

# 线下交互的动态社交网络研究进展：挑战与展望

梁 迪<sup>1,2)</sup> 崔 靖<sup>1)</sup> 李 翔<sup>1,2)</sup>

<sup>1)</sup>(复旦大学信息科学与工程学院电子工程系自适应网络与控制研究室 上海 200433)

<sup>2)</sup>(复旦大学信息科学与工程学院智慧网络与系统研究中心 上海 200433)

**摘 要** 移动互联网技术的发展和智能可携带设备的普及为感知人类社会行为提供了新的契机。线下交互的动态社交网络研究作为人类动力学和社会计算研究领域的重要组成部分,在推荐系统、传染病免疫、机会网络路由中都有重要的研究意义。该文对动态社交网络中的线下交互研究进行了回顾和梳理,从数据、特性(拓扑、时间、地理)、模型(基于轨迹、基于地点、基于拓扑)、应用四个方面入手,总结了线下交互动态社交网络的研究现状,也提出了目前线下交互研究中尚待解决的开放性问题,为相关的研究者提供借鉴。

**关键词** 动态社交网络;线下交互;交互数据;交互建模;人类动力学

**中图法分类号** TP399 **DOI号** 10.11897/SP.J.1016.2018.01598

## A Review on Offline Interaction in Dynamical Social Networks: Challenge and Prospect

LIANG Di<sup>1,2)</sup> CUI Jing<sup>1)</sup> LI Xiang<sup>1,2)</sup>

<sup>1)</sup>(Laboratory of Adaptive Network and Control, Department of Electronic Engineering, School of Information Science and Engineering, Fudan University, Shanghai 200433)

<sup>2)</sup>(Research Center of Smart Networks & Systems, School of Information Science and Engineering, Fudan University, Shanghai 200433)

**Abstract** Social network analysis, as an important part of complex network theories, attracts the attention of a large number of researchers. Different from the study of traditional social networks constructed by nodes of individuals and edges of social relationships, the study of modern social networks concerns not only static social relationships, but also dynamical interactions between people. Nowadays, the development of mobile internet and increasing popularity of smart mobile devices make it easier to sense human behaviors and interactions, whose temporal dynamics are mapped into dynamical social networks with variable topologies. In our study, we mainly focus on the temporal social networks constructed by human offline interactions, which require the geographical closeness between pairwise individuals in the physical world. As an important part of human dynamics and social computing, human offline interactions in dynamical social networks play a crucial role in lots of applications such as recommendation systems, epidemical immune strategies and opportunistic routings. In this paper, we review the offline interactions in dynamical social networks through four main facets: empirical data, characteristics, models and applications. Empirical data is essential to understand and analyze human interactions. Therefore, we discuss the features of different data sources and data acquisition methods such as WiFi, Bluetooth and GPS. In these data acquisition methods, both the latitude and longitude

收稿日期:2016-10-08;在线出版日期:2017-05-26. 本课题得到国家杰出青年科学基金(61425019)、国家自然科学基金(71731004)、国家社科重大项目(12&ZD218)、上海市曙光学者计划(14SG03)资助。梁迪,女,1991年生,硕士研究生,主要研究方向为时效网络与人类行为动力学。E-mail: 13210720030@fudan.edu.cn. 崔靖,女,1990年生,硕士,主要研究方向为时效网络与人类行为动力学。李翔(通信作者),男,1975年生,博士,教授,博士生导师,主要研究领域为复杂网络与网络科学的理论与应用。E-mail: lix@fudan.edu.cn.

information identified by GPS, the wireless hotspot physical address in the WiFi connection log, the POS machines in the stores as the geographical tags are the proxy of human interactions. We classify interaction characteristics into three categories: topological, temporal and spatial, whose characteristics such as temporal clustering coefficients, temporal motifs and inter-contact time are described in detail. For human encounter modeling, we review three types of models based on traces, locations and topology, and introduce the state-of-art models in each type. Modeling offline interactions is a key step for simulating applications of offline interactions by making supplements of real world datasets. The artificial models, on the one hand, preserve the basic characteristics of real data, allowing the various parameters (e. g., the number of interactions, the number of locations, the length of time, etc.) to adjust in a more flexible way in the simulation situations. On the other hand, modeling offline human interactions helps to explore their essential mechanisms to provide theoretical support for the control and prediction in practical processes. With the abundance of data and the deepening of research, the study of dynamical social networks with offline interactions has brought a new perspective to the application of many fields such as inferring social ties, community detection, public health strategies, and recommendation systems. We finally outlook several directions with some open questions for future study.

**Keywords** dynamical social network; offline interaction; interaction data; interaction modeling; human dynamics

## 1 引 言

社交网络(social network)研究历来就是一个倍受关注的话题. 1988年, Wellman作为最早一批提出成熟社交网络理论的社会学家, 把社交网络视为连接行动者的一系列社会关系<sup>[1]</sup>, 并将人或团体作为节点, 将社会关系作为节点间的连边. 随着研究的发展, 社交网络中的连边不再局限于狭义的社会关系(用亲疏程度定义连边的权重), 而是扩展到各类社会交互行为(的总和). 不论是面对面的交互, 还是在互联网络上、电话中的交互, 都成为连接人类个体的纽带.

人们过去常常将社交网络看作一个静态结构, 社会交互行为的动态属性在研究中往往被忽略. 然而, 社交网络研究是人类动力学<sup>[2-4]</sup>研究的重要组成部分, 其节点自身具有动力学, 这导致节点间的关系(社会关系、交互行为)不再是一成不变的. 例如, 交互行为的时效特性使得网络的连边并不是持续存在的, 连边的发生遵循一定的顺序. 因此, 在真实社交网络上研究网络特性(例如社交网络的可达性<sup>[5]</sup>)和动力学过程(例如信息传播过程和传播规模<sup>[6]</sup>), 所得到的结论与静态网络相比相差甚远. 越来越多的研究不仅仅局限于网络的集聚拓扑结构, 更多致力

于探索节点或连边的动态特性及其所构成的动态社交网络. 不同于对人类个体层面动力学的研究, 动态社交网络更关注个体间的关系与相互作用, 以及其对疾病/信息传播等人类动力学过程的影响<sup>[7]</sup>.

动态社交网络中的连边由节点间的各种社会交互行为构成. 根据交互途径的不同, 我们可以将交互行为分为线上交互和线下交互<sup>[8]</sup>. 微博互动、手机通话等行为均为线上交互. 它借助互联网、无线通信等技术手段实现个体间的互动, 因此不受个体间实际地理距离的影响. 然而, 尽管突破了地理空间的限制, 大部分线上交互仍然发生在熟人或具有共同兴趣爱好的个体之间. 线下交互则要求发生交互的两个个体在一定时间窗口内出现在同一地理区域, 并且两个个体间的实际地理距离足够邻近, 例如, 同乘一辆公交车、同在一个收银台排队结账等. 需要注意的是, 线下交互并不要求个体间必须发生交谈或肢体接触, 而是指不同个体在地理位置上足够靠近, 只要在个体间能够实现诸如信息传递、疾病传播等过程. 因此从熟人到陌生人都可能成为我们线下交互的对象. 与此同时, 交互行为发生的时间和地点信息则是描述线下交互的重要维度.

通常我们用图 1 所示的五元组来记录人类线下交互行为. 该五元组包括了两个交互个体 A 和 B、交互开始时间、交互持续时间和交互地点. 在过去的

二十年间,人们采用了许多方式来同时采集时间和地理两种行为信息,例如,射频识别设备、手机蓝牙、GPS、WiFi 等等.在这些采集方法中,GPS 所标识的经纬度信息、WiFi 连接日志中的无线热点物理地址、商店 POS 机编号等等都可以作为个体行为发生地点的地理位置标识.

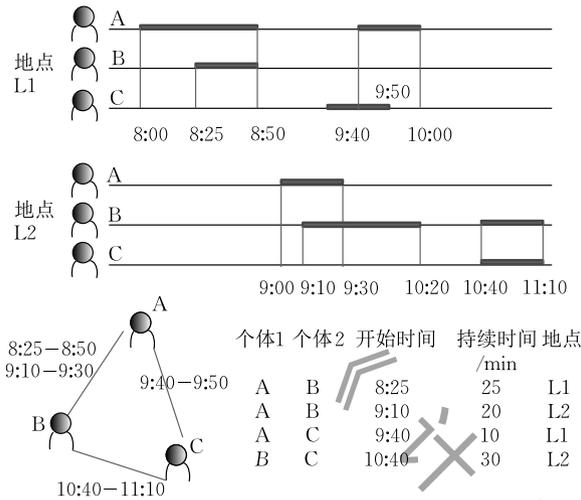


图 1 线下交互行为的表示示例

研究线下交互的动态社交网络具有重要意义.

首先,线下交互驱动了社会关系的形成<sup>[9]</sup>.研究发现,共享同一幢宿舍的学生间的邮件来往次数比不在同一幢楼的学生要多,而在同一层甚至同寝室居住的学生邮件来往又更加频繁;在公司中,同一职能部门的员工间的电话留言、邮件来往数量较不在同一部门的员工高了 1000 多倍.尽管数字通信使得世界变得扁平化,人们仍然倾向于与那些地理位置更为接近的个体交互,而线下交互往往也反映了现实世界中最真实的人际关系.其次,线下交互不受个体社交能力和熟人圈子大小的限制,每一个个体都可能拥有大量的邻居.当我们乘坐公交车时,在健身房健身时,或者在餐厅吃饭时,我们都与许许多多个体发生线下交互.尽管很多时候作为网络中节点的我们往往忽略了这些“弱连接”的存在,但这种关系却潜移默化地影响着我们的生活<sup>[10]</sup>.挖掘这种弱连接将有助于识别潜在的重要联系,促进朋友推荐、位置预测、商品推荐等实际应用.另外,现实中很多传播现象本身也依赖于人们的线下交互行为,比如疾病的传播(通过飞沫或者接触)<sup>[11-13]</sup>,机会网络中的信息传递<sup>[14]</sup>等等.近年来伴随着移动互联网的高速发展,各种便携式设备迅速普及,计算通讯能力也快速提升,这使得移动机会网络成为学术界持续关注的热点,其不需要网络全联通的特点更符合实际

环境下自组网的需求.对个体之间线下交互行为的分析正是移动机会网络众多实际应用的基础<sup>[15]</sup>.

无论是在理论上研究线下交互网络对疾病传播等动力学过程的影响,还是研究如何将其应用于实际生活,其根本都在于对线下交互网络本身的特性和驱动机制有一个充分的理解.如何分析和刻画线下交互动态社交网络特性,如何设计机制来建立线下交互模型正是本综述关注的重点.

我们通常从三个方面来刻画线下交互的动态社交网络特性:静态/动态拓扑特性、时间特性、地理特性.拓扑特性既反映了在一定时间内的集聚社交网络特性,包括度分布、聚类系数、平均路径长度等,也包括了随时间窗口推移所得到的动态拓扑特性.时间特性与网络中节点或边的时变特征密切相关,例如连边的持续时长、两次交互的间隔时长等.地理特性反映的则是线下交互发生的地理模式,交互地点的数目、范围等特征均是我们关心的范畴.近年来已经有很多工作尝试挖掘这些特性,并在不同数据集上得到相同或不同的特性,因此对这些工作的结论和贡献进行总结是十分必要的.

然而真实数据集是有限的,其规模也是固定的.大多数的数据都采集自一个有界的环境,比如校园、会议场所等.我们需要通过建模的方式来构建更具扩展性和灵活性的社交网络,同时保留实际数据的关键统计特性.通过建模得到的社交网络须具有拓扑、时间、空间的关联性,同时能够通过调节模型参数来满足不同应用场景的需求.可见,交互数据、特性、建模与应用四个方面是相辅相成缺一不可的(图 2).

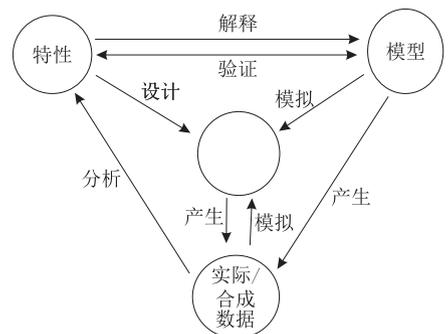


图 2 线下交互研究内容关系概览

本文我们在第 2 节中总结该领域研究的困难与挑战,并在接下来的章节里从四个方面总结目前对线下动态社交网络的研究工作:线下交互数据、线下交互特性、交互模型与应用;在第 3 节中,我们对五类线下交互数据采集方式进行介绍,并总结了已有的数据集的特征;在第 4 节中,我们从拓扑、时间、地

理特性三个角度来分析线下交互特性;我们在第 5 节总结已有的线下交互的人工模型,并依据应用场景和时空尺度对其进行分类;我们在第 6 节中对线下交互在热门研究领域的实际应用进行阐述.最后,我们在第 7 节对该领域研究方向进行了展望.

