在线社交网络中群体影响力的建模与分析

孟 青[1)] 刘 波[1)] 张恒远[1)] 孙相国[1)] 曹玖新[2)] 李嘉伟[3)]

[1)](东南大学计算机科学与工程学院 南京 211189)

[2)](东南大学网络空间安全学院 南京 211189)

[3)](萨斯喀彻温大学计算机科学系 萨斯卡通市 S7N 5C9 加拿大)

摘 要 移动互联网技术的飞速发展,给社交网络平台带来了新的颠覆性的转变,也不断地改变着人们的生产、生活和交流方式.在线社交网络由于其特有的注册开放性、发布信息自由性、用户兴趣趋同性等特点,已经超越传统媒体,成为人们传播消息、获取新闻和接收实时信息的主要途径.同时,社交网络中用户之间的各种关系类型多样、相互交织、相互影响,促使用户生活在复杂的在线群体网络环境中,使得用户的在线行为时刻都受到所属的多种群体环境的影响作用.现有的针对在线群体环境影响的研究大多依据静态的、单一的网络结构对社交网络进行建模,而网络中通常存在多种类型的、动态的社会关系,较少研究能同时考虑多种类型的用户关系,建模社交网络中复杂环境下用户受到的影响作用.本文对用户所处的多类在线群体环境进行分析,挖掘用户所能感知的不同类型的群体环境,建模多维群体环境下用户所受的影响作用.首先,从用户间的社交关系类型出发,对在线社交网络中复杂的网络拓扑关系进行分类挖掘,分析用户可能感知的不同维度的在线群体环境,并提出静态群体环境和动态群体环境的定义和挖掘方法.其次,在不同的在线社交群体环境下,从宏观角度量化环境中用户所感知的群体结构特征,并从微观角度建模并模拟用户间的影响机制,提出了基于图注意力网络的融合多维在线群体环境的影响力模型.最后,以在线社交网络中用户的转发行为为例,研究多维群体环境影响下的用户行为模式,并在真实数据集上,基于群体影响力模型预测个体转发行为状态,验证模型的合理性和有效性.实验结果表明,本文提出的群体影响力模型能够更有效地描述在线社交网络中用户所属群体对用户的影响作用,并且在用户转发行为状态预测方面,比现有的群体影响力模型在综合评价指标 $F1$ 值方面最高可以提升 33%,在 AUC 值方面可提升 16%.

关键词 在线社交网络;群体;群体影响力;环境感知;图注意力网络

中图法分类号 TP393 **DOI 号** 10.11897/SP.J.1016.2021.01064

Multi-Relational Group Influence Modeling and Analysis in Online Social Networks

MENG Qing[1)] LIU Bo[1)] ZHANG Heng-Yuan[1)] SUN Xiang-Guo[1)] CAO Jiu-Xin[2)] LEE Roy Ka-Wei[3)]

[1)](*School of Computer Science and Engineering, Southeast University, Nanjing* 211189)

[2)](*School of Cyber Science and Engineering, Southeast University, Nanjing* 211189)

[3)](*Department of Computer Science, University of Saskatchewan, Saskatoon, Canada, S7N* 5C9)

Abstract The rapid development of Mobile Internet technology has brought new and subversive changes to online social network platforms, and also has changed the styles of people's production, lifestyle, and the way of communication. The characteristics of online social networks (OSNs), such as the openness of registration, the freedom of information diffusion, and the homophily of users' interests, have made OSNs overtaken traditional media (e.g. newspapers, TV, magazines)

收稿日期:2020-04-30;在线发布日期:2020-12-18.本课题得到国家重点研发计划项目(2017YFB1003000,2019YFC1521403)、国家自然科学基金项目(61972087,61772133,61632008)、国家社会科学基金项目(19@ZH014)、江苏省自然科学基金项目(SBK2019022870)、江苏省网络与信息安全重点实验室(BM2003201)、江苏省计算机网络技术重点实验室(BE2018706)、教育部计算机网络与信息集成重点实验室(东南大学)(93K-9)资助.**孟 青**,博士研究生,中国计算机学会(CCF)学生会员,主要研究方向为群体影响力、推荐系统、数据挖掘.E-mail:qmeng@seu.edu.cn.**刘 波**(通信作者),博士,副教授,中国计算机学会(CCF)会员,主要研究方向为在线社交网络、大数据分析、数据挖掘.E-mail:bliu@seu.edu.cn.**张恒远**,硕士研究生,主要研究方向为群体影响力、推荐系统与数据挖掘.**孙相国**,博士研究生,主要研究方向为社交网络分析和用户建模.**曹玖新**,博士,教授,主要研究领域为社交网络分析、隐私保护和数据安全.**李嘉伟**,博士,助理教授,主要研究方向为社会计算和自然语言处理.

as one of the main channels for users to reading news, show their daily life and receive their friends' dynamic information. At the same time, various types of relationships among users in social networks are interwoven and influence each other, which makes the online social network environments more complex and informative. Furthermore, users' online social behaviors are also affected by the diverse online social group environments that they belong to, which brings us a lot of challenges to social influence analysis. Recently, most of the previous studies that focus on the social influence of online group environment only use single type of relationship and assume that the relationships among users are static, and few studies can consider diverse types of relationships and the dynamics at the same time when quantifying the influence impacted on users in the complex environment of the social network. In this paper, we mine the diverse types of group environments that users perceive in OSNs and model the influence of users' multi-relational group environments. At last, we evaluate the influence of the online group environment that impacted on individuals and propose a Multi-Relational Group Influence model (MRINF) to predict the status of users' online retweet behaviors. Specifically, First, we analyze the diverse types of online social relationships and mine structural characteristics of complex network in online social networks platforms. Then, we give a deep analysis of users' perceived social group environment and propose two formalized definitions and give the group detection methods separately to find the potential group environments in social networks, which include the static social group environment and the dynamic social group environment. Secondly, we quantify the macroscopic structural features of two types of users' perceived groups. Moreover, using convolutional operation in graph attention network to simulate the information diffusion in OSNs, a multi-relational group influence model is proposed that combines the microscopic influence process among users and the macroscopic perceptions features. Finally, taking users' retweet behavior status as the main explicit expression of group environment influence in social networks, the proposed model is applied to the application of predicting individual retweet behaviors on two datasets and compared with the state-of-the art algorithms to verify the rationality and effectiveness of existing models. The extensive experimental results show that the MRINF model that is proposed in this paper could effectively describe the influence of users' perceived groups that impact individuals in online social networks from the static and dynamic perspective. In terms of the task of users' retweet behavior prediction, the MRINF outperforms other state-of-the-art algorithms on the evaluation metrics. Specifically, our model has an improvement of over 33% in $F1$-value and 16% higher in Area Under Curve (AUC) value compared with the existing state-of-the-art social influence model.

Keywords　online social network; groups; group influence; environment-aware; graph attention networks

1　引　言

在线社交网络中用户产生的大量关系数据，为社交影响力分析[1]、社交关系分析[2]、推荐系统[3]等领域的研究提供了新的可能. 而在线社交网络中用户的关系通常具有类型多样、错综复杂、动态变化等特点，给在线社交网络的研究带来了巨大的挑战. 例如，用户在社交平台中，通常会跟现实生活中的朋友互相关注，另外还会关注一些感兴趣的网络用户，可以建立用户间的关注(被关注)关系. 通过这种关系，用户能接收到所关注用户在线上发布的各种即时信息，与用户周围好友形成了相对稳定的在线社交群体环境. 同时，用户在社交网络中也会与其他用户进行互动. 例如，转发其他用户的推文或对其他用户推文进行留言、转发、点赞等. 这类交互关系和交互对象随着时间、用户兴趣偏好等因素而动态变化，组成了用户周围的动态社交群体环境. 用户周围的静态

以及动态社交群体环境相互重叠、相互交织、相互影响,共同影响着用户的在线社交行为.

群体环境影响用户的行为、偏好和情感等方面.有研究者[4]指出群体环境的影响作用所造成的用户行为改变无法用普通的社交关系影响来解释.而衡量群体社交环境的影响力为研究在线用户行为产生机制提供了新的思路.近年来,研究者们在研究信息传播、影响力最大化等问题中逐渐开始从不同角度涉及到群体影响力的衡量方法.我们将学术界对群体影响力的相关研究方法概括为以下几类:(1)将群体影响力量化为个体影响力的总和.此类研究通常利用社区发现算法发现网络中的社团结构并抽象为"超级节点(supernode)",基于信息传播模型,在群体层面计算各个超级节点的影响力[5-6].但是研究中缺少对群体层次的特征量化和对成员个体的细粒度研究;(2)通过聚类用户抽象层次的特征(观点、兴趣等)计算群体影响力.基于聚类群体的影响力研究主要考虑在线社交网络中的用户属性,对用户进行抽象层次的画像与聚类,挖掘同质的聚类群体[7],使用社会心理学理论中的群体一致性、群体规范等概念,量化个体在多个维度中受到的社会影响力.此类研究注重用户在属性特征方面的聚类性,较少考虑用户之间的实际连接关系;(3)以用户为中心研究个体受到周围群体环境的影响.通过分析在线用户的局部群体环境,挖掘用户所属中心网络的拓扑结构、内容等属性特征,衡量并量化用户所属局部群体环境对个体用户的社交影响作用[8-9],继而通过影响力排名或者行为预测任务衡量和分析模型的有效性.

由此可见,当前的研究存在以下问题:(1)现有群体影响力研究缺少对网络结构和用户的网络感知能力的细粒度建模与分析;(2)针对群体层次的研究大多为手工提取特征,依赖专家领域知识,具有一定的局限性;(3)目前的研究还较少注意到群体环境的动态性.

