

社交网络结构预测健康和健康

Suwen Lin^{1,2}, Louis Faust^{1,2}, Pablo Robles-Granda^{1,2}和Nitesh V. Chawla^{1,2,*}

¹美国圣母大学圣母大学计算机科学与工程系, IN 46556, USA

²美国圣母大学圣母大学网络科学与应用跨学科中心, IN 46556, USA

*通讯作者。 电子邮件: 露营 “nd.edu WLA

摘要

社交网络影响健康相关的行为, 如肥胖和吸烟。 虽然研究人员已经将社交网络作为影响和行为传播的驱动因素进行了研究, 但人们对网络的结构或拓扑本身如何影响个人的健康行为和健康状况却知之甚少。 在本文中, 我们调查社交网络的结构或拓扑是否为个人的健康和健康提供了额外的洞察力和可预测性。 我们开发了一个名为网络驱动的健康预测器 (NetCARE) 的模型, 该模型利用代表社交网络结构的功能。 使用在圣母大学NetHealth研究中注册的大量学生纵向数据集, 我们证明NetCARE模型比使用人口统计学和物理属性的基线模型提高了整体预测性能38%, 65%, 健康状态的55%和54% - 压力, 快乐, 积极态度和自我评估的健康状况 - 在本文中考虑。

介绍

社交网络在行为, 态度, 品味和信仰的传播中发挥着重要作用。 一些研究表明, 这些特征利用了现有的社会联系和扩散联系。 这种现象证明了网络中节点之间的相似性或同质性 (例如, 自我和改变), 但也影响了影响人们的社会影响。 这种扩散过程的一些例子包括: 不同健康状况对社交网络的传播机制 - 如肥胖¹和吸烟², 社交网络对个人心理特征的影响 - 如情感³和幸福感 (TF21), 以及通过社交网络传播健康行为⁴。 人们通过社交网络或社交媒体平台进行的互动也被用于发现个体⁵, 精神疾病⁶ 和活动模式⁷所经历的情绪方面。 不同的社交网络类型, 如友谊或非友谊网络, 也可以提供预测成人心理健康的见解⁸。

虽然这些研究表明社会网络效应在行为, 态度和信仰的传播中普遍存在, 但还有其他自我选择和外部因素影响行为和态度 (背景, 文化, 倾向) 的相似性。 考虑网络效应假设。 它表明生活方式和健康实践中的相似之处, 包括健康行为, 情绪, 情绪, 文化规范等^{9, 10}, 也是其网络中通过关系影响和传播的结果。 此外, 自我选择假设表明人与人之间的联系是由对态度或信仰或行为的类似预先处置所驱动的, 因此这些因素甚至可能推动关系的形成^{11, 12}。 但是, 网络和自我选择假设如何帮助解释一个人的健康和健康状况? 我们假设两者对于解释和预测个体的健康状态至关重要。 存在自我选择和外部因素 (例如人口统计学和个体活跃倾向) 以及网络因素 (通过通信网络获得的关系), 这些因素共同导致个体的健康状况改善。 我们将这个假设分解为本文中的两个研究问题, 这些问题也推动了核心贡献。

RQ1: 社交网络结构是否表明健康行为? (分析)

我们认为社交网络结构由社交网络属性表示, 例如节点度, 聚类系数和中心性。 我们将健康和保健行为视为从可穿戴设备捕获的数据 - 心率, 每日步数和活动状态 - 以及调查数据。 我们定量和定性分析社交网络结构与上述健康信息之间的关系。 图中显示了这种关系的一个例子 [1a](#)。 该图显示了网络上的节点度 (如虚线框图所示) 如何与心率的变化相关 (用常规线表示)。 该图显示这些值似乎正相关, 因为随着时间的推移, 当均值或中位数时, 节点度的平均值和中位数 (分别显示为蓝线和绿色三角形) 增加或减少

心率（分别显示为橙色线和深绿色三角形）也增加或减少，其中相应的归一化互相关为0.84。在本文中，我们将提供证据表明社交网络结构包含捕获健康行为统计变化的信息。

RQ2: 融入社交网络结构的健康状况如何可预测？（预测）

虽然之前的研究表明，从可穿戴设备中捕获的健康行为数据是疾病或疾病症状的指示^{2, 3}，但在本文中，除了可穿戴设备使用预测的机器学习框架捕获的健康行为数据外，我们还结合了社交网络结构特征。健康和保健方面（NetCARE）。我们考虑各种健康和健康状态，如压力，快乐，积极态度和自我评估的健康指标。数字 1b 通过涉及网络结构信息，总结了整体F1-测量和类内F1-测量的积极态度预测的改进。显然，与仅使用可穿戴设备和/或个人的入口统计数据相比，社交网络结构的知识提供了显著的改进。

