

基于移动通信数据的城市可视分析研究

王桂娟^{1,2}, 周锐¹, 蔡梦杰¹, 汤勇^{2,3}, 李茸茸¹, 陈华容^{1,2}, 吴亚东⁴

1. 西南科技大学计算机科学与技术学院, 四川 绵阳 621000;

2. 西南科技大学信息工程学院, 四川 绵阳 621000;

3. 四川轻化工大学自动化与信息工程学院, 四川 自贡 643000;

4. 四川轻化工大学计算机科学与工程学院, 四川 自贡 643000

摘要

随着移动电话的深入普及,大规模通信数据给人们提供了前所未有的观测城市微观结构和动态的机会,而如此大规模的高维异构时空关系数据又给高效数据解读带来了挑战。作为重要的大数据分析手段,可视化被越来越多地应用到这一领域。回顾近年来基于通信数据的城市可视分析研究工作,首先归纳了移动通信数据的主要来源、特征和常用的数据处理方法,然后从通信数据的内在对象“人”“通信设备”和“城市空间”3个方面阐述相应的可视化方法,并对基于通信数据的城市可视分析面向的任务、方法和特点进行了梳理,最后对基于通信数据的城市可视分析进行了展望。

关键词

通信数据;可视分析;智能城市感知;人类行为模式;城市动态

中图分类号:TP311

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2021013

A survey on mobile communication data based urban visual analysis

WANG Guijuan^{1,2}, ZHOU Rui¹, CAI Mengjie¹, TANG Yong^{2,3}, LI Rongrong¹, CHEN Huarong^{1,2}, WU Yadong⁴

1. School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621000, China

2. School of Information Engineering, Southwest University of Science and Technology, Mianyang 621000, China

3. School of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong 643000, China

4. School of Computer Science and Engineering, Sichuan University of Science and Engineering, Zigong 643000, China

Abstract

With the popularity of mobile phones, large-scale communication data has brought people unprecedented opportunities to observe the urban microstructure and dynamics. However the complexity of such large-scale high-dimensional heterogeneous spatio-temporal data has brought challenges to effective data interpretation. Accordingly, visualization, as

an important means of big data analysis, is being applied to this field more and more. The recent research work on urban visual analysis based on communication data was reviewed. Firstly, the major sources, characteristics and common data processing methods of the mobile phone data were summarized. Secondly, the corresponding visualization methods from three aspects were elaborated, namely, the internal objects of mobile communication data such as human, communication equipment and urban space, and the frequent urban visual analysis task and features based on communication data were summarized. Finally, the prospect of urban visual analysis based on communication data was presented.

Key words

communication data, visual analysis, intelligent urban sensing, human behavior pattern, urban dynamics

1 引言

移动电话的深入普及给人们带来了前所未有的观测市民流动性的机会,可以利用数据驱动的洞察力更好地规划城市和优化城市服务^[1]。目前,全世界93%的人口能够访问移动宽带网络,大部分国家50%以上的人拥有手机,发达国家拥有手机的人数占比甚至达到90%^[2]。通信运营商的计费需求和与通信相关的数据记录需求带来了大量的通信数据。这些数据记录了移动电话用户在通信过程中与基础设施交互的痕迹,可近乎实时地反映手机持有者的地理位置。与传统的基于问卷调查的城市分析相比,通信数据采集成本低廉,具有更大规模、更细的时空分辨率,而且能反映城市的动态变化。通信数据包含的丰富的人与人、人与空间的时变关系信息给城市人类行为研究和城市微观动态研究提供了基础,基于通信数据的城市分析已成为一种重要的城市智能感知途径。

然而,由于城市数据的异构性、高复杂性和大规模性,经常需要在城市分析过程中融合人的感知,这促进了可视化的广泛应用^[3]。可视化将各种类型的数据翻译为相应的可交互图形,使得分析者可以借助高效的视觉通道直观地与数据进行沟通。进一步地,可视分析通过丰富的交互设计

让领域专家和分析者有机会参与数据分析的全流程。分析结果的可视化成为人与机器合作的桥梁^[4]。可视分析系统的用户能够在不失去全局信息的同时,从多个角度和不同尺度观察数据^[5]。因此,许多研究者开始采用可视分析方法解读通信数据中蕴含的丰富的城市信息,这逐渐形成了一个较有成效的交叉研究领域。

本文调研了近年来基于通信数据的城市可视分析方面的研究工作,总结了通信数据的特点,从通信数据城市分析的内在对象——人、通信设备和城市空间3个方面阐述了相应的可视化方法,对基于通信数据的城市可视分析的任务和方法进行了梳理,并对基于通信数据的城市可视分析进行了展望。