针对以上问题,本文从用户的静态社交环境和动态社交环境角度出发,对在线社交网络中群体环境影响力进行建模与分析.首先,对用户所属多种群体环境进行隐含特征学习,减少对人工特征的依赖.在此基础上,构建多粒度的影响力分析框架,研究用户在多种群体影响力的作用下产生个体行为的机制,并通过预测目标用户未来的行为状态,验证群体对个体的影响作用.主要贡献总结为以下三点(1)本文综合考虑了静态社交环境和动态社交环境,提出了以用户为中心的群体环境挖掘算法,检测出了符合影响力传播的群体环境;(2)本文提出了融合静态网络和动态网络的深度学习框架来计算群体影响力.该框架融合宏观层次、微观层次和时间演化特性,能够更好地刻画群体影响力;(3)建立了融合多维复杂群体环境的社交影响力模型,并在多个真实数据集上进行实验,实验证明本文提出的模型有较好的表现.

本文在第1节引言主要介绍相关背景和研究现状;第2节主要介绍相关工作;第3节描述所研究的问题,并给出形式化定义;第4节提出模型的基本思想,并且给出各模块详细描述;第5节描述实验设置和实验过程,并对实验结果进行分析;最后,第6节对文章进行总结和展望.

2 相关工作

群体环境的影响力主要基于用户所在的群体,利用在线社交网络中的行为表现,量化群体对个体的影响作用.Tang 等人[9]基于用户关注网络的拓扑结构,把用户周围群体环境划分为邻居用户、所属社区,对个体在群体环境中所表现出的从众行为进行了量化,并在用户转发行为预测任务中获得了良好的表现.张静[10]基于用户社交关系网络提出局部社交影响力(social influence locality)的概念,并利用概率图模型对用户所属的群体环境影响力进行量化.Jia 等人[11]综合社交网络中的信息和网络拓扑结构,利用个体的交互权重、影响力、传播意愿等,提出了基于社区一致性的影响力分析模型,分析群体对用户的影响作用.Jiang 等人[12]则重点关注用户的社会行为,提出用户之间的消息传播能力、交互活跃度等指标并进行线性融合,以此表示周围社会环境对用户的影响作用.

以上研究主要基于社会学理论,通过分析在线社交网络拓扑结构及用户社会行为信息,以人工定义特征的方式来量化社会影响力并用于用户行为预测.而通常在模型中,特征选择的合理性将对模型性能产生直接影响.同时,这类方法对于群体环境特征融合方法大多是线性方法,难以对用户间高阶的、非线性关系进行精准量化与描述.

随着神经网络的发展,深度学习开始用于社交影响力的分析过程中,相比传统人工特征的方法,深度学习模型的表达能力更强,能够挖掘更多潜在信息.Luceri 等人[13]基于深度神经网络来建模社交影

响力，通过学习用户所处群体环境中其他活跃用户的历史行为，来探究对目标用户是否产生实际影响.在网络嵌入表示学习层面，Wang 等人[14]基于 Node2vec 的方法提取用户在社交网络所处群体环境的拓扑信息，描述相似用户之间的社会关系，从而生成当前用户节点的嵌入向量表征，并应用于多标签社会行为预测，从而对模型有效性进行验证.而 Feng 等人提出的 Inf2vec 模型[15]，首先从局部群体环境出发，通过随机游走模拟社交影响的传播过程，并在全局范围内挖掘兴趣相似的用户群体，最后以网络嵌入表示学习的方式形成用户节点的向量表示，并用于多类用户行为预测任务.

近年来，图神经网络(Graph Neural Networks，GNNs)展示出处理网络结构数据的优越性，并在节点分类[16]、社会化推荐[17]等任务中取得了很好的效果.有学者利用图注意力网络[18](Graph Attention Networks，GATs)对用户的局部影响力展开了探索性研究.Qiu 等人[8]利用在线社交网络中用户之间的好友关系，基于图神经网络对用户所处的群体环境进行描述，并利用注意力机制量化了群体环境中成员对目标用户的影响力，提出了深度影响模型 DeepInf，并应用于用户转发行为预测，验证了模型的合理性和有效性.Wang 等人[19]同样基于图注意力网络，对社交影响力进行建模，通过图注意力机制融合周围亲密好友的影响作用，并结合网络表示学习思想提取群体结构的拓扑特征，将两部分特征融合后用于社交行为预测.以上基于图神经网络的社交影响力研究，均取得比传统模型更好的效果.

然而现有研究仍存在不足，主要体现在以下两个方面：首先，现有研究在研究群体环境影响时，大多只考虑用户与所属群体环境中成员之间的相互作用，较少考虑用户受到的来自群体层面的整体影响，而群体层面的影响不是个体间影响的简单叠加，它的作用不可忽视.其次，尽管部分研究在群体层面考虑了网络中用户所属的局部群体影响作用，但是大多工作是基于静态社交关系进行建模，缺少考虑在动态交互过程中形成的动态群体环境的影响，难以从多维群体环境角度，全面地分析用户所受到的社交影响力.

3　相关概念及问题定义

3.1　相关概念

在线社交网络中，用户可以通过关注其他用户，及时地获得关注用户的社交即时行为动态或者新闻信息；同时，当用户浏览到某些内容，由于即时的上下文环境的影响作用，可能会对部分推文进行转发、点赞等交互操作.以新浪微博为例，用户会根据自己的兴趣，关注微博平台中的新闻媒体、自媒体、好友等账号，这类用户的即时社交行为(发微博、转发等)会推送并显示在粉丝用户的微博页面中.当用户在浏览微博时，页面中展示的邻居用户发布或转发的微博，会对用户造成整体的信息冲击，使得用户也有概率对某些推文进行转发、点赞等.为了量化在线用户所属群体环境对其行为的影响作用，在此首先对相关概念进行定义.

定义 1.　关注网络.给定 $\bar{G}=(V,\bar{E})$，其中 V 为社交网络中的用户集合，$\bar{E}$ 为用户之间的关注关系集合，$\bar{E}=(u_i,u_j)$ 代表用户 u_i 关注了用户 u_j.

定义 2.　交互网络.给定 $\tilde{G}=(V,\tilde{E})$，其中 V 为社交网络中的用户集合，$\tilde{E}$ 为用户之间的转发关系集合，$\tilde{E}=(u_i,u_j,t)$ 代表用户 u_i 在时间 t 转发用户 u_j 的推文.

定义 3.　在线网络群体影响力.在线社交网络用户间通过多种社交行为(关注关系、发布推文、转发推文等)关联在一起，使得网络环境中形成一定群体氛围，从而对用户造成的影响作用，这种影响作用一般通过交互行为(转发行为)表现.

在线社交网络中，相比于交互关系，大多数用户的关注关系变更周期较长，本文中假设用户之间的关注关系为静态关系.所以，关注网络又称为静态网络.而通过转发关系而形成的用户之间的交互关系，具有较强的时间敏感性.所以，在本文中把交互网络又称为动态网络.

3.2　问题定义

给定网络环境 $G=\bar{G}\cup\tilde{G}=(V,E)$，其中 V 为社交网络中的用户集合，$E=\bar{E}\cup\tilde{E}$ 为网络中用户之间的有向边，代表用户间社交关系，包含用户之间的关注关系($\bar{E}$)和用户之间的转发关系($\tilde{E}$)；假设 $En_{v_i}\subseteq G$ 表示在线网络中用户 u_i 所属的群体环境.则问题可以定义为，给定在线网络中用户历史时间的多维群体环境 En_{v_i}，预测用户当前的行为状态.

4　群体影响力模型

社交网络中存在静态的网络结构和用户之间的动态交互关系，综合运用社交网络中静态结构属性和动态的交互关系网络能够更好地刻画用户在网络中所属的在线群体环境[20]，从而建模用户所受到的群体影响作用.

本文从在线社交网络中用户所处的网络环境出发，提出融合用户感知的多维度群体影响力量化模型(Multi-Relational based Group INFluence Model，MRINF)，模型的整体架构如图 1 所示. 模型包括静态融合模块、动态融合模块和群体影响力融合预测模块. 在静态融合模块和动态融合模块，首先，从宏观角度，通过嵌入表示方法获得用户对网络结构的感知特征；其次，设计群体发现算法挖掘用户所属的两个维度的群体环境，并对群体特征进行量化；然后，利用图注意力网络融合用户的感知特征向量，建模群体影响力的传播过程，从微观角度融合两个模块中的属性信息；最终，融合用户静态群体环境以及动态群体环境的信息，得到用户的表征向量. 在群体影响力融合预测模块中，利用前馈神经网络及多维群体环境中的表征属性，计算在线社交网络中用户所受到的群体影响力.