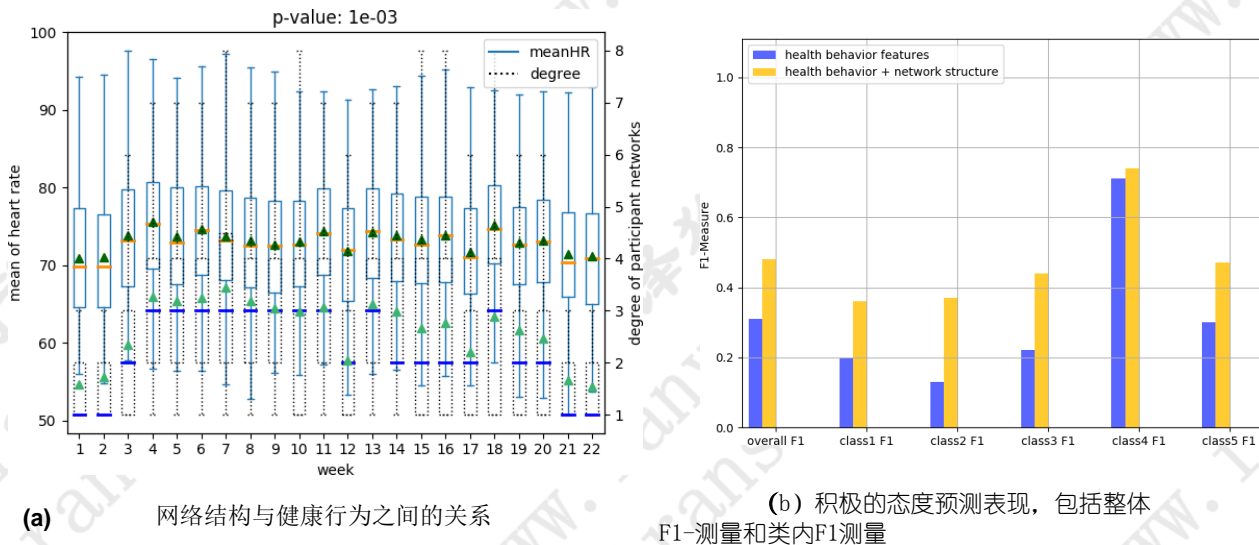


图1. 研究问题的主要结果

方法

数据

我们使用NetHealth研究²的数据，该研究是圣母大学正在进行的项目，收集了2015年秋季入学的698名一年级学生²的初步队列中的调查，电话和Fitbit数据。所有程序均由Notre Dame IRB完全批准。图中提供了招聘流程和学生样本编号的概述²。向参与者提供了Fitbit Charge HR，并在他们的手机上安装了一个应用程序，用于根据通信模式（呼叫，消息）构建社交网络。每个参与者都是匿名的，并给出了唯一的hashID。为了研究的目的，数据分为三个部分：调查数据，社交网络数据和Fitbit数据。

学生在到达校园之前必须完成入学调查，并在每学期后进行后续调查。调查包含一系列关于个人入口统计和自我报告的心理和身体健康评估的问题。应该指出的是，每个学期的调查问题都不同，并非所有学生都参与了所有的调查。我们考虑NetHealth的以下三个数据集：

社交网络数据. 学生的社交网络是使用他们的沟通活动建立的，包括通过应用程序捕获的文本和电话。该应用程序可以自动收集其通信活动的时间，来源和目的地。至于电话，应用程序还可以记录通话的持续时间以及是否接听了电话。

健康行为和入口统计数据. 该物理数据来自Fitbit指标，其包括与健康相关的行为变量，例如心率，步数和活动状态。除了每分钟的原始心率和步数外，Fitbit还分离和跟踪四个心率区（超出范围，脂肪燃烧，有氧运动和峰值）²。至于活动记录，Fitbit基于METS计算每分钟的活动状态，METS是一种体重不可知的活动度量，代表久坐，轻度活跃，相当活跃和非常活跃的活动状态²。我们还在分析中考虑参与者的性别（唯一的人口统计特征）。

