一种基于深度学习的动态社交网络用户对齐方法

王飞扬,冀鹏欣,孙 笠,危 倩,李 根,张忠宝
(北京邮电大学计算机学院,北京 100876)

摘 要: 社交网络对齐旨在从不同的社交网络中识别出属于同一自然人的社交账户. 现有的相关研究大多着眼于静态社交网络的对齐上,然而,社交网络是动态发展的. 本文观察到,这种动态性可以揭示出更多的决定性模式,从而更有利于社交网络的对齐,这种现象促使本文在动态场景中重新思考这个问题. 于是,本文利用社交网络的动态性,设计一个深度学习架构来解决动态社交网络的对齐问题,其称为 DeepDSA(Deep learning based Dynamic Social network Alignment method). 首先设计一个深度序列模型来分别捕捉社交网络结构和属性的动态性;其次,对于每一个社交网络,通过保持相同用户结构和属性之间的相关性来融合二元动态,得到原始的综合嵌入表示;最后,以半监督的方式进行空间变换学习,并将每个网络的原始嵌入投影到一个目标子空间中,在该子空间中自然人是唯一表示的. 本文在真实世界的数据集上进行大量的实验,证明 DeepDSA 方法相较于目前的主流算法提升了 10% 的对齐效果.

关键词: 社交网络对齐; 动态性; 深度学习; 特征融合; 子空间学习
中图分类号: TP311 **文献标识码:** A **文章编号:** 0372-2112(2022)08-1925-12
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20201436

A Deep Learning Based Dynamic Social Network Alignment Method

WANG Fei-yang, JI Peng-xin, SUN Li, WEI Qian, LI Gen, ZHANG Zhong-bao
(*Department of Computer Science, Beijing University of Posts and Telecommunications, Beijing* 100876, *China*)

Abstract: Social network alignment aims to identify social accounts belonging to the same natural person from different social networks. Most of the existing related researches focus on the alignment of static social networks. However, social networks are dynamically evolving. We observe that dynamics can reveal more discriminative patterns and thus can benefit social network alignment. This phenomenon motivates us to rethink this issue in dynamic scenarios. Therefore, we propose to leverage the dynamics of social networks and design a deep learning architecture to address the dynamic social network alignment problem, termed as DeepDSA. Specifically, we first design a deep sequence model to capture the dynamics of social network structure and attributes respectively. For each social network, we merged binary dynamics by maintaining the correlation between structure and attributes of the same user to obtain the original comprehensive embeddings. We finally perform spatial transformation learning in a semi-supervised manner, and project the original embedding of each network into a target subspace in which a natural person is uniquely represented. We conduct extensive experiments on real-world datasets and demonstrate the proposed DeepDSA achieves 10% improvement of precision against the current mainstream algorithm.

Key words: social network alignment; dynamics; deep learning; feature fusion; subspace learning

1 引言

近年来,随着社交网络的不断涌现和普及,人们逐渐开始参与多个社交网络平台,享受多样的社交服务. 社交网络对齐问题应运而生,即在不同的社交网络中识别属于同一自然人的社交账户. 社交网络对齐可以支持广泛的应用,如链接预测、异常检测和跨域推荐等. 总体而言,社交网络的对齐为跨社交网络的广度学习铺平了道路,越来越受到学术界和工业界的关注. 大多现有的方法[1~18]研究的是静态社交网络间的对齐,它们仅考虑一个快照中的网络特性. 然而,社交网络本质上是高度动态,并且不断发展的. DNA(Dynamic Network Alignment)[19]考虑了网络拓扑结构的动态演化,但

收稿日期:2020-12-15;修回日期:2021-04-11;责任编辑:王天慧
基金项目:国家自然科学基金联合培育基金(No.U1936103)

忽略了同样很有价值的属性特征. 事实上,一个人在社交网络中会不断地改变他的朋友列表并更新文本或位置,这种动态性揭示了静态场景中被忽略的用户的行为模式. 在同一时段,不同社交网络中的同一自然人往往会表现出相似的行为,根据行为出现的时期,本文可以把静态网络中难以区分的用户区别开来. 图1给出了一个动态网络对齐与静态网络对齐的对比实例. 不同颜色的图表示不同的社交网络,每个结点都是一个社交帐号. 连接不同网络间结点的垂直黑线表示结点之间的对应关系,即a和a'是已知的对齐结点. 靠近每个结点的虚线框表示该用户的主题或关键字,展示了社交帐户的属性. 箭头展示了社交网络随时间的动态演化. 仅考虑第一个快照,网络A中的b,c和网络B中的b',c'行为均相同,静态对齐方法无法区分网络B中b和c的对齐结点. 然而,在第二个快照中,b和c在结构上表现不同. 进一步地,它们在第三个快照中的属性方面仍存在差异. 同时,b的行为总是与b'相同,c和c'也一致. 因此,本文可以非常肯定地说,账户b和b'属于同一自然人,而账户c和c'属于另一自然人. 静态社交网络是一个仅关注用户在某一阶段的结构和属性的快照,而动态社交网络则是考虑整个时间轴上的信息,从而使用户的行为模式更加清晰. 通过捕获用户的动态行为模式,本文可以使个体的表示更加精确和有判别性,从而对实现对齐有所裨益.

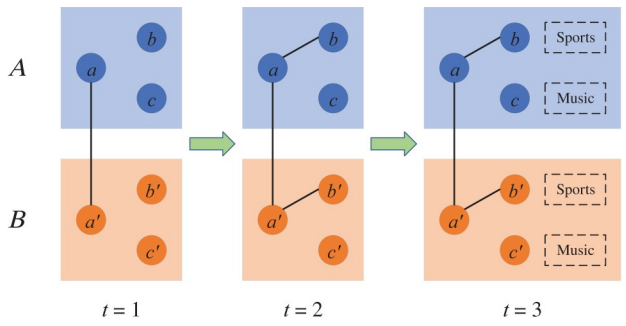


图1 利用结构和属性的动态社交网络对齐样例

上述观察结果促使本文重新思考动态情境下的社交网络对齐. 然而,在利用社交网络的动态性进行对齐方面仍然存在3个主要挑战.

(1)如何对每个用户的动态模式进行建模? 用户的行为总是随着动态社交网络的演化而变化,因此用户的动态模式是复杂的,而浅层模型由于其有限的表现形式很难捕捉到网络的高度非线性的动态模式.

(2)如何有效融合结构和属性的动态性? 社交网络通常有两个特性,即结构和属性. 结构描述了社交网络用户之间的关系,包含了用户间社交互动的规律. 属性,例如名称、文本和位置等,通常展示用户的主题或偏好. 因此,此挑战的关键在于如何有效地融合结构和属性的动态性,以获得全面的网络嵌入表示.

(3)如何设计对齐算法? 在静态网络中亦很难获取一组具有相同使用者的账号,更不用说在动态场景中了. 此外,一个用户很可能在不同的网络中与不同的好友联系,并在社交网络中为了不同的目的做出不同的行为. 缺乏标签数据和网络间普遍存在的差异性,使得这一问题更具挑战性.

为了解决上述问题,本文设计了一个深度架构来解决动态社交网络对齐问题,称为DeepDSA(Deep learning based Dynamic Social network Alignment method).

在DeepDSA中,本文提出了一种用于复杂动态性建模的深度神经模型. 用户在不同快照上具有特定的特征,这些快照自然地形成一个时间序列. 因此,本文提出了一个基于门控循环单元(Gated Recurrent Unit, GRU)的序列模型,分别用于对每个社交网络的结构和属性的动态性进行建模. 具体地说,本文设计了一个GRU编码器来对结构或属性的动态特征进行建模,并将每次的输出反馈到解码器中对结构或属性进行预测. 序列预测任务总是存在一个交叉依赖的问题,即当前特征可能主要受间接的历史而不是直接前继的影响[20]. 因此,针对这一问题,本文在神经模型的核心部分设计了一个注意力机制,并将每个时间阶段的隐藏状态作为相应的动态性表示.