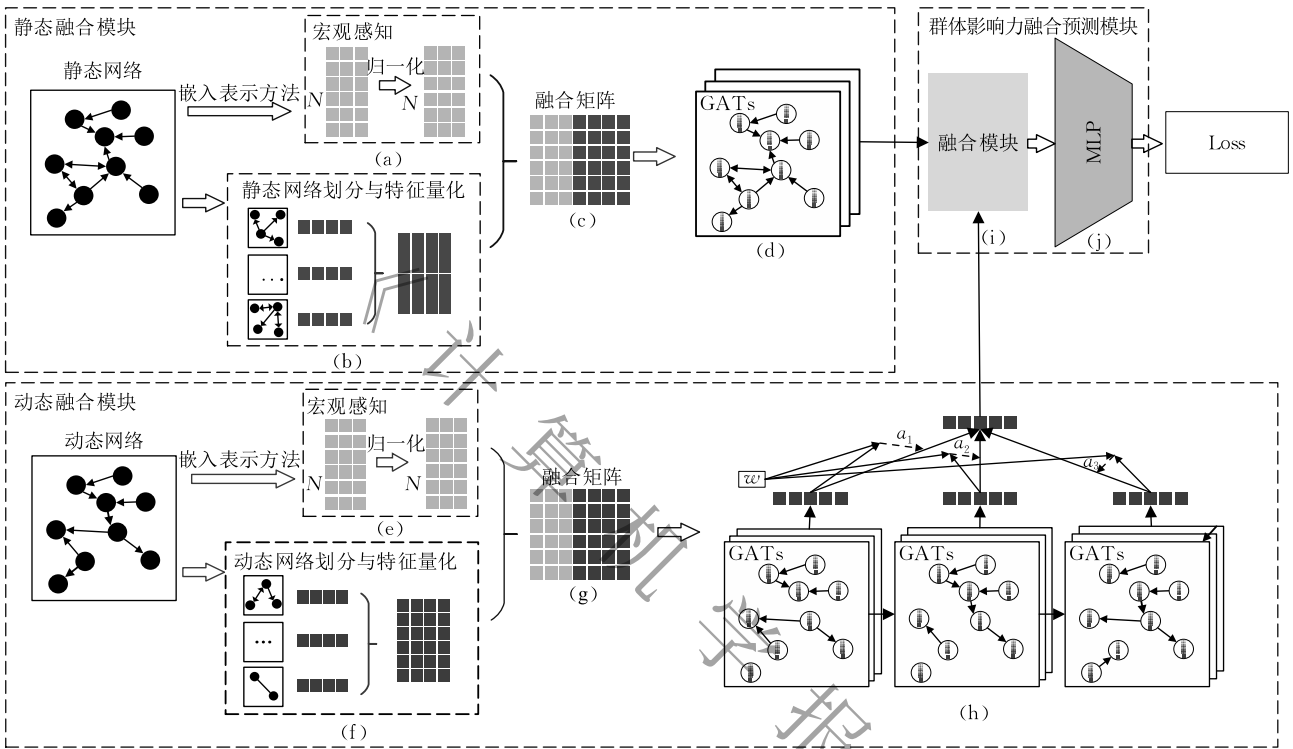

图 1 群体影响力模型框架

4.1 静态融合模块

4.1.1 静态结构属性嵌入层

网络嵌入通过设计特定的随机游走方法，构建能反映网络结构的游走序列集合，借鉴词嵌入的思想，把复杂的、高维的网络结构特征表示为低维的向量表示. 近些年网络表示学习方法发展迅速，诸多模型例如 Deepwalk[21]、LINE[22]、Node2Vec[23]、SDNE[24]等被提出，给社交网络分析中拓扑结构的特征提取带来极大的便利. 在模型的静态结构属性嵌入层，如图 1 中(a)所示，使用嵌入表示方法，提取节点的局部拓扑结构，表征用户对群体环境的感知. 利用网络嵌入表示学习方法，把静态关注网络中的用户节点 u 映射到低维向量表示空间，得到表征向量，表示为 $\boldsymbol{e}_{s,u}\in\mathbb{R}^D$，$D$ 为向量的维度.

4.1.2 静态群体挖掘模块

社交网络是由诸多用户和用户之间的关系组成的复杂系统，是真实社会环境中人际关系的写照. 根据邓巴数字理论[25]，人类的智力允许人拥有的稳定社交关系的人数大约为 150. 尽管通信技术的发展，拉近了人与人之间的距离、加快了信息流动速度，但是这个数字仍然没有大幅度的提升[26]. 另一方面，根据三度影响力理论，个体在社交网络中产生的影响，超过三度分隔，影响力就会大幅度减弱甚至消失[27]. 所以，在线社交网络中，对目标用户产生影响的群体环境在范围和数量方面是具有限制的.

定义 4. 影响限定原则. 根据社会心理学中的邓巴数字理论，在社交网络中对目标用户产生影响的用户数量是有上限的.

定义 5. 影响就近原则. 根据社会心理学三度影响力理论，在社交网络中对目标用户产生影响的用户群大多位于 3 跳以内的网络中.

根据定义 4 影响限定原则，建模中选择影响用户的候选用户群时，需要限定用户数量. 选取 Weibo 数据集(具体信息详见 5.1 节)进行统计，发现在线

社交网络中用户的一阶中心网络内，69.5%的用户的中心网络成员小于 150 人，而有 87.3%的用户的中心网络成员的数量小于 300. 而在用户的二阶中心网络中，仅仅有 1.8%的用户的二阶中心网络成员不足 300. 同样，在 Twitter 数据集中(具体信息详见 5.1 节)，大部分用户的二阶中心网络中的用户均超过 300 人. 因此，结合定义 5 影响就近原则，在选择影响用户的候选用户群时，采用目标用户的二阶中心网络中的用户作为影响用户的静态群体成员候选集合.

在目标用户的二阶中心网络，目标用户对网络中的成员的感知程度也具有一定的差别. 所以，基于二阶中心网络，以目标节点为固定起点，使用重启随机游走算法(Random Walk with Restart，RWR)对目标节点周围的用户进行随机游走，并累计周围节点的采样次数，从而模拟用户对周围用户的感知度($\mathcal{P}_i$). 然而，直观来说，在线社交网络中距离用户越远的个体，目标用户对他们的感知就越弱. 在此我们设计参数为节点之间最短距离(dis)的衰减函数 $f(dis)$. 当游走到节点 u_i 时，以 $f(dis)$ 代替原来的游走计数，之后再累计到节点的游走总次数中. 最终，依据目标用户对网络成员的感知程度，对候选节点进行排序，并选取 top-k 个节点作为用户在在线社交网络中的感知群体成员节点，得到用户感知的静态群体环境.

4.1.3　静态群体属性提取模块

静态群体是对个体用户所属网络环境的细粒度描述，静态群体在结构、活跃度等方面所具有的不同特征属性会影响个体行为状态. 在线社交网络中，每个个体都犹如在线的传感器，能感知所属的群体环境氛围，从而会受到不同的影响作用. 在此，根据静态群体的挖掘方法，对用户所属的群体环境进行特征提取，用平均感知度描述用户对群体结构的感知，用交互概率来描述用户所属群体的活跃度.

定义 6.　平均感知度($\overline{\mathcal{P}}$). 在静态群体的挖掘过程中，通过随机游走模拟用户对周围节点的感知程度. 而对于不同用户而言，由于在线社交群体环境的不同，或由于自身的个性化属性的不同，对周围环境的感知程度也有一定的差异. 在此，我们使用个体与所属群体环境中用户感知度的平均值，来衡量中心用户对群体环境的感知或熟悉程度，并描述用户所属群体环境的结构信息. 其形式化定义如下：

$$\overline{\mathcal{P}}_u=\frac{1}{|S_u|}\sum_{v\in S_u}\mathcal{P}_v \tag{1}$$

其中，$\mathcal{P}_v$ 为用户 v 对周围社交群体环境的感知度，$|S_u|$ 为用户所属静态群体环境中成员用户的数量.

定义 7.　交互概率(f). 静态群体环境相对稳定，所以用户所属的静态群体网络结构可能十分相似，从而降低了模型的预测能力. 为了进一步区分不同的静态群体环境，使用个体与所属群体环境中成员的交互频率，即用户所转发的源推文的发布者数量占群体成员数量的比例，来量化静态群体环境的动态变化，反映用户与所属群体环境的动态性. 其形式化定义如下：

$$f_u=\frac{\sum_{v\in S_u}\mathbb{I}(u,v)}{|S_u|} \tag{2}$$

其中，$\mathbb{I}(u,v)$ 为指示函数，如果用户 u 转发了用户 v 的推文，则 $\mathbb{I}(u,v)=1$. 否则，$\mathbb{I}(u,v)=0$.

4.1.4　静态群体属性融合层

在上述模块中，利用在静态结构属性嵌入层中得到的描述用户 u 的局部网络拓扑结构的表征向量 $\boldsymbol{e}_{s,u}$，以及用户所属静态群体中对静态群体的平均感知程度的量化特征 $\overline{\mathcal{P}}$ 和 f，衡量了每个用户对周围环境的感知程度. 而根据气泡过滤理论(filter bubble effect)[28]，用户周围邻居对用户所接收到的信息具有过滤效应. 所以，周围邻居用户在感知所属群体氛围的同时，也会将环境对自己的影响作用以影响力的形式传递到目标用户，进而影响目标用户的行为状态.

图神经网络通过卷积操作，可以直接在图结构数据(Graph-structured Data)上进行计算，实现对邻居节点属性的融合，在节点分类任务中获得了良好的表现[18]. 随后，图注意力网络考虑用户与周围邻居节点的重要性关系，可以对邻居节点的属性以及网络结构进行融合，能更细粒度地建模邻居用户对中心用户的影响力. 所以，我们基于图注意力网络，综合考虑用户对群体环境的感知过程，建立在线社交网络中群体影响力模型. 在宏观角度，拼接融合用户对网络结构的感知特征向量和用户所属静态群体的整体属性向量，得到 $\boldsymbol{s}_u=\{\boldsymbol{e}_{s,u}\|\overline{\mathcal{P}}_u\|f_u\}$，描述用户 u 对周围环境的感知程度. 其中，$\boldsymbol{s}_u\in\mathbb{R}^{|F|}$，$|F|$ 为用户所属在线群体环境的属性向量维数. 在微观角度，通过注意力机制建模用户所属群体中，群体成员用户与目标用户的影响关系，细粒度地融合用户所属静态群体成员的结构和属性信息.