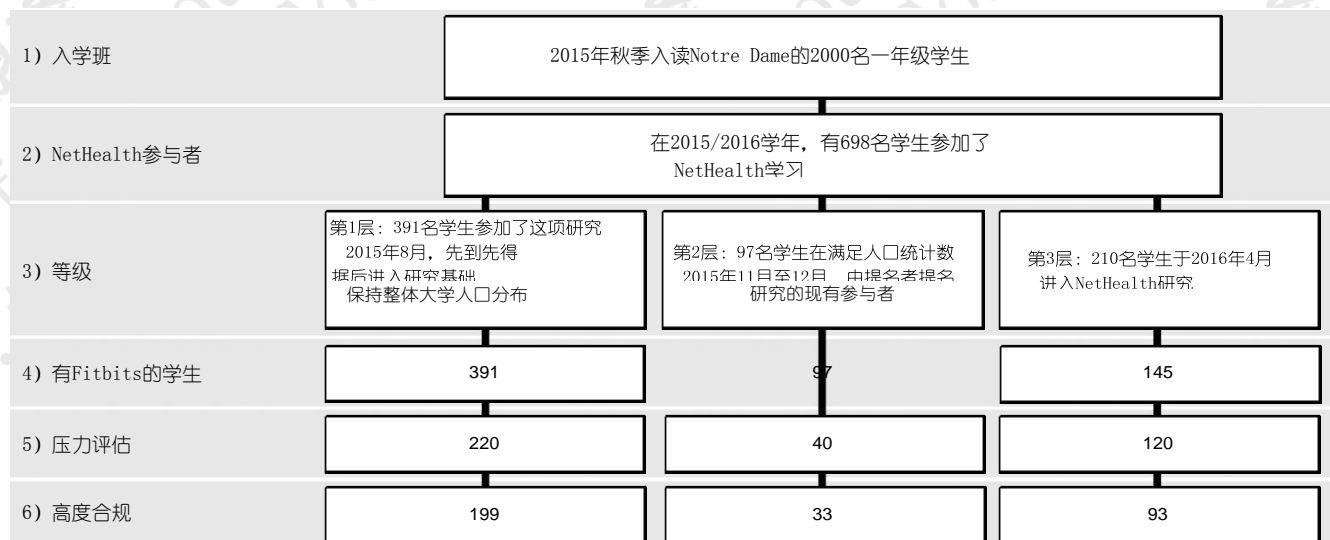


图2. 本文中为此分析选择的NetHealth招聘和学生的联合图。

人口 性别	男	# 数据点 146 (45%)
	女	179 (55%)
种族	白色	227 (70%)
	拉丁美洲人	36 (11%)
	亚洲	29 (9%)
	黑色	18 (6%)
	国外	14 (4%)
年龄	17	36 (11%)
	18	182 (56%)
	19	11 (3%)

表1. 数据样本中的人口统计数据摘要。 样本总数为325。

健康状态。 这些数据来自每个学期的参与者回答的调查。 由于整个学期的调查问题不同, 我们无法共同分析所有调查。 出于这个原因, 我们选择了2016年秋季进行的调查, 其中包含有关健康状态的问题 - 压力, 快乐, 积极态度和自我评估的健康状况 - 并涵盖了大多数参与者 (380名受试者), 因为这是第一学期这三个层次都存在于研究中。 因此, 我们从2016年8月至2016年12月考虑了来自健身追踪器和社交互动的当代数据。我们排除了47名参与者因缺少Fitbit或社交网络数据。 因此, 我们的数据涵盖325名参与者。 表 1 从人口统计学的角度展示了如何在NetHealth研究中选择这些数据样本, 其中一些缺失的数据与种族和年龄特征相对应。 表 2 介绍调查中的问题, 他们可能的答案以及我们研究的四个感兴趣领域的相应构成。 压力分为4类, 从1到4, 类别1表示最低水平的压力, 类别4表示最高水平的压力。 同样, 幸福和自我健康报告也分为四类, 从第1类到第4类的数量增加。最后, 鉴于积极态度的分布, 我们将其分为五类, 第1类表示最不正面, 第5类表示最多。

数据预处理

数据预处理有两个步骤。 首先, 为了确保每日Fitbit数据稀疏程度不同的学生之间没有偏见, 我们在分析中删除了每日合规性低于80%的样本, 因为80%的日常合规性提供了对学生活动的合理估计⁹。 其次, 我们使用一周的窗口将2016年8月至2016年12月的数据划分, 每个数据点包括一周的Fitbit数据, 社交数据和2016年秋季调查的相应调查数据。 此数据分区的原因是该调查是在2016年秋季进行的, 但健康行为和社交网络数据跨越了2016年8月至12月的五个月。随后, 必须选择时间范围

问题压力	水平	#Participants (#男, #女)
感到紧张和压力	1	14 (12,2)
	2	95 (58,37)
	3	134 (52,82)
	4	82 (24,58)
幸福	1	40 (14,26)
	2	84 (38,46)
	3	144 (64,80)
	4	57 (30,27)
积极的态度	1	2 (0,2)
	2	23 (8,15)
	3	65 (28,37)
	4	160 (72,88)
	5	75 (38,37)
健康	1	7 (4,3)
	2	56 (18,38)
	3	200 (95,105)
	4	62 (29,33)

表2. NetHealth研究中与健康相关的调查问题摘要。 选定的参与者总数为325。

来自Fitbit或社交网络的相应数据流。 由于一个人的每日或每小时社交互动是有限的, 因此对于社交网络的测量而言, 对于一天或一小时的时间间隔太精细化是无效的。 如果使用无向和未加权的图表来表示网络, 则由于可能的信息丢失, 太粗粒度的间隔也会产生负面影响。

特征提取

我们从数据流中提取了几个特征, 为学习算法构建了一个合适的特征向量, 如下所述。 我们还将性别信息纳入特征向量, 因为我们想要研究性别对健康行为可预测性的影响。

性别信息

世界卫生组织认识到压力相关综合征存在性别差异⁹。 例如, 女性的压力发生率比男性高得多。 基于这种洞察力, 我们从调查数据中提取性别信息, 以用作特征向量中的附加特征。 表 2 显示了男性和女性不同水平调查变量的人口分布。 特别地, 考虑到压力, 快乐和健康的情况, 男性主要分为2级和3级, 大多数女性属于3级。我们在分析中使用性别作为自变量 (预测因子)。

健康行为数据

我们将从Fitbit捕获的物理属性 (心率, 步数和活动状态) 分类为健康行为数据。 该数据被分段为上一节中讨论的每周间隔。 然后, 在这些时间段上计算均值和方差 (或标准偏差) 的汇总统计。