## 2 线下交互研究的困难与挑战

(1) 数据采集精度与规模限制. 在研究中,我们希望通过采集人类行为数据来实证分析.我们一方面希望能够尽可能精确地在三维空间内刻画人类行为轨迹,同时又希望尽可能完整地保留观察区域内所有人类的线下交互行为,这对于采集技术的精度要求和大范围部署能力都是巨大挑战.常用的手机 GPS 数据、WiFi 热点接入数据虽然能够满足大规模观测的需求,但其地点定位精度不高,且通常只能在用户自主使用手机或上网时才有记录,因此对行为轨迹的刻画并不连续.为了实现更加精确而连续的观测,研究者们也尝试使用定制的射频标签(RFID)或者手机蓝牙对用户间的近距离交互进行连续观测.然而这类技术的缺点也十分明显:首先,传统的射频标签或蓝牙技术均不具备交互地点识别的功能;其次,对设备的依赖性使得实验只能在很小的人群范围内进行.可以说,迄今为止的人类行为数据集都只是对人类行为的截面式刻画.

(2) 线下交互行为模式的多样性. 目前对人类线下交互行为研究的所有结论都是基于已有数据集

而得到的.不同的数据采集场景、采集方法、人群特性各有特点,这导致了我们在分析时难以得到普适性的结论.而人类行为往往复杂多样,不同场景下可能表现出截然不同的行为方式,不同社会关系的人之间也往往呈现不同的交互模式.例如,同事总是在工作时间有规律地交互,朋友则常常不定期约会,陌生人有可能在各种公共场所不期而遇.各种各样的交互模式混合在一起构成了复杂的人类线下交互网络,不论是挖掘交互模式,还是研究交互驱动机制,多样性都是必须面对的挑战.

(3) 线下交互建模的复杂性. 由于交互行为模式多种多样且互相融合,挖掘人类交互的驱动机制并由此建立交互模型也十分困难.许多简单的模型仅能模拟单一场景的交互行为,而一些模型为了还原更复杂的人类线下交互,对每一种状态和场景都分别考虑,这也导致模型变得复杂低效并且难以扩展.此外,对线下交互行为的建模研究缺乏有效而充分的衡量标准.

## 3 线下交互数据

要了解人类的线下交互行为,大量的经验数据必不可少.信息技术的发展使得通过数据实证研究人类交互行为成为可能.这里我们总结了几种常用的数据采集手段和已有的数据集.表 1 总结了文中提到的常用真实数据集,并描述了它们的基本特征.利用这些数据,我们可以从中提取出人类线下交互行为.

表 1 真实数据集概述(R(RFID),B(Bluetooth),W(WiFi),E(Ecard),L(LBS),NA(不存在))

数据集	类型	采集地点范围	采集时间/年	邻近范围/m	时间跨度/天	节点数目	是否存在外部节点	是否为多地点	时间分辨率	地点分辨率
HT09 <sup>[18]</sup>	R	会议	2009	1	3	100	是	否	20 s	NA
SG <sup>[18]</sup>	R	展览	2009	1	10	25	是	否	20 s	NA
SFHH <sup>[17]</sup>	R	会议	2009	1	2	575	是	否	20 s	NA
Hosp <sup>[18]</sup>	R	医院	2010	1	5	84	是	否	20 s	NA
Infocom <sup>[19]</sup>	B	会议	2005	4	3	41	是	否	20 s	NA
Reality mining <sup>[20]</sup>	B	不限	2006	10	180	100	是	否	5 min	NA
High school <sup>[21-22]</sup>	R	高中	2012	1	7	180	是	否	20 s	NA
USC <sup>[16]</sup>	W	大学	2006	>10	>365	4548	否	是	1 s	楼
Dartmouth <sup>[16]</sup>	W	大学	2006	>10	>365	5582	否	是	1 s	AP
Fudan09 <sup>①</sup> [15]	W	大学	2009	>10	120	14000	否	是	1 s	AP
Fudan13	W	大学	2013	>10	120	10000	否	是	1 s	AP
Singapore <sup>[23]</sup>	E	城市公交	2013	5	7	289570	否	是	1 s	公交车
Sjtu <sup>②</sup>	E	校园卡	2014	5	92	14000	否	是	1 s	POS机
Gowalla <sup>[24]</sup>	L	不限	2010	>10	120	380000	否	是	>1day	POI

(1) WiFi. 如今 WiFi 的覆盖范围越来越广,学校、商场、住宅、办公场所几乎都被无线信号覆盖.我们通过用户的无线连接时间及接入点(AP)的地理信息,可以获取用户的 WiFi 使用行为日志数据.由

于一个无线接入点的覆盖范围有限,我们考察同时接入某接入点的用户,就可以推断出用户间的地理

① <http://www.can.fudan.edu.cn/data/>

② <https://www.kesci.com>

邻近关系<sup>[16]</sup>. WiFi 数据具有较高的时间分辨率和一定的地点识别能力,普及率高,覆盖面广,在人类行为分析中颇具应用潜力.

(2) RFID. 射频标签是一种非接触式的自动识别技术,通过射频信号自动识别目标对象并获取相关数据. RFID 拥有较高的时空分辨率,通过调节射频标签的发射功率,可以使识别半径固定在 1 m 左右. 文献[17]通过这种方式来观测两个人的面对面交互(射频标签需佩戴在胸前). 同时,RFID 的时间分辨率也是可以调节的,研究者可以根据实际情况选择时间的粗粒化程度,使射频标签按照一定的频率持续发包,在不影响数据采集效果的情况下达到减少功耗的目的.

(3) Bluetooth. 大多数的手机、电脑中都装备有蓝牙,每一台蓝牙设备都可以发现周围 5~10 m 范围内的其它蓝牙设备. 通过软件可以持续扫描并记录与被观察个体邻近的其它个体所携带的蓝牙设备<sup>[20]</sup>. 传统的蓝牙技术并不能获取准确的交互地点信息. 近年来低功耗蓝牙技术使得人们可以在固定地点部署诸如 Ibeacon 这样的通信模块,该模块可以向周围的设备推送信息,同时也作为周围设备的地理定位依据.

(4) Ecard. 利用既有的电子卡类生活工具,也可以从某些特定角度反映出线下交互行为. 例如校园内学生卡的使用数据,城市内公交车卡刷卡数据<sup>[23]</sup>. 固定一个较小的时间窗口,就可以推测出在同时间窗口内同一地点刷卡的人有一次地理邻近.

(5) 其它. 除了上述数据采集手段外,还有一些技术可以粗略地获取人类的行为轨迹. 但需要注意的是,由于地理分辨率低,并不能从这些数据中找到常规意义上的线下交互. 但是从更广泛的角度来说,如果数据所覆盖的地理范围足够大,例如一个城市、国家甚至是全球范围,那么我们可以将地理邻近的标准降低. 在一个手机基站范围内,或是在 1 度的经纬度范围内即可视为地理邻近. 相应的时间尺度也可以放宽,比如一天或是一周,而不要求有完全的时间重合. LBS(Location-Based Service)网站数据和手机 GPS 数据都是这样的数据类型. LBS 社交网站的用户签到记录<sup>[25]</sup>,照片分享网站的照片地理信息<sup>[24]</sup>等都可以作为挖掘人类地理邻近或线下交互的数据源. 这些网站往往以“签到”的方式来获取用户的地理移动信息,其规模可以达到全球范围,但是由于签到行为是个体的自发行为,其时间连续性非

常低. GPS 数据通过对用户随身携带的移动电话定位,实现地理位置的经纬度标注. 由于经纬度坐标的连续性,关于地理邻近并没有统一的标准,很多时候要根据具体的场景和研究的问题进行粗粒化的选择和地点的合并.

## 4 线下交互的特性分析

人类线下交互行为在时间、空间上的模式描述,不仅启发了对动态社交网络中动力学过程的思考,也为建立人工模型提供了验证依据. 人们试图从中找到规律并解释其背后隐藏的机制. Holme 在其 2013 年《时效网络》一书<sup>[26]</sup>中,总结了动态网络的度量指标. 我们在本文中介绍了更多度量线下交互网络特性的指标,并总结了在研究线下交互所需要关注的问题. 下面我们将从拓扑、时间、地理三方面特性来分别介绍.

### 4.1 拓扑特性

在研究线下交互的动态社交网络时,我们以交互事件作为连边,交互个体作为节点构建线下交互网络,当且仅当两节点间至少有一次交互时,连边存在. 如果将网络在一定的时间窗口内聚集起来,便得到静态网络. 如果我们保留连边的时间或地点信息,则得到时效网络. 根据这两种情况,我们将拓扑特性分为集聚拓扑特性和时效拓扑特性.

#### 4.1.1 集聚拓扑特性度量指标

(1) 度分布. 度分布是网络最基本的度量指标之一,它衡量了一个随机挑选的节点度为  $k$  的概率. 度分布一方面能够描述网络的密度,另一方面也体现各节点在邻居数目上的差异性. 一般认为,节点的度越大,则节点的重要性越大.

无标度网络作为复杂网络中的典型网络结构,在自然界中大量存在. 其特征就在于,网络的度分布符合幂律分布,即少数的节点往往拥有大量的邻居,而大部分节点却很少. 对于线下交互网络来说,度分布随场景变化和网络构建方式的不同,既可能呈现幂律,也可能呈现非幂律的特征. 文献[16-17,27]中的数据采集自会议、医院、校园 WiFi 等场景,它们的集聚网络度分布均表现为指数分布,并且对于不同的集聚时间窗口具有鲁棒性. 这意味着在同样的时间窗口内,单个节点所能接触到的邻居数目具有同质性,并不存在无标度网络中的“超级传播节点”. 在公交车数据 Singapore<sup>[23]</sup>中,当网络连边定义为至

少发生一次交互时,其度分布接近正态分布;然而当网络连边定义为至少发生一次以上交互时,网络度分布则呈现幂律分布. 这种现象在线下交互中并不罕见(Fudan13, Sjtj 数据集均如此),其原因在于,线下交互中充斥着大量的偶然交互(如路过的陌生人),这使得整个交互网络更接近随机网络. 然而,当我们仅将交互一次的连边去除时,相当于去除了大量的随机交互,剩余的交互行为很大程度上取决于个体规律的行为模式,而个体的行为异质性便可能导致网络结构的异质性.

(2) 聚类系数. 聚类系数(亦或称为集聚系数)源于社会学中的“可传递三元组比率”<sup>[28]</sup>,它描述网络中个体的邻居节点也互为邻居的可能性. 节点  $i$  的聚类系数  $c_i$  基本定义如下:

$$c_i = \frac{\sum_{i,j,k} a_{i,j} a_{i,k} a_{j,k}}{\sum_{i,j} a_{i,j} a_{i,k}} \quad i \neq j \neq k \quad (1)$$

其中,  $a_{i,j}$  表示集聚网络中节点  $i, j$  之间的连边,若连边存在,则  $a_{i,j} = 1$ ; 否则  $a_{i,j} = 0$ . 通常而言,聚类系数的大小影响了网络上的传播动力学,当网络其它参数不变时,平均节点聚类系数越大,传播越慢.

聚类系数是描述小世界网络<sup>[29]</sup>的重要参数,聚类系数大、平均最短路径短的网络被称之为小世界网络. 对于线下交互而言,研究者们猜想,由于地理空间的隔离性,线下交互可能具有较大的聚类系数,同时也有较长的路径长度. 然而线下交互数据的实证分析否定了这一观点<sup>[23,30]</sup>. 研究者们发现,即使在城市尺度下,人类个体间的交互并没有限制在有限的地理区域或社团中,整个线下交互网络仍然符合小世界的特性.

(3) 平均节点强度. 我们定义节点强度  $s$  为一个节点与其所有邻边的交互总时长之和,节点强度与节点度的关系用平均强度  $\langle s(k) \rangle$  表示,  $s(k)$  为度为  $k$  的节点的强度. 通常研究者们关心  $\langle s(k) \rangle$  与  $k$  之间的相关关系<sup>[17-18]</sup>形式. 线上交互(如大规模电话通讯网络)的  $\langle s(k) \rangle$  与  $k$  往往表现出亚线性的相关关系,说明更加活跃或拥有更多邻居的人,在每一次交互中分配给邻居的时间更少,个体社交能力有限. 然而在线下交互(例如会议场景)中,该相关关系表现为超线性,暗示了某些“超级传播节点”在诸如信息传播的过程中起到了主导作用. 这些节点不仅具有大量的邻居,同时与每个邻居都能维持较长的交互时间. 平均节点强度与节点度的不同相关关系是线

下交互与线上交互的一大区别.

#### 4.1.2 时效拓扑特性度量指标

(1) 时效可达性与时效路径长度. 在时效网络中,网络路径存在时效限制,只有顺着时间而延伸的路径才标志着可达性,我们将这样的路径称之为时效路径. 对于两个节点来说,时效最短路径  $d_{i,j}$  表示连接节点  $i$  和  $j$  的总时长最短的路径,则在整个交互网络上,平均时效最短路径的长度  $L$  表示如下:

$$L = \frac{1}{N(N-1)} \sum_{i,j} d_{ij} \quad (2)$$

其中  $N$  为网络的节点数目,平均时效最短路径长度衡量了一个网络的传播能力. Zhang 等人<sup>[5]</sup>通过分析校园 WiFi 数据,发现时效交互网络的时效路径长度长于静态网络路径长度,时效可达性小于静态网络,并根据节点位于时效路径中的位置提出了衡量节点重要性的方法.