2 数据和数据处理

数据是开展可视化研究的基础。通信数据是指人在使用通信网络进行交流和资源访问的过程中产生的记录。通信数据包含的信息可大致分为人、通信设备和城市空间3个维度的时变数据。

2.1 通信数据

从国内外的研究来看,通信运营商是通信领域开展可视化研究最主要的数据来

源,通信运营商可以提供通话记录数据、基站数据、移动网络数据等方面的信息。此外,一些研究人员通过招募志愿者,开发特定的手机应用程序或者科研开放项目进行通信数据采集,能够获取更细粒度的数据,如智能手机的事件log、App使用记录。通信领域主要数据的来源、类型、典型属性以及代表文献见表1。同时,在通信数据可视分析中,部分研究结合其他领域数据进行融合和关联分析,如结合社交网络数据、交通数据和兴趣点(point of interest, POI)数据等。

(1) 通话记录详单

通话记录详单(call detail record, CDR)是通信领域可视化研究中使用非常频繁的数据之一,CDR数据是通信运营商为了对用户通话进行计费而产生的记录。当用户拨打电话时,接入基站会把该用户的通话记录下来用于计费。CDR记录的信

息包括主被叫用户ID、通话开始时间、通话持续时间、关联基站等。国内外研究人员基于CDR数据集开展了大量的通信领域可视化的研究,取得了一系列的成果。Jiang H Y等人^[6]通过CDR提取用户行为特征、识别用户关系,基于用户行为特征对运营设施进行优化,能够完成基站选址、基站网络微调等分析任务。Dong Y X等人^[7]从CDR数据中检测异常的人群事件。Andrienko G等人^[8]利用通话记录开展了人口流动性分析。Mai W M等人^[9]利用基于密度的含噪声应用空间聚类(density-based spatial clustering of applications with noise, DBSCAN)方法,从CDR数据中提取用户轨迹,挖掘交通特点,描绘关键线路。李艳妮等人^[10]基于通话记录数据开发了用户行为分析系统。黄文彬等人^[11]分析了移动用户的行为模式,并建立了用户模型,利用K-means进行4类用户划分。向峰^[12]开展了移动用户行为研究,研究了移动性对用户关系的影响和城市地理区域的感知。

(2) 基站数据

基站数据包含基站编号、基站地址、基站经纬度坐标等信息,对于研究用户的行为轨迹、城市人群流动有重要意义。Li Z H等人^[13]利用基站轨迹数据开发了一套城市移动模式可视分析系统,可有效地帮助用户分析移动人群与城市区域间关系的动态变化规律。WANG F等人^[14]根据基站实际位置,用Voronoi图划分城市域,从而通过空间结构有效地管理数据。Zhang S H等人^[15]基于基站Erlang测量对基站行为进行分析,其观测结果可以用于蜂窝优化、资源规划等。

(3) 手机网络数据

手机网络数据是用户使用移动网络产生的记录。Chen L B等人^[16]基于网络流量数据进行未来流量预测,提出了一种距离约束互补感知(DCCA)算法,以优化容量

表1 通信领域常见数据集

数据	属性	代表文献	
来源	类型		
通信运营商	通话记录详单	主被叫用户ID	[2][6-12]
		通话开始时间	
		通话持续时间	
	基站数据	关联基站	[13-15]
		基站编号	
		基站地址	
手机网络数据 (信令数据、基站切换)	基站经纬度坐标	[5][16]	
	用户ID		
	访问网址		
	接入起止时间		
	关联基站位置		
机构和个人	定制App/开放项目	App使用情况	[17-18]
		智能手机事件 log	
		Wi-Fi	
		GPS	
		POI	
		源-目的地(origin-destination, OD) 数据	

利用率和部署成本为目标,寻找最优的基站分簇方案。Sagl G等人^[15]从移动网络流量中分析了大规模的集体城市流动行为,有助于更好地理解动态城市系统的“脉搏”。

(4) 定制App

在通信数据分析中,有研究采用定制App的方式采集更细粒度的通信网络使用信息。Yadav K等人^[17]通过在志愿者手机上安装专门用于数据采集的App,扫描并记录时间戳、MCC、MNC、LAC、Cell ID和RSSI信息;同时,扫描可见的Wi-Fi接入点,并用时间戳记录它们的SSM和BSSID信息,从而研究参与者的行为模式,并和其他数据集进行比对。Slingsby A等人^[18]根据志愿者18个月的手机使用情况,探索了参与者的社交关系,提供了一种社交网络的构建方法。