心率。 我们计算了每个参与者每周心率的均值和方差。 我们还应用ANOVA测试来检查不同压力水平, 幸福水平, 健康水平和积极态度水平之间的心率差异。 结果显示, 不同应激水平 ($p < .001$), 幸福水平 ($p < .001$), 健康水平 ($p < .001$) 和积极态度水平 ($p < .001$) 的心率存在显著差异。

脚步。 步骤的原始数据也是逐分钟记录的, 但由于行走的性质, 一周内的大部分时间更可能为零。 因此, 我们首先将原始步骤数据转换为每天的步骤总和。 然后, 我们计算每周和每个人的每日步骤的平均值和标准差作为特征。

活动状态。 Fitbit跟踪用户的活动并每分钟记录相应的预定义状态。 有4种可能的状态: 久坐, 轻度活跃, 相当活跃和非常活跃。 因此, 首先计算每天每个州的分钟总和, 然后计算每周内每个州的这些每日总和的均值和标准差。

社交网络数据

社交网络是根据电话和短信的通信模式构建的。 为了避免虚假连接 (例如垃圾邮件), 我们消除了频率低于三倍的通信边缘

5个月的时间。NetHealth研究不仅收集了所有参与者内部的通信数据，还收集了参与者和研究外部人员之间的通信数据。结果，我们有两种类型的社交网络：一种包括所有数据（整个网络），一种仅包括研究中参与者（参与者网络）的通信数据的数据。参与者网络仅包括朋友或同学，因为所有参与者都是具有相同班级或同一年级的同学的本科学生。我们将参与者网络视为朋友网络，这是可影响健康的五种社交网络之一²。但是，我们在分析中研究了整个网络和参与者网络。

正如我们之前提到的，我们的社交网络分析中的每一步都包含一周。社交网络是每周通信模式的无向和未加权的表示。然后，我们推导出代表社交网络结构的几个特征，包括网络中的每个人的网络度²，三角形数量，聚类系数²，中介中心性²和亲密度中心性²。

分析框架

健康行为关系分析

为了回答RQ1，我们调查社交网络结构和健康行为之间是否存在关联，以及社交网络结构属性是否可以预测健康行为。具体来说，我们检查社交网络拓扑属性之间的关系，包括程度，三角形数量，聚类系数，亲密度中心性和参与者网络中的每个节点（个体）和整个网络的中介中心性以及健康行为变量，包括心率，步骤和活动状态。

我们使用箱形图对所有22周进行可视化，以定性方式显示网络结构变量与健康行为统计之间的关系。我们使用互相关系数²来捕获每个行为变量和每个网络结构变量之间的相关性。应该注意的是，从物理和行为变量到社交网络变量的链接因人而异。因此，在本研究中，我们计算每个个体的相关性，然后我们总结显示大于0.5的值的系数总数。

健康状况预测

为了回答RQ2，我们提出NetCARE，一种网络驱动预测架构，充分利用健康预测问题中的社交网络结构特征。数字3显示NetCARE的架构。该算法将社交网络结构，可穿戴数据和人口统计数据作为机器学习模型的独立变量。这种算法架构允许我们根据需要选择网络功能并添加其他数据源。它还确保了修改，扩展或添加分类器的灵活性。

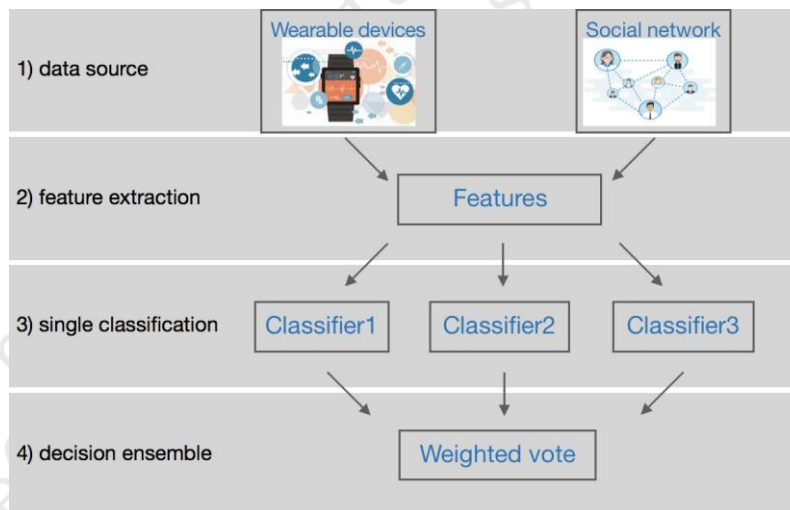


图3. 网络驱动预测架构NetCARE。

让我们以压力检测为例详细解释这种架构。我们将问题表述为4类分类之一，其中特征从健康行为和网络结构中提取，类别是压力水平。然后我们使用了五个流行的分类器来解决这个问题：K-最近邻（KNN），分类和回归树（CART），支持向量机（SVM），Logistic回归（LR）和随机森林（RF）。数据分为75%用于培训，25%用于测试。我们使用5倍交叉验证来找到超参数