(2) 时效聚类系数. 作为静态网络聚类系数的时效扩展,Cui 等人提出了时延聚类系数和时长聚类系数两种方法<sup>[31]</sup>. 节点  $i$  的时延聚类系数  $C_i^d$  定义如下:

$$C_i^d = \frac{\sum_{j,k \in N_i} c_{i,j,k}^0 e^{-t_{i,j,k}^0}}{k_i(k_i-1)/2} \quad (3)$$

其中,  $k_i$  为节点  $i$  的度,分子  $\sum_{j,k \in N_i} c_{i,j,k}^0 e^{-t_{i,j,k}^0}$  为时效网络中节点  $i$  与其邻居节点  $j, k$  之间的连边实际构成的“时效三角形”的个数. 节点  $i$  的时长聚类系数  $C_i^w$  定义如下:

$$C_i^w = \frac{\sum_{j,k \in N_i} t_{\Delta i,j,k}}{\sum_{j,k \in N_i} t_{\wedge i,j,k}} \quad (4)$$

其中,  $t_{\Delta i,j,k}$  为节点  $i$  与其邻居节点  $j, k$  构成三角形的三条边的共同持续时间,  $t_{\wedge i,j,k}$  为节点  $i$  与节点  $j, k$  分别交互的时长的乘积. 在整个网络上,动态社交网络的时延聚类系数  $C_{net}^d$  与时长聚类系数  $C_{net}^w$  分别为

$$C_{net}^d = \frac{1}{N} \sum_{i=1}^N C_i^d \quad (5)$$

$$C_{net}^w = \frac{1}{N} \sum_{i=1}^N C_i^w \quad (6)$$

文章指出,相比于静态聚类系数,时效聚类系数能有效反映出网络时效特性改变带来的影响. 并且在传播仿真中,网络的时效聚类系数越大,有效传播阈值也越低,传播越容易爆发.

(3) 时效模体. 时效模体反映了动态社交网络

的微观拓扑特性,它不仅需要考虑连边的拓扑关系,还要考虑连边发生的时间先后关系和连边发生的时间间隔——通常间隔时间小于一定阈值的连边才会被视作一个时效模体<sup>[32]</sup>(如图 3 所示).通过分析动态社交网络中不同类型的时效模体所占的比例,可以获知该类交互行为的主要衍化模式.文献[33]探讨了会议、展览、校园 WiFi、手机短信交互数据集中时效模体的构成形式,并发现了线下交互(会议、展览、校园 WiFi)中存在“星形”和“链状”两种主导模体,他们分别对应于“领导”和“队列”两种线下交互模式.然而在手机短信的线上交互数据中,主导模体则以“ping-pong”为主,这说明在线上交互中个体更倾向于在短时间内与同一对象反复交互.

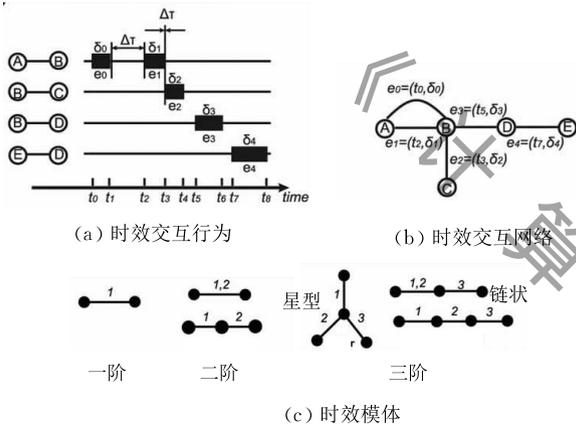


图 3 构建时效模体<sup>[33]</sup>

综上所述,线下动态社交网络在时间、空间两个维度上的复杂性使得其不可简单地用传统的基于静态网络的拓扑度量指标来刻画.尽管在集聚网络的层面上,一些经典的网络模型(无标度网络、小世界网络)仍然适用,但只考察集聚特性则会丢失重要的动态信息.例如 Zhang 等人<sup>[34]</sup>提出的动态社交网络新表示形式——时效线图,节点在时效线图上的度排序与节点在集聚网络中的度排序相反.这就意味着,即便集聚网络为无标度网络,度大的节点也未必是真正重要的节点.

对网络动态拓扑特征的刻画依然处于探索阶段,时效路径长度、时效聚类系数等指标的定义依然没有统一的标准.但可以确定的是,对线下交互网络的研究将涉及更加丰富的网络特性,以适应更加复杂而多变的实际场景.

## 4.2 时间特性

在时间特性的研究中,我们将从节点和边的层面来考虑交互特性,有的工作将其归纳为节点动力学和连边动力学,即所观察到的性质是节点固有的

性质或连边固有的性质.这些性质在时效网络的研究中非常重要,随着网络所描述的场景不同,这些性质也不尽相同.到目前为止,这个领域依然缺少统一的框架来完整的描述这些性质.

### 4.2.1 时间特性度量指标

(1)交互时间(Contact Time,简称 CT).交互时间表示个体之间单次交互的时长.在疾病传播中,感染者与非感染者的接触时长,决定了非感染者被感染的概率大小,交互时间越长,则越容易被感染.在机会网络信息传播中,交互时间代表着信息传递的容量.在我们所接触到的大部分实际数据集中,交互时间呈幂律分布,即大部分的交互时长都比较短,同时也存在少量时间很长的交互,这种现象在 RFID、Bluetooth 数据中均有体现<sup>[17]</sup>(如图 4(a)).这种特性常被用“惰性”机制进行解释,即一个人在一个地点停留或与某人交互的时间越长,他就越不想改变自己当前的状态.然而在 Singapore 公交车数据和 Fudan13 校园 WiFi 数据中,单次交互时间分布呈现指数分布(如图 4(c)、(e)),这是因为人们在公交车上的交互时间取决于公交车站点间的运行时间,在一辆车上乘坐超过两个小时的情况是很少的,因此这套数据也成为类似场景的典型代表,其交互时间表现出指数分布特征.除了用单一分布就能描述的数据集外,Li 等人<sup>[35]</sup>对出租车间的交互时长数据的拟合则呈现指数头、幂律尾的分段分布函数,进一步体现了线下交互的复杂模式.

(2)累计交互时间(Accumulated Contact Time,简称 ACT).与交互时间类似,个体对间的累计交互时长也能反映出信息/疾病在这两个个体间传播可能性的大小.在链路预测时,累计交互总时长也常常作为两个个体亲密程度的度量,累计时长越长,则认为两个个体越亲密.在构建时效网络的集聚静态网络时,也往往将累计交互时长作为连边的权重.可见该指标总是作为衡量两个节点关系的重要度量.累计交互时间常常呈现幂律分布<sup>[18,21]</sup>(如图 4(a)、(c)、(e)),这样的异质性说明该指标常取决于个体对的交互模式.

(3)个体交互间隔时间(Inter-Event Time,简称 IET).个体交互间隔时间定义为个体 A 连续先后与 B、C 两个不同个体发生交互的时间间隔.这个指标在信息/疾病传播等动态过程中,决定了一个个体从接收到信息/病毒后将其传到下一个个体的时间长度,因此它与传播过程本身时间尺度的相互作用对于传播过程至关重要.2005 年,Barabási 在

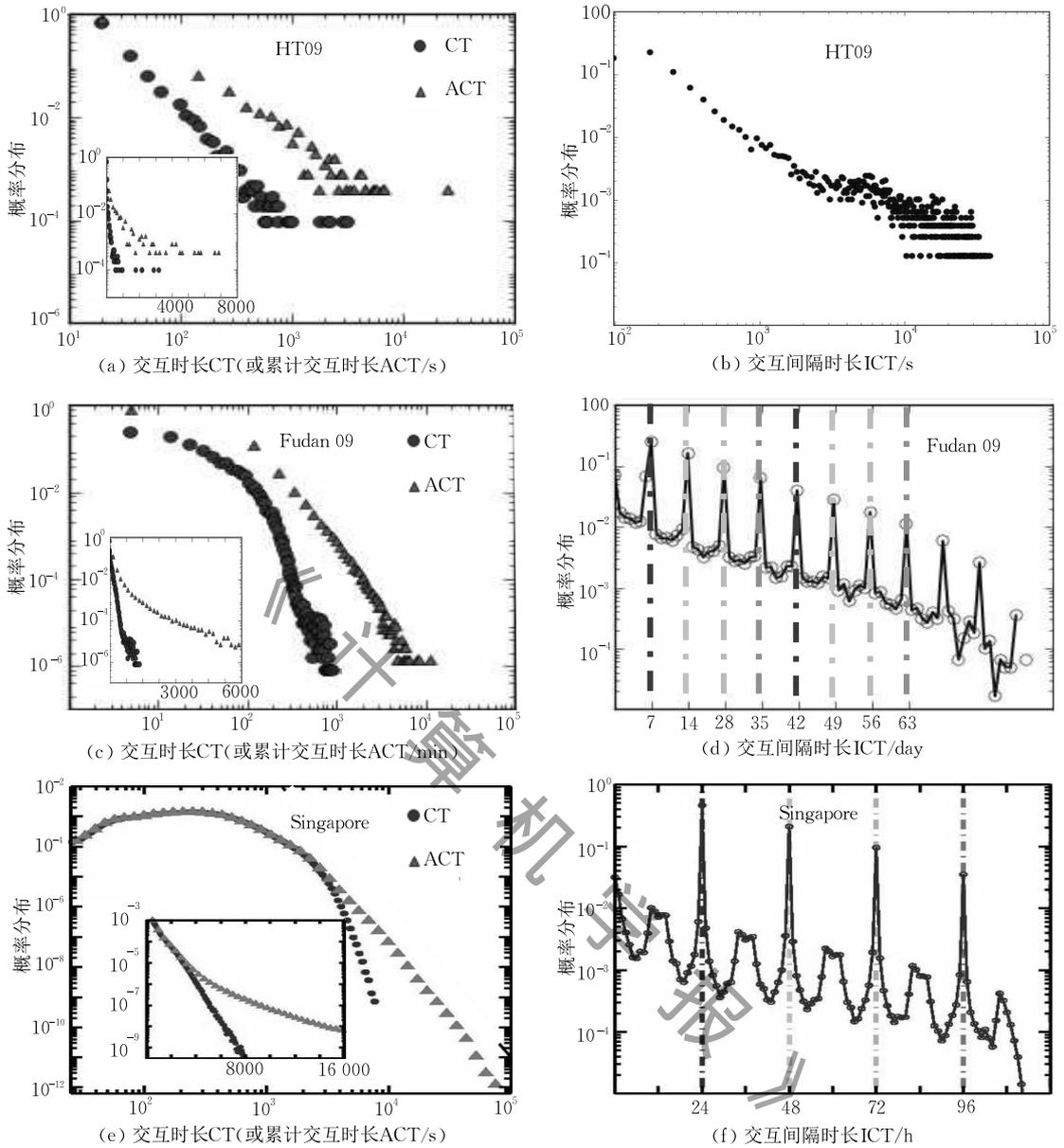


图 4 不同类型数据集的时效特性((a)、(b)以 HT09 数据集为代表的 RFID 数据集的 CT、ACT、ICT 概率分布；(c)、(d)以 Fudan09 数据集为代表的 WiFi 数据集的 CT、ACT、ICT 概率分布；(e)、(f)Singapore 公交车刷卡数据集的 CT、ACT、ICT 概率分布<sup>[23]</sup>)

《Nature》上的文章<sup>[36]</sup>引起了“Burst(阵发)”研究的热潮,文中统计了个体发邮件的间隔时间分布,与常见的泊松假设不同,个体发邮件的时间间隔分布为幂律分布.2006年,Barabási 组的 Vázquez 等人<sup>[37]</sup>进一步计算了多个数据的幂律分布指数,并根据幂律指数数值的不同对行为种类进行了划分.在众多的线下交互数据集中,个体线下交互的间隔时间都表现出了“阵发”的现象<sup>[17,38]</sup>.

