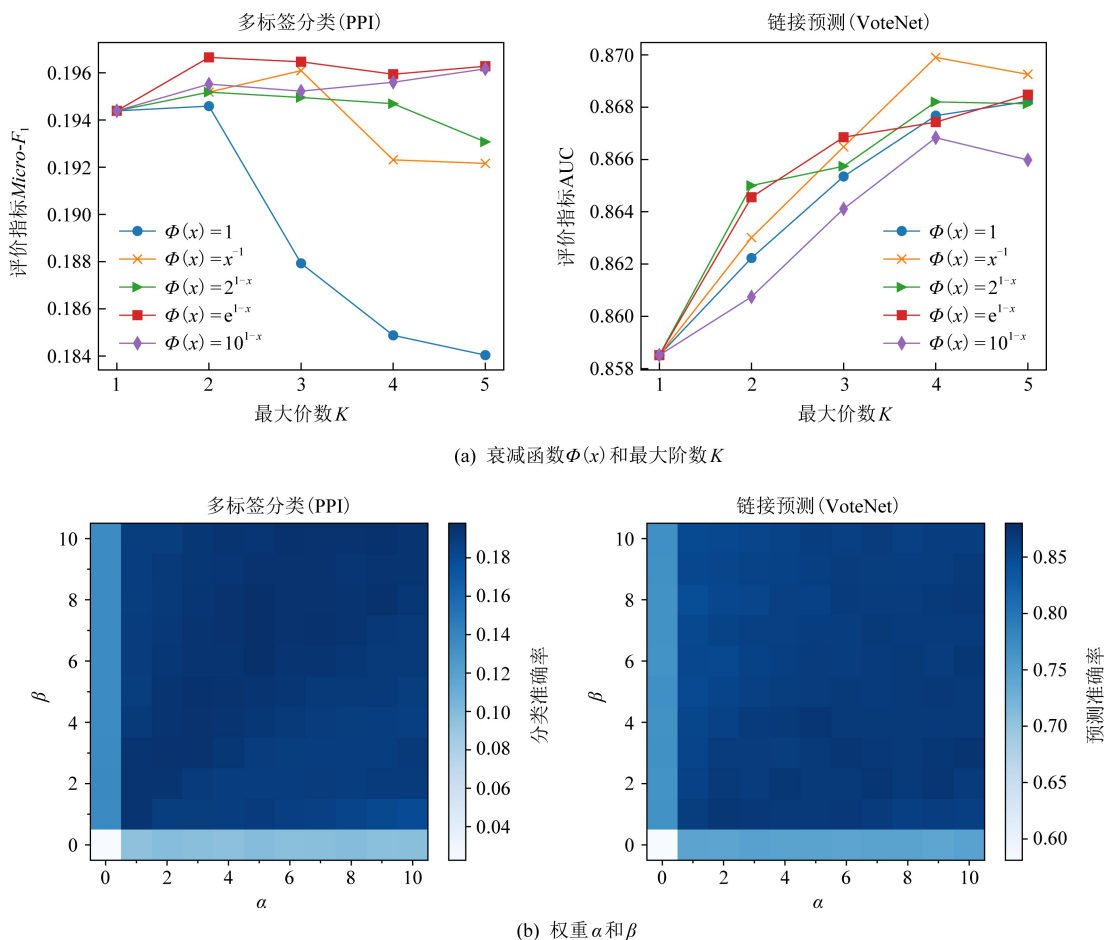


Fig. 6 Parameter study on two different tasks
图 6 在 2 种不同任务上的参数学习结果

损失的权重.我们以 0.1 为步长,将 α 和 β 分别从 0.0 增加到 1.0,并用热力图展示不同参数组合对多标签分类和链接预测效果的影响,结果如图 6(b)所示.总体上,PMI 模型对于两者的取值并不是很敏感,但当 α 和 β 分别取 0 时,方法有明显的下降,这也从侧面反映位置信息和结构信息对于节点表征都很重要.另一方面,我们也发现 2 类任务对于超参的偏好也是不一样的.对于多标签分类任务而言,颜色较深的区域往往靠近对角线附近,这说明对于分类任务,位置信息和结构信息同等重要;而对于链接预测任务,右下角的区域颜色较深,这意味着一个较大的 α 值会取得更好的效果,也说明节点的位置信息对于链接预测而言更加重要.

5 总 结

本文主要针对现有网络表征学习方法中位置信息缺失的问题,提出一种位置感知的网络表征学习

PMI 模型.在本文方法中,我们最大化每个中心节点与其各阶邻居之间的互信息,从而在训练的过程中激励每个节点记住其 K 阶邻居的信息,我们在 4 个不同领域的真实数据集上进行了多标签分类、链接预测、网络重构实验,实验结果表明本文所提 PMI 模型较现有表征学习模型不仅在各类下游网络分析任务上有所提升,并且通过在邻居对齐任务上的进一步分析,验证了 PMI 模型学到的表征可以有效地获取节点的位置信息.