静态群体属性融合层由多层图注意力网络组成，对于图注意力网络中的第 l 层，输入为 $\boldsymbol{H}^{(l)}=$

$\{\boldsymbol{h}_1^{(l)},\boldsymbol{h}_2^{(l)},\cdots,\boldsymbol{h}_n^{(l)}\}$，$\boldsymbol{h}_i^{(l)}$ 为第 l 层中用户 u_i 对在线群体环境感知的融合表征向量，$\boldsymbol{h}_i^{(l)}\in\mathbb{R}^F$，初始输入时 $\boldsymbol{h}_i^{(0)}=\boldsymbol{s}_i$，$n$ 为用户节点数量. 输出为 $\boldsymbol{H}^{(l+1)}=\{\boldsymbol{h}_1^{(l+1)},\boldsymbol{h}_2^{(l+1)},\cdots,\boldsymbol{h}_n^{(l+1)}\}$，$\boldsymbol{h}_i^{(l+1)}$ 为用户 u_i 融合了所属在线群体中用户感知属性和微观感知角度的感知向量，$\boldsymbol{h}_i^{(l+1)}\in\mathbb{R}^{F'}$，$F'$ 为输出的节点感知向量的维度. 具体而言，首先，计算邻居节点对目标节点的重要性，设计注意力函数 $attn$：$\mathbb{R}^{F'}\times\mathbb{R}^{F'}\to\mathbb{R}$，表示为 $e_{ij}=attn(\boldsymbol{W}\boldsymbol{h}_i,\boldsymbol{W}\boldsymbol{h}_j)$. 其中 $\boldsymbol{W}\in\mathbb{R}^{F'\times F}$ 为可学习的权重参数矩阵. 借鉴图注意力网络[18]，我们仅针对在线用户所属静态群体中的成员，计算其对目标用户的影响权重 e_{ij}. 一方面，可以增强目标节点对局部邻居信息的考虑；另一方面，减少了计算的时间复杂度和空间复杂度. 然后，利用 softmax 函数归一化用户与所属静态群体成员用户的影响权重，计算如下：

$$\alpha_{ij}=\mathrm{softmax}(e_{ij})=\frac{\exp(e_{ij})}{\sum_{k\in S_i}\exp(e_{ik})} \tag{3}$$

其中，S_i 为用户所属的静态群体. 特别地，对于注意力函数，使用一个权重参数为 $\boldsymbol{a}$ 的前馈神经网络，其中 $\boldsymbol{a}\in\mathbb{R}^{2F'}$，并且使用 LeakyReLU 激活函数对输出的值进行非线性变换. 最终，得到注意力机制的系数计算如下：

$$\alpha_{ij}=\frac{\exp(\mathrm{LeakyReLU}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{W}\boldsymbol{h}_i\,\|\,\boldsymbol{W}\boldsymbol{h}_j]))}{\sum_{k\in S_i}\exp(\mathrm{LeakyReLU}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{W}\boldsymbol{h}_i\,\|\,\boldsymbol{W}\boldsymbol{h}_k]))} \tag{4}$$

其中，$\|$ 为向量之间的拼接操作. 经过多层图注意力网络，得到静态群体属性融合输出：

$$\boldsymbol{h}_i^{(l+1)}=\sigma\Big(\sum_{j\in S_i}\alpha_{ij}\boldsymbol{W}_s\boldsymbol{h}_j^{(l)}\Big) \tag{5}$$

为了使得注意力机制在模型中能够发挥稳定的作用，在此同样使用多头注意力机制，即在学习过程中，使用 K 个相互独立的注意力结构进行并行学习，最后再整合各个结构的输出结果. 最终，在静态群体网络属性融合层中，得到融合用户的感知特征结果向量，表示为

$$\boldsymbol{h}_{\mathrm{static}}=\sigma(Aggregate_{i=1,\cdots,K}(head_i)) \tag{6}$$

其中，$Aggregate(\cdot)$ 是整合函数，一般利用平均值函数对各头的输出结构进行整合，$head_i$ 是第 i 个注意力机制的输出结果.

在静态模块中，首先，通过对用户周围静态环境整体的量化，获得目标用户在宏观角度对周围环境的感知表征向量. 其次，再通过图注意力网络，从微观角度融合表征向量，结合目标用户所属的静态群体环境的影响机制，获得目标用户对所属静态群体环境的感知向量表征. 最终输出向量表示为 $\boldsymbol{h}_{\mathrm{static}}$，融合了用户所属静态群体环境在不同层次的影响作用.

4.2 动态融合模块

4.2.1 动态结构属性嵌入层

在静态群体属性模块中，基于社交网络中相对稳定的关注关系，利用网络表示学习获得了用户静态群体环境的结构属性表示. 然而，用户在在线社交网络中的交互关系（例如，转发关系）由于其特殊的时间敏感性，更能反映用户所在的短期上下文环境和即时的用户间影响作用. 所以，在动态群体结构属性嵌入层，需要充分考虑并建模用户短期内的上下文环境. 首先，选取当前时刻之前一段时间内所有用户转发行为构建动态网络快照. 其次，类似静态群体结构属性嵌入层，基于已建立的动态交互网络快照，通过重启的随机游走构建节点序列集合，利用网络表示学习方法将节点映射到低维向量表示空间，表示成 D 维的向量 $\boldsymbol{e}_{d,u}\in\mathbb{R}^D$. 该向量表征了宏观角度下，动态关系网络中对用户 u 产生影响作用的网络结构特征.

4.2.2 动态群体划分模块

与关注关系不同的是，在线社交网络中用户的交互关系具有一定的时效性. 例如，在 Weibo 数据集中（具体信息详见 5.1 节），由于社交平台的显示限制和用户本身的浏览习惯，用户历史转发行为对用户当前影响作用是随时间逐渐递减的. 对数据集中用户相邻的转发行为进行统计，绝大部分（98.8%）的用户平均转发行为间隔不会超过 5 天，几乎所有（99.7%）的用户的转发行为间隔均在 10 天以内. 从用户的行为模式中可以发现，超过一定时间阈值的用户历史交互行为，影响目标用户的可能性极小. 所以，我们在动态群体的划分过程中，选择距离当前时间最近的一段时间（10 天）内目标用户转发过的用户对象及其转发关系所组成的一阶用户中心网络作为该用户的动态群体成员候选集合，建模在线社交网络中用户的动态群体环境.

4.2.3 动态群体属性提取模块

在动态结构属性嵌入层，通过网络表示学习把用户所属的动态群体环境的特征映射到低维向量空间中. 而通过对用户所属的动态群体环境整体属性进行描述，可以更为细致并且补充性地量化用户对群体环境的宏观感知. 因此，定义个体交互程度来描述用户所属动态群体环境的特点.

定义 8. 个体交互程度（c）. 个体交互程度指

在动态群体网络中，用户具有转发行为的次数. 转发次数描述了在最近的时间段内，用户与所属动态群体中成员用户的亲密程度以及用户本身的活跃程度. 具体表现为用户在动态群体所在时间窗口内的行为数量总和，其定义如下：

$$c = action_u^T \tag{7}$$

其中，T 为用户当前所处的时间窗口.

4.2.4　动态群体属性融合层

对于用户动态群体环境，在对用户感知的拓扑结构属性和用户动态属性进行融合后，得到用户 i 的表征向量 $\boldsymbol{h}_i = \{\boldsymbol{e}_{d,i} \| c\}$. 在用户间的微观融合方面，类似于静态群体环境属性的融合方法，用图注意力网络对个体所属动态群体环境属性进行融合. 同样，计算动态群体中具有交互关系的一对节点之间的影响权重 e_{ij}，并通过 softmax 函数进行归一化，最终结合激活函数 LeakyReLU 计算注意力机制的系数，具体过程不再赘述，注意力系数计算方式为

$$\beta_{ij} = \frac{\exp(\mathrm{LeakyReLU}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{W}\boldsymbol{h}_i \| \boldsymbol{W}\boldsymbol{h}_j]))}{\sum_{k \in D_i} \exp(\mathrm{LeakyReLU}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{W}\boldsymbol{h}_i \| \boldsymbol{W}\boldsymbol{h}_k]))} \tag{8}$$

其中，D_i 为用户 u_i 所属的动态群体环境中的用户集合.

经过多层图注意力网络，得到动态群体的属性融合输出为

$$\boldsymbol{h}_i^{(l+1)} = \sigma\Big(\sum_{j \in D_i} \beta_{ij} \boldsymbol{W}_d \boldsymbol{h}_i^{(l)}\Big) \tag{9}$$

其中，D_i 为用户 u_i 所属的动态群体环境用户集合. 最后，结合多头注意力，从而提高模型的稳定性和泛化能力. 集合多个头的输出得到节点对动态群体环境信息的融合向量为

$$\boldsymbol{h}_{\mathrm{dynamic}} = \sigma(Aggregate_{i=1,\cdots,K}(head_i)) \tag{10}$$

其中，$\sigma(\cdot)$ 为激活函数，$Aggregate(\cdot)$ 为整合函数，在模型中取均值函数. 但是，在动态网络的构建过程和动态群体环境的划分过程中，没有区分用户近期的动态交互关系随着时间变化的重要性. 然而，对于在线社交网络中的用户，个体行为对自身的影响在时间方面具有马尔科夫性质，即用户在某个时间点前的行为会对该时间点后的行为产生影响作用，而且距离当前时间越近的行为，对用户产生的影响越大. 所以，对于用户周围邻居节点所组成的动态群体，需要划分时间窗口，进行更加细粒度的分析. 最终，综合考虑用户交互关系的时间敏感性和交互数据的稀疏性，把用户的动态群体环境，根据用户交互行为的时间顺序，划分得到 $|T|$ 个动态子群体环境. 在每个时间窗口的子群体环境中，保持用户的群体环境中的用户节点不变，但仅保留在此时间片内的用户之间的交互关系，以此组成当前时间片内的动态子群体环境. 在不同的时间窗口内，分别并行使用多头注意力网络层对动态群体环境属性进行融合，如图 1 中(h). 具体而言，在每个时间窗口(T_N)中，把用户初始表征与前一时间窗口(T_{N-1})的输出向量 $\boldsymbol{h}^{T_{N-1}}$ 进行拼接，使得在不同窗口的模型具有一致的输入输出维度，实现模型参数共享，挖掘潜在的融合模式. 其中，在第一个时间窗口，初始输入向量 $\boldsymbol{h}^{T_0}$ 为全零向量.

在各个时间窗口得到的输出结果表示为 $\boldsymbol{h}_{\mathrm{dynamic}}^{T_1}, \cdots, \boldsymbol{h}_{\mathrm{dynamic}}^{T_i}, \boldsymbol{h}_{\mathrm{dynamic}}^{T_N}, \boldsymbol{h}_{\mathrm{dynamic}}^{T_i} \in \mathbb{R}^{d_k}$，$d_k$ 为输出向量的维度. 对于不同时间窗口的输出，引入注意力机制，量化各时间窗口的输出结果对用户行为的影响程度.

$$e_i = \boldsymbol{W}_f \boldsymbol{h}_{\mathrm{dynamic}}^{T_i} \tag{11}$$

$$\gamma_i = \mathrm{softmax}(e_i) \tag{12}$$

其中，$\boldsymbol{W}_f$ 为权重参数，γ_i 为第 i 个时间窗口的注意力权重. 最终，动态群体属性融合的最终输出 $\boldsymbol{h}_{\mathrm{dynamic}}$ 的计算表达式可以表示为

$$\boldsymbol{h}_{\mathrm{dynamic}} = \sum_{i \in N} \gamma_i \boldsymbol{h}_{\mathrm{dynamic}}^{T_i} \tag{13}$$

在本层得到的动态群体的属性融合向量 $\boldsymbol{h}_{\mathrm{dynamic}}$ 中，$\boldsymbol{h}_{\mathrm{dynamic}} \in \mathbb{R}^{d_k}$，包含用户所属动态群体结构的宏观感知和周围邻居节点影响作用的微观感知. 结合时间融合模块，在最终的表征中综合考虑了在不同时间窗口内交互关系的重要性.