(4) 个体对交互间隔时间(Inter-Contact Time, ICT).个体对交互间隔时间是另一个描述动态社交网络的重要特性,在线下交互中,个体对交互间隔时

间是同一对个体连续两次线下交互之间的间隔时间.在信息传播中,ICT 刻画了个体间传输数据的频率,这在机会网络中是十分重要的衡量指标,它体现了在特定时间段内与一个信息接受者或者携带者发生接触的概率.同样的概念也可以推广到疾病传播中.在研究 ICT 的工作中,通常分为两种:从单一个体对的角度进行研究或者从集聚的个体对角度进行研究.从单一个体对的角度进行的研究<sup>[39-42]</sup>发现同一套数据中个体对所呈现出的 ICT 分布不尽相同,有的呈现幂律分布,有的呈现指数分布.介于对单一个体对的处理往往受到采集数据量多少的影响,并

且个体对间千差万别,因此一些文献[43-44]转为对于集聚个体对的 ICT 进行分析,而他们大多得到一个带指数截断的幂律分布来描述 ICT(图 4(b)),这样的结果在一定程度上也解释了 ICT 的阵发现象. Passarella 等人<sup>[45]</sup>提出了异质性假设,并证明了集聚个体对的阵发可以由不同接触率的泊松分布的单个个体对 ICT 来进行拟合.然而 ICT 的阵发并不是普遍现象,在 Singapore 公交车数据和校园 WiFi 这类人类周期性显著的数据集中,其间隔时间具有明显的泊松特征或周期性特征(图 4(d)、(f)).这提醒我们,在集聚个体对上的 ICT 的表现可能是各种交互(即连边)的动力学叠加.此外, Singapore 公交车数据的研究表明,ICT 表现出的显著周期性表征了极具规律性的交互行为,而这正是公交车环境下“熟悉的陌生人”的产生机制.

(5) 时序性. 时序性是时效网络的特有特性,它代表了连边(或事件)发生的先后顺序,导致了时效网络拓扑结构上时效路径与时效模体的出现.而对于时序特性本身,我们可以通过分析节点参与时序事件的情况来衡量节点或连边的特性. Takaguchi 等人<sup>[46]</sup>提出,节点的上一个交互对象和下一个交互对象之间存在某种未知的联系,基于 Takaguchi 等人的工作, Cui 等人创新性的提出了“圈边”的概念<sup>[47]</sup>,并在建立社交网络时,在拥有固定时序行为两个个体之间建立连边. Zhang 等人构建的时效线图<sup>[34]</sup>同样关注时效交互网络中事件发生的先后关系.在时效线图中节点表示两个或多个个体的交互事件,当两个交互事件先后发生,且共享至少一个个体时,两个交互事件所代表的节点间存在连边.作者发现在线下交互数据集里,时效网络中度小的节点,在时效线图中很可能参与到度很大的交互事件中.这表明由于时序性的存在,即使是邻居数目很小的个体,在时效影响大的各种传播行为中却可以起到十分重要的作用.

(6) 周期性. 在长期人类线下交互网络中,由于人类行为自身的固有生活节律,周期性是时效线下交互网络的固有特征.个体的移动、连边的出现、网络结构的变化等均具有周期性<sup>[5,21,23]</sup>.以周期长度为分界,同周期内的交互特性与周期间的交互特性往往存在一定区别.正确识别周期长度对挖掘长期行为及交互模式至关重要.

综上所述,时间特性是动态社交网络的研究者们最为关心的特性之一,对时间特性研究指导了大量的实际应用.但是由于时间特性对场景的依赖性

和时间尺度的敏感性,对不同数据集在不同时间尺度下进行研究将得到不同的结论.这些时间特性的产生机制及其对网络动力学过程产生的影响依然是有待深入研究的课题.例如网络的时序性问题,尽管研究者们已经大量观测了时序性对其它特性(例如网络可达性,路径长度等)造成的影响,但研究工作依然停留在相关性分析的层面,如何对网络时序性进行更有效地表达和建模尚有待研究.

### 4.3 地理特性

由于大部分线下交互数据的采集都限制于较小的范围内,例如会场、中学等.而传统的 RFID, Bluetooth 一类的采集设备并不能够提供地理位置信息,因此大部分工作仅从时效的方面来讨论,而不考虑地点信息.在实际生活中,人们的行为不可能只局限于一个有界的区域,区域内的交互模式与区域间的交互模式也会体现出差异.不同的地点所带来的语义信息能帮助我们理解交互发生的动机<sup>[20]</sup>,更多地点的交互也许暗示了更加亲密的社会关系<sup>[25]</sup>.同时,地理上的移动使得一个城市乃至一个国家中的人们构成一个小世界网络,并不会因为地理的隔离形成很多孤立的团体<sup>[23]</sup>.

#### 4.3.1 地理特性度量指标

(1) 交互地点数目. 交互地点数目指的是在一定的地理分辨率下,两个个体共同出现的地点数目的多少.这个指标与地理位置的粗粒化程度关系密切,例如在一个大学校园里,如果分辨率精确到每栋楼,我们也许可以辨别出不同专业的同学;而如果将整个学校视为一个地点,我们则只能知道出现在这个地点的那些人可能是同校同学.因此在实际应用中,我们需要根据具体的情况和目的来选择地点的粗粒度. Crandall 等人<sup>[25]</sup>使用交互地点数目作为推断朋友关系的依据,其基本原理来自于“同质性原则”<sup>[48]</sup>,作者认为如果两个个体共同出现的地点数目越多,则他们越相似,也就越有可能互为朋友/熟人关系.

(2) 地点熵. 地点熵描述地点的受欢迎程度,它与一个地点的历史访问人数以及访问者对该地点的偏爱程度均有关. Scellato 等人在文献[24]中对地点熵做了详细的定义并由其衍生出一系列指标用以衡量两个个体地理位置相关的相似性.作者认为两个人共现地点越多,且这些地点的地点熵越小,则两个个体更有可能相识.

### 4.4 其它的特性

以上部分概括了常用于衡量线下交互的三方面

特性的度量指标, 还有一些工作从更深层次以及更具实际语义的角度描述了交互特征. 研究者们借助机器学习的模型通过提取交互特征, 训练学习出特定模式的交互. Do 等人<sup>[49-50]</sup>通过改进 LDA 模型并将其应用于线下交互数据, 其结果不仅仅描述两个个体间的交互模式, 还考虑了每一个时间片刻所有个体形成的交互网络的交互模式, 从拓扑结构的层面描述了交互模式的实际物理含义. 例如根据在一

定的时间窗口内的交互拓扑, 可以将交互分为工作交互、家庭/私人交互、集体活动交互等等. 但是这些工作都只局限于较小的数据样本, 且模式种类繁多, 难以扩展到更大规模的数据.

#### 4.5 数据特性总结与比较

表 2 中总结了各类数据集及其对应的交互特性. 随数据采集的场景、精度、年份不同, 各交互特性也会呈现不同的特征.

表 2 数据集的线下交互特性及适用模型  
(PW(幂律分布), EXP(指数分布), Norm(正态分布), NA(不存在), - (未讨论))

数据类型	数据集	CT	ACT	ICT	度分布	度与节点强度关系	集聚网络聚类系数	路径长度	交互地点数目分布(最大值)	适用模型
RFID	HT09	PW	PW	PW	EXP	超线性	0.46	1.89	NA	单日: LEVY <sup>[51]</sup> .
	SFHH	PW	PW	PW	EXP	超线性	0.28	2.20	NA	SBP <sup>[52]</sup> .
	SG	PW	PW	PW	EXP	亚线性	0.43	2.50	NA	ESBP <sup>[53]</sup> .
	hosp	PW	PW	PW	EXP	超线性	0.74	1.59	NA	STEPS <sup>[54]</sup> .
	High school	PW	PW	周期内:PW 整体:周期衰减	EXP	超线性	0.50	2.15	NA	ECMM <sup>[56]</sup> , SIMPS <sup>[57]</sup> .
蓝牙	Infocom	PW	PW	PW	EXP	超线性	0.65	2.39	NA	多日: TVC <sup>[58]</sup>
	Reality mining	PW	PW	周期内:PW 整体:周期衰减	EXP	超线性	0.73	1.47	NA	
WiFi	USC	PW	PW	周期内:PW 整体:周期衰减	EXP	超线性	0.62	2.25	PW (12)	单日: SWIM, SAME <sup>[59]</sup> .
	Dartmouth	PW	PW	周期内:PW 整体:周期衰减	EXP	超线性	0.65	4.00	PW (10)	IPAPM <sup>[60]</sup> .
	Fudan09	EXP	PW	周期内:PW 整体:周期衰减	EXP	超线性	0.20	2.05	PW (23)	多日: TVC, WDM <sup>[61]</sup> .
	Fudan13	EXP	PW	周期内:PW 整体:周期衰减	Norm(连边交互次数 $\geq 1$ ) PW(连边交互次数 $> 1$ )	超线性	0.30	1.96	EXP (30)	SLAW <sup>[44]</sup> , GeSoMo <sup>[62]</sup>
E-card	Sjtu	NA	PW	周期指数衰减	Norm(连边交互次数 $\geq 1$ ) PW(连边交互次数 $> 1$ )	超线性	0.20	1.73	PW (31)	WDM, GeSoMo
	Singapore	EXP	PW	周期指数衰减	Norm(连边交互次数 $\geq 1$ ) PW(连边交互次数 $> 1$ )	超线性	0.19	2.95	-	-
LBS	Gowalla	NA	PW	PW	PW	超线性	0.50	7.69	PW (364)	SPoT <sup>[63]</sup>

场景: RFID 和蓝牙数据在 CT 和 ACT 两个特性上均为幂律分布, 但是校园环境的 WiFi 数据和公交一卡通数据(Singapore)则有可能出现指数的 CT 分布, 这个现象在 4.2 小节中有详细解释. 此外, 对于不同场景, 在交互周期性上的有显著区别. 通常由 RFID 采集的数据都是持续时间较短的数据, 例如会议、展览. 该类数据不具备明显周期性, 因此在考虑 ICT 等特性时, 通常只考虑一天内的情况. WiFi、智能卡等数据则适合长时间大规模的采集, 因此包含了人类行为的周期性特征. 在我们考察的几套 WiFi 数据中, ICT 特性在一天内的分布和在整个时间轴上的分布是有明显区别的: 前者同 RFID 数据一样, 呈幂律分布; 但后者则表现出显著周期特征. 这暗示了幂律分布也许是短时间 ICT 分

布的固有特性, 而长期的 ICT 分布才能真正体现个体的交互规律.

精度: 数据采集的精度主要体现在对人类行为刻画连续性与准确性上. 例如, 同为大学校园 WiFi 数据的 USC 数据与 Fudan13 数据, USC 数据的地点精度为教学楼, Fudan13 为无线接入点, 这导致了两者在最大交互地点数目上有较大区别. 此外, 对于 Gowalla 这样的 LBS 数据, 它需要用户主动记录, 因此时间精度也大不如另外几类数据. 所以在分析 LBS 类数据时, 通常将时间粗粒化到“天”的层面来考察, 以避免引入不必要的噪声.

年份: 近十年中, 数据采集的精度和规模随着信息技术的进步都有相应的提高, 导致一些类似场景的数据在以前和现在出现不同的分析结论. 例如采

集自同一所大学的 Fudan09(2009 年)和 Fudan13(2013 年)WiFi 数据集. 由于 2013 年校园内无线热点覆盖率以及智能手机普及率都远高于 2009 年,因此 2013 年采集的数据对个体行为的刻画更为细致,交互网络的密度、交互事件数量都大量增加. 在交互网络的度分布上,不同于 2009 年的指数分布,2013 年的数据呈现正态分布. 类似的,同样采集于 2013 年之后的 SjtU 和 Singapore 数据也都表现出相同的正态分布特征.

当然,除了由于客观条件改变所带来的区别之外,线下交互也有其普遍存在的特性,例如度与节点强度之间的超线性关系(SG 数据除外),集聚交互网络的小世界特性等. 在研究分析时需充分了解这些共性和特性,避免得到片面的结论.

## 5 线下交互的建模

由于实际线下交互数据规模的限制,缺乏扩展性和灵活性,研究者们通过建模的方式来人工合成交互数据. 人工模型一方面保留了真实数据的基本特性,一方面使得各种交互参数可调节(例如交互人数、地点数目、时间长度等),有利于我们在更加灵活的场景中研究网络动力学问题. 其次,研究者们通过建模,在复现人类交互行为的同时,也逐步探索驱动人们交互行为的内在机制,从而为进一步的控制与预测提供支持. 在目前的研究中,通常通过对比真实数据特性来进行验证所建立的模型是否有效. 在接下来的部分中,我们根据模型的基本特征归纳了几类比较经典的线下交互模型:基于轨迹的模型、基于地点的模型和基于拓扑的模型. 在大多数的模型中,交互行为都是由于节点(个体)动力学所造成的,模型倾向于构建每个节点的行为规则,通过节点的移动、行为选择以及节点间的地理邻近来得到交互行为. 因此,对交互行为的建模离不开对个体行为机制的探索,个体移动模型与交互模型在某种程度上是密不可分的.

### 5.1 基于轨迹的模型

基于轨迹的交互模型着眼于模拟连续的人类移动轨迹,使节点按照一定的规则在二维平面上连续移动. 在规定交互半径范围内,节点按照一定的概率与该范围内的其它节点产生交互.