2.2 数据处理

(1) 数据清洗

数据清洗是数据处理的第一步。通信领域原始数据中存在大量空间不确定性数据和噪声数据,不能直接进行分析,必须进行数据清洗。对于异常数据,可以设定一定的规则,使用有效性验证的方法进行过滤。对于数据缺失,可以采取删除或者填充的方法进行处理。同时,数据的清洗还包括重复值删除以及数据类型检查等。以最常见的CDR数据为例,可以按照基站位置划分,将用户分配到特定位置,然后根据用户在特定时间段内停留位置的合理性进行数据清洗。

(2) 数据校准

基于通信数据的定位服务是通信领域可视化研究的一个重要方面,然而由于手机基站覆盖范围较大、分布不均匀,利用基站获取到的用户行为轨迹精度较低,需要进行数据校准。基站数据需要转换为轨迹

数据,按时间顺序聚集每个用户的移动电话记录,然后形成连续的旅行点。使用该方法计算定位点(即记录)之间的时间间隔和空间距离,从而利用时间阈值和空间阈值对定位点进行聚类。同时,也有许多研究者结合基站、网络、Wi-Fi、蓝牙、GPS等数据,利用多源数据融合的方法进行数据校准,从而提高轨迹数据的精度。

(3) 数据聚类

通信领域常用的CDR数据属于时间序列数据,不能直接聚类,需要进行特征提取。传统特征提取方法有:基于基本统计方法的特征提取、基于模型的特征提取、基于快速傅里叶变换(fast Fourier transform, FFT)的特征提取。深度学习是时间序列特征提取的新方向。Liu C F等人^[19]以手机数据为基础,融合特征提取与聚类分析的方法,对城市结构进行了分析。牛国庄^[20]使用模糊聚类算法进行用户行为分析。黄诗瑶^[21]使用模糊聚类算法研究用户群体的细分、移动性和群聚现象。

3 通信数据的城市对象可视化

移动通信数据表征用户在何时何地及与谁发生了联系,通信数据可视化可以从基础的数据类型和内在的数据对象两个层面刻画通信数据。在基础的数据类型层面,通信数据因包含通话时间和通话地点而同时具有典型的时空特征,可采用时间和空间类型的可视化方法(如采用时间线图、时间流图)表达时间,采用点、线、面地图表达空间信息,采用地理轨迹展示空间的移动性。参考文献[22-24]已分别从时间序列、地理空间和轨迹分析方面系统地综述了时空可视化的流程、技术和方法,因此本节着眼于通信数据的内在数据对象层面,即通信数据城市可视分析中关注的

人、通信设备、城市空间3个对象。下面根据城市可视分析的内在数据对象和数据特征对基于通信数据的城市可视化方法进行分门，并总结了常见的可视化方法，见表2。

3.1 城市用户可视化

人是城市生活的主体，深入理解城市空间中的人类行为是理解和优化城市运营的基础。在基于通信数据的城市可视分析中，很大一部分工作是围绕人类行为分析展开的。根据人在时间和空间上的通话行为习惯，对人的通话模式、用户角色、用户关系以及用户群体和社交网络等方面进行可视化。

(1) 用户通话模式可视化

用户的通话模式是通信数据分析的起点和基础。通话模式包括时间上的多尺度特征、空间分布、频繁模式以及多维通话特征的刻画。在通话模式时间特征可视化方面，用户在工作日/周末，以及一天内各小时的通话频率具有不同的热度，Shen Z Q等人^[25]采用两组日历图分别以“天”和“小时”为尺度，可视化了一天24小时以及一周不同工作日的通话热度分布，如图1(a)所示。考虑到用户通话时间分布的周期特性，另一种使用较多的可视化方法是环形可视化布局。蒋宏宇等人^[26]通过多时间窗口聚集环展示了一个月内的一天各个小时、一周各

天以及总时间线的通话分布，如图1(b)所示。在通话模式的空间分布可视化方面，一般采用空间节点的面积来编码事件发生的频率，采用节点间的连接展示地点之间的联系情况。向峰^[12]采用在地图上叠加圆形节点和连接的方式可视化一个用户的空间通话模式，可直观地展示用户通话的频繁空间地点及地点之间的起迄点(origin-destination, OD)关联强度，如图1(c)所示。此外，根据研究目标的不同，研究者一般会基于通话记录提取不同的通话特征，如社交方面的联系人数量、主被叫比率、通话平均时长，出行方面的移动距离、回旋半径、访问点个数、移动方向熵等，这使得通话模式具有了典型的多维数据特征。对于多维通话特征的可视化，常用的方式是雷达图和平行坐标图。Pu J S等人^[27]使用平行坐标直观地展示了用户的在线状态、时间、呼叫次数、呼叫比例等，如图1(d)所示，平行坐标多维通话模式展示也支持在各个数据轴的过滤操作。