算法，并使用网格搜索找到那些达到最佳性能的参数的组合。然后，我们考虑所有级别的平均F1-度量和每个级别内的那些作为性能度量。具体来说，我们调整了KNN的邻居数量和CART的叶子大小。对于SVM，我们在三个不同的内核上进行了实验：多项式，线性和径向基函数（RBF）内核，其中还考虑了不同程度的多项式内核。对于LR，我们搜索正则化系数的不同值和优化算法的学习率。对于RF，我们使用不同数量的树从10到100进行实验，增量为5。我们使用35棵树来获得本文档中报告的结果。

此外，我们应用加权投票方案的集合方法来改善个体的表现分类。我们选择了三个单一分类器作为基本分类器。我们使用符号 $p_{i,j}$ 来表示分类器 i 将输入实例 x 分类为类 j 的概率。因此，给定权重 $w_{i,j}$ 分配给概率 $p_{i,j}$ ，得到输出 $y_{投票}$ 的集合规则可以表示为 (1)。对于重量选择，我们使用对训练数据的交叉验证来从权重 $w_{i,j}$ 的所有可能组合中选择最佳权重 $w_{i,j}^{最佳}$ ，从0到1，间隔为0.1。因此，最终的决定是由标签做出的，该标签最大化了方程式加权投票。(2).

$$y_{投票}(\{w_{i,j}\}) = \underset{j}{\operatorname{argmax}} \sum_{i=1}^3 w_{i,j} p_{i,j} \quad \text{ST} \quad \sum_{j=1}^4 w_{i,j} = 1, j \in \{1, 2, 3, 4\}, \text{ 和 } i \in \{1, 2, 3\} \quad (1)$$

$$y_{final} = \underset{j}{\operatorname{argmax}} \sum_{i=1}^3 w_{i,j}^{最佳} p_{i,j} \quad \text{st } j \in \{1, 2, 3, 4\}, \text{ 和 } i \in \{1, 2, 3\} \quad (2)$$

结果

为了测试我们的框架，我们进行了两组实验。首先，我们评估了与社交网络结构相关的变量与健康行为相关的变量之间的相互作用。此分析的目的是验证这些交互是否有意义。其次，我们使用我们的框架来预测各种健康变量。我们将框架的性能与基线进行了比较，该基线单独考虑了健康行为数据和社交网络功能。我们这样做是为了验证我们的假设，即结合网络效应和自相似性将导致更好的预测。

社会网络结构与健康行为的关系分析

我们分析了物理和社交网络变量之间的关系。为了实现这一目标，我们将2个网络（整个网络和参与者网络）中的社交网络的所有可能的5对结构特征和6个物理和行为特征与箱形图进行可视化，从而产生了60个箱形图。对于图中的每个框，我们为22周内的所有参与者呈现物理行为特征和社交网络结构特征的分布。请注意，每周的健康行为特征被提取为原始数据的平均值。例如，图 4a 显示所有参与者的平均心率分布以及参与者网络22周内的节点度分布。数字 4a 和 4b 表示参与者网络结构的关系，而整个网络的关系如图所示 4c 和 4d。其余的箱形图可以在补充材料中找到。

如图所示 4，每周健康行为数据的中位数和平均值（图中的深橙色线条和深绿色三角形）随时间变化，网络属性的中位数和平均值（图中的深蓝色线条和海绿色三角形，分别）随着时间的推移遵循类似的模式。

我们进行了测试，以验证健康行为特征的分布与高值和低值范围内的网络结构特征之间是否存在统计学上的显著差异。具体来说，使用t检验我们测试了代表行为数据的特征值是否随着与网络数据范围的关系强度而变化。例如，考虑每日步骤和网络程度之间的关系。得出的p值为0.0003表明，每日步骤与较高网络度相比，每日步骤与较低网络度相结合存在显著差异。人物 4b 显示结果。在校正多个测试之后，我们的结果显示社交网络属性在60次中有43次具有显著关系，支持社交网络指示健康行为变化的假设。

此外，我们使用互相关系数来量化网络结构特征可以反映健康行为信息流的范围。在计算每周健康行为和网络结构的每个特征的均值后，我们计算了健康行为特征的均值系数和结构特征的均值（60对）。结果显示，60对中有43对绝对相关系数不小于0.5，60对中有28对，绝对相关系数不小于0.7。表 3 显示所有行为特征和网络结构特征对的结果。

从网络结构特征到非常活跃状态的相关性通常比从结构到久坐，相当活跃或轻度活跃的状态的平均强度大约0.3。我们注意到结构特征中没有一个与轻度活动状态的数据有很强的关系。在排除了轻度活动状态的结果后，我们发现参与者网络和整个网络中的三角形数量可以很好地指示其他健康行为数据的变化。此外，除了轻微活动之外，每个行为特征可以与绝对系数不低于0.7的至少一个结构特征相关。特别是，整个网络的程度与步长之间的相关系数几乎为0.9。我们每周对每个特征的中位数进行相同的实验。我们发现60对中有37对具有不低于0.5的相关性。将参与者网络中的指标与表中整个网络指标的结果进行比较，我们无法确定两个网络之间的主要差异。例如，参与者网络中相当活跃的状态数据和紧密度中心性之间的系数几乎是整个网络中相当活跃的数据和紧密度中心性之间的系数的一半，而参与者网络中相当活跃的数据和聚类系数之间的系数几乎是使整个网络中相当活跃的数据和聚类系数之间的系数加倍。这一发现表明，在两种网络类型中存在不同的影响，因此有必要在我们的分析中包含这两种网络。总之，表3表明网络结构似乎捕捉到了健康行为的变化 - 尽管在轻微活跃状态方面的程度较小。