对于第二个挑战,本文通过分别保持每个社交网络相同用户结构和属性之间的相关性来融合此二元动态性. 具体来说,首先将前一个问题得到的每个时间阶段的结构表示通过时间衰减效应进行合并得到结构嵌入表示,属性特征亦然. 然后,对于每个社交网络,通过保持同一用户的结构和属性之间的相关性,对每个用户施加约束. 结构特征和属性特征分别描述了用户的不同部分,它们相互补充,使用户更具可判别性. 为了将结构和属性结合在一起,对每个用户结构和属性的联合概率进行最大似然估计. 最后,将结构嵌入和属性嵌入结合起来,作为每个用户的原始的综合嵌入表示.

对于第三个挑战,在子空间学习的启发下,本文以半监督的方式进行空间变换学习,并将每个网络的原始嵌入投影到一个目标子空间中,在此子空间中每个自然人均有唯一的表达. 由于普遍存在的网络间差异性,直接在原始嵌入空间中进行相似性度量可能是不合适的. 每一个特定的网络都容易产生不同的偏移,每个用户的本质可能在原始空间中被隐藏. 因此,空间转换可以展现用户的本质,以便更好地对齐. 具体地说,利用一个前馈神经网络来拟合变换,并利用部分的已知标签用户以半监督的方式进行空间变换学习.

本文所提出的DeepDSA方法的优点在于,它具有强大的非线性学习能力,能够捕捉复杂的结构和属性

的动态性,并对整体结构进行统一优化,以自然地对用户的本质特征进行建模.

为了评估所提出的DeepDSA方法,本文首先通过构建一系列网络快照(通过收集的时间戳)来仿真现实世界中的动态社交网络. 然后,本文在真实数据集上进行了广泛的实验,并证明了所提出的DeepDSA方法的明显优势.

本文的主要创新点如下:

(1)同时运用结构和属性特征,解决动态场景下社交网络对齐问题;

(2)设计了一个统一的深度神经结构DeepDSA,提出了一个来捕捉用户结构和属性的复杂动态性的深度神经模型,从而实现社交网络的对齐,并对整个框架联合优化学习;

(3)在真实数据集上进行了大量的实验,实验结果证明了所提出的DeepDSA方法存在明显的优越性.

2　相关工作与预备知识

2.1　相关工作

社交网络用户对齐问题,是指在不同的社交网络中识别出属于同一个人的社交账号,由Zafarani等人[14]提出. Zhang等人[16]提出了基于能量的模型COSNET(Connecting Heterogeneous Social Networks),通过考虑本地和全局一致性来链接用户身份. COSNET首先提取基于距离的轮廓特征和基于邻域的网络特征,然后使用聚合算法获得局部一致性. Su等人[7]设计了一个约束的双重嵌入模型,将社交网络的对齐问题转化为一个统一的优化问题,将多个社交网络嵌入到一个公共的潜在空间中进行协调. Mu等人[21]提出"潜在用户空间"的概念,以建模潜在真实用户与其在各种社交平台上观察到的投影之间的关系,从而使真实用户越相似,则其在潜在用户空间中的画像越接近. Man等人[6]采用基于嵌入的网络特征将社交网络结构映射到低维空间,基于用户身份的潜在特征提出了一种投影方法. Liu等人[5]提出的IONE(Input-Output Network Embedding)方法提取基于嵌入的网络特征,同时学习每个用户的拓扑结构,种子的对齐用户对约束着社交关系网络结构的上下文传递. Zhou等人[3]提出了半监督的端到端方法DeepLink,对网络进行采样,并学习将网络节点编码成矢量表示,以捕获局部和全局网络结构,进而利用这些结构通过深度神经网络对结点进行对齐.

几乎所有现有方法均聚焦在静态场景下的社交网络用户对齐而忽略了社交网络固有的动态性. Sun等人[19]嵌入了结构的局部和全局动态性并提出了DNA方法,通过矩阵分解,提出了通过嵌入空间相互作用的统一优化方法来构造公共空间,并设计了交替优化算法来逼近局部最优,但是其忽略了同样有价值的属性动态性.

2.2　问题定义

在动态场景中,本文考虑带有时间戳的社交网络的结构和属性. 如图1所示,本文将网络分割成片,并在时域中构造一系列的图快照. 每个快照都反映了当前时间片中网络的特征. 把一个动态的社交网络定义为$G=(V,S,A)$,其中,$V=\{v_1,v_2,\cdots,v_n\}$,$S=\{\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_T\}$和$A=\{\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_T\}$,分别表示结点集、结构集和属性集. 对于每个$i\in[1,T]$,$\boldsymbol{S}_i\in\mathbb{R}^{n\times d_s}$和$\boldsymbol{A}_i\in\mathbb{R}^{n\times d_a}$分别是第$i$时间片中每个用户的结构特征矩阵和属性特征矩阵.

考虑到两个动态的社交网络G^s和G^t,在不失一般性的前提下,已知两个网络间的部分对齐账号作为标签数据. 本文引入了一个对齐结点对的集合$P_{st}=\{(v_s,v_t)|v_s\in V^s,v_t\in V^t\}$,其中每个对齐结点对$(v_s,v_t)$表示$G^s$中的账号$v_s$和$G^t$中的账号$v_t$在现实中属于同一自然人.

本文总结了论文中的主要符号(见表1)并将研究问题正式定义如下.

表1　主要符号和定义

符号	定义
G	动态社交网络
V	用户集合
S	结构特征集合
A	属性特征集合
P	已知对齐结点集合
$\boldsymbol{U}$	嵌入矩阵
$\boldsymbol{Q}$	空间映射矩阵
$\boldsymbol{I}$	指示矩阵
$\boldsymbol{M}$	负指示矩阵
β,γ	模型参数

定义1　动态社交网络用户对齐问题. 给定两个动态社交网络,即源网络G^s和目标网络G^t以及一组已知对齐结点P_{st},动态社交网络对齐的问题旨在为每个社交账户找到两个映射函数Φ_s和Φ_t,其将社交账号映射至真实归属自然人,即$\Phi_s(v_s)=\Phi_t(v_t)$当且仅当(v_s,v_t)在P_{st}中存在.

2.3　门控循环单元

如图2所示,门控循环单元(GRU)是循环神经网络(Recurrent Neural Networks,RNN)的另一个变体,与基本的长短期记忆网络(Long Short-Term Memory networks,LSTM)相比,它能更好地将序列数据历史的信息连接到本文的问题[22]中. GRU将输入和遗忘门耦合到更新门中,以避免长期依赖性问题. 最后,它的输出门

(称为重置门)只对块输入的循环连接进行门控. 具体来说,本文将每个时间切片的结构或属性特征依次输入GRU单元,并获得相应的输出. GRU单元在第t步的计算流程如下:

$$z_t=\sigma\left(W_z\left[h_{t-1},x_t\right]\right)\ r_t=\sigma\left(W_r\left[h_{t-1},x_t\right]\right)$$

$$\tilde{h}_t=\tanh\left(W\left[r_t\odot h_{t-1},x_t\right]\right)\ h_t=\left(1-z_t\right)\odot h_{t-1}+z_t\odot\tilde{h}_t \quad (1)$$

其中,x_t是第t时刻的特征输入,h_{t-1}和h_t是隐藏状态同时也是上一步和当前步骤的输出,σ表示sigmoid函数,$\odot$表示Hadamard积. GRU可以通过逐步地更新具有历史和当前特征的单元状态,自然地对序列输入进行建模并捕捉动态性. 在每个步骤t中,细胞接收先前的隐藏状态h_{t-1}和当前的输入x_t,以获得当前的输出h_t. 重置门通过r_t确定h_{t-1}对新内存$\tilde{h}_t$的权重. 另外,更新门通过z_t确定h_{t-1}到下一步的传输量是多少.