有限二维平面上的随机游走(RWM)模型早在 2000 年就被提出<sup>[64]</sup>. 单纯的随机游走无法体现出阵发的交互时间特征,不论在节点层面(IET)还是连

边层面(ICT)均产生指数分布的形态. 在文献[65]中作者证明了正是由于二维平面的有界性,导致了 ICT 呈现指数分布,这可以用一种记忆截断机制去解释:当节点在有界平面中进行移动时,每当节点触碰边界(导致反射或者停止),其效应相当于对节点的位置记忆进行了擦除,这种记忆截断导致了 ICT 分布的截断而生成指数分布. 当我们去除边界条件后,同样是随机游走,就可以得到幂律分布的 ICT. 同时,个体对的交互时长 CT 也呈幂律分布. 除了随机游走,无边界二维平面上的 levy 飞行<sup>[51]</sup>也可以产生带重尾的幂律分布 ICT. 当个体在一个点停留时间取自不同系数的经验 levy 分布时,还可以得到相应的幂律分布 CT.

在实际生活中,场景不同,人们的行为方式也不同. 例如在一个会场中,由于场地小,个体密度大,且场地没有明显的语义区别,因此个体是相对均匀地混合的,使用简单的随机游走就可以刻画一些特征. 但是对于更大的地理范围,例如一个社区或一个校园,当所考察范围内的地点存在明显语义差别时(住址、餐厅、体育馆、娱乐区等等),人们在地点间的移动就会存在偏向性,每个个体存在自己活动更加频繁的区域. 在 Hsu 的 TVC<sup>[58]</sup>模型中,作者在随机游走的基础上,为每个个体预设了地点偏向,即在平面中划分出一些方形区域作为某个节点的偏向区域. 当个体选择了某个区域后,它就在该区域内随机游走. 当个体根据某种概率切换区域时,就在备选区域内随机选点并从当前位置匀速移动过去. 考虑到个体的周期性行为,区域的偏向概率是一个周期性变化的函数. 在模型仿真数据的时效特性中,作者得到了 ICT 满足幂律分布,个体的移动行为也呈现周期性. 在拓扑方面,模型得到的线下社交网络平均度与实际数据相符.

类似于上文中时效区域偏向的思路,另一系列模型采取了更为细致的刻画手段. Zhu 等人提出的 SAME<sup>[59]</sup>刻画了大学校园交互场景,它根据在校生活的 5 种状态定义了 5 个子模型,并规定每个模型使用的起始时间、持续时间、活动范围等. 这一类模型在物理意义上易于理解,符合生活常识,但模型通常比较繁琐,也需要预设大量的参数.

与上述模型的繁琐截然不同, Michele 等人提出了一个十分简单的 SBP 模型<sup>[52]</sup>(名称 SBP 取自三位作者名字首字母). 该模型没有对个体的移动进行特别的前提假设,而仅对每一个个体本身预设了吸引力因子及个体活跃性. 个体根据周围其它个体

的吸引力进行移动,并由活跃因子决定是否与邻近个体交互.建模的结果很好地符合了几套室内交互(会议、展览、医院)数据的经验统计结果,即 ICT 与 CT 的幂律分布,以及个体对累计交互时间 ACT 的幂律分布.但是同随机游走系列的模型一样,SBP 模型导致个体交互在时间空间维度下都是均匀的.Zhang 等人<sup>[53]</sup>则观察到,即便是在会议这样简单的场景中,个体交互的时间也并非完全均匀,某些时间段(例如会议中的茶歇)个体间的交互更加密集.这种现象将导致整个交互网络在不同的时间段内其时效可达性也是不一样的,而网络上的传播速度也是时变的.因此 Zhang 等人基于 SBP 模型提出了基于公共活跃度和个体可达性机制的新模型(这里称为 Enhanced-SBP,简称 ESBP),在由外部事件驱动的公共活跃期,所有个体的活跃性集体增大,个体移动速度也根据活跃性作相应调整.Zhang 等人的模型不仅在时效特性和时效拓扑特性上与经验数据一致,其所产生的动态社交网络相比于原始 SBP 模型,在传播仿真中也得到了与真实数据更接近的传播速度.

## 5.2 基于地点的模型

在基于地点的模型中,个体不是在一个平面上连续移动,而是按顺序访问一系列离散的地点,两个个体同时访问同一个地点则发生线下交互.这类模型能够十分容易地克服人群均匀混合的问题,使得个体地点偏向这个特征能够很好地表达出来.

WDM<sup>[61]</sup>模型是最早能同时刻画幂律 ICT 和 CT 的模型.它采用子模型组合的机制模拟实际的场景,根据时间地点建立不同规则的模型,例如上班时间模型、午饭时间模型、夜晚休闲时间模型.不同模型有不同的移动交互规则,组合模型具有可扩展性,根据需要添加新的子模型和模型转换规则.

受启发于上一部分中提到的 TVC 模型, Kim 等人提出的 SLAW 模型<sup>[44]</sup>可看成是 TVC 的一种扩展.对应于 TVC 中每个个体预设的地理偏向区域,SLAW 中将所有的地点进行分簇,每个簇由临近的一组离散地点组成.其中,每个簇具有权重,个体根据权重选择自己偏爱的簇,并从中挑选出一定比例的地点作为每个周期必须访问的地点,同时再随机挑选一个簇及其中的一些地点作为随机访问的地点.为了刻画经验数据中的幂律分布 ICT,个体在每个地点的停留时间取自指数截断的幂律分布.该模型不仅很好地刻画了交互时间特性,也保留了交互发生的地点范围,其离散化的地点特性常用于对

WiFi 数据集进行建模.

SWIM<sup>[54]</sup>模型对 SLAW 做出了进一步的改进,地点权重静态变为时变,而时变权重取决于当前地点的访问人数.随机地为每个个体安排一个 home 地点,个体的地点选择取决于地点权重以及地点与自己 home 的距离.这个模型相对简单,参数很少易于调整.类似的还有 STEPS<sup>[55]</sup>模型,每个个体存在预设的偏向地点且倾向于在偏向地点及其周围移动.特别地, Ramiro 等人在文献<sup>[55]</sup>中还着重讨论了经验数据及模型仿真数据中所体现出的小世界特性.

以上的几个模型在 ICT 这个指标上都能够得到带指数截断的幂律分布,他们的共同点是都假设了每一对个体对之间的交互时间间隔均满足幂律分布,由此形成的集聚 ICT 也满足幂律分布.然而个体对 ICT 均满足幂律分布并不是集聚 ICT 呈幂律分布的必要条件, Zhang 等人提出的 IPAPM<sup>[60]</sup>模型就是很好的例证. IPAPM 模型使用了个体活跃度和地点偏向机制,文中证明了个体的泊松行为使得单个个体对的 ICT 呈指数分布,但集聚 ICT 的却呈幂律(阵发)特性,同时 CT 也呈幂律分布.这对我们研究交互行为的建模是一个很好的启发,也提供了新的思考和思路.

## 5.3 基于拓扑的模型

前面两节中提到的模型着重于关心时间和地点特性,这一节中所提到的模型则希望将社会结构这个因素也放入模型当中.研究者们相信,交互行为的产生,某种程度上受到其社会关系的影响.因此在基于拓扑的模型中,人们移动及交互行为的发生受到社会关系网络的驱动,静态的社会网络往往作为模型的参数或者输入.加入了社会因素,研究者们希望通过模型模拟出更多的交互特征,例如拓扑特征、一致性特征.所谓一致性,就是人们认为在社会网络中关系更加亲密的个体间应该拥有更长的累计交互时长.

Amorim 等人提出的 SIMPS 模型<sup>[57]</sup>基于平面随机游走,并将社会网络放入其中,个体在移动时受到存在社交连边的个体的吸引或排斥. CMM<sup>[66]</sup>系列模型是一类典型的以社团结构为基础网络拓扑结构的离散地点模型,后期很多工作对原始的 CMM 模型进行改进. CMM 模型事先将个体划分为多个不重叠社团,社团内部连接(有权重),社团间无连边.每个社团在初始时被安排在一个地点上,社团内部的节点能以一定概率自发移动到其它地点,每个节点在移动时都会参考同社团内部其它成员的位

置,以此来确定每个地点对其的吸引力. HCMM 模型<sup>[43]</sup>针对普通 CMM 模型的弱点进行改进,除了在原始社团间重连一些边使其构成小世界之外,还为每个个体定义 home 地点. 个体在选择地点时参考属于一个 home 的所有个体的地理位置及社会网络连边权重. ECMM 模型<sup>[56]</sup>也在 CMM 基础上加以改进,除了增加停留时间的幂律分布外,还加强了非同社团好友对个体吸引的能力. 两个改进工作都实现了 ICT 的带指数截断幂律分布,后者还考虑了 CT 的幂律特性,同时实现了较高的一致性.

另一部分基于拓扑的模型并没有特别规定拓扑形式,甚至适用于任何结构的社交网络. GeSoMo<sup>[62]</sup>模型在其构建过程中,个体的行为与社交网络拓扑是分离的,静态社交网络是模型的一个输入. 模型中包括了地点偏向机制、地点周期吸引机制以及个体间时效的吸引机制. SPoT<sup>[63]</sup>则是一个综合考虑了时间、空间、社会三方面特性的模型框架,其基本思想是将一个社交网络转化为一个个体-地点的“到达网络”,个体-地点连边存在与否取决于对输入社会

网络的社团分析(可重叠、有层级),同一个社团的个体连接到相同的地点,而个体-地点连边的通断可以视为一个随机点过程,实验发现伯努利过程可以很好地逼近实际数据的 ICT 分布(LBS 数据). 选择不同的随机过程将影响所生成分布的形状,为模型提供了很好的扩展性.

此外,还有一类模型在脱离地点的条件下单纯考虑交互网络连边的时效特性<sup>[67-69]</sup>,这类模型通常使用记忆机制来实现 ICT 和 CT 的阵发特性. 在文献[68]中的模型 BSZ(模型名称取自三个作者姓氏缩写)中,个体可以是独立的,也可以位于一个群组中. 个体有参与交互和不参与交互两种状态,某种状态保持时间越久,则状态越不容易改变. 同样的,个体参与群组与离开群组的概率也与个体状态保持的时间相关.

#### 5.4 线下交互模型总结

我们在表 3 中总结了上述所提模型的主要构建机制和其可以描述的交互特性,并在表 2 中总结了各数据适用的建模方法.

表 3 线下交互模型总结(√(满足特性),×(不满足特性),-(未提及))

类别	模型	机制	特性				
			ICT	CT/ACT	周期性	地点特性	拓扑特性
基于轨迹的模型	LEVY	列为飞行,无界平面	√	—	×	×	—
	TVC	时效地点偏爱,随机游走	√	—	√	—	√
	SAME	子模型组合	√	√	×	—	—
	SBP	个体吸引力,个体活跃度,随机游走	√	√	×	×	—
	ESBP	个体吸引力,个体/集体活跃度,随机游走	√	√	×	×	√
基于地点的模型	WDM	子模型组合	√	√	√	×	—
	SLAW	地点分簇,地点偏爱	√	—	√	√	—
	SWIM	地点偏爱动态更新	√	√	×	√	—
	STEPS	地点偏爱,地点吸引力	√	—	×	√	√
	IPAPM	地点偏爱,个体活跃度	√	√	×	√	—
基于拓扑的模型	HCMM	地点偏爱,个体吸引力,社团结构拓扑	√	—	×	√	√
	ECMM	地点偏爱,个体吸引力,社团结构拓扑	√	√	×	√	√
	GeSoMo	时效个体吸引力,地点周期偏向	√	—	√	√	√
	SPoT	社团结构拓扑,个体-地点连边随机过程	√	—	×	√	√
	SIMPS	网络拓扑,个体吸引力	√	—	—	×	√
	BSZ	网络拓扑,记忆性	√	√	×	×	√

根据所建模场景的个体行为轨迹是否连续、是否具有周期性、场景范围大小、时间精确程度,应选用合适的模型. 同时,我们发现在现有研究工作中,时间特性是大部分模型最为关心的部分,尤其是针对 ICT 和 CT 这样的基础特性. 部分模型在满足时间特性的基础上,也融入了对地点特性或拓扑特性(如度分布)的考虑. 然而在前文对交互特性的概述中,我们知道,对线下交互网络更深入的研究需要通过更复杂的特性进行刻画,例如各种时效拓扑特

性、网络时序性等. 现有的模型对复杂交互特性的建模还十分缺乏,例如 SBP 模型,尽管其在宏观上对于 ICT 和 CT 的拟合十分准确,但在微观的时序结构,网络的时效可达性上却与实际数据相去甚远. Zhang 等人的 ESBP 模型正是考虑到了这一点而对 SBP 模型进行了改进,使模型具有与实际数据更接近的时效可达性,并实现了与真实情况更加接近的传播过程. 可见,如何进一步对复杂交互特性进行建模的意义重大.