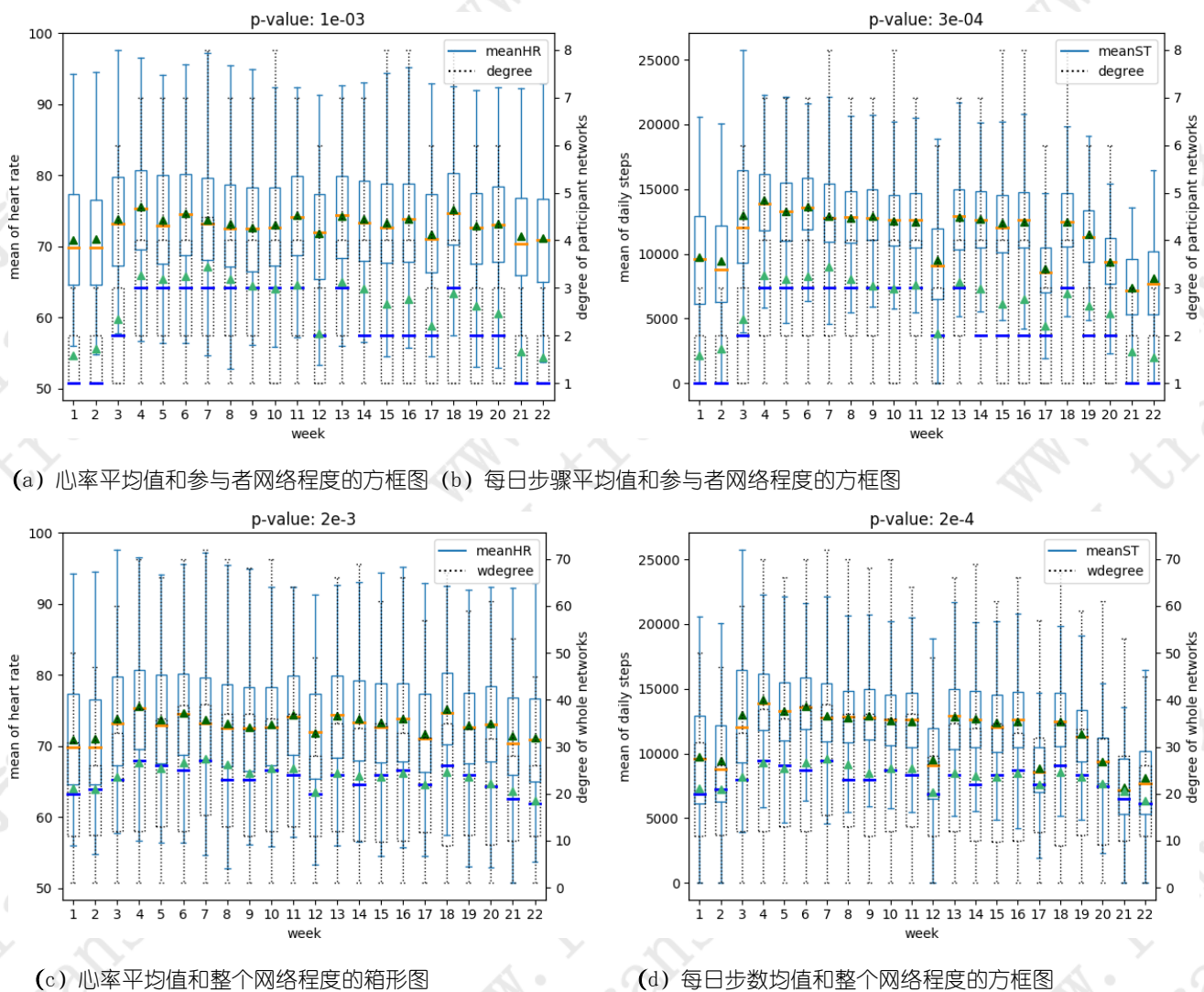


图4. 社交网络结构与健康行为之间的关系

我们还评估了数据集中每个参与者的变量交互。这样做是为了评估每个人健康行为的变化。每个分析的样本点是每人一周的行为的平均值。我们数了一下

	心率	脚步	久坐不动	轻微活跃	相当活跃	很活跃
参与者网络的学位	0.84	0.89	-0.44	-0.014	0.49	0.87
参与者网络中的三角形数量	0.74	0.83	-0.61	0.24	0.68	0.79
参与者网络中的聚类系数	0.65	0.75	-0.51	0.15	0.59	0.66
参与者网络中的中介性	0.78	0.68	-0.19	-0.20	0.20	0.72
参与者网络中的亲密度	0.83	0.85	-0.32	-0.14	0.35	0.86
整个网络的学位	0.81	0.90	-0.57	0.15	0.62	0.88
整个网络中的三角形数量	0.79	0.89	-0.62	0.23	0.69	0.85
整个网络中的聚类系数	0.83	0.79	-0.32	-0.12	0.35	0.79
整个网络中介中心性	-0.76	-0.85	0.65	-0.28	-0.71	-0.79
整个网络的紧密程度	0.75	0.78	-0.59	0.24	0.65	0.71