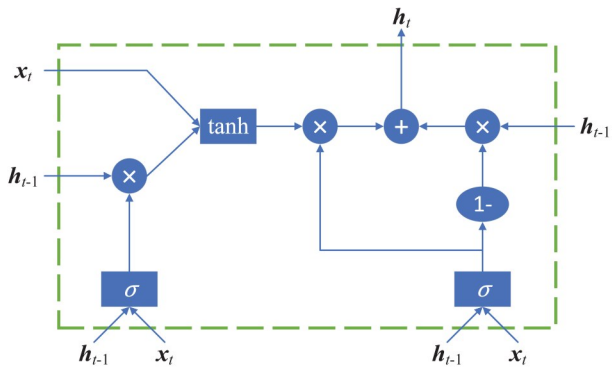


图2 GRU细胞的内部结构

3 DeepDSA

如图3所示,为了解决动态社交网络用户对齐问题,本文提出DeepDSA模型. 首先,将动态网络划分为T个切片,分别构造每个切片的结构和属性特征. 其次,设计一个基于GRU的序列模型,利用注意力机制分别对结构和属性的动态性进行建模. 然后,对结构和属性进行融合,得到每个用户的嵌入表示. 最后,利用空间变换,将用户嵌入由正交初始化矩阵Q控制的子空间中.

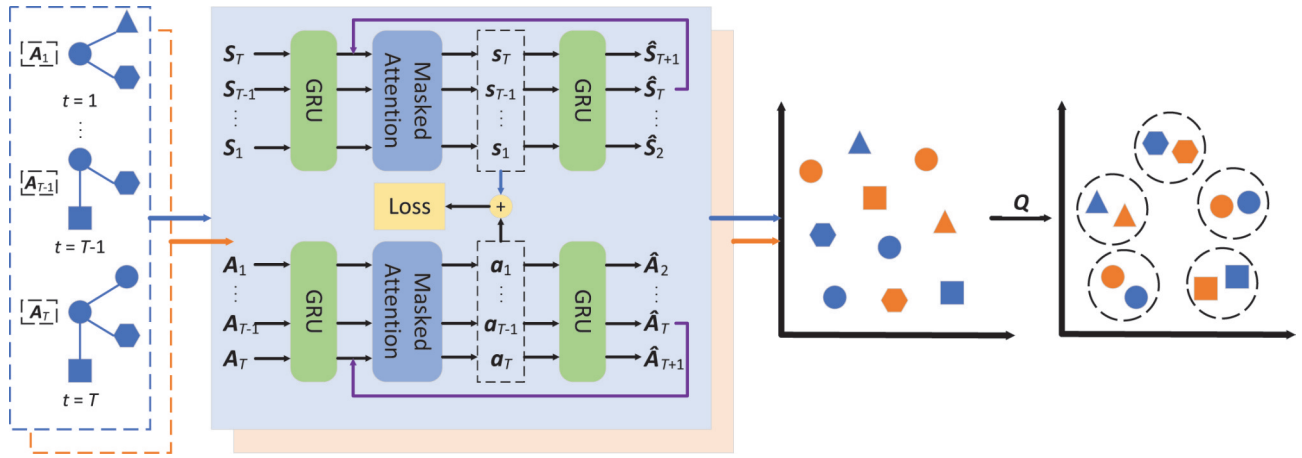
3.1 动态性建模

如图3(a)所示,社交网络中的一个结点,通常有两种特征:结构和属性. 结构显示用户之间的链接. 如果存在关注/被关注关系,则两个用户是链接的,这意味着他们彼此很接近. 另外,属性揭示了社交网络的部分固有特征. 特定于某个用户,属性特征的覆盖范围很广,例如,名称、文本和位置. 此外,社交网络的结构和属性都表现出高度的动态性,因为用户总是在添加/删除彼此之间的链接同时更新文本或位置,这会体现出有价值的附加信息,并且蕴含了静态社交网络中忽略的动态行为模式. 因此,本文提出的模型旨在捕捉结构和属性的丰富动态性.

本文首先将动态网络划分为T个切片,分别构造每个切片的结构特征和属性特征. 对于结构特征,一个简单而直接的方法是利用二进制邻接矩阵,其中对于每个$M\in\mathbb{R}^{n\times n},i,j\in[1,n]$,如果用户$i$和用户$j$之间存在链接,则元素$M_{i,j}$为1,否则为0. 进一步地,利用DeepWalk中的随机游走[23,24]来保持高阶的临近性. DeepWalk已经被证明实际上是一种近似特征矩阵S_i的采样方法:

$$S_i=\log\frac{\overline{R}_i+\overline{R}_i^{\,2}+\cdots+\overline{R}_i^{\,k}}{k} \quad (2)$$

其中,$\overline{R}_i$是第i个时间片的行规范化邻接矩阵,k表示阶数,每个R_{ij}表示顶点v_i以固定步数随机走到顶点v_j的平均概率的对数.



(a) 输入 (b) 建模 (c) 对齐

图3 DeepDSA模型图

对于属性特征,本文使用用户的文本并对文本进行矢量表示. 本文中亦将位置名称视为文本. 一种直接的方法是词袋(Bag Of Words,BOW)模型,但其表示效果通常很差,因为BOW忽略了许多文本的表示信息,如单词的顺序. 潜在狄利克雷分配(Latent Dirichlet Allocation,LDA)也是一种常见的主题建模技术(从文本中提取主题/关键词),但它很难训练,且结果也很难评估. 因此,本文采用Doc2vec[25],一种无监督的学习算法,它可以学习可变长度的文本片段(如句子和文档)的矢量表示. 具体来说,本文利用分布式词袋(Distributed Bag Of Words version of Paragraph Vector,PVDBOW)算法,如图4所示,文本/文档由内容词和位置词组成,文本向量用于预测其自身的单词. Doc2vec受到单词表示学习方法的启发,对于一个文本,它的向量表示被用来预测其自身的单词. 实际上,这意味着在每次梯度下降的迭代中,本文都会采样一个文本窗口,然后从文本窗口中随机抽取一个单词,并在给定文本向量的情况下形成一个分类任务[25]. 值得一提的是,本文总是会采样到文本的位置词,因为位置信息对于用户画像来说是非常简明和决定性的特性. 在每个时间片中,通过取每个用户的文本向量的平均值来构造特征矩阵$\boldsymbol{A}$.

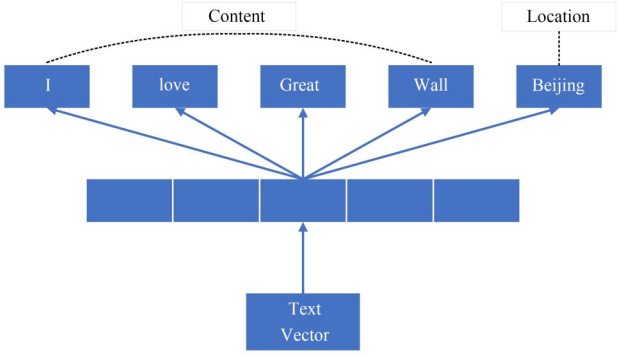


图4　学习文本向量的框架

对于每个社交网络,本文从结构和属性2个角度对网络的动态性进行建模. 由于结构与属性建模过程基本相同,本文仅以结构动态性建模部分为例,属性动态性建模过程同理可得.

如图3(*b*)所示,在将动态网络分割成T片后,构造一组结构特征$S=\{\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_T\}$. 对于$i\in[1,T]$和$m\in[1,n]$,$\boldsymbol{s}_i^m$是$\boldsymbol{S}_i$的第$m$行,表示第$i$时间片上第$m$个用户的特征. 鉴于GRU内在的对序列的动态建模能力,本文使用GRU编码器来收集每个用户m的序列特征$S^m=\{\boldsymbol{s}_1^m,\boldsymbol{s}_2^m,\cdots,\boldsymbol{s}_T^m\}$,并保留相应的输出$H^m=\{\boldsymbol{h}_1^m,\boldsymbol{h}_2^m,\cdots,\boldsymbol{h}_T^m\}$. 然后将$H^m$输入到GRU解码器中,对结构进行预测,即$S^m$可以由$\{\boldsymbol{h}_1^m,\boldsymbol{h}_2^m,\cdots,\boldsymbol{h}_T^m\}$重构. 其背后的依据在于,个人在社交网络中的结构演化是确定性的而不是随机的,并且当前的结构受到历史因素的影响[26,27]. 此建模模型对于属性特征同样适用,依据在于用户的主题或关键字总是平滑地发展并且历史信息被用来动态地建模用户画像[28].