参 考 文 献

[1] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online learning of social representations [C] //Proc of the 20th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2014: 701-710

[2] Tang Jian, Qu Meng, Wang Mingzhe, et al. LINE: Large-scale information network embedding [C] //Proc of the 24th Int Conf on World Wide Web. New York: ACM, 2015: 1067-1077

- [3] Liu Zhiyuan, Sun Maosong, Lin Yankai, et al. Knowledge representation learning: A review [J]. *Journal of Computer Research and Development*, 2016, 53(2): 247-261 (in Chinese)
(刘知远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展[J]. *计算机研究与发展*, 2016, 53(2): 247-261)
- [4] Tsoumakas G, Katakis I. Multi-label classification: An overview [J]. *International Journal of Data Warehousing and Mining*, 2007, 3(3): 1-13
- [5] Yang Shuanghong, Long Bo, Alexander J S, et al. Like like alike: Joint friendship and interest propagation in social networks [C] //Proc of the 20th Int Conf on World Wide Web. New York: ACM, 2011: 537-546
- [6] Backstrom L, Leskovec J. Supervised random walks predicting and recommending links in social networks [C] //Proc of the 4th Int Conf on Web Search and Web Data Mining. New York: ACM, 2011: 635-644
- [7] Leskovec J, Huttenlocher D P, Kleinberg J M. Predicting positive and negative links in online social networks [C] //Proc of the 19th Int Conf on World Wide Web. New York: ACM, 2010: 641-650
- [8] Long Qingqing, Wang Yiming, Du Lun, et al. Hierarchical community structure preserving network embedding: A subspace approach [C] //Proc of the 28th Int Conf on Information and Knowledge Management. New York: ACM, 2018: 409-418
- [9] Wang Xiao, Cui Peng, Wang Jing, et al. Community preserving network embedding [C] //Proc of the 31st AAAI Conf on Artificial Intelligence. Palo Alto: AAAI Press, 2017: 203-209
- [10] Grover A, Leskovec J. Node2vec: Scalable feature learning for networks [C] //Proc of the 22nd Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 855-864
- [11] Wang Daixin, Cui Peng, Zhu Wenwu. Structural deep network embedding [C] //Proc of the 22nd Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1225-1234
- [12] Cao Shaosheng, Lu Wei, Xu Qionghai. GraRep: Learning graph representations with global structural information [C] //Proc of the 24th Int Conf on Information and Knowledge Management. New York: ACM, 2015: 891-900
- [13] Belghazi M I, Baratin A, Rajeswar S, et al. Mutual information neural estimation [C] //Proc of the 35th Int Conf on Machine Learning Research. Cambridge MA: JMLR, 2018: 530-539
- [14] Hjelm R D, Fedorov A, Samuel L M, et al. Learning deep representations by mutual information estimation and maximization [C/OL] //Proc of the 7th Int Conf on Learning Representations. 2019 [2021-05-31]. <https://openreview.net/forum?id=Bklr3j0cKX>
- [15] Mikolov T, Chen Kai, Corrado G, et al. Efficient estimation of word representations in vector space [C/OL] //Proc of the 1st Int Conf on Learning Representations. 2013 [2021-05-31]. <http://arxiv.org/abs/1301.3781>
- [16] Ribeiro L F R, Saverese P H P, Figueiredo D R. struc2vec: Learning node representations from structural identity [C] //Proc of the 23rd Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2017: 385-394
- [17] Chang Shiyu, Han Wei, Tang Jiliang, et al. Heterogeneous network embedding via deep architectures [C] //Proc of the 21st Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2015: 119-128
- [18] Zhu Zhihua, Fan Xinxin, Chu Xiaokai, et al. HGCN: A heterogeneous graph convolutional network-based deep learning model toward collective classification [C] //Proc of the 26th Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2020: 1161-1171
- [19] Chu Xiaokai, Fan Xinxin, Yao Di, et al. Cross-network embedding for multi-network alignment [C] //Proc of the 28th Int Conf on World Wide Web. New York: ACM, 2019: 273-284
- [20] Xu Linchuan, Wei Xiaokai, Cao Jiannong, et al. Multi-task network embedding [C] //Proc of the Int Conf on Data Science and Advanced Analytics. Piscataway, NJ: IEEE, 2017: 571-580
- [21] Wang Suhang, Aggarwal C C, Tang Jiliang, et al. Attributed signed network embedding [C] //Proc of the 26th Int Conf on Information and Knowledge Management. New York: ACM, 2017: 137-146
- [22] Wang Suhang, Tang Jiliang, Aggarwal C C, et al. Signed network embedding in social media [C] //Proc of the Int Conf on Data Mining. Piscataway, NJ: IEEE, 2017: 327-335
- [23] Yuan Shuhan, Wu Xintao, Xiang Yang, et al. SNE: Signed network embedding [C] //Proc of the 21st Advances in Knowledge Discovery and Data Mining Pacific-Asia Conf. Berlin: Springer, 2017: 183-195
- [24] Liu Jie, He Zhicheng, Wei Lai, et al. Content to node: Self-translation network embedding [C] //Proc of the 24th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2018: 1794-1802
- [25] Tu Cunchao, Liu Han, Liu Zhiyuan, et al. CANE: Context-aware network embedding for relation modeling [C] //Proc of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2017: 1722-1731
- [26] Cavallari S, Zheng V W, Cai Hongyun, et al. Learning community embedding with community detection and node embedding on graphs [C] //Proc of the 23rd Int Conf on Information and Knowledge Management. New York: ACM, 2017: 377-386