(2) 城市用户画像可视化

出于对用户隐私的保护，研究者面对的通信数据一般已经经过匿名化处理。而对于精准用户营销、用户安全分析和城市规划等细分领域的研究，需要对城市用户或者用户群体进行画像。目前针对移动通信数据的用户画像的研究相对较少。已有的研究主要根据用户的通话特征以及频繁

表2 城市对象可视化方法

可视化对象	可视化焦点	数据特征	可视化方法
城市用户	用户通话模式	时间、空间、数量	折线图、日历图、径向图、地图、平行坐标、雷达图
	城市用户画像	时间、空间、关系	甘特图、线图、词云、ego画像
	城市用户关系	关系、数量	甘特图、节点-连接图、矩阵图、星形图、ego网络、社会网络图
通信设备	基站模式	时间、空间、数量	柱状图、像素矩阵、地图
	基站覆盖范围	空间、关系	维诺图、饼图
城市空间	城市空间热度	空间、数量	热力地图、渐变色地图
	城市区域划分	空间、时间、数量	六边形分隔地图、正方形分隔地图、维诺图、节点-连接图
	城市区域动态	时间、空间、轨迹	时空立方体

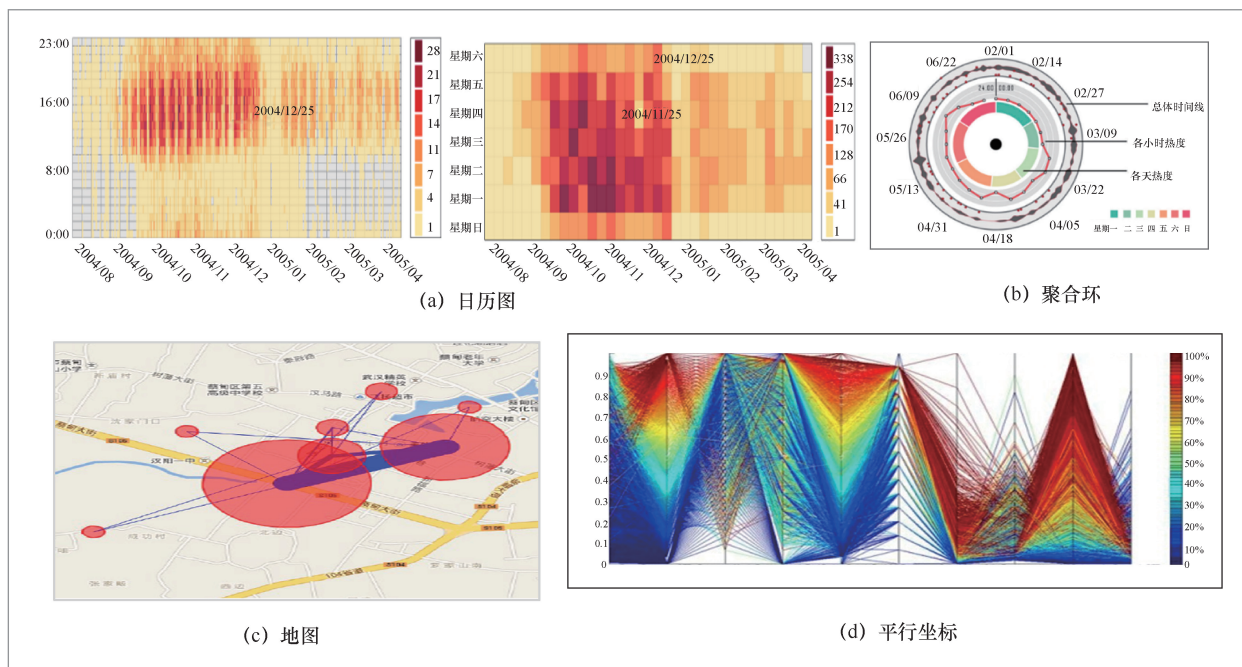


图1 通话模式可视化

出现的位置进行学习,为用户添加相应的标签,进而对用户进行分类,并大致推断用户的社会角色。

Jiang H Y等人^[28]使用甘特图展示用户的社会角色。如图2(a)所示,在甘特图中,横坐标表示日期,纵坐标表示地点,根据时间和空间的特点对用户的社会角色进行识别。此外,还能使用甘特图展示用户间的关系。王峰等人^[29]基于与通信类似的基于位置服务(location based service, LBS)的微博数据,根据用户ID、签到时间(check-in time)和GPS坐标等提取用户的移动规律,从而建立用户角色与城市地域结构的互推断模型。如图2(b)所示,城市用户角色用不同的色点表示,色点上的不同符号表示用户在工作或休息,连线表示不同角色用户的活动轨迹。张海旭等人^[30]构建了移动距离、回旋半径、访问点个数、移动方向熵、通话时长、主叫比率社交熵等通信特征词库,然后采用词云的方式对通信用户进行画像。图2(c)为一个随机用户

的画像,可以看出该用户移动距离大、访问地点多,同时通话时间长、联系人比较多。基于此可以推测用户可能是在较大城市区域内从事联系交流工作的室外工作者。Wang Q等人^[31]使用散点图来展示用户自我中心网络(ego network)的统计特征,并为用户的自我中心网络设计了一种新型的视图来展示群体用户画像,如图2(d)所示。

(3) 城市用户关系可视化

用户关系分析是用户分析的重要组成部分,用户关系的识别可以帮助相关人员认知用户的社交模式和社交规模。用户关系研究包括两两用户关系研究、一对多用户关系研究和多对多用户关系研究。

在两两用户关系可视化方面,Slingsby A等人^[18]采用可缩放的时间线可视化用户和用户关系,如图3(a)所示。每行编码一个用户,第一列用不同色块描述用户的性别、年龄和社交活跃度,右侧的列编码不同时间的通话数据,红色的垂直方向的弧线编码了用户64和用户123之间的通话关系。此