此外,序列预测通常存在交叉依赖问题[20]. 当前的特征可能主要是受间接历史而不是直接前继历史所影响. 例如,社交网络用户可能在某个时间点结交新朋友并对某个新话题感兴趣,这会受到其老朋友而不是最近结识的朋友的影响. 因此,本文利用注意力机制来解决交叉依赖问题. 在第i步,构造$\boldsymbol{e}_i$替代$\boldsymbol{h}_i$作为特征,这是通过学习所有历史信息的综合效应而获得的:

$$\boldsymbol{e}_i=\sum_{j=1}^{i}\tilde{\alpha}_{i,j}\boldsymbol{h}_j$$
$$\text{s.t.}\sum_{j=1}^{i}\tilde{\alpha}_{i,j}=1 \tag{3}$$

其中,$\tilde{\alpha}_i=\{\tilde{\alpha}_{i,1},\tilde{\alpha}_{i,2},\cdots,\tilde{\alpha}_{i,i}\}$衡量历史信息对当前第$i$时刻的贡献. 之后$\boldsymbol{e}_i$被输入解码器以进行结构预测. 如图3(*b*)所示,本文还应用了掩码[29],以确保$\tilde{\alpha}_{i,j}$仅由先前的输入和解码器的上一状态p_{i-1}确定,具体公式如下:

$$\alpha_{i,j}=\boldsymbol{v}^{\mathrm{T}}\tanh\left(\boldsymbol{W}\left[\boldsymbol{p}_{i-1},\boldsymbol{h}_j\right]\right),j\in[1,i] \tag{4}$$
$$\tilde{\alpha}_i=\text{softmax}\left(\alpha_i\right) \tag{5}$$

其中,$\alpha_i=\{\alpha_{i,1},\alpha_{i,2},\cdots,\alpha_{i,i}\}$衡量当前和历史之间的联系,而$\boldsymbol{W}$是参数矩阵. 序列预测约束了自动编码器模型捕捉结构的动态模式.

如图3(*b*)所示,给定输入S和A以及相应的解码器预测结果$\hat{S}=\left\{\hat{\boldsymbol{S}}_2,\hat{\boldsymbol{S}}_3,\cdots,\hat{\boldsymbol{S}}_{T+1}\right\}$和$\hat{A}=\left\{\hat{\boldsymbol{A}}_2,\hat{\boldsymbol{A}}_3,\cdots,\hat{\boldsymbol{A}}_{T+1}\right\}$,本文将GRU自动编码器的重建误差最小化如下:

$$L_d=\sum_{i=2}^{T}\left\|\boldsymbol{S}_i-\hat{\boldsymbol{S}}_i\right\|_F^2+\left\|\boldsymbol{A}_i-\hat{\boldsymbol{A}}_i\right\|_F^2 \tag{6}$$

3.2　二元动态性融合

利用上述模型,本文分别捕获了每个社交网络结构和属性的动态性. 然而,接下来关键的任务是要有效地融合这两个方面的特征,以考虑互补信息,从而得到每个用户的全面嵌入表示.

本文首先将序列模型捕捉到的动态性与时间衰减效应相结合,分别构造结构($\boldsymbol{u}_{\mathrm{s}}$)和属性($\boldsymbol{u}_{\mathrm{a}}$)的嵌入表示. 对于结构嵌入,具体表示为

$$\boldsymbol{u}_{\mathrm{s}}=\sum_{k=1}^{T}\exp\left(-(T-k)\right)\boldsymbol{e}_k \tag{7}$$

其中,时间最临近的特征由于包含了更多的信息而获得了更大的贡献值. $\boldsymbol{u}_{\mathrm{a}}$同理可得.

为了融合结构和属性,一个简单的方法是直接拼接$\boldsymbol{u}_{\mathrm{s}}$和$\boldsymbol{u}_{\mathrm{a}}$,但不能保证这两种特性之间的一致性[30]. 此

外,本文期望通过有效地融合结构信息和属性信息,更准确地描述用户,并强调用户间的独特性和差异,以服务于接下来的对齐任务. 因此,本文利用结构和属性之间的联合概率,最大化同一用户的两种特征的似然概率. 联合概率和目标函数给定如下:

$$p\left(\boldsymbol{u}_{\mathrm{s}}^{i},\boldsymbol{u}_{\mathrm{a}}^{j}\right)=\frac{1}{1+\exp\left(-\boldsymbol{u}_{\mathrm{s}}^{i\,\mathrm{T}}\cdot\boldsymbol{u}_{\mathrm{a}}^{i}\right)} \tag{8}$$

$$\begin{aligned}L_f=&-\sum_{v_i\in V}\log p\left(\boldsymbol{u}_{\mathrm{s}}^{i},\boldsymbol{u}_{\mathrm{a}}^{i}\right)\\&+\sum_{v_i\in V}\sum_{v_j\in V\setminus\{v_i\}}\log p\left(\boldsymbol{u}_{\mathrm{s}}^{i},\boldsymbol{u}_{\mathrm{a}}^{j}\right)\end{aligned} \tag{9}$$

本文最大化同一用户的对数似然,同时最小化不同用户间的结构-属性的对数似然. 抑制所有不同用户间的对数似然会施加一个过于严格的约束,即具有高度临近性的用户会被疏离,并带来大量的计算消耗. 因此,本文采用负采样[27]方法. 根据结点的度分布$p_n(v)\propto d_v^{3/4}$,其中d_v是v的度,抽样到当前为止从未与v有过链接的负采样用户. 重写目标函数如下:

$$L_f=-\sum_{v_i\in V}\log p\left(\boldsymbol{u}_{\mathrm{s}}^{i},\boldsymbol{u}_{\mathrm{a}}^{i}\right)+\sum_{v_i\in V}\sum_{k=1}^{K}\mathbb{E}_{v_k\sim p_n(v)}\log p\left(\boldsymbol{u}_{\mathrm{s}}^{i},\boldsymbol{u}_{\mathrm{a}}^{k}\right) \tag{10}$$

其中,k是根据$p_n(v)$抽样的负采样用户数. 利用目标函数将$\boldsymbol{u}_{\mathrm{s}}$和$\boldsymbol{u}_{\mathrm{a}}$连接起来,作为每个网络中每个用户的综合嵌入$\boldsymbol{u}$.

3.3 利用动态性对齐

对源网络G^{s}和目标网络G^{t}进行上述的动态嵌入表示后,一种简单的方法是直接测量网络间结点表示的相似度. 该方法的实质是将嵌入结点看作原始空间中的结点,并最小化属于同一自然人的结点之间的距离. 然而,一个人在社交网络中出于不同的目的可能会有不同的行为. 在一个网络中,某项特征可能得到增强,但在另一个网络中却被隐藏. 例如,一个人可能总是在源网络中分享他对音乐的兴趣,而在目标网络中很少或从不分享音乐. 此外,由于社交网络的高度复杂性,某些特征信息可能会在整个空间中弥散. 例如,有人在源网络分享了他的长城之旅和他品尝的饺子. 同时期,他在目标网络中关注了一个中国账号,并分享了一段京剧. 这些行为虽然不尽相同,但都显示出他与中国有关,蕴含了潜在的相似性. 网络间普遍的差异使得直接的相似性度量变得困难和不可行.