- [27] Chen Jifan, Zhang Qi, Huang Xuanjing, et al. Incorporate group information to enhance network embedding [C] //Proc of the 22nd Int Conf on Information and Knowledge Management. New York: ACM, 2016: 1901-1904
- [28] Ma Xuwei, Qin Geng, Qiu Zhiyang, et al. RiWalk: Fast structural node embedding via role identification [C] //Proc of the Int Conf on Data Mining, Piscataway, NJ: IEEE, 2019: 478-487
- [29] Rossi R A, Ahmed N K, Koh E, et al. A structural graph representation learning framework [C] //Proc of the 13th Int Conf on Web Search and Web Data Mining. New York: ACM, 2020: 483-491
- [30] Hamilton W L, Ying Zhitao, Leskovec J, et al. Inductive representation learning on large graphs [C] //Proc of the 30th Annual Conf on Neural Information Processing Systems. San Francisco: Morgan Kaufmann, 2017: 1024-1034
- [31] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [C/OL] //Proc of the 5th Int Conf on Learning Representations. 2017 [2021-05-31]. <https://openreview.net/forum?id=SJU4ayYgl>
- [32] Velickovic T, Cucurull G, Casanova A, et al. Graph attention networks [C/OL] //Proc of the 6th Int Conf on Learning Representations. 2018 [2021-05-31]. <https://openreview.net/forum?id=rJXMpikCZ>
- [33] Hjelm R D, Fedorov A, Marchildon S L, et al. Learning deep representations by mutual information estimation and maximization [C/OL] //Proc of the 7th Int Conf on Learning Representations. 2019 [2021-05-31]. <https://openreview.net/forum?id=Bklr3j0cKX>
- [34] Velickovic P, Fedus W, Hamilton W L, et al. Deep graph infomax [C/OL] //Proc of the 7th Int Conf on Learning Representations. 2019 [2021-05-31]. <https://openreview.net/forum?id=rklz9iAcKQ>
- [35] Peng Zhen, Huang Wenbing, Luo Minnan, et al. Graph representation learning via graphical mutual information maximization [C] //Proc of the 29th Int Conf on World Wide Web. New York: ACM, 2020: 259-270
- [36] Kipf T N, Welling M. Variational graph auto-encoders [C/OL] //Proc of the 29th Annual Conf on Neural Information Processing Systems. 2016 [2021-05-31]. <http://arxiv.org/abs/1611.07308>
- [37] Zhu Dingyuan, Cui Peng, Wang Daixin, et al. Deep variational network embedding in wasserstein space [C] //Proc of the 24th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2018: 2827-2836
- [38] Hinton G E, Osindero E, Teh Y W, et al. A fast learning algorithm for deep belief nets [J]. *Neural Computing*, 2006, 18(7): 1527-1554
- [39] Wu C Y, Beutel A, Ahmed A, et al. Explaining reviews and ratings with PACO: Poisson additive co-clustering [C] //Proc of the 25th Int Conf on World Wide Web. New York: ACM, 2016: 127-128
- [40] Crammer K, Singer Y. On the learnability and design of output codes for multiclass problems [J]. *Machine Learning*, 2002, 47(2/3): 201-233



Chu Xiaokai, born in 1994, PhD candidate. His main research interests include network alignment, graph neural networks, network embedding.

储晓恺, 1994年生, 博士研究生, 主要研究方向为网络对齐、图神经网络和网络表征学习。



Fan Xinxin, born in 1984, PhD. His main research interests include machine learning security, trust computing, network security, etc. (fanxinxin@ict.ac.cn)

范鑫鑫, 1984年生, 博士, 主要研究方向为机器学习安全、可信计算、网络安全等。



Bi Jingping, born in 1974, PhD, professor and PhD supervisor. Senior member of CCF. Her main research interests include intelligent data processing, network intelligence, etc.

毕经平, 1974年生, 博士, 研究员, 博士生导师, CCF高级会员, 主要研究方向为智能数据处理、网络智能化等。