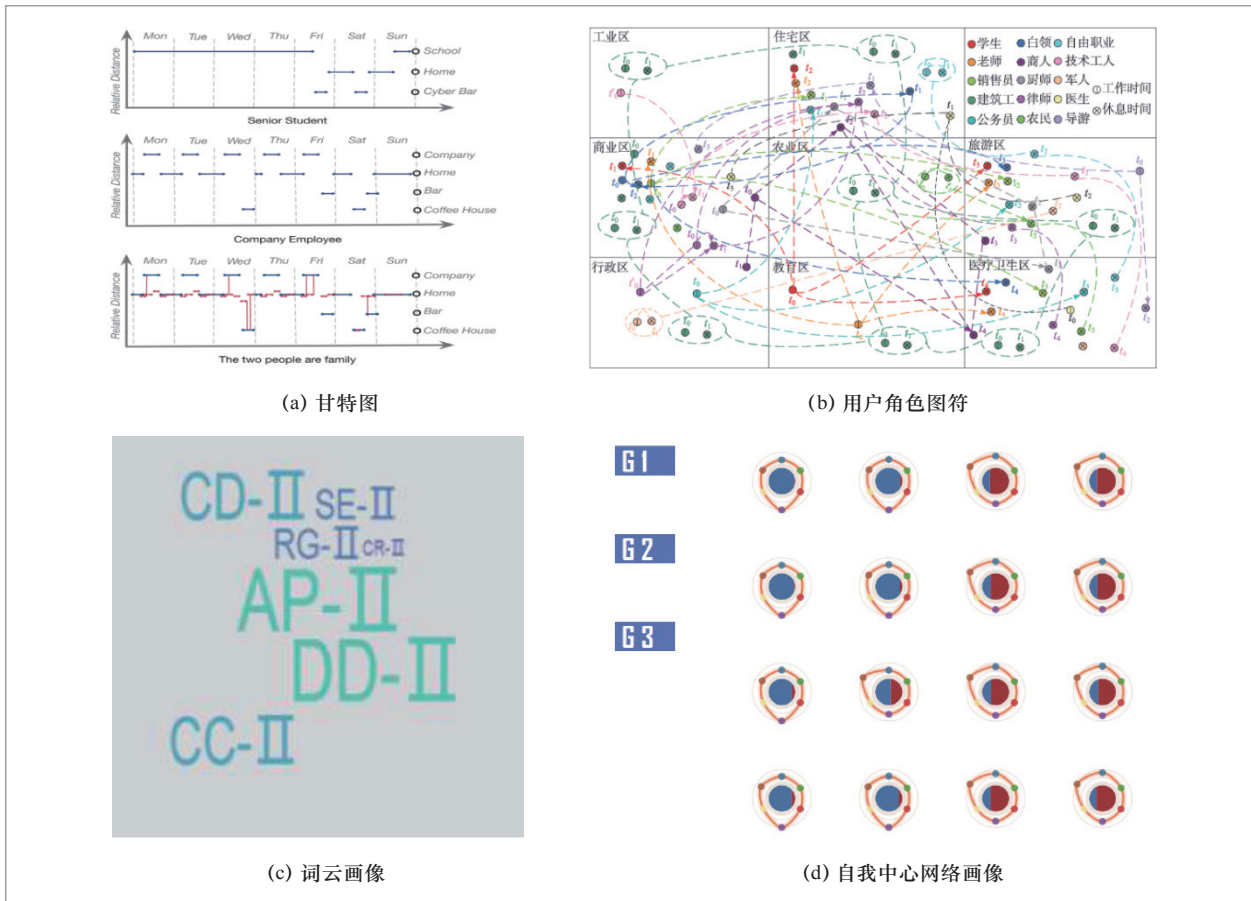


图2 用户画像

可视化设计可以直观地展示这两个用户间的通话时间分布和频度。Riegler V等人^[32]根据美学标准和定量用户实验进行了一项通话记录可视化的评估,研究采用了如图3(b)所示的可视化设计,该设计用于显示在26个用户中的24条两两连接关系。横坐标轴标识的是数,顶部是通话ID,底部是按照降序排列的通话强度,在每个通话旁边增加了通话双方的名字。

当用户研究的侧重点是特定个体的关系时,用户关系的视角转变为一对多的用户关系研究。Golban O等人^[33]使用加权的星形图表示个体用户的社交网络关系,其仅展示以单一用户为中心的网络,节点间的距离编码了用户之间的亲密度,如图3(c)所示。Han M等人^[34]采用复杂网络中的自我中心

网络概念描述选定用户与其他用户间的一对多关系,如图3(d)所示,中心节点表示分析人员关注的自我(ego),周边节点表示邻居(alter),周边节点的颜色编码其他节点的方向,包括出节点、入节点和双向节点,边编码了联系的强度。

当用户关系研究转向整个网络时,需要展示多对多的用户关系。根据用户关系的复杂程度,常用的方式有节点-连接图和矩阵图。节点-连接图是社交关系最直观的表现方式,节点表示关系中的实体(即城市用户),连接表示实体之间的联系(即用户联系),用节点间距离、节点的大小、连接的宽度等视觉通道编码用户关系的其他属性,如用户之间的亲密度等。Shen Z Q等人^[25]采用节点-连接图可视化了周六晚上通