如图3(c)所示,为了解决这个问题,本文首先学习空间变换,将原始空间变换为目标子空间. 然后将原始数据投影到目标子空间,利用部分监督信息度量网络结点间的相似性. 具体地说,找到一个空间变换$\boldsymbol{Q}$,并将各网络的原始嵌入空间投影到一个公共的目标子空间中,在该子空间中,属于同一自然人的账号表示尽可能接近. 因此,目标函数给定如下:

$$\boldsymbol{X}_{\mathrm{s}}=\boldsymbol{Q}\boldsymbol{U}_{\mathrm{s}},\ \boldsymbol{X}_{\mathrm{t}}=\boldsymbol{Q}\boldsymbol{U}_{\mathrm{t}} \tag{11}$$

$$L_a=-\left\|\boldsymbol{M}^{\mathrm{st}}\boldsymbol{X}_{\mathrm{s}}-\boldsymbol{M}^{\mathrm{ts}}\boldsymbol{X}_{\mathrm{t}}\right\|_F^2+\left\|\boldsymbol{I}^{\mathrm{st}}\boldsymbol{X}_{\mathrm{s}}-\boldsymbol{I}^{\mathrm{ts}}\boldsymbol{X}_{\mathrm{t}}\right\|_F^2 \tag{12}$$

其中,$\boldsymbol{U}_{\mathrm{s}}$和$\boldsymbol{U}_{\mathrm{t}}$分别是源网络和目标网络的嵌入矩阵. 从已知对齐结点对集合U_{st}中引入一对指示矩阵$\{\boldsymbol{I}^{\mathrm{st}},\boldsymbol{I}^{\mathrm{ts}}\}$,其中$\boldsymbol{I}^{\mathrm{st}}\in\mathbb{R}^{|U_{\mathrm{st}}|\times N^{\mathrm{s}}}$,$\boldsymbol{I}^{\mathrm{ts}}\in\mathbb{R}^{|U_{\mathrm{st}}|\times N^{\mathrm{t}}}$,$|U_{\mathrm{st}}|$表示对齐结点对的数量. 指示矩阵$\boldsymbol{I}^{\mathrm{st}}$定义为$\left[I_1\ I_2\ \cdots\ I_{|U_{\mathrm{st}}|}\right]^{\mathrm{T}}$,其行向量$I_i\in\mathbb{R}^{N^{\mathrm{s}}}$是一个只有一个元素、值为1的二进制向量. 此元素的位置是指示要对齐的相应行,相同的规则适用于$\boldsymbol{I}^{\mathrm{ts}}$. 类似地,本文采用负采样,对于每个源网络账号,选择目标网络中概率较小的一个对齐结点作为负采样,并引入负指示矩阵$\{\boldsymbol{M}^{\mathrm{st}},\boldsymbol{M}^{\mathrm{ts}}\}$.

通过空间变换$\boldsymbol{Q}$,本文减少了网络间行为不同的特性的权重. 此外,通过原始坐标基的复杂旋转和融合,揭示了自然人与网络平台无关的真实特征,对实现对齐大有裨益.

3.4 总体目标

算法1总结了DeepDSA方法的总体流程. 整个损失函数的定义如下:

$$\min L=\left(L_d^{\mathrm{s}}+L_d^{\mathrm{t}}\right)+\beta\left(L_f^{\mathrm{s}}+L_f^{\mathrm{t}}\right)+\gamma L_a \tag{13}$$

其中,L_d^{s}和L_d^{t}分别表示源网络和目标网络的目标函数,如式(6)所示,并且相同的规则适用于式(10)所示的L_f^{s}和L_f^{t}. β和γ是控制权衡的超参数. 为了稳定地训练,首先分别预训练L_d和L_f. 然后利用$\boldsymbol{Q}$进行正交初始化,以统一的方式学习整个框架. 本文使用RMSProp(Root Mean Square Prop)优化器优化整个框架.

算法1 DeepDSA方法

输入: 源动态网络G_{s},目标动态网络G_{t},对齐结点对集合P_{st},时间片数t,模型参数β和γ,学习率r

输出: 用于对齐的结果矩阵$\boldsymbol{X}_{\mathrm{s}}$和$\boldsymbol{X}_{\mathrm{t}}$

```
FOR 对每个动态网络 G DO
    将网络分为 T 个时间片;
    为每个切片 i 生成结构特征 S_i 和属性特征 A_i 作为模型输入;
    FOR G 中每个用户 v DO
        采样关于 v 的负样本;
    END FOR
    通过预训练 L_d 和 L_f 生成嵌入矩阵 U;
END FOR
生成指示矩阵 I 和负指示矩阵 M;
用正交初始化生成映射矩阵 Q;
WHILE 尚未收敛 DO
    训练式(13)
通过 Q 生成用于对齐的结果矩阵 X;
```

3.5 时间复杂度

假设源网络和目标网络的账号个数为n,网络分为T个时间片,嵌入维度为d. 首先进行动态性建模,GRU细胞包含三个门,故编码器的时间复杂度为$O(3Tnd^3)$,解码器相对于编码器加入了注意力机制,故其时间复杂度为$O(3Tnd^3+T^2nd)$. 其次是二元动态性融合,其首先将动态性与时间衰减效应相结合,然后计算联合概率密度,同时进行k次负采样,时间复杂度为$O\left(Tnd^2(1+k)\right)$. 最后在子空间进行对齐,具体步骤为空间变换和缩小距离,其时间复杂度为$O(nd^3+2nd)$. 省略常数后,总的时间复杂度为$O(Tnd(T+d^2+kd))$.

4 实验

4.1 数据集

本文的数据集基于经典数据集Twitter-Foursquare数据集(TF)[3,5,7,31]和自采样数据集豆瓣Online-Offline. Twitter和Foursquare网络分别有5 167和5 240个账号,2 858组已知的对齐结点对作为标签数据,然而,TF数据集不包含动态特征,并且获取同时具有动态信息和对齐信息的社交网络数据并不是易事. 本文以TF用户列表作为种子,对具有动态的结构和属性特征的账号列表进行抓取. 此外,为了验证算法在大规模社交网络中的有效性,本文还爬取了豆瓣社区网站. 豆瓣用户包括线上活动账号(Online)和线下活动账号(Offline),其中线上活动账号34 737个,线下活动账号34 076个,同时在线上和线下活动的账号有33 158个,本文将豆瓣用户间的关联作为结构信息,将豆瓣用户的评论作为属性信息. 在收集数据的同时将数据产生的时间记录下来作为时间戳. 在构建网络快照时,根据时间戳把处于同一时段的数据放到同一个快照中,从而还原真实世界中动态的社交网络. 例如,如果两个用户在某一时间段内有联系,则他们在对应快照中存在着边,最终这些边构成了快照中的结构信息;同时,用户在同一时间段内的所有评论将生成快照中的属性信息. 数据集的统计信息见表2.

表2　数据集统计信息

数据集	节点	链接	文本	对齐数
Twitter	7 440	72 658	3 467 472	3 896
Foursquare	6 771	69 350	289 674	
Online	34 737	1 294 814	155 296	33 158
Offline	34 076	1 228 485	101 648	

4.2 实验设置

对于所提出的DeepDSA方法,本文以3个月为间隔生成5个快照,并根据时间戳将信息分配到相应的快照中. 将式(2)中的结构特征的阶数和式(10)中的每个账号的负样本数目分别设置为4和5. 使用学习率为0.001的RMSProp优化器对整个框架进行优化,式(13)中的超参数β和γ分别为0.5和1. 将预训练迭代次数(算法1中的第7行)设置为1 000以稳定训练. 使用一个正交初始化的一层10个神经元的前向神经网络作为空间变换器$\boldsymbol{Q}$. 本文选择COSNET[16],MASTER[7],ULink[21],PALE[6],IONE[5],DeepLink[3]和DNA[19]作为对比方法. 所有对比方法的实验设置都是在原始文献的基础上实现的. 为了公平对比,所有对比方法的嵌入维度为128,DeepDSA中结构和属性的嵌入维度均为64.

通过随机删除账号生成不同的重叠率的数据集. 重叠率λ用$\frac{2N_{\mathrm{S}}}{N_{\mathrm{T}}+N_{\mathrm{F}}}$衡量,其中$N_{\mathrm{S}}$,$N_{\mathrm{T}}$和$N_{\mathrm{F}}$分别是标签数据、Twitter用户和Foursquare用户的数量.

4.3 性能指标

本文使用2种主要指标.

•Precision@k:其由$\frac{1}{N_A}\sum_{i=1}^{N_A}\mathbb{I}_i\{\mathrm{success@}k\}$求得,其中$\mathbb{I}_i\{\mathrm{success@}k\}$指示前$k$项的候选列表中是否存在(命中)真实对齐账号,$N_A$表示测试集中的标签数据数目.