4.3　群体影响力融合预测模块

对用户静态群体和动态群体的输出结果进行拼接 $\boldsymbol{h}_{\mathrm{final}} = [\boldsymbol{h}_{\mathrm{static}} \| \boldsymbol{h}_{\mathrm{dynamic}}]$，$\boldsymbol{h}_{\mathrm{final}} \in \mathbb{R}^{2d_k}$，如图 1 中(i)所示. 通过权重参数为 $\boldsymbol{W}_l$ 的前馈神经网络，其中 $\boldsymbol{W}_l \in \mathbb{R}^{2d_k \times 2}$，得到融合用户所属在线群体环境的表征向量，计算用户受到的影响力作用，从而预测目标用户的转发行为状态，如图 1 中(j)所示.

4.4　模型损失函数

模型最终输出为一个二维向量，代表用户在当前时间不同转发行为状态的概率. 最终，与数据集中的真实值进行对比，构建模型损失函数. 在此，使用对数似然函数作为模型的损失函数：

$$\mathcal{L}(\theta) = -\sum_{i=1}^{|V|} \log(p(a_i | Env_i; \theta)) \tag{14}$$

其中，a_i 为用户 i 的行为状态，当用户在当前时间具有转发行为时 $a_i = 1$，否则 $a_i = 0$. 在神经网络模型优化方面，采用固定批次大小的 Adam 优化方法[29]优化预测模型和更新参数.

5 实验结果与分析

5.1 数据集介绍

实验中所使用的数据集为从国内广泛使用的社交平台新浪微博①所采集的数据.具体采集方式为:首先,随机选择100个种子节点,基于关注关系使用广度优先遍历的方法采集距离种子节点为3度之内的用户节点集合.其次,对于所采集到的用户集合,爬取其发布的微博信息和转发信息.最终,随机选择连续一段时间(2015年11月)内具有转发行为的用户作为目标用户,作为实验数据集中的用户集合.另外,实验中还增加了国外社交平台Twitter②的公开数据集[30],数据集中包含用户之间的好友关系和用户之间的转发关系.该数据集是Twitter中关于"希格斯玻色粒子"发现事件的从2012年7月1日到7月7日的数据集.与Weibo数据集不同的是,该数据集中仅包含事件相关的用户和推文,所以用户间的关系和历史行为密度更加稀疏.各实验数据集的统计信息如表1所示.

表1 实验数据集统计

名称	数量	
	Weibo	Twitter
用户数量	18789	23066
关注关系	8774463	208564
转发关系	791825	45997
关注关系密度	2.48e-2	3.92e-4
转发关系密度	2.24e-3	8.64e-5

针对Weibo实验数据集和Twitter实验数据集,对用户的关注数量、用户转发数量分布进行统计.由图2、图3可以看出,在Weibo数据集和Twitter数据集中,用户的关注数量及其转发数量均呈幂律分布.例如,统计用户关注数量及其出现频次,可以发现绝大多数用户关注人数比较少,而只有少数用户会关注较多网络用户.以上统计结果证明了实验数据中的用户行为符合社交网络中用户行为的一般规律,证明了实验所选数据集的合理性和可用性.

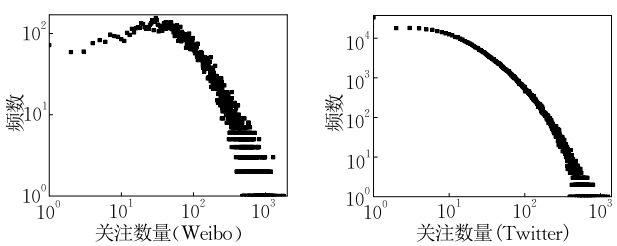

图2 用户关注关系数量频次统计

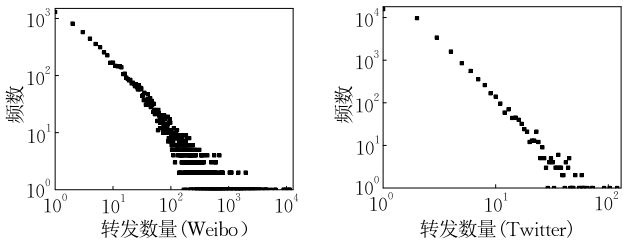

图3 用户交互关系数量频次统计

在实验中,首先选择日期t的数据作为用户行为的真实值,选择日期t之前的数据提取用户和网络特征.把实验数据集按照比例8∶1∶1进行划分,分为实验的训练集、验证集和测试集.其中训练集用来训练模型参数,验证集用来选择并确定能使模型达到最优性能的超参数.最后利用测试集的数据来测试模型的最终效果.

5.2 实验设置和评价方法

在实验中,设置用户所属静态群体规模为50,Weibo数据集中设置动态群体的规模为20,Twitter数据集中设置的动态群体规模为50.网络中节点嵌入表征向量的维度分别为16(Weibo)和64(Twitter).实验过程中拼接的手动提取的群体特征向量均为一维向量,例如由式(1)得到的平均感知度($\bar{\mathcal{P}}$)、由式(2)得到交互概率(f)、由式(3)得到的交互程度(c).在训练过程中使用Adam优化算法[29]进行优化和参数更新,学习率(learning rate)为0.0002,训练轮数(epoch)为1000,采用早停法(early stopping)防止模型过拟合,当模型的$F1$值在验证集上连续30轮不再增长时停止训练.设置衰减函数为指数函数$f(dis)=X^{dis}$,dis为用户节点间的最短距离,通过网格搜索算法确定$X=0.8$.在模型中图注意力网络的层数设置为3层.在动态群体属性融合层,首先,选择2个时间窗口对用户历史行为记录进行建模;其次,在Weibo数据集中选择3天(即$T=3$)为一个时间窗口的长度,在Twitter数据集中选择2(即$T=2$)天为一个时间窗口的长度.在每个时间窗口,使用3层图注意力网络层进行特征融合.对于对比模型均对参数进行调优使得获得最好表现.硬件环境为Ubuntu16.04系统,Intel® E5-2683@2.0 GHz CPU,显卡配置为GTX 1080 Ti.

对于模型的实验结果,使用广泛应用于分类问题中的评价标准:精准度(*Precision*)、召回率(*Recall*)、$F1$值($F1$-value)进行评估,并使用ROC

① http://www.weibo.com
② https://twitter.com

曲线下与坐标周围成的面积(Area Under Curve，*AUC*)评价模型的训练效果. 各个评价指标的具体计算方法如下：

$$Precision=\frac{TP}{TP+FP} \tag{15}$$

$$Recall=\frac{TP}{TP+FN} \tag{16}$$

$$F1=\frac{2\times Precision\times Recall}{Precision+Recall} \tag{17}$$

其中，*TP* 为真正例，为将正样本预测为正样本的数量. *FP* 为假正例，为将负例样本预测为正例样本的数量. *FN* 为假反例，为将正样本预测为负样本的数量. *AUC* 值为 ROC 曲线与坐标轴围成的面积. 衡量标准的值越大，证明模型的效果越好.

5.3　对比模型

为了验证所提 MRINF 模型的有效性，本文与现有研究中群体影响力建模的最新模型(state-of-the-art)进行对比，算法特点及具体介绍如下：

DeepInf. 文献[8]提出的模型，模型基于在线社交网络中用户间的关注网络结构，利用社会心理学中的从众效应，结合图注意力网络，对用户的局部群体影响进行建模，并利用用户的转发行为对模型进行评估与验证.

WL+GAT. 文献[19]提出的模型，该模型基于用户的社交关注关系和交互关系对网络结构进行建模，利用 Weisfeiler-Lehman 算法思想对网络中节点的拓扑结构进行向量表征，并结合图注意力网络对目标用户的属性进行融合. 最终，综合考虑多个方面的用户特征属性，量化在线社交网络中目标用户所受到的群体影响力. 通过利用用户间的多种关系构建网络模型，在一定程度上缓解了在线社交网络中数据稀疏性的问题.

MRINF. 本文提出的模型，综合在线社交网络中用户所属的静态群体环境和动态群体环境，基于图注意力网络设计的多维群体环境影响力模型，最终利用用户转发行为对模型进行评估与验证. 由于历史行为对目标用户的影响具有时间衰减性，在模型的动态群体属性融合层，对于用户的历史交互关系划分时间窗口，并在此基础上使用自注意力机制学习各个时间窗口对用户所属群体环境的影响作用. 在节点的嵌入表示方面，在静态网络结构中和动态网络结构中，均使用嵌入表示方法 LINE 作为节点表征向量的预训练模型.

5.4　实验结果分析

基于模型的设计思想和实验参数设置，对文中提出的模型和对比算法进行实现，并在 Weibo 社交网络数据集以及 Twitter 数据集中进行实验验证. 最终，各模型在两个数据集上的实验结果如表 2 所示.