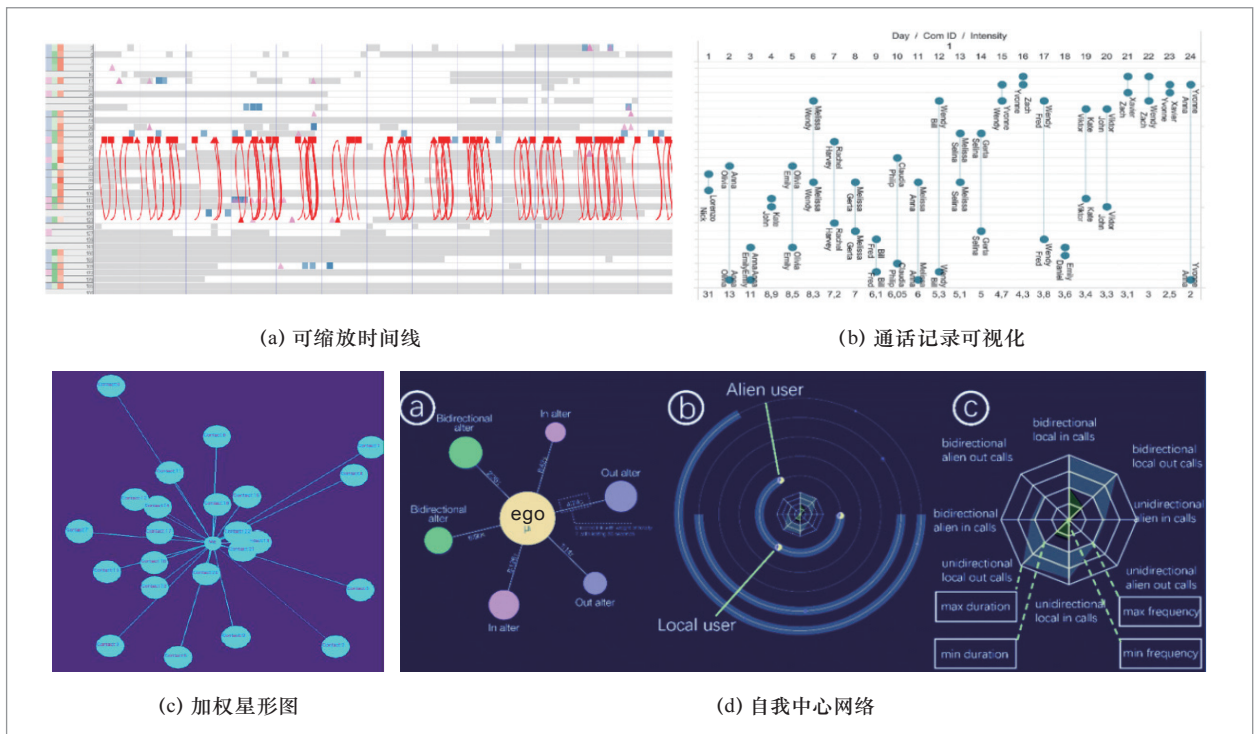


图 3 用户关系可视化

信用户的朋友关系,如图4(a)所示。此外,除了节点-连接图,矩阵也是表示关系的常用方式之一^[32]。使用矩阵的方式表示关系数据可以解决节点-连接图中连边交叉而产生视觉混乱的问题,但是由于受到网格分辨率的限制,矩阵图可视化方法仅适用网络规模较小的情况。Slingsby A等人^[18]

采用矩阵图的形式表示社区中人与人之间的联系,如图4(b)所示。然而,随着社交网络规模的变大,网络中的节点和连边越来越多,网络结构也越来越复杂,一种较新的方式是采用