表3. 每对健康行为特征平均值和社交网络结构特征平均值的归一化互相关系数。

绝对互相关系数超过0.5的总人数，以显示结构特征可以捕捉健康行为变化的程度。表4列出每对健康行为特征和网络特征具有中到强相关性的参与者数量。在该表中，每个健康行为特征可以与至少20%的参与者的整个网络或参与者网络的结构特征之一相关。该表显示整个网络或参与者网络中的Closeness Centrality捕获与超过42%的样本的步骤的关系。这些结果意味着社交网络结构与健康行为之间的潜在关系。

表5总结了参与者网络，整个网络或两个网络中具有中等到强相关性的参与者与健康行为特征和至少一个图形结构特征的比例。特别地，百分比是其健康行为数据与任何网络结构特征具有不小于0.5的互相关系数的人的比例。例如，对参与者网络的5个特征中的任何一个与系数不小于0.5的步骤数据相关的人进行计数。根据表中的第一和第二列5，来自各种社交网络的三个指标可以密切捕捉约50%的参与者的一些健康行为的变化。具体而言，74%的参与者在步骤和结构特征的一个方面之间具有更高的相关性。表中的最后一列5表明，对于超过50%的参与者，两种类型的社交网络都保持每个活动状态的时变心率平均值，步长平均值和平均分钟数的信息。另外，在整个网络或参与者网络中步数和结构特征之间具有相关系数不小于0.7的人的比例是40%，并且心率和结构特征之间的平均值是26%。这些结果意味着整个网络包含的健康行为信息比参与者网络更多，但鉴于最后一栏中的数字增加，这两者都是必不可少的部分。

	心率	脚步	久坐不动	轻微活跃	相当活跃	很活跃
参与者网络的学位	52	86	43	39	42	47
参与者网络中的三角形数量	31	38	29	24	37	26
参与者网络中的聚类系数	28	37	27	24	35	27
参与者网络中的中介性	38	38	34	35	32	34
参与者网络中的亲密度	99	145	81	79	82	98
整个网络的学位	78	94	81	58	70	68
整个网络中的三角形数量	63	100	69	43	62	60
整个网络中的聚类系数	47	66	46	37	51	39
整个网络中介中心性	50	52	57	54	53	49
整个网络的紧密程度	91	137	98	83	94	80

表4. 健康行为与社交网络结构具有中等至强相关性的人数。

总之，这些实验验证了网络结构变量和健康行为变量之间的相互作用。具体来说：1) 我们可视化网络结构对健康行为的影响，并发现网络结构可以定性地捕捉行为变量的变化。2) 我们进行了t检验，以检查健康行为变量的较高值是否与较高的结构变量值相对应，并且与两种变量的较低值不同。我们的结果显示，在Bonferroni校正后，60个中有43个具有显著差异。3) 我们使用归一化互相关系数来描述网络结构在统计中的作用。结果显示了60对行为特征中的43对

与健康有关的数据	参与者网络 (%)	全网 (%)	两个网络 (%)
心率	133 (41)	157 (48)	193 (59)
脚步	182 (56)	204 (63)	239 (74)
久坐不动	125 (38)	166 (51)	202 (62)
轻微活跃	132 (41)	145 (45)	199 (61)
相当活跃	122 (38)	164 (50)	194 (60)
很活跃	133 (41)	143 (44)	186 (57)

表5. 与社交网络结构具有中度到强相关性的主题的总结。百分比是其健康行为数据与任何网络结构特征具有不小于0.5的互相关系数的人的比例，其中数据中的总人数是325。

结构特征的相关系数不小于0.5，大约一半的系数不小于

0.7。4) 我们分析了个体层面结构和物理特征的可变相互作用。我们发现，325名参与者中有多达145人表现出他们的网络和步骤的亲密度中心之间的高度相关性，高达74%的参与者表明了聚合网络功能和步骤之间的相似关系。

预测健康状况

在实施了五个单一分类器之后，我们选择了SVM，KNN和RF来创建我们的集成学习模型。表的前五行6显示我们的集合分类器在压力预测中的表现。我们报告所有压力水平和每个水平的F分数。该表显示单独的社交网络变量与整体F分数和压力水平3的健康行为数据相比甚至更好，而其他变量则更差。因此，我们怀疑社会网络结构包含与健康行为变量相比的补充视角下的压力信息，即社交网络结构与压力状态之间似乎存在潜在的关系。该表还显示，加入社交网络和健康行为的特征可以改善预测，并通过F1综合绩效和每个压力水平的个人绩效进行评估。最显著的改进对应于压力水平1。