表 2　在 Weibo 数据集及 Twitter 数据集上的实验结果

数据集	模型	*AUC*	*Precision*	*Recall*	*F*1
Weibo	DeepInf[8]	0.6905	0.4092	0.6564	0.5041
	WL+GAT[19]	0.7355	0.4362	0.6695	0.5282
	MRINF(ours)	**0.8535**	**0.7074**	**0.7060**	**0.7067**
	p-value	5.6e-5	6.5e-5	9.9e-3	1.0e-4
Twitter	DeepInf[8]	0.6797	0.3523	0.5167	0.4189
	WL+GAT[19]	0.7421	0.4108	**0.6691**	0.5091
	MRINF(ours)	**0.7522**	**0.4868**	0.6167	**0.5441**
	p-value	2.6e-2	4.8e-2	1.8e-1	2.0e-2

首先，从表 2 中可以看出，在两个数据集中，我们的模型在各综合评价指标均能达到最优，并具有显著提升. 其中，DeepInf 模型基于用户之间的关注关系网络进行建模，结合目标用户邻居的行为状态信息，预测目标用户的行为状态. 但是，对于网络中的很多用户而言，关注关系网络十分稀疏，而且仅基于关注关系对网络进行建模会忽略用户交互关系所组成的网络拓扑结构内存在的潜在影响关系. 而 WL+GAT 模型通过把用户的关注关系和用户间的交互关系融合到统一的社交交互网络(Social Interaction Network，SIN)中，并且对目标用户的拓扑信息进行编码，从而缓解了在线社交网络中用户间关系的稀疏性，提高了模型的性能. 本文提出的 MRINF 模型，对在线社交网络中的用户关系进行分类，建立用户所属的不同群体环境，并在静态群体环境和动态群体环境方面分别进行建模，能清晰地刻画多维群体环境对用户的影响作用. 而且模型在动态群体融合模块，对用户历史交互关系组成的社交环境在时间维度进行划分，且在此基础上利用注意力机制学习在不同的时间窗口内所属群体环境对用户当前行为的影响作用. 从而在 *AUC* 值、*F*1 值等综合评价指标方面都能达到最优的实验结果. 在实验中对文中所提模型与表现最优的对比算法在各评价指标上进行单样本 *t* 检验，当 $p<0.05$ 时，表明本文所提的模型在该评价指标下具有显著提升的效果.

其次，在 Twitter 数据集中，文中提出的模型与 WL+GAT 模型明显优于其它算法，并且与在 Weibo 数据集上的性能表现相比，两种模型在 Twitter 数据集上的表现差距变小，具体表现在 WL+GAT

模型在 *AUC* 值上与文中提出的模型接近，并且在 *Recall* 方面超过文中提出模型. 原因在于两个方面，一方面，融合用户的关注关系和交互关系对网络结构进行建模，在一定程度上能较为全面地量化在线社交网络中目标用户所受到的群体影响力. 本文所提模型和 WL＋GAT 模型均考虑了用户的关注关系和交互关系网络的拓扑结构，不同点在于本文所提的模型对两种关系进行分开建模，而 WL＋GAT 模型把两种关系建模为单一网络. 因此，两种模型在所有指标上性能均优于其它算法. 另一方面，相比于 Weibo 数据集，Twitter 数据集中用户的关注关系与交互关系更加稀疏. 在相对稠密的数据集中，区分在线社交网络中用户的群体环境能更加清晰全面地描述用户的复杂的在线社交群体环境信息，从而提高并凸显模型性能. 而在相对稀疏的数据集中，对群体环境进行分开建模会受到数据稀疏性的影响，从而影响模型性能的显著提升. 而 WL＋GAT 模型由于把各种关系建模为单一网络，可以缓解用户间单一类型关系的稀疏性，受到的影响较小. 因此在关系相对稀疏的 Twitter 数据集上，WL＋GAT 模型性能表现与本文所提的模型差距较小，甚至在 *Recall* 值上超出了本文所提的模型. 但是对于相对稠密的数据中，例如在关系相对稠密的 Weibo 数据集中，把各种关系建模为一个融合的网络会模糊目标用户周围不同类别的群体环境带来的影响作用，从而会降低模型的性能. 在 Weibo 数据集上的实验结果表明，文中所提的模型分别建模不同类别群体环境下的影响作用会有更好的表现，在各评价指标方面均明显优于 WL＋GAT 模型.

5.5 嵌入表示方法分析

图嵌入表示学习通过假设规则下的随机游走构建语料，把网络中的复杂拓扑结构属性的节点映射到低维的向量表示空间中，从而可以方便地进行节点分类、聚类、链路预测等下游任务. 实验模型中节点的属性主要来源于图嵌入表示方法，为了验证模型对各类图嵌入表示方法的敏感性以及其在模型中的作用，基于广泛使用的节点嵌入表示方法 Deepwalk(简称 DW)、LINE(其中使用一阶相似性时，表示为 L1，使用二阶相似性时，表示为 L2)以及 Node2Vec(简称 N2V)，分别构造相应的模型变种 MRINF_DW、MRINF_L1、MRINF_L2 以及 MRINF_N2V. 由于文献[8]中所提 DeepInf 模型也使用同类嵌入表示方法对网络中的节点进行了预训练，因此也加入进行对比分析. 各模型在两个数据集上的实验结果如图 4 所示.

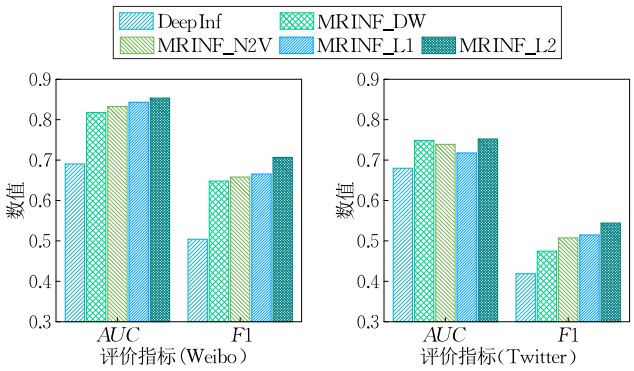

图 4 各嵌入表示方法实验对比

DeepInf 模型和文中所提模型变种 MRINF_DW 均使用嵌入表示算法 Deepwalk 预训练得到网络节点的嵌入表示向量作为节点的主要输入特征. 但是从图 4 中可以看出，在不同数据集中，MRINF_DW 模型在 *F*1 值和 *AUC* 值指标均有明显提高. 这是因为模型变种 MRINF_DW 仍然保留了对用户所属的静态群体环境和动态群体环境的建模，能更加全面地细粒度地描述在线社交网络中用户所属的多种维度的群体，较为完整地刻画了在线社交网络中目标用户所受到的多个维度的群体影响力.

对比其他图嵌入表示学习方法，在本文的任务中网络表示学习算法 LINE 和 Node2Vec 对用户网络结构属性的嵌入表示要比 Deepwalk 更好，使得模型能达到更好的性能. 而 LINE 的实验效果在各种网络嵌入表示方法中表现都要好，这也说明 LINE 能更好的表示用户的局部信息，而 DeepWalk 算法和 Node2Vec 算法则更着重表示节点在网络中的全局信息. 同时也能说明，相比于距离用户较远的用户集合，距离目标用户近的用户所组成的局部群体环境对用户有更强的影响作用. 另外，对比模型 MRINF_L2 模型和 MRINF_L1 模型，前者在各项评价指标均优于后者. 这可能由于 LINE 算法中的二阶相似性对于量化用户动态交互关系网络中直接连接的两个用户的相似性更加合理. 具体而言，在线社交网络中用户与其转发的源用户的行为模式可能会有较低的相似性，这对节点的向量表示造成影响，并且会进一步对模型的最终性能产生干扰. 例如，在社交网络中用户 A 转发了用户 B 的信息，并不能说明用户 A 与用户 B 在用户动态行为模式方面相似. 用户 B 是内容的发布者，通常是在线社交网络的大 V 用户，有较强的影响力；而用户 A 可能只是用户

B 千万粉丝当中的一位，本身在网络中的角色是不一样的. 所以，在基于动态转发关系组成的网络中，仅基于用户 A 用户 B 的一阶相似性，并不能较好地描述用户间的相似关系. 而 LINE 模型中的二阶相似性，建模了用户共同邻居的相似性，而此类相似性可能在动态转发网络中，更能描述用户之间的相似性. 例如，用户 A 和用户 B 所转发的博文，大多都来自相同的用户集合，则用户 A 和用户 B 是相似用户的可能性会增加.

图嵌入表示算法对向量表示的维度有可能会影响模型的性能. 不失一般性，在此选择 Tang 等人[22]提出的 LINE 模型对动态交互网络中的用户进行表示. 而静态交互网络中的用户表示对结果的影响类似，由于篇幅有限，在此不做赘述. 网络中节点向量的维度分别设置为 16、32、64、128、256. 具体实验结果如图 5 所示，随着节点向量的表征维度逐渐增加，模型在各综合评价指标方面的表现并没有大幅度变化. 一方面，这可能是由于高维的向量使得模型的参数增多，在建模用户影响力和预测用户行为方面会造成过度拟合，而向量维度的增加在某些数据集中也会使得模型难以训练，所以并不能有效提升模型效果. 例如，在 Twitter 数据集中评价指标会有一定幅度的下降. 另一方面，可能对于图神经网络模型而言，网络结构是决定模型性能的关键因素，所以节点属性的维度并不能显著地影响基于图神经网络模型的性能. 根据以上分析，本文在 Weibo 数据集中设置节点的向量表征维度为 16，在 Twitter 数据集中设置节点表征维度为 64，使得模型在综合评价指标 $F1$ 上有更好的表现.

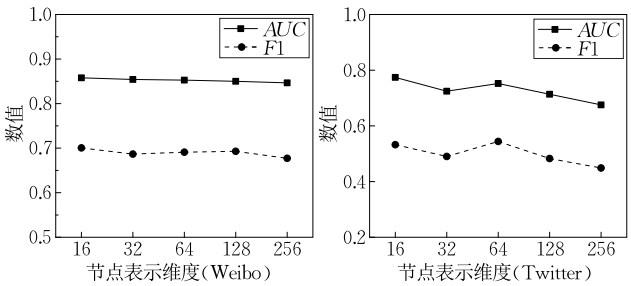

图 5　不同节点表征向量维度实验对比

5.6　消融实验

实验中 MRINF 模型的主要部分为静态模块(S)、动态模块(D). 其中在静态模块包含提取的静态群体特征(SF)，动态模块包含提取的动态群体特征(DF)、注意力机制(ATTN)核心组件. 在本节把去除静态模块和动态模块的模型称为模型的主体部分，表示为 $\mathrm{MRINF}_{\mathrm{base}}$；静态模块中去除静态群体特征部分称为静态网络基础部分，表示为 $\mathrm{S}_{\mathrm{base}}$；动态模块中去除动态群体特征和注意力机制部分称为动态网络基础部分，表示为 $\mathrm{D}_{\mathrm{base}}$. 在本节基于以上组件和模块设计消融实验验证各个组件和模块在模型中的作用.