大规模网络图布局算法展示用户的社团关系,图4(c)为CDR通话社会网络的一个子网^[35]。为了更好地展示网

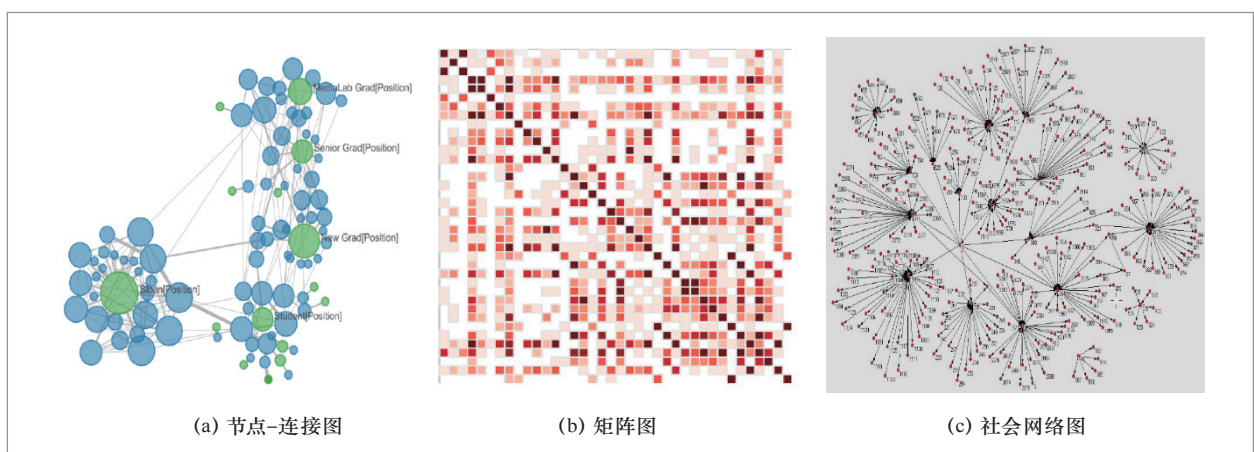


图 4 多用户关系可视化

络的结构, 学者们针对社交网络设计了众多网络布局算法^[36-40]。

3.2 城市通信基站可视化

用户通过基站设备及其后端的无线网络系统使用无线通信服务, 作为离用户最近的通信终端, 基站被用来指示通信应用中用户的大致位置, 同时也是通信系统中的重要一环。因此, 在基于通信数据的城市可视分析中, 存在较多的以基站为中心的研究。