此外，我们对其他健康和健康变量执行相同的分析。特别是，我们评估结合社交网络结构变量和健康行为变量来预测幸福感，积极态度和自我评估健康的效果（表2）。这些结果也显示在表中6。与压力情况一样，该表显示使用社交网络结构可以提高这3个健康和健康变量的预测性能。表6，表明我们的NetCARE框架提供了以下改进：1) 整体F1-Measure和内部F1-幸福度量分别为65%和高达617%；2) 整体F1-测量值和内部F1-积极态度测量值分别为55%和高达185%；3) 54%和高达200%的整体F1-测量和内部F1-测量自我评估的健康状况，分别。这些结果提供的证据表明，不仅结构特征可能有助于健康状态预测和健康感知的应用，它们还显示健康变量和社交网络结构之间的潜在关系。

讨论

本文的主要贡献可归纳如下：

- 我们发现社交网络结构与从可穿戴设备捕获的健康行为数据相关联，并且可以捕捉趋势。社交网络结构与健康行为之间的这种关系具有统计学意义。
- 我们证明了社交网络结构对于提高健康状态的可预测性至关重要。这一结果非常重要，因为仅依靠可穿戴设备和人口统计数据得出的数据并不能表达关于个人的完整情况，而且一个人的社交网络是理解和预测健康与健康的基本要素。

社交网络分析已被用于健康相关问题，包括心理健康^{2, 3}，身体健康^{4, 5}和疾病^{6, 7}。大多数工作主要集中在社交网络作为健康的扩散机制^{8, 9, 10, 11}或情绪^{12, 13, 14}。本文提供了一个新的视角，社会网络结构的价值不仅在于了解我们的健康行为，而且在预测健康状态，超出可穿戴设备或人口统计数据告诉我们的数据。显然，社交网络是关于我们健康和健康的难题的重要组成部分。我们通过包含来自社交网络的特征表明，准确性显著提高，有时仅使用社交网络功能增加了更多的可预测性。具体而言，我们发现除了健康行为和人口统计数据之外，当使用社交网络结构特征时，快乐和积极态度具有最显著的跳跃。这清楚地表明它是紧的

压力预测	F1	1级	级别2	级别3	级别4	
性别+健康行为数据	0.42	0.18	0.53	0.64	0.34	
社交网络结构	0.34	0.05	0.43	0.63	0.26	
性别+健康行为数据+社交网络	0.58	0.46	0.63	0.70	0.55	
起色	38%	155%	19%	9%	62%	
幸福预测	F1	1级	级别2	级别3	级别4	
性别+健康行为数据	0.31	0.06	0.31	0.62	0.24	
社交网络结构	0.21	0.00	0.2	0.60	0.02	
性别+健康行为数据+社交网络	0.51	0.43	0.52	0.67	0.44	
起色	65%	617%	68%	8%	83%	
积极态度预测	F1	1级	级别2	级别3	级别4	求等级 5
性别+健康行为数据	0.31	0.20	0.13	0.22	0.71	0.30
社交网络结构	0.40	0.70	0.08	0.23	0.70	0.25
性别+健康行为数据+社交网络	0.48	0.36	0.37	0.44	0.74	0.47
起色	55%	80%	185%	100%	4%	57%
自我评估的健康预测	F1	1级	级别2	级别3	级别4	
性别+健康行为数据	0.35	0.29	0.13	0.77	0.20	
社交网络结构	0.21	0.00	0.05	0.77	0.00	
性别+健康行为数据+社交网络	0.54	0.6	0.39	0.79	0.37	
起色	54%	107%	200%	3%	85%	

表6. 幸福，积极态度和自我评估健康的预测结果

结合自我的社会和健康行为，从而提高对自我健康状态的理解和可预测性。

我们的研究还可能收集到其他见解。考虑结构变量和健康行为变量之间的相关性（参见例如表格 3）。我们观察到聚类系数与心率，步数和高活动状态之间存在中度到强相关性，这可能反映了参与校园体育。这些活动为参与的学生提供了大量的身体活动和紧密结合的社会群体，这些因素之前已被证明与心理健康有关^{7, 8}。此外，当一个人处于非活动状态或至少相当活跃时，似乎社交网络结构更容易捕获活动状态，而不是该人处于轻度活动状态。它可以表明活动与群居或个人外向之间的关系。此外，由于活动时间较短，包括步行，学生日常生活中涉及的宿舍，课程和其他必要目的地的位置可能会对此级别的活动产生重大影响。未来的研究方向是研究更精细的数据和时间窗口，以了解与活动状态相关的通信模式的即时性。

致谢

本文报道的研究得到了美国国立卫生研究院国家心脏，肺和血液研究所（NHLBI）的支持，奖学金编号为R01HL117757。内容完全由作者负责，并不一定代表国立卫生研究院的官方观点。

作者贡献声明

SL 和 NVC 构思了这项研究。SL, PRG 和 NVC 开发了建模框架。SL 实施了模型并进行了实验。SL, LF, PRG 和 NVC 进行了分析。SL, LF, PRG 和 NVC 撰写了手稿。

附加信息

补充信息：评估论文结论所需的所有材料均可在论文和/或补充材料中找到。可能会要求作者提供与本文相关的其他信息。

数据可用性：可以向作者请求与论文相关的所有数据。

竞争利益：作者声明他们没有竞争利益。