首先，为了验证模型中各子模块在模型中的重要性以及对实验结果和模型性能的影响，根据各模块与子模块、特征之间的包含关系，在不同数据集中，通过控制变量的方法，在现有模型的基础上，对模型的各个子模块部分进行消融实验分析. 变种模型包括，去除模型的静态群体特征得到模型变种 MRINF-SF、去除模型的动态群体特征得到模型变种 MRINF-DF、去除模型的静态模块得到模型变种 MRINF-S、去除模型的动态模块得到模型变种 MRINF-D 以及去除模型的注意力机制得到模型变种 MRINF-ATTN. 其次，为了验证模型中基础模块在模型中的作用，组合以上基本模块得到模型的变种模型，包括保留静态网络基础部分得到模型变种 $\mathrm{MRINF}_{\mathrm{base}}+\mathrm{S}_{\mathrm{base}}$，保留动态网络基础部分得到模型变种 $\mathrm{MRINF}_{\mathrm{base}}+\mathrm{D}_{\mathrm{base}}$，以及保留上述两个基础模块得到的模型变种 $\mathrm{MRINF}_{\mathrm{base}}+\mathrm{S}_{\mathrm{base}}+\mathrm{D}_{\mathrm{base}}$. 最终，利用以上多个模型变种在实验中所用的两个数据集上进行实验，并对实验结果进行对比和分析，结果如图 6 所示.

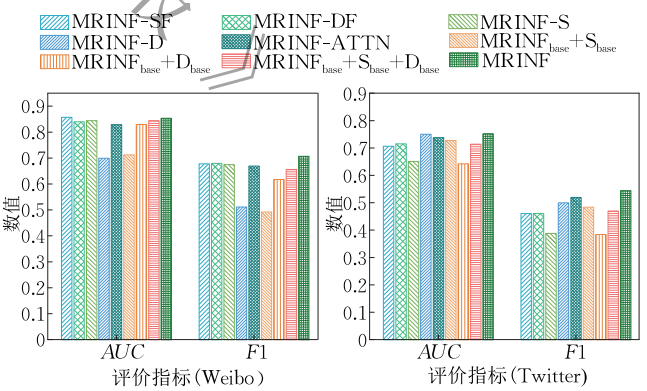

图 6　模型模块消融实验对比

从图 6 中可以看出，在不同的数据集中，去除模型的部分组件都会使得模型的实验结果造成一定幅度的下降，从而实验性地证明各个组件在模型中的重要性. 具体而言，首先，在群体的属性部分，对比 MFINF-SF 模型、MFINF-DF 模型与原始模型 MFINF，可以看出在两个数据集中模型的性能都存在下降的现象，且在 Twitter 数据集中尤为明显. 这

可能是因为 Twitter 数据集的稀疏性相对更高，所以通过外部知识所提取的特征能给模型带来更明显的提升.这也说明现有的深度模型在自动提取数据特征的时候仍然具有很多缺陷，加入含有领域知识的人工特征仍然能在一定程度上提升模型的性能.而在 Weibo 数据集中，可以看到去除静态群体特征部分，模型的 *AUC* 会有小幅度的上升，这可能由于 *AUC* 对正负样本的相对排序较为敏感，去除静态群体属性特征会让正负例样本的相对排序变得相对合理，但是却损害了部分样本的预测性能.其次，对比去除静态模块的模型(MRINF-S)和去除动态模块的模型(MRINF-D)，可以看出在两个数据集中表现出了不同的现象：在 Weibo 数据集中的实验结果显示，模型中动态模块的作用要远高于静态模块的作用.而在 Twitter 数据集中，模型中的静态模块的作用超过了动态模块.我们认为主要存在两方面的原因：一方面，社交网络中用户所属的静态群体环境对目标用户的行为有一定的影响作用，但是用户行为动因十分复杂且动态变化，而其所属的静态群体并不是用户所受影响力的唯一来源.具体地，静态群体环境中的用户由相对稳定的关注关系连接，能在一定程度上反映用户的行为习惯.但是，环境中包含的用户类型、活跃用户数量等参差不齐，容易给模型的预测带来噪音，从而影响模型性能进一步提高.另一方面，用户所属的动态群体成员均与目标用户在近期存在直接的交互关系，体现了目标用户近期的行为模式和主题偏好，能反映当前时间段内对用户行为造成的群体影响力.在 Weibo 数据集中，通过用户转发关系构建的网络结构相对稠密，能够为挖掘用户在动态网络中潜在的行为模式从而提供较为充足的信息.所以，用户所处的动态群体环境对用户的影响作用远超其所属的静态群体环境.相反，在 Twitter 数据集中，由于转发网络的稀疏性较高，不能为模型提供充足的信息，所以模型只能更多地根据用户所属的静态环境信息预测用户将来的行为状态.类似的结论在社会化推荐领域中也有相应的体现，社会化推荐算法通常基于用户动态的交互关系(用户-物品关系网络)来建模，可以为用户产生较好的推荐结果.但是当动态交互网络相对稀疏时，则需要借助用户间的静态社交关系为模型补充额外信息，从而缓解数据稀疏性.另外，对比原始模型与其变种模型 MRINF-ATTN，可以看出注意力机制同样也给模型带来性能上的提升.这由于社交网络中的用户通常具有不同的行为模式，不同时间片中的动态群体环境对目标用户产生的影响作用也不同，所以针对各个时间窗口内群体环境，给与不同的影响权重，并在此基础上进行差异性融合，可以在时间粒度方面更加细粒度地描述和追踪群体影响力的来源，从而进一步提高模型的性能.

对比 $\text{MRINF}_{\text{base}}+\text{S}_{\text{base}}$ 与 $\text{MRINF}_{\text{base}}+\text{D}_{\text{base}}$ 模型可以看出：在 Weibo 数据集中，基于用户动态转发关系构建的网络结构在模型中的作用本身要高于基于相对稳定的关注关系构建的网络结构.在 Twitter 数据集中，由于用户转发关系过于稀疏，而且缺少人工特征对群体环境信息的补充，所以通过转发关系构建的网络结构难以提供充足的信息来描述用户所属的动态群体环境，使得模型预测能力大幅度降低.与 $\text{MRINF}_{\text{base}}+\text{S}_{\text{base}}+\text{D}_{\text{base}}$ 模型对比，发现在 Twitter 数据集上，组合后的模型的性能会降低.这可能由于动态转发网络结构过于稀疏，在没有人工特征信息的补充的情形下，简单地组合动态网络基础模块与静态关注网络基础模块不能很好地建模用户群体环境，从而使得模型预测性能降低.

5.7 模型参数分析

模型中存在部分重要超参，包括群体规模、时间窗口的大小、时间窗口的数量，会影响模型的性能.针对上述超参，本节在两个实验数据集中展开对此类参数的实验及分析.首先，静态群体和动态群体的规模都会影响模型的表现，两者对于模型的影响较为类似，这里仅列出动态群体规模对模型的性能影响，如图 7 所示.

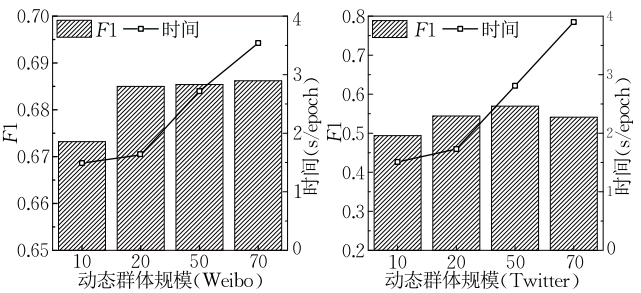

图 7 动态群体规模参数分析

在图 7 中可以看到随着动态群体的人数规模增加，模型在综合评价指标 *F*1 值方面，开始有明显提升，但是随着群体中选取的人数增加，模型的性能提升幅度开始变小，甚至在 Twitter 数据集中，还会出现一定幅度的下降.说明在用户所处的群体环境中，选择目标用户群体规模越大，对量化所受到的群体影响力带来的信息增益越大.但随着群体规模的继

续增加，信息增益的边际效用逐渐递减，而且由于新加入的用户节点很有可能带来噪音的干扰，模型的预测性能难以提升，甚至会呈现出下降趋势. 同时，随着群体规模的增大，预处理的时间成本、空间成本以及模型的时间复杂度都呈指数级增长，造成模型每轮训练时间也显著增加，从而降低了模型的整体性能.

在时间模块，对于超参数时间窗口的数量、时间窗口的大小等超参数分别进行实验分析，实验结果如图 8、图 9 所示. 首先，固定时间窗口大小设计实验分析超参数时间窗口数量的设置对模型的影响. 由于数据集限制，在 Weibo 数据集中时间窗口大小为 3 天，在 Twitter 数据集中为 2 天，具体实验结果如图 8 所示. 可以看出，在两个数据集中当时间窗口数量为 2 时，模型在综合评价指标 *AUC* 和 *F*1 值方面都具有最优的表现. 而当时间窗口的数量继续增大时，模型的性能并不会随之增高，反而可能还会出现下降趋势. 这说明距离用户一定时间内的动态群体环境对目标用户的影响相对较大，超出此范围的动态信息可能会提供额外信息，但是同时带来干扰信息的可能性也随之增大. 这与前文在 Weibo 数据集中的统计结果“98.8% 的用户的转发间隔都在 5 天内，99.98% 的用户的转发间隔都在 10 天内”相一致. 其次，固定用户时间窗口数量为 2，对超参数时间窗口大小进行分析，实验结果如图 9 所示. 从折线图中可以看出，在 Weibo 数据集中，当时间窗口大小为 3 天时，模型能得到最佳结果. 而 Twitter 数据集相对稀疏，而被预测用户的行为通常具有时间聚集性，所以选择时间跨度较小时，即当时间窗口大小为 2 天时，模型性能最优. 之后，随着时间窗口的继续增大，模型在 *AUC* 和 *F*1 值的综合衡量指标方面会出现不同幅度的波动或者下降. 说明动态群体在影响用户行为时，考虑到信息的丰富程度和噪音干扰等因素，在模型中选择合适的时间范围内的用户行为作为参考，能提高模型的在下游任务上的性能表现.