## 6 线下交互研究的应用

挖掘人类线下交互模式并建立模型,在实际生活中有着广泛的用途。随着数据的丰富和研究的深入,线下交互研究为诸多领域的应用带来了新的视角,在创新应用技术中具有潜在的重要意义。

### 6.1 社会关系推断

社会关系是连接人与人的纽带,准确推断人与人之间的社会关系在检测遗漏信息、辨别虚假交互、评估用户关系、建立推荐系统上都有重要应用。线下交互是真实世界中个体之间最直接的交互,对线下交互特性(例如交互频率、时长、次数等)的分析往往被作为推断社会关系的直接依据<sup>[24-25]</sup>。

Crandall 等人<sup>[25]</sup>利用照片分享网站的照片信息(包括拍摄时间、地点),统计一定时间窗口内两个个体线下交互的不同地点数目,并以该数目作为时空巧合次数指标,推测网站用户的社会关系。巧合次数越多,则用户间互为好友关系的可能性越大。在上述工作的基础上,Scellato 等人<sup>[24]</sup>认为除了考虑交互地点数目外,地点本身的特性也是决定用户关系的重要因素,因此设计了地点熵指标来衡量一个地点的随机性。公众地点比如餐厅、广场具有较大的随机性,而一个小的书店、咖啡厅常常有固定的访客,因此具有较小的随机性。随机性越小,则该交互地点对亲密社会关系的贡献越大。

利用线下交互的时效特性,Cui 等人提出了利用“圈边”推断个体好友关系的方法<sup>[47]</sup>。该方法通过观察被选节点与其交互对象的交互时序特征,推测其交互对象之间的好友关系。该方法只需要选取部分节点进行观测,即可推断出整个网络中所有节点间的社会关系,有效地减少观测成本。

由于线下交互是时间、地理两个维度共同作用的结果,不同时间、地点具有不同的语义,他们所刻画场景也不相同,不同场景下发生的线下交互也暗示着不同的社会关系。Eagle 等人通过分析包含蓝牙数据和 GPS 数据的 Reality mining 数据集,对用户社会关系类型进行了更加细致的划分<sup>[20]</sup>。Eagle 等人<sup>[20]</sup>利用 HMM 模型识别不同时间地点所代表的语义(例如办公、家、外出娱乐),并识别出家人、朋友、办公室熟人三种社会关系。

Melo 等人<sup>[70]</sup>通过建立线下交互网络得到刻画交互模式的两个指标:社会相似性和交互频率,通过随机化交互网络得到上述两个指标的随机阈值。通

过比较两个用户间的真实交互指标与随机阈值之间的关系,将用户社会关系划分为熟人、陌生人、桥、朋友四种关系。不同于好友/非好友的二元关系推断,线下多关系推断更加符合实际生活,且不同关系背后蕴含的交互模式为研究交互网络上的信息传播、疾病传播等动力学过程带来了新的启发。

### 6.2 社团挖掘

在社交网络的应用中,社团挖掘作为一个重要方向而备受关注。许多工作专注于研究诸如 Facebook、Twitter 等线上社交网络的社团结构<sup>[71-72]</sup>以此更好地推动虚拟社区建设、广告营销等活动。然而线上社交网络往往只能反映人们在虚拟空间的行为,并不代表在真实世界中的社会关系与社交活动。在线下交互网络中,网络连边不再是传统的社会关系(例如好友关系),而是交互行为特性(如交互时长、交互地点数目等)经过处理转换而得的权重,对线下交互网络的社区挖掘有助于改善真实生活的诸多方面,例如机会路由、城市交通规划、企业组织建设等。同时,线下交互的复杂性与动态特性也为传统的静态社团挖掘方法带来了新的挑战,需要研究者们提出更加灵活的解决方法。

Jones 等人设计了 GPI 方法<sup>[73]</sup>,充分利用地理信息来挖掘社团结构,指出社团出现所对应的地理位置。然而该方法是集中式处理的方法,且需要用户报告自己的行为轨迹,无法保证用户的隐私。作为改进,Jones 等人设计了 GDC 社团挖掘方法<sup>[74]</sup>,通过分析蓝牙交互数据的交互频率、交互时长等指标,排除线下偶遇的情况,找到显著稳定的社团结构(允许社团重叠)。由于仅需要用户保存和交换线下交互的人员列表及交互时长等必要信息,该方法能较好保护用户行为隐私,并可分布式实现。

由于线下交互的时效特性,网络结构在每个时刻都有可能发生变化,网络社团也因时间和交互场景的不同而发生改变。近几年,张量成为了表示这种时效交互网络的流行方式,而张量分解也成为挖掘时效社团的有效方式。Gauvin 等人<sup>[75]</sup>分析小学校园内的学生蓝牙交互数据,将不同时刻的交互网络视为时变的邻接矩阵,并用一个三向的张量表示,其中两个维度均为用户个体,第三个维度为时间。通过非负张量分解,作者得到两个矩阵,分别描述个体与社团间的关系、社团与时间的关系,从而得到鲁棒的社团结构以及相关的社团活跃时间。类似的,Zhang 等人<sup>[76]</sup>也使用张量分解的方法,但与上述方法不同的是,作者以行为活动为张量的第三个维度,而非使

用时间,例如“星期四下午在操场踢足球”是一种行为活动.作者通过这种构建方式得到了个体与社团的关系、社团与具体活动的关系.

对线下交互的研究除了改进社团挖掘算法本身之外,也拓宽了社团挖掘的应用场景.在企业员工行为研究中,通过分析社团内部交互行为,可以对社团组织的行为节律进行建模<sup>[77-78]</sup>,依次建立效率更加高效的团队.通过分析旅游公司游客出行信息,建立游客线下交互网络,挖掘游客社会网络局部社区,可以提升旅游公司客户管理方式,提供创新的个性化的服务,降低营销成本<sup>[79]</sup>.在智慧城市建设背景下,大规模的人群线下交互分析与社团演化分析有助于动态感知城市功能区域,挖掘公共交通使用模式,实现更合理的交通规划,交通资源分配与车辆调度<sup>[80-81]</sup>.

### 6.3 机会网络中的信息传播

机会网络为未来城市中自组织移动设备间的通信提供了新的范式,数据的交换需要用户之间实际距离足够接近,且维持足够长的时间.线下交互作为机会网络中信息传递的基础,其交互特性与信息传播的路径、效率、有效性均密切相关.

考虑到线下交互并非均质的,人们总是会与一部分的社会关系亲密(社会相似性高)的个体间发生更为紧密的接触.通常,研究者们通过用户自报告的社会关系数据,线上社交数据或历史线下交互数据来衡量用户间的社会相似性,并将机会路由协议的设计与个体间的社会相似性一起.在 Mtibaa 等人设计的 PeopleRank 机会路由协议中<sup>[14]</sup>,个体在转发信息时,将考虑转发对象与目标节点间的相似性来决定是否转发. Hsu 等人设计的协议<sup>[82]</sup>则设定了初始的相似性阈值,当转发对象与目标节点相似性大于阈值时则进行转发.相似性的衡量可以来自于自报告的社会关系、线上社交数据或历史线下交互数据.由 Hui 等人设计的 BUBBLE Rap 协议<sup>[19]</sup>结合了社团挖掘与网络中心性两方面,将每个节点的网络介数中心性(Betweenness)排名分为全局排名和社团内部排名.信息携带节点将信息转发给排名高于自己的节点,其中,当转发对象与目标对象处于同一社团时,采用社团内部排名,否则采用全局排名.该方法既保留了前面两种方法的优点,又通过中心性排名的方式提高了传播效率.

与上述的基于社会相似性的方法不同,另一类路由协议则是基于个体间的历史交互时间间隔与剩余时间<sup>[83-86]</sup>.该类方法通过分析两个个体的历史交

互时间规律,并结合最近一次相遇时间推断出距离下次相遇的剩余时间,从而找到最快的传输路径.

更多的机会网络路由协议相关的工作参见前人综述<sup>[87]</sup>,这里就不再赘述.但需要强调的是,路由协议的设计与线下交互网络的建模密切相关,为了验证路由协议在各种网络特性参数下的性能,只利用有限且固定的真实数据集是远远不够的.因此需要首先采用适当的模型生成交互网络,并调节模型参数模拟不同的情形.

### 6.4 疾病传播与公共卫生健康

线下交互模式影响着疾病传播的过程,Salathé 等人<sup>[12]</sup>在高中的交互数据上进行了传染病模型的仿真,发现其与流感季的真实传染特征高度一致.同时,作者也强调了在研究免疫策略时经验交互数据的重要性. Lee 等人<sup>[88]</sup>设计了基于交互模式熟人免疫策略,实验表明考虑了交互时效特性的免疫策略优于传统的熟人免疫.因此,在时效线下交互背景下,我们需要对网络中的“超级传播者”做出新定义.传统的针对静态网络的节点重要性衡量指标(例如度中心性,介数中心性,邻近中心性, $k$ -壳等)在动态的线下交互中效果并不理想,研究者们需要根据实际情况设计新的重要性指标<sup>[89-93]</sup>,例如时效介数中心性,时效邻近中心性<sup>[90]</sup>等.

除了传染病之外,压力、焦虑甚至肥胖等人类心理或生理状态都受到人与人之间交互的影响<sup>[94]</sup>. Aharony 等研究者们基于面对面的交互,对实验对象的健身活动进行干预<sup>[95]</sup>,并采用不同的激励机制来试图提高实验对象的健身活动次数.研究发现,当每个个体的健身奖励与其交互邻居的健身行为绑定时,群体的健身活动次数有所增加,也验证了基于线下交互行为进行个体行为干预的有效性.

### 6.5 推荐系统

移动互联网发展与便携式智能设备的普及大大推动了基于地理位置的应用服务(LBS),连接了人们的线上和线下活动.按照推荐的对象进行划分,基于人类线下活动和交互的推荐可以分为好友推荐、地点/活动推荐、商品推荐.

好友推荐将向用户推荐其感兴趣的其它用户,以增加应用或服务的社交特性,增强用户粘度.在前文的社会关系推断一节中,已经列举了一些社会关系推断的方法用于好友推荐,然而使用这些方法仅局限于向用户推荐已经熟识的好友.在基于位置服务的应用中,我们则希望向用户推荐那些在将来有可能成为好友的用户,以扩大用户的社交圈.同时在

这些应用中, 线下交互不再严格地限定于同时同地出现, 而是将其定义扩展为线下行为的相似性。Li 等人提出了一种层次化相似性指标<sup>[96]</sup>, 利用 GPS 得到的人类移动轨迹数据, 根据不同地理粗粒化层次上两个个体的轨迹重合程度给出相似性得分, 并依据得分高低推荐好友。Xiao 等人<sup>[97]</sup>则对上述移动轨迹相似性指标进行扩展, 作者认为个体移动轨迹的重合不必拘泥于地理位置的重合, 也可以是地理语义的重合。如果两个个体访问地点的语义(类别)是相同的, 例如博物馆、电影院、咖啡厅等等, 那么可以认为个体有相似的兴趣爱好。如果他们的轨迹在时序上也是相同的, 那么就可以认为他们在语义地点上的轨迹是相似的。

基于线下交互的地点/活动推荐常常采用基于用户的协同过滤方法, 通过上一段中提到的各种方法找到与当前用户相似的用户集合, 根据该用户级集合中用户喜爱的地点和活动进行推荐<sup>[98-100]</sup>。

基于线下交互的商品推荐是近几年的新兴模式。例如我国的“淘身边”应用, 卖家在发布商品的同时, 为商品标注地理标签, 当用户出现在附近时, 便可看到这个商品。其优势在于让送货/自提更方便, 节省时间成本也更低, 这样就可以有效地避免网购难以规避的风险。另外诸如基于低功耗蓝牙 Ibeacon 一类的技术使得基于位置的广告推送、智能车位寻找等成为可能。

此外, 拼车已经成为当今出租车行业重要组成部分。智能拼车技术可以看作是特殊的基于线下交互的推荐系统, 为用户匹配最合适的共乘伙伴和路线。尤其是当今国内几大打车软件运营商的合并, 为分析海量用户出行数据和租车模式带来了新的机遇。拼车技术中最关键的步骤在于匹配用户路线, 静态的匹配方法通过分析用户历史地理移动轨迹, 为用户建立移动模型, 通过模型预测未来的移动及地理路线的重合性, 并以此预测潜在的拼车行为<sup>[101-102]</sup>。动态匹配则通过集中式的处理服务从即时的用户行车路线数据中找到合适的拼车用户, 使得乘车时间最小化, 同时减少用户乘车费用。集中式的处理在计算上存在巨大的开销, 针对这种弊端, Lin 等人设计了基于邻域线下社交网络的方法<sup>[103]</sup>, 将备选的拼车用户范围限制在当前用户邻近的地点范围内, 以减少路线匹配的计算代价。

## 7 展 望

基于线下交互的动态社交网络的研究是“小荷

才露尖尖角”, 很多研究也只是管中窥豹。展望未来, 我们认为以下几个方面将是研究的主要方向:

### (1) 多元化行为数据采集与分析。

现有的各类行为数据在数据精度与规模方面都各有劣势, 而人类行为大数据时代的到来使得数据资源多元化的特点。未来的多元数据研究应涉及以下几个方面:

① 利用多元数据提高线下行为的位置定位精度。目前常用的智能手机+WiFi 定位模式具有覆盖范围广、成本低廉、无需专用设备支持并易与用户交互等优点, 但在室内定位的应用中, 依然具有地理位置精度低、跨楼层识别准确率低的缺陷。因此在手机 WiFi 这类泛在传感器的基础上结合类似 iBeacon 等高精度低功耗定位技术<sup>[104]</sup> 将为精度更高的用户行为数据采集提供必要的技术支撑和数据保障。