•MAP@k:其由$\frac{1}{N_A}\sum_{i=1}^{N_A}\frac{1}{\mathrm{Rank}_i}$求得,其中$\mathrm{Rank}_i$表示前$k$项的候选列表中存在真实对齐账号的排名,$N_A$的定义同Precision@$k$. MAP以非线性方式突出强调了命中候选的排名.

指标值越高表示该方法的对齐效果越好.

4.4 实验结果

本文重复每个实验10次并计算平均结果,实验结果如下.

对齐效果. 图5展示了50%重叠率的TF数据集k值从1到30的比较结果,图6展示了50%重叠率的豆瓣数据集k值从1到100的比较结果. 可以发现,在TF数据集$k=30$时,DeepDSA的精度接近80%,而其他方法的精度为30%~66%,这表示本文的方法比最好的对比方法精度仍提高了10%,在豆瓣数据集中同样如此. 图7展示了TF数据集$k=5$时重叠率λ值从10到50的比较结果. 在这些对比中,DeepDSA和DNA在精度和MAP方面始终优于其他比较方法,这是动态性作用的有力证明. 所有的对比方法均存在对社交账户表示的局限性,不可避免地会丢失网络动态中丰富的信息. DeepDSA构造了一个目标子空间来建模潜在的真实自然人画像,把不同社交账户映射在同一目标子空间中进行对齐,而其他方法,例如IONE中的链接亦或PALE和DeepLink中的映射,则无法显示用于对齐的现实自然

人的真实画像. DeepDSA 以及 MASTER 和 COSNET 在大多数情况下由于同时利用了结构信息和属性信息而表现得更好,而其他方法则仅利用了结构(如 IONE)或属性(如 ULINK). 此外,与传统方法相比,DeepLink 由于具有深度模型的表示能力而表现得相对较好. 需要强调的是,本文固定了图 5 和图 6 中的 λ 值以及图 7 中的 k 值. 在接下来的实验中,仍采用 λ 和 k 的固定值. 事实上,DeepDSA 在不同的设置下总是表现得更好. 综上所述,DeepDSA 在考虑动态性的基础上,很好地对社交网络的综合信息进行了建模.

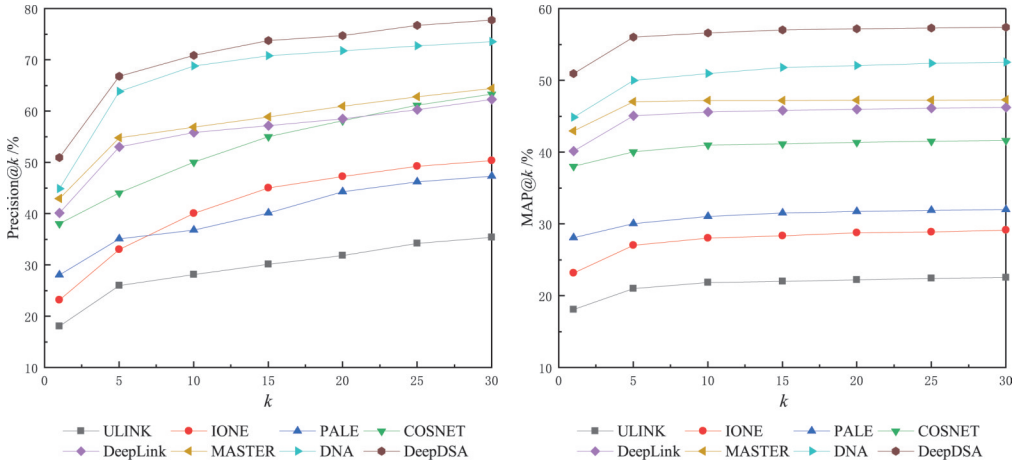


图 5 TF 数据集不同 k 值下的实验结果

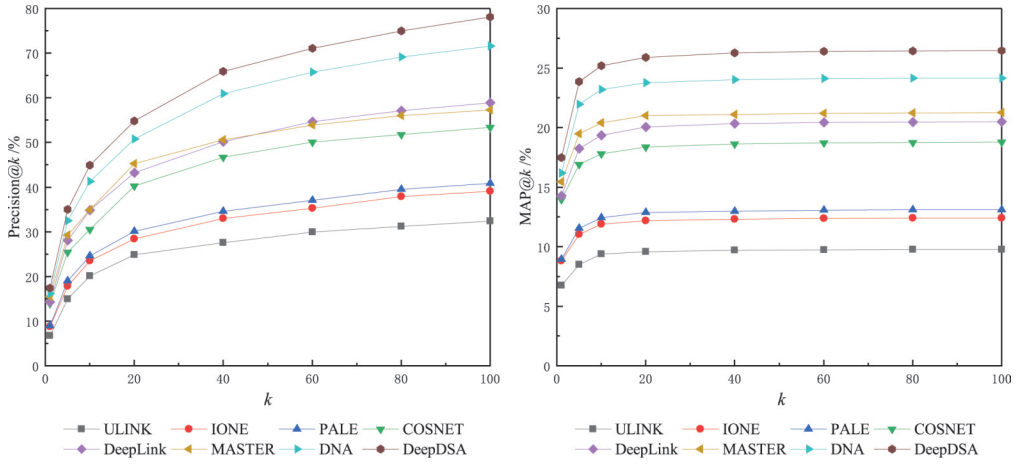


图 6 豆瓣数据集不同 k 值下的实验结果

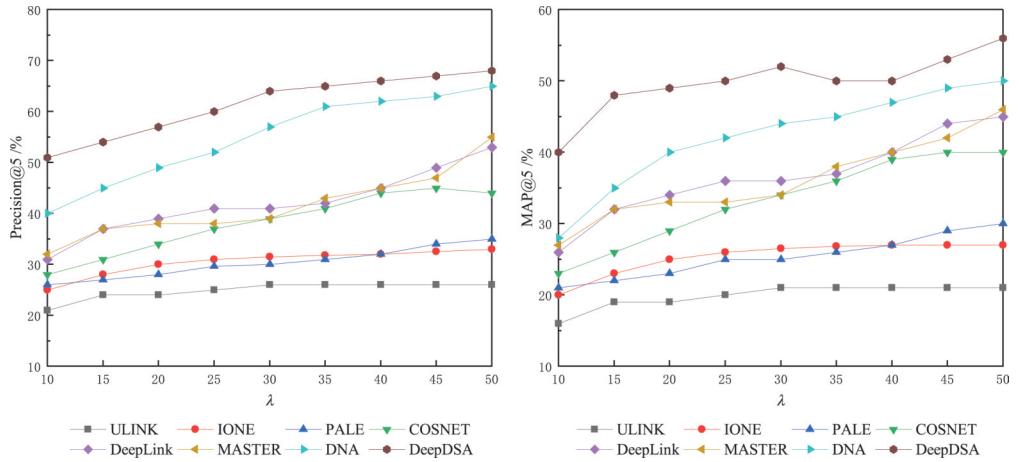


图 7 TF 数据集不同重叠率 λ 下的实验结果

训练率的影响. 图8展示了TF数据集训练率η对对齐结果的影响. 当η的值从1%增加到3%时,DeepDSA的性能会显著提高,随后当η超过3%时会饱和. 在重叠率为50%的TF数据集上,仅考虑少量的监督信息,DeepDSA在Precision和MAP方面都表现得更好. 此结果是符合期望的,因为在DeepDSA中,动态性是以无监督的方式编码,也就是说,本文可以在没有监督的情况下获得嵌入空间. 另外,通过少量的标签数据即可将原始嵌入空间进行对齐,学得空间变换. 总之,实验结果验证了DeepDSA强大的学习能力.