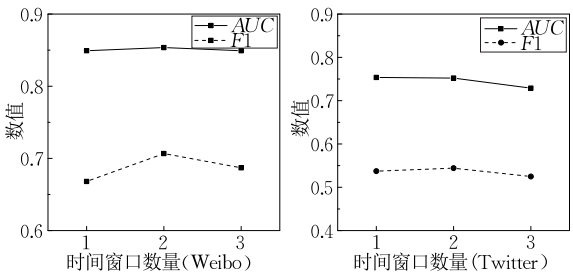

图 8　时间窗口数量参数分析

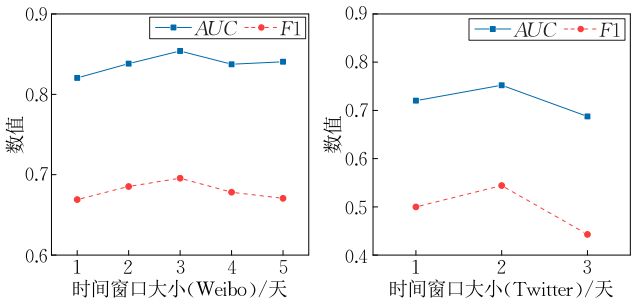

图 9　时间窗口大小参数分析

6　总结与展望

本文基于多个真实社交网络的数据，首先分析了在线社交网络中用户所处的复杂群体环境；其次，通过引入社会心理学的概念，量化了网络中多个维度下的群体影响力，提出了一种在线社交网络中群体影响力建模与分析方法；最终，应用于用户转发行为预测中，并通过丰富的实验证明了该模型的合理性和有效性. 实验证明，针对在线社交网络中用户所属的复杂群体环境进行细粒度地挖掘和建模，可以更加全面、清晰地衡量目标用户所受到的多个维度的群体影响力作用，并预测用户的在线行为. 而且由于用户之间的连接关系的形成原因和本质的不同，用户在网络结构中所表现出的性质也不能一概而论. 未来，在社交网络群体的主题方面，可以进一步对用户所受到的局部群体影响力进行细致的划分与描述，并分析影响用户的关键路径及其行为动因，从而能增强用户行为的可解释性. 另外，在用户动态关系稀疏性较高的情况下，也可以进一步挖掘用户间多种类型的隐式关系来补充额外信息，缓解稀疏性对模型的影响，从而能提高模型的通用性.

参　考　文　献

[1] Cha M, Haddadi H, Benevenuto F, et al. Measuring user influence in Twitter: The million follower fallacy//Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. Washington, USA, 2010: 10-17

[2] Zhao Shu, Liu Xiao-Man, Duan Zheng, et al. A survey on social tie mining. Chinese Journal of Computers, 2017, 40(3): 535-555(in Chinese)
(赵姝，刘晓曼，段震等. 社交关系挖掘研究综述. 计算机学报，2017，40(3)：535-555)

[3] Yin H, Wang Q, Zheng K, et al. Social influence-based group representation learning for group recommendation// Proceedings of the 2019 IEEE 35th International Conference

on Data Engineering. Macao, China, 2019: 566-577

[4] Jurgens D, McCorriston J, Ruths D. An analysis of individuals' behavior change in online groups//Proceedings of the International Conference on Social Informatics. Springer, Cham, 2017: 473-498

[5] Eftekhar M, Ganjali Y, Koudas N. Information cascade at group scale//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago, USA, 2013: 401-409

[6] Mehmood Y, Barbieri N, Bonchi F, et al. CSI: Community-level social influence analysis//Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin, Germany: Springer, 2013: 48-63

[7] Yu J, Gao M, Li J, et al. Adaptive implicit friends identification over heterogeneous network for social recommendation//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino, Italy, 2018: 357-366

[8] Qiu J, Tang J, Ma H, et al. DeepInf: Social influence prediction with deep learning//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK, 2018: 2110-2119

[9] Tang J, Wu S, Sun J. Confluence: Conformity influence in large social networks//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago, USA, 2013: 347-355

[10] Zhang Jing. Modeling and Measuring Social Influence[Ph. D. dissertation]. Tsinghua University, Beijing, 2016(in Chinese)
(张静. 社会网络影响力建模与度量[博士学位论文]. 清华大学, 北京, 2016)

[11] Jia X, Li X, Du N, et al. Influence based analysis of community consistency in dynamic networks//Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). San Francisco, USA, 2016: 1-8

[12] Jiang B, Liang J, Sha Y, et al. Retweeting behavior prediction based on one-class collaborative filtering in social networks//Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. Pisa, Italy, 2016: 977-980

[13] Luceri L, Braun T, Giordano S. Social influence (deep) learning for human behavior prediction//Proceedings of the International Workshop on Complex Networks. Cham, Springer, 2018: 261-269

[14] Wang X, Guo Z, Wang X, et al. NNMLInf: Social influence prediction with neural network multi-label classification//Proceedings of the ACM Turing Celebration Conference-China. Chengdu, China, 2019: 1-5

[15] Feng S, Cong G, Khan A, et al. Inf2vec: Latent representation model for social influence embedding//Proceedings of the 2018 IEEE 34th International Conference on Data Engineering (ICDE). Paris, France, 2018: 941-952

[16] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks//Proceedings of the 5th International Conference on Learning Representations. Toulon, France, 2017

[17] Fan W, Ma Y, Li Q, et al. Graph neural networks for social recommendation//Proceedings of the World Wide Web Conference. San Francisco, USA, 2019: 417-426

[18] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks//Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada, 2018

[19] Wang H, Meng Q, Fan J, et al. Social influence does matter: User action prediction for in-feed advertising//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020: 246-253

[20] Song W, Xiao Z, Wang Y, et al. Session-based social recommendation via dynamic graph attention networks//Proceedings of the 12th ACM International Conference on Web Search and Data Mining. Melbourne, Australia, 2019: 555-563

[21] 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

[22] 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

[23] 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

[24] 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

[25] Dunbar R I. Neocortex size as a constraint on group size in primates. Journal of Human Evolution, 1992, 22(6): 469-493

[26] Wang Q, Gao J, Zhou T, et al. Critical size of ego communication networks. Europhysics Letters, 2016, 114(5): 58004

[27] Christakis N A, Fowler J H. Connected: The surprising power of our social networks and how they shape our lives. Little, Brown Spark, 2009

[28] Rowland F. The filter bubble: What the Internet is hiding from you. Portal: Libraries and the Academy, 2011, 11(4): 1009-1011

[29] Kingma D P, Ba J. Adam: A method for stochastic optimization//Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA, 2015

[30] De Domenico M, Lima A, Mougel P, et al. The anatomy of a scientific rumor. (Nature Open Access) Scientific Reports, 2013, 3: 2980

MENG Qing, Ph. D. candidate. His main research interests include social influence, recommender systems, and data mining.

LIU Bo, Ph. D., associate professor. Her main research interests include social networks, big data analysis, and data mining.

ZHANG Heng-Yuan, M. S. candidate. His main research interests include social influence, recommender systems, and data mining.

SUN Xiang-Guo, Ph. D. candidate. His main research interests include online social networks and user modeling.

CAO Jiu-Xin, Ph. D., professor. His main research interests include social network analysis and privacy protection.

LEE Roy Ka-Wei, Ph. D., assistant professor. His main research interests include social computing, and natural language processing.

Background

Social influence depicts the phenomenon that users' emotions, opinions, and behaviors could change due to the social environment. It has been studied for a decade due it is widely applied in the area of user modeling, recommender systems and information diffusion analysis, etc. From the perspective of the granularity of influencers and influencees, social influence could be classified into four taxonomies: the influence among individuals, the individual influence on the group, the group influence on individuals, the influence among groups. In this paper, we mainly focus on the group influence on individuals. In the previous studies, the researchers have modeled the group influence in various applications, like social phenomenon, influence maximization, and information diffusion. Most studies mainly focus on the macro diffusion pattern in the network and neglect the interacting details among individuals. On the other hand, some researchers regard the group as a super-node, to reduce the model's complexity. In this paper, from the perspective of the entire group, we model the multi-dimensional social group influence on individuals, which is also simply called group influence in social psychology. Specifically, first, we analyze the users' online social behaviors and propose group detection methods to find the potential groups in the complex social network. The online social group environments are classified into two categories: the static group environment and the dynamic group environment. Secondly, based on this, we model the online social environments from both the macro perspective and the micro perspective. At the macro level, an embedding algorithm is used to catch the global network structure, and on the micro-level stuff, graph attention networks are leveraged to simulate the influence diffusion via information propagation. At last, we apply our model to the application of users' behavior prediction, and extensive experiments are conducted on two datasets. The experimental results show that the proposed model in our paper could describe the overall influence of group environments, and can achieve a better performance in most evaluation metrics.

This work is supported by the National Key R&D Program of China (Grant Nos. 2017YFB1003000, 2019YFC1521403), the National Natural Science Foundation of China (Grant Nos. 61972087, 61772133, 61632008), the National Social Science Foundation of China (Grant No. 19@ZH014), the Natural Science Foundation of Jiangsu Province (Grant No. SBK2019022870), the Jiangsu Provincial Key Project (Grant No. BE2018706), the Key Laboratory of Computer Network Technology of Jiangsu Province, the Jiangsu Provincial Key Laboratory of Network and Information Security (Grant No. BM2003201), and the Key Laboratory of Computer Network and Information Integration of Ministry of Education of China (Grant No. 93K-9).