② 线上线下数据互联。当今大量的应用如团购网站、打车软件等都将人类的线上线下行为紧密联系在一起。文献<sup>[10]</sup>提出了一个感知大规模社交网络的应用框架并投入实验, 在其 2012、2013 年的实验中均成功地从多个数据源获取到人类行为数据, 包括问卷调查、手机数据、WiFi 数据、社交网站数据等。研究发现人类的手机通话行为与线下移动交互行为密切相关<sup>[105]</sup>, 利用通话记录可以提高对个体线下移动交互预测的准确性<sup>[106]</sup>。但是, 多种类型的数据包含了用户的行为模式信息, 处理不当则会导致用户隐私的暴露。因此如何在保证用户隐私的同时综合多种数据来推动理论研究和实际应用是一个具有挑战性的问题。

③ 不完备数据分析。由于数据采集不可能做到百分之百完整, 因此如何对不完备数据进行分析也是一个重要的课题。Barrat 等人<sup>[107]</sup>通过对线下交互数据集的重采样和重建, 估算数据不完整性对交互网络中传播过程的影响, 以此对实验结果进行修正。这也为我们在不完整或小规模数据集上研究人类行为带来了新的启发。

### (2) 复杂行为模式分析与可扩展的建模框架。

尽管现在已经有大量工作对人类线下交互进行刻画, 但是大部分工作都着重于分析个别数据集, 得到的交互模式片面而缺乏普适性。当面对更复杂的社会系统时, 传统的单个刻画指标无法衡量整个系统。例如我们在文中提到的“熟悉的陌生人”<sup>[23, 108]</sup> 这种交互模式, 它是我们在社会生活的特点场景中一定会发生的交互模式, 如果仅仅以交互时长、交互频率等简单指标来衡量, 我们往往会将这种重要

模式忽略. 在大数据背景下, 对复杂人类行为的分析不能再仅仅依靠统计技术, 而需要对多维特征进行刻画, 并结合机器学习、数据挖掘等技术手段, 从多角度进行分析, 这样才能挖掘发现更真实的人类交互模式.

当前大部分交互建模工作均是建立在对一类数据集的相似特征的刻画上, 缺乏可扩展性. Karamshuk 等人<sup>[66]</sup>提出的 SPoT 模型综合考虑了时间、空间、社会结构三方面的特征, 提出了灵活的模型框架以适应不同的交互特性. 这种框架式的建模或将成为未来的趋势, 在建立人类交互模型与仿真行为交互数据时, 我们寻求类似“复杂特征+简单逻辑”的方法, 在统一框架中, 根据不同的特征组合, 添加不同的驱动机制, 形成复杂的场景模拟.

(3) 智慧城市中的创新应用. 城市的智慧化以及行为数据的爆发式增长需要我们将眼光投入新兴的应用领域. 例如城市智慧交通系统, 通过对线下交互数据进行分析, 我们可以实现定制的公交线路来减轻交通压力和交通资源的动态分配. 在大量兴起的 LBS 位置服务中, 线下交互数据可以作为各种定位工具的补充, 为用户提供更精准的室内或室外定位; 在智能物流方面, 线下交互的研究十分具有应用潜力. 例如新近提出的众包物流概念, 正如同机会网络中的信息路由一样, 基于交互预测实现快递路由, 可以充分利用社会闲散运力以节省人力成本, 运输成本. 以上列举了一些新环境下的创新应用, 而对这些应用的思考和尝试也是我们研究最终追求的目标.

## 8 结 论

在便携式移动设备快速普及、移动互联网迅猛发展的今天, 感知人类行为、探测人类线下交互变得更加容易而精确, 分析人类线下交互行为模式, 理解其背后的驱动机制已经成为诸如疾病免疫、机会路由、推荐系统等实际应用的基础. 本文从人类线下交互的数据、特性、建模及应用四个方面入手, 回顾了相关领域的新近发展, 从拓扑、时间、空间三个维度详细梳理了我们所关心的线下交互特性以及基于这些特性的交互模型, 探讨了当前研究应用场景, 最后展望了未来的主要研究方向. 总之, 目前对人类线下交互动态社交网络的研究才刚刚起步, 尚需深化和拓展, 技术的发展也将为其更深入的探索和应用带来新的契机.

致 谢 感谢复旦大学信息办公室协助采集的校园 WiFi 数据, 感谢 CRAWDAD、Sociopatterns、Kesci 平台提供的数据支持!

## 参 考 文 献

- [1] Wellman B, Berkowitz S D. Social structures: A network approach. *American Journal of Sociology*, 1989, 94(6): 83
- [2] Barabási A. The architecture of complexity. *IEEE Control Systems*, 2007, 27(4): 33-42
- [3] Fan Chao, Guo Jin-Li, Han Xiao-Pu, et al. A review of research on human dynamics. *Complex Systems and Complexity Science*, 2011, 8(2): 1-17(in Chinese)  
(樊超, 郭进利, 韩筱璞等. 人类行为动力学研究综述. 复杂系统与复杂性科学, 2011, 8(2): 1-17)
- [4] Zhou Tao, Han Xiao-Pu, Yan Xiao-Yong, et al. Statistical mechanics on temporal and spatial activities of human. *Journal of University of Electronic Science and Technology of China*, 2013, 42(4): 481-540(in Chinese)  
(周涛, 韩筱璞, 闫小勇等. 人类行为时空特性的统计力学. 电子科技大学学报, 2013, 42(4): 481-540)
- [5] Zhang Y, Wang L, Zhang Y Q, et al. Towards a temporal network analysis of interactive WiFi users. *Europhysics Letters*, 2012, 98(6): 68002
- [6] Min B, Goh K I, Vazquez A. Spreading dynamics following bursty human activity patterns. *Physical Review E*, 2010, 83(3): 515-529
- [7] Fu X, Small M, Chen G. *Propagation Dynamics on Complex Networks: Models, Methods and Stability Analysis*. Chichester, UK: Wiley Higher Education Press, 2014
- [8] Simmel G, Wolff K H. *The Sociology of Georg Simmel*. New York, USA: Free Press, 1950: 397-397
- [9] Rivera M T, Soderstrom S B, Uzzi B. Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annual Review of Sociology*, 2010, 36(1): 91-115
- [10] Stopczynski A, Sekara V, Sapiezynski P, et al. Measuring large-scale social networks with high resolution. *PLoS One*, 2014, 9(4): e95978
- [11] Read J M, Eames K T D, Edmunds W J. Dynamic social networks and the implications for the spread of infectious disease. *Journal of the Royal Society Interface*, 2008, 5(26): 1001-1007
- [12] Salathé M, Kazandjieva M, Lee J W, et al. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107(51): 22020-22025
- [13] Zhang Y Q, Li X. When susceptible-infectious-susceptible contagion meets time-varying networks with identical infectivity. *Europhysics Letters*, 2014, 108(2): 28006

- [14] Mtibaa A, May M, Diot C, et al. Peoplerank: Social opportunistic forwarding// Proceedings of the IEEE INFOCOM. San Diego, USA, 2010; 1-5
- [15] Ma Hua-Dong, Yuan Pei-Yan, Zhao Dong. Research progress on routing problem in mobile opportunistic networks. *Journal of Software*, 2015, 26(3): 600-616(in Chinese)  
(马华东, 袁培燕, 赵东. 移动机会网络路由问题研究进展. *软件学报*, 2015, 26(3): 600-616)
- [16] Hsu W J, Helmy A. On nodal encounter patterns in wireless LAN traces. *IEEE Transactions on Mobile Computing*, 2010, 9(11): 1563-1577
- [17] Cattuto C, Wouter V D B, Barrat A, et al. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS One*, 2010, 5(7): e11596
- [18] Barrat A, Cattuto C, Colizza V, et al. Empirical temporal networks of face-to-face human interactions. *European Physical Journal Special Topics*, 2013, 222(6): 1295-1309
- [19] Hui P, Crowcroft J, Yoneki E. Bubble rap: Social-based forwarding in delay-tolerant networks. *IEEE Transactions on Mobile Computing*, 2011, 10(11): 1576-1589
- [20] Eagle N, Pentland A. Reality mining: Sensing complex social systems. *Personal & Ubiquitous Computing*, 2006, 10(4): 255-268
- [21] Barrat A, Fournet J. Contact patterns among high school students. *PLoS One*, 2014, 9(9): e107878-e107878
- [22] Mastrandrea R, Fournet J, Barrat A. Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS One*, 2015, 10(9): e0136497
- [23] Sun L J, Axhausen K W, Der-Horng L, et al. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, 110(34): 13774-13779
- [24] Scellato S, Noulas A, Mascolo C. Exploiting place features in link prediction on location-based social networks// Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego, USA, 2011; 1046-1054
- [25] Crandall D J, Backstrom L, Cosley D, et al. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107(52): 22436-22441
- [26] Holme P. *Temporal Networks*. New York, USA; Springer, 2014
- [27] Zhang Y Q, Li X. Characterizing large-scale population's indoor spatio-temporal interactive behaviors//Proceedings of the ACM SIGKDD International Workshop on Urban Computing. Beijing, China, 2012; 25-32
- [28] Wasserman S, Faust K. *Social Network Analysis: Method and Application*. Cambridge, UK; Cambridge University Press, 1994
- [29] Watts D J, Strogatz S H. Collective dynamics of "small world" networks. *Nature*, 1998, 393(6684): 440-442
- [30] Moreno B N, Vespignani A, Vucelja M, et al. A spatio-temporal network model to represent and analyze LBSNs//Proceedings of the IEEE Pervasive Computing and Communication International Conference Workshops. Saint Louis, USA, 2015; 142-147
- [31] Cui J, Zhang Y Q, Li X. On the clustering coefficients of temporal networks and epidemic dynamics//Proceedings of the IEEE International Symposium on Circuits and Systems. Beijing, China, 2013; 2299-2302
- [32] Kovanen L, Karsai M, Kaski K, et al. Temporal motifs in time-dependent networks. *Journal of Statistical Mechanics Theory & Experiment*, 2011, 2011(11): 1293-1307
- [33] Zhang Y, Li X, Xu J, et al. Human interactive patterns in temporal networks. *IEEE Transactions on Systems Man & Cybernetics Systems*, 2015, 45(2): 214-222
- [34] Zhang Y Q, Li X. Temporal dynamics and impact of event interactions in cyber-social populations. *Chaos*, 2013, 23(1): 013131
- [35] Li Y, Jin D, Wang Z, et al. Exponential and power law distribution of contact duration in urban vehicular ad hoc networks. *IEEE Signal Processing Letters*, 2013, 20(1): 110-113
- [36] Barabási A L. The origin of bursts and heavy tails in human dynamics. *Nature*, 2005, 435(7039): 207-211
- [37] Vázquez A, Oliveira J G, Dezső Z, et al. Modeling bursts and heavy tails in human dynamics. *Physical Review E*, 2006, 73(3): 80-98
- [38] Zhou T, Han X P, Yan X Y, et al. Statistical mechanics on temporal and spatial activities of human. *Journal of the University of Electronic Science & Technology of China*, 2013, 44(4): 481-540
- [39] Chaintreau A, Hui P, Crowcroft J, et al. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 2007, 6(6): 606-620
- [40] Karagiannis T, Boudec J Y L, Vojnović M. Power law and exponential decay of inter contact times between mobile devices//Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Network. Montréal, USA, 2007; 183-194
- [41] Conan V, Leguay J, Friedman T. Characterizing pairwise Intercontact patterns in delay tolerant networks//Proceedings of the ACM International Conference on Autonomic Computing and Communication Systems. Rome, Italy, 2007; 19
- [42] Gao W, Li Q, Zhao B, et al. Multicasting in delay tolerant networks: A social network perspective//Proceedings of the ACM International Symposium on Mobile Ad Hoc Networking and Computing. New Orleans, USA, 2009; 299-308
- [43] Boldrini C, Passarella A. HCMM: Modelling spatial and temporal properties of human mobility driven by users' social relationships. *Computer Communications*, 2010, 33(9): 1056-1074