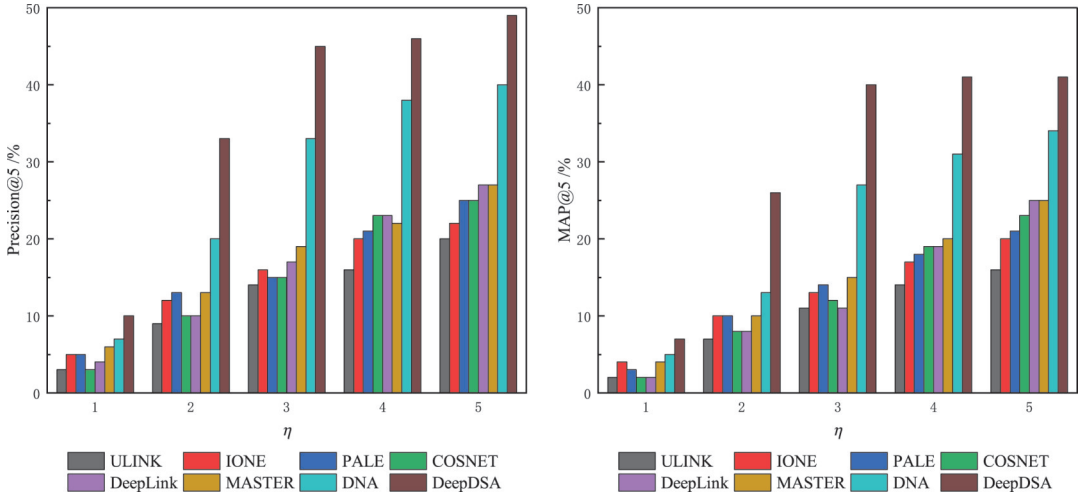


图8　TF数据集不同训练率η下的实验结果

动态性的影响. 图5展示了TF数据集下DeepDSA采用5个快照(间隔3个月)的对齐效果. 进一步,改变快照的时间频率和数量来聚焦于社交网络的动态特性,以研究动态特性如何影响网络对齐. 图9展示了DeepDSA在各种动态网络设置下的性能. 当快照的频率或数量超过一定的阈值时,Precision和MAP性能趋于饱和甚至下降. 其关键在于,从结构和属性信息中提取的用户动态行为模式具有高度的辨别能力,因为它主要与用户跨社交网络行为的动机和模式相关,这一论点在社会心理学研究中已得到了支撑[26]. 例如,一个用户通常会在几周内扩大或缩小一次他的朋友列表,而社交网络可能会在几个月内揭示出这种行为. 因此,当快照的时间频率与这种行为模式一致或者快照的数量足以捕获该行为模式时,性能往往更好. 然而,过于冗余的信息并不能进一步促进对齐,还可能引入噪声,导致性能的下降.

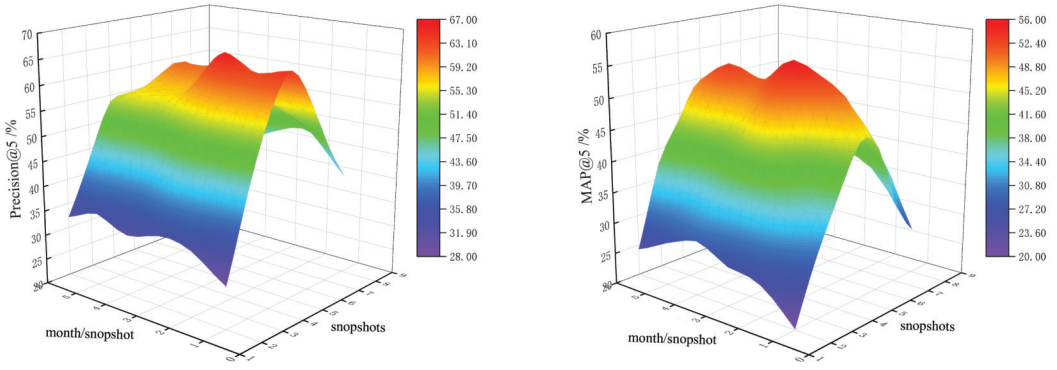


图9　TF数据集不同快照设置下的实验结果

嵌入维度的影响. 合适的嵌入维度对结果也有着不可或缺的影响. 本文将维度从10升至300,并在图10中展示了Precision和MAP. COSNET在本图中没有被列出,因为其是在不对用户进行嵌入表示的情况下实现的对齐. 当维度较低时,每种方法对这两种度量的性能都是不理想的,并且随着维度的增加而趋于上升. 当维度超过某个阈值时,性能开始下降. 然而,其中DeepDSA始终优于所有对比方法.

结构、属性和时间的影响. 图11展示了结构、属性和时间3个方面对用户嵌入和用户对齐产生的影响.

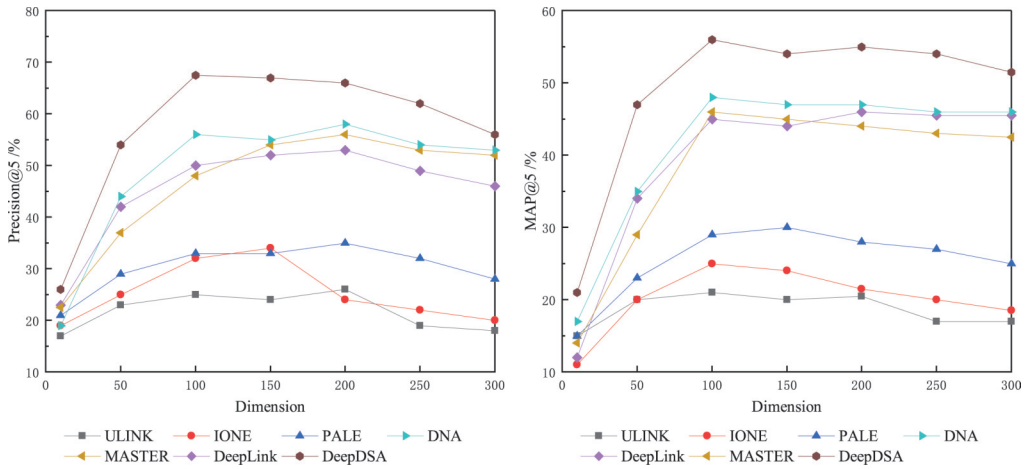


图 10 TF 数据集不同维度下的实验结果

DeepDSA 模型的输入包含 n 个切片以及各个切片的结构特征和属性特征，本文略微修改模型，分别去除模型输入中的结构信息、属性信息以及时间信息（不按照时间进行切片），得到的结果如图 11 所示．可以发现，包含结构、属性和时间信息的模型表现最好，没有属性信息的模型次之，没有时间信息的模型表现中等，没有结构信息的模型表现最差．这同样也暗合了图 6 中各方法的实验结果，DNA 没有属性信息，其表现次于 DeepDSA；MASTER 和 COSNET 没有时间信息，其表现中等；IONE 和 PALE 只有结构信息，表现较差；ULINK 只有属性信息，表现最差．

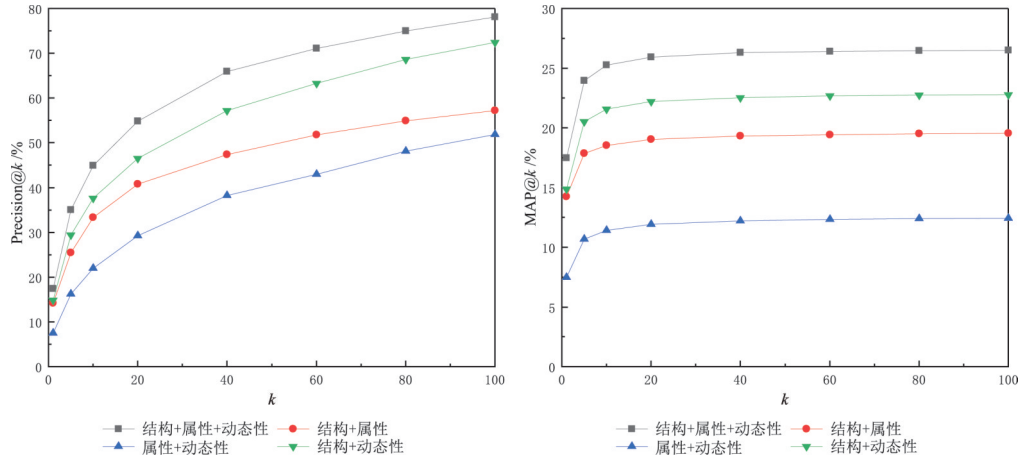


图 11 豆瓣数据集结构、属性和时间信息对实验结果的影响

5 结论

本文同时利用结构信息和属性信息解决动态社交网络对齐问题．事实上，社交网络中的动态性蕴含了一种有很强判别性的模式，这对实现网络对齐大有裨益．然而，在网络中揭示这种动态模式是一个巨大的挑战．为了填补这一研究上的空白，本文设计了 DeepDSA 方法来模拟社交网络中的丰富动态性，利用结构和属性信息来实现网络对齐．具体来说，在 DeepDSA 方法中，为了捕捉网络的动态特性，本文设计了一个深度神经模型来嵌入和融合结构动态性和属性动态性，以获得更精确的网络表示．为了缓解网络间普遍存在的差异性，本文学习了一个由少量监督信息引导的空间变换，在目标子空间中属于同一自然人的账号彼此临近．本文在真实的数据集上进行了大量的实验，实验结果表明 DeepDSA 明显优于现有的对齐方法．

参考文献

[1] CAO X Z, YU Y. BASS: A Bootstrapping approach for aligning heterogenous social networks[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Riva del Garda: Springer, 2016: 459-475.

[2] ZHOU J Y, FAN J X. TransLink: User identity linkage across heterogeneous social networks via translating embeddings[C]//IEEE INFOCOM 2019-IEEE Conference on Computer Communications. Paris: IEEE, 2019: 2116-2124.

[3] ZHOU F, LIU L, ZHANG K P, et al. DeepLink: A deep learning approach for user identity linkage[C]//IEEE INFOCOM 2018-IEEE Conference on Computer Communications. Honolulu: IEEE, 2018: 1313-1321.

[4] LI C Z, WANG S Z, YU P S, et al. Distribution distance minimization for unsupervised user identity linkage[C]//CIKM'18: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino: ACM, 2018: 447-456.

[5] LIU L, CHEUNG W K, LI X, et al. Aligning users across social networks using network embedding[C]//IJCAI'16: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. New York: AAAI Press, 2016. 1774-1780.

[6] MAN T, SHEN H, LIU S, et al. Predict anchor links across social networks via an embedding approach[C]//IJCAI'16: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. New York: AAAI Press, 2016. 1823-1829.

[7] SU S, SUN L, ZHANG Z B, et al. MASTER: Across multiple social networks, integrate attribute and structure embedding for reconciliation[C]//IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: AAAI Press, 2018: 3863-3869.

[8] TAN S L, GUAN Z Y, CAI D, et al. Mapping users across networks by manifold alignment on hypergraph[C]//AAAI'14: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. Québec City: AAAI Press, 2014: 159-165.

[9] WANG Y Q, SHEN H W, GAO J H, et al. Learning binary hash codes for fast anchor link retrieval across networks [C]//WWW'19: The World Wide Web Conference. San Francisco: ACM, 2019: 3335-3341.

[10] SHU K, WANG S H, TANG J L, et al. User identity linkage across online social networks[J]. ACM SIGKDD Explorations Newsletter, 2017, 18(2): 5-17.

[11] ZHANG S, TONG H H, MACIEJEWSKI R, et al. Multi-level network alignment[C]//WWW'19: The World Wide Web Conference. San Francisco: ACM, 2019: 2344-2354.

[12] XIE W, MU X, LEE R K W, et al. Unsupervised user identity linkage via factoid embedding[C]//2018 IEEE International Conference on Data Mining. Singapore: IEEE, 2018: 1338-1343.

[13] ZHOU F, WEN Z J, TRAJCEVSKI G, et al. Disentangled network alignment with matching explainability[C]//IEEE INFOCOM 2019-IEEE Conference on Computer Communications. Paris: IEEE, 2019: 1360-1368.

[14] ZAFARANI R, LIU H. Connecting corresponding identities across communities[C]//AAAI International Conference on Web and Social Media. Palo Alto: AAAI Press, 2009. 354-357.

[15] ZHANG J, CHEN B, WANG X M, et al. MEgo2Vec: Embedding matched ego networks for user alignment across social networks[C]//CIKM'18: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino: ACM, 2018: 327-336.

[16] ZHANG Y T, TANG J, YANG Z L, et al. COSNET: Connecting heterogeneous social networks with local and global consistency[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney: ACM, 2015: 1485-1494.

[17] ZHONG Z, CAO Y, GUO M, et al. Colink: An unsupervised framework for user identity linkage[C]//AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018. 5714-5721.

[18] ZAFARANI R, LIU H. Connecting users across social media sites: A behavioral-modeling approach[C]//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago: ACM, 2013: 41-49.

[19] SUN L, ZHANG Z B, JI P X, et al. DNA: Dynamic social network alignment[C]//2019 IEEE International Conference on Big Data(Big Data). Los Angeles: IEEE, 2019: 1224-1231.

[20] WANG Y Q, SHEN H W, LIU S H, et al. Cascade dynamics modeling with attention-based recurrent neural network[C]//IJCAI'17: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne: AAAI Press, 2017: 2985-2991.

[21] MU X, ZHU F D, LIM E P, et al. User identity linkage by latent user space modelling[C]//KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016: 1775-1784.

[22] GREFF K, SRIVASTAVA R K, KOUTNIK J, et al. LSTM: A search space odyssey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(10): 2222-2232.

[23] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: Online learning of social representations[C]//Proceedings of

the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2014: 701-710.

[24] YANG C, LIU Z, ZHAO D, et al. Network representation learning with rich text information[C]//International Joint Conference on Artificial Intelligence. Buenos Aires: AAAI Press, 2015. 2111-2117.

[25] LE Q, MIKOLOV T. Distributed representations of sentences and documents[C]//International Conference on Machine Learning. Beijing: JMLR.org, 2014: 1188-1196.

[26] UTZ S, TANIS M, VERMEULEN I. It is all about being popular: The effects of need for popularity on social network site use[J]. Cyberpsychology, Behavior, and Social Networking, 2012, 15(1): 37-42.

[27] ZUO Y, LIU G N, LIN H, et al. Embedding temporal network via neighborhood formation[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018: 2857-2866.

[28] LIANG S S, ZHANG X L, REN Z C, et al. Dynamic embeddings for user profiling in twitter[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018: 1764-1773.

[29] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017: 5998-6008.

[30] GAO H C, HUANG H. Deep attributed network embedding[C]//IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: AAAI Press, 2018: 3364-3370.

[31] KONG X N, ZHANG J W, YU P S. Inferring anchor links across multiple heterogeneous social networks[C]//CIKM'13: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management. San Francisco: ACM, 2013: 179-188.

作者简介

王飞扬 男,1999年生,山东德州人. 北京邮电大学博士研究生. 2020年获得北邮计算机科学与技术学士学位. 主要研究方向为机器学习、社交网络分析、图神经网络.
E-mail: fywang@bupt.edu.cn

冀鹏欣 男,1995年生,河北邯郸人. 2017年获得计算机科学与技术学士学位,2020年获得北邮计算机科学与技术硕士学位. 主要研究方向为大数据、机器学习、社交网络分析.
E-mail: jpx@bupt.edu.cn

孙 笠 男,1994年生,河北唐山人. 北京邮电大学博士研究生. 2016年获得北邮物联网工程学士学位. 主要研究方向为机器学习、社交网络分析、图神经网络、非欧几何机器学习.
E-mail: l. sun@bupt.edu.cn

危 倩 女,1997年生,湖北荆门人. 北京邮电大学硕士研究生. 2019年获得北邮通信工程学士学位. 主要研究方向为大数据与智能信息处理.
E-mail: wei_qian@bupt.edu.cn

李 根 男,1991年生,山东菏泽人. 2017年获得应用物理学士学位,2020年获计算机科学与技术硕士学位. 主要研究方向为机器学习、社交网络分析、图神经网络.
E-mail: genli@bupt.edu.cn

张忠宝(通讯作者) 男,1985年生,山东菏泽人. 博士. 北京邮电大学网络与交换技术国家重点实验室副教授、博士生导师. 主要研究方向为大数据、人工智能、社交网络分析.
E-mail: zhongbaozb@bupt.edu.cn