- [44] Lee S, Hong S, Kim S J, et al. SLAW: A mobility model for human walks//Proceedings of the IEEE INFOCOM. Rio de Janeiro, Brasil, 2009: 855-863
- [45] Passarella A, Conti M. Analysis of individual pair and aggregate intercontact times in heterogeneous opportunistic networks. *IEEE Transactions on Mobile Computing*, 2013, 12(12): 2483-2495
- [46] Takaguchi T, Nakamura M, Sato N, et al. Predictability of conversation partners. *Physical Review X*, 2011, 1(1): 3382-3391
- [47] Cui J, Zhang Y Q, Li X. Contextual relationship in temporal social networks: Circle link//Proceedings of the International Symposium on Nonlinear Theory & Its Applications. Luzern, Switzerland. 2014: 260-263
- [48] McPherson M, Smith-Lovin L, Cook J M. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 2001, 27(1): 415-444
- [49] Do T M T, Gatica-Perez D. Groupus: Smartphone proximity data and human interaction type mining//Proceedings of the 15th Annual International Symposium on Wearable Computers. San Francisco, USA, 2011: 21-28
- [50] Han Y J, Cheng S B, Park S Y, et al. Finding social interaction patterns using call and proximity logs simultaneously//Proceedings of the IEEE International Conference on Advances in Social Networks Analysis and Mining. Beijing, China, 2014: 399-402
- [51] Rhee I, Shin M, Hong S, et al. On the levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking*, 2011, 19(3): 630-643
- [52] Michele S, Andrea B, Romualdo P S. Modeling human dynamics of face-to-face interaction networks. *Physical Review Letters*, 2013, 110(16): 1-9
- [53] Zhang Y Q, Cui J, Zhang S M, et al. Modeling temporal networks of human face-to-face contacts with public activity and individual reachability. *The European Physical Journal B*, 2016, 89(2): 1-8
- [54] Kosta S, Mei A, Stefa J. Large-scale synthetic social mobile networks with SWIM. *IEEE Transactions on Mobile Computing*, 2014, 13(1): 116-129
- [55] Ramiro V, Diaz M, Nguyen A D, et al. STEPS—An approach for human mobility modeling. *Lecture Notes in Computer Science*, 2011, 6640(1): 254-265
- [56] Vastardis N, Yang K. An enhanced community-based mobility model for distributed mobile social networks. *Journal of Ambient Intelligence & Humanized Computing*, 2014, 5(1): 65-75
- [57] Amorim M D D, Fdida S, Borrel V, et al. SIMPS: Using sociology for personal mobility. *IEEE/ACM Transactions on Networking*, 2009, 17(3): 831-842
- [58] Hsu W J, Spyropoulos T, Psounis K, et al. Modeling spatial and temporal dependencies of user mobility in wireless mobile networks. *IEEE Transactions on Networking*, 2009, 17(5): 1564-1577
- [59] Zhu X, Bai Y, Yang W, et al. SAME: A students' daily activity mobility model for campus delay-tolerant networks//Proceedings of the 18th IEEE Asia-Pacific Conference on Communications. Kota Kinabalu, Malaysia, 2012: 528-533
- [60] Zhang Y Q, Li X, Liang D, et al. Characterizing bursts of aggregate pairs with individual poissonian activity and preferential mobility. *IEEE Communications Letters*, 2015, 19(7): 1225-1228
- [61] Ekman F, Keränen A, Karvo J, et al. Working day movement model//Proceedings of the 1st ACM SIGMOBILE Workshop on Mobility Models, MobilityModels 2008. Hong Kong, China, 2008: 33-40
- [62] Fischer D, Herrmann K, Rothenmel K. GeSoMo—A general social mobility model for delay tolerant networks//Proceedings of the IEEE 7th International Conference on Mobile Adhoc and Sensor Systems. San Francisco, USA, 2010: 99-108
- [63] Karamshuk D, Boldrini C, Conti M, et al. SPoT: Representing the social, spatial, and temporal dimensions of human mobility with a unifying framework. *Pervasive & Mobile Computing*, 2014, 11(6): 19-40
- [64] Davies V A. Evaluating Mobility Models Within an Ad Hoc Network [M. S. dissertation]. Colorado School of Mines, Golden, USA, 2002
- [65] Cai H, Eun D Y. Crossing over the bounded domain: From exponential to power-law intermeeting time in mobile ad hoc networks. *IEEE/ACM Transactions on Networking*, 2009, 17(5): 1578-1591
- [66] Musolesi M, Hailes S, Mascolo C. An ad hoc mobility model founded on social network theory//Proceedings of the ACM International Symposium on Modeling Analysis & Simulation. Volendam, The Netherlands, 2004: 20-24
- [67] Juliette S, Alain B, Ginestra B. Dynamical and bursty interactions in social networks. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2010, 81: 417-422
- [68] Kun Z, Juliette S, Ginestra B, et al. Social network dynamics of face-to-face interactions. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2011, 83: 1447-1448
- [69] Vestergaard C L, Mathieu G, Alain B. How memory generates heterogeneous dynamics in temporal networks. *Physical Review E*, 2014, 90(4): 042805-042805
- [70] Melo P O S D, Vaz D, Viana A C, et al. Recast: Telling apart social and random relationships in dynamic networks. *Performance Evaluation*, 2015, 87: 19-36
- [71] Chen Y L, Chuang C H, Chiu Y T. Community detection based on social interactions in a social network. *Journal of the Association for Information Science and Technology*, 2014, 65(3): 539-550
- [72] Lam A J. Improving Twitter community detection through contextual sentiment analysis of tweets//Proceedings of the ACL 2016 Students Workshop. Berlin, Germany, 2016: 30-36

- [73] Gupta A, Paul S, Jones Q, et al. Automatic identification of informal social groups and places for geo-social recommendations. *International Journal of Mobile Network Design and Innovation*, 2007, 2(3-4): 159-171
- [74] Mardenfeld S, Boston D, Pan S J, et al. GDC: Group discovery using co-location traces//*Proceedings of the IEEE 2nd International Conference on Social Computing*. Minneapolis, USA, 2010: 641-648
- [75] Gauvin L, Panisson A, Cattuto C. Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach. *PLoS One*, 2014, 9(1): e86028
- [76] Zhang Y, Chen M, Mao S, et al. CAP: Community activity prediction based on big data analysis. *IEEE Network*, 2014, 28(4): 52-57
- [77] Eagle N, Pentland A S, Lazer D. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(36): 15274-15278
- [78] Eagle N, Pentland A S. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 2009, 63(7): 1057-1066
- [79] Fan Wei, Chen Xu-Guang. Algorithm, of mining local communities in tourists social network. *Computer Engineering and Design*, 2016, 37(6): 1505-1509(in Chinese)  
(樊玮, 陈旭光. 旅游游客社会网络局部社区挖掘方法. *计算机工程与设计*, 2016, 37(6): 1505-1509)
- [80] Sun Cong. Data Gathering Technique in Urban Sensing Network Based on Public Transport [M. S. dissertation]. Beijing University of Posts and Telecommunications, Beijing, 2015(in Chinese)  
(孙聪. 基于公共交通的城市感知网络中数据收集技术研究[硕士学位论文]. 北京邮电大学, 北京, 2015)
- [81] Zhao Wei-Dan, Research of Urban Subway Traffic Pattern Based on Big Data Platform [M. S. dissertation]. Jilin University, Jilin, 2016(in Chinese)  
(赵卫丹. 基于大数据平台的城市地铁交通模式研究[硕士学位论文]. 吉林大学, 吉林, 2016)
- [82] Hsu W, Dutta D, Helmy A. Structural analysis of user association patterns in university campus wireless lans. *IEEE Transactions on Mobile Computing*, 2012, 11(11): 1734-1748
- [83] Srinivasa S, Krishnamurthy S. CREST: An opportunistic forwarding protocol based on conditional residual time//*Proceedings of the 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*. Rome, Italy, 2009: 1-9
- [84] Zhu H, Fu L, Xue G, et al. Recognizing exponential inter-contact time in VANETs//*Proceedings of the IEEE INFOCOM*. San Diego, USA, 2010: 1-5
- [85] Chen H, Lou W. On using contact expectation for routing in delay tolerant networks//*Proceedings of the IEEE International Conference on Parallel Processing*. Taipei, China, 2011: 683-692
- [86] Jia Jian-Bin, Chen Ying-Wen, Xu Ming. Prediction based relay selection method in opportunistic vehicular networks. *Journal of Software*, 2015, 26(7): 1730-1741(in Chinese)  
(贾建斌, 陈颖文, 徐明. 基于预测的机会车载网络中继选择策略研究. *软件学报*, 2015, 26(7): 1730-1741)
- [87] Mota V F S, Cunha F D, Macedo D F, et al. Protocols, mobility models and tools in opportunistic networks: A survey. *Computer Communications*, 2014, 48: 5-19
- [88] Lee S, Rocha L E C, Liljeros F, et al. Exploiting temporal network structures of human interaction to effectively immunize populations. *PLoS One*, 2010, 7(5): 1496-1520
- [89] Lee S, Rocha L E C, Liljeros F, et al. Exploiting temporal network structures of human interaction to effectively immunize populations. *PLoS One*, 2012, 7(5): e36439
- [90] Tang J, Mascolo C, Musolesi M, et al. Exploiting temporal complex network metrics in mobile malware containment//*Proceedings of the IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks*. Lucca, Italy, 2011: 1-9
- [91] Masuda N, Holme P. Predicting and controlling infectious disease epidemics using temporal networks. *F1000 Prime Reports*, 2013, 5(1): 6
- [92] Starnini M, Machens A, Cattuto C, et al. Immunization strategies for epidemic processes in time-varying contact networks. *Journal of Theoretical Biology*, 2013, 337: 89-100
- [93] Holme P. Information content of contact-pattern representations and predictability of epidemic outbreaks. *Scientific Reports*, 2015, 5(1): 14462
- [94] Madan A, Cebrian M, Moturu S, et al. Sensing the health state of a community. *IEEE Pervasive Computing*, 2012, 11(4): 36-45
- [95] Aharony N, Pan W, Ip C, et al. Social fMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, 2011, 7(6): 643-659
- [96] Li Q, Zheng Y, Xie X, et al. Mining user similarity based on location history//*Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Irvine, USA, 2008: 34
- [97] Xiao X, Zheng Y, Luo Q, et al. Inferring social ties between users with human location history. *Journal of Ambient Intelligence and Humanized Computing*, 2014, 5(1): 3-19
- [98] Ye M, Yin P, Lee W C. Location recommendation for location-based social networks//*Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. San Jose, USA, 2010: 458-461
- [99] Zheng V W, Zheng Y, Xie X, et al. Collaborative location and activity recommendations with GPS history data//*Proceedings of the 19th International Conference on World Wide Web*. Raleigh, USA, 2010: 1029-1038
- [100] Zheng Y, Xie X. Learning travel recommendations from user-generated GPS traces. *ACM Transactions on Intelligent Systems and Technology*, 2011, 2(1): 2

- [101] Liu N, Feng Y, Wang F, et al. Mobility crowdsourcing: Toward zero-effort carpooling on individual smartphone. *International Journal of Distributed Sensor Networks*, 2013, 9(2): 615282
- [102] Baldacci R, Maniezzo V, Mingozzi A. An exact method for the car pooling problem based on Lagrangean column generation. *Operations Research*, 2004, 52(3): 422-439
- [103] Lin Y, Shen H. VShare: A wireless social network aided vehicle sharing system using hierarchical cloud architecture// *Proceedings of the IEEE 1st International Conference on Internet-of-Things Design and Implementation*. Berlin, Germany, 2016: 37-48
- [104] Deepesh P C, Rath R, Tiwary A, et al. Experiences with using iBeacons for indoor positioning//*Proceedings of the 9th India Software Engineering Conference*. New York, USA, 2016: 184-189
- [105] Toole J L, Herrera-Yaque C, Schneider C M, et al. Coupling human mobility and social ties. *Journal of the Royal Society Interface*, 2015, 12(105): 20141128
- [106] Zhang D, Zhang D, Xiong H, et al. NextCell: Predicting location using social interplay from cell phone traces. *IEEE Transactions on Computers*, 2015, 64(2): 452-463
- [107] Géniois M, Vestergaard C L, Cattuto C, et al. Compensating for population sampling in simulations of epidemic spread on temporal contact networks. *Nature Communications*, 2015, 6(1): 8860
- [108] Liang D, Li X, Zhang Y Q. Identifying familiar strangers in human encounter networks. *Europhysics Letters*, 2016, 116(1): 18006



**LIANG Di**, born in 1991, M. S. Her current research interests include temporal network, social network analysis and human dynamics.

**CUI Jing**, born in 1990, M. S. Her current research interests include temporal network, social network analysis and human dynamics.

**LI Xiang**, born in 1975, Ph. D., professor, Ph. D. supervisor. His research interests include theories and application of complex network and network science.

## Background

The development of mobile internet and increasingly popularity of smart mobile devices make it easier to sense human behaviors and interaction. Offline interaction, as an important part of human social networks, tell that who is close to whom at what time, linking human individuals in both spatial and temporal dimensions. Recent studies demonstrate an increasing interest in offline interaction for their importance in various dynamic process and applications, such as designing epidemical immune strategies, creating efficient routing protocols and friend recommendations. There are a lot of work trying to characterize human offline interaction patterns and uncover the hidden mechanism driving human interaction. This paper provides a thorough and comprehensive overview of the current study on human offline interaction in

dynamical social networks. We summarize the difficulties and challenges in the data collection, analysis and modeling, and review the existing achievements. We classify the interaction characteristics into topological, temporal and spatial categories, and review the human interaction modeling with trace-based, location based and topological based models. We address the applications and outlook the future directions.

This work was partly supported by the Natural Science Fund for Distinguished Young Scholar of China (No. 61425019), the National Natural Science Foundation of China (No. 71731004), the Key Project of National Social Science Fund of China (No. 12&ZD218), and the Shanghai SMEC-EDF Shuguang Project (No. 14SG03).