对基站的可视分析是对基站自身利用率和关联用户情况等工作模式的可视化。基站因其所处位置不同、服务群体不同而体现出不同的工作模式, 如基站的热度、基站的时间动态趋势、基站的用户熵等。这些分析类似第3.1节中的用户可视化, 本节不再展开。唐楷等人^[41]使用柱状图来展示不同伪基站的时间特性, 并使用热力地图的方式展示伪基站的停留情况。蒲誉文等人^[42]设计了雷达玫瑰图, 并结合地图、热力图对位置基站发送的垃圾短信进行分析,

以发现垃圾短信的分布情况及伪基站的移动轨迹情况。Andrienko G等人^[8]使用像素矩阵来编码基站的呼叫次数, 将基站天线的呼叫次数映射为像素点的颜色, 通过像素矩阵直观地展示不同基站的呼叫模式, 如图5(a)所示。同时, 将像素矩阵映射到地图上, 以实现对不同基站模式间的对比。

同时, 作为无线通信系统中的独立个体, 基站具有设备自身独有的特征。因此, 基站可视分析的另一方面是对基站特有特征的可视化, 如基站的覆盖范围分析。为了展示基站的覆盖范围, 学者们多采用维诺图(Voronoi diagram)。维诺图又称泰森多边形, 其根据平面内的多个点对空间平面进行多边形剖分, 特点是多边形内的任何位置离该多边形内样点的距离小于离相邻多边形内样点的距离, 且每个多边形内包含且仅包含一个样点。维诺图很切合基站的特性, 将基站所处的位置点作为区域的几何中心建立维诺图, 基站所处的几何区间即该基站覆盖的范围。Mai W M等人^[9]采用维诺图对广州市的基站进行了可视化, 如图5(b)所示, 其中黑色实线勾画

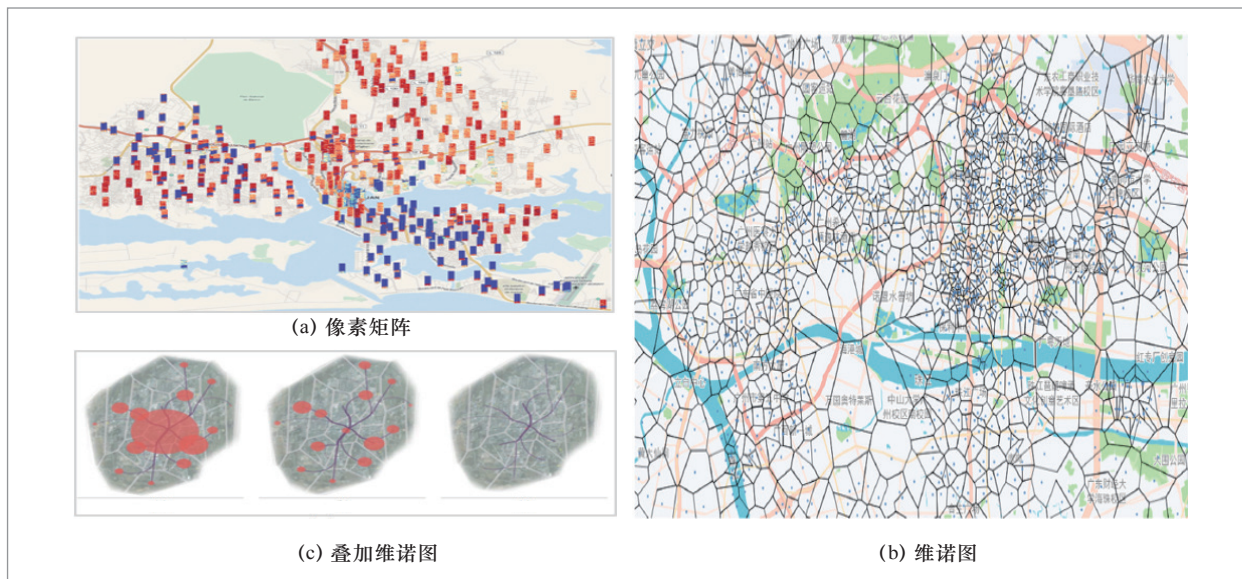


图5 通信基站可视化

的多边形表示基站的覆盖区域,多边形内的蓝色点表示基站的位置,可以看出在城市边缘和区域交界处存在覆盖范围偏大的区域。维诺图也可以作为基站分析的基础图层,研究人员可在其上展开进一步的分析。Pu J S等人^[27]在维诺图的基础上分析了3个用户组的不同移动模式。如图5(c)所示,白色的多边形编码了基站的通话总量,红色圆形的半径编码了用户组在该基站下的通话量。

3.3 城市空间可视化

在现实生活中,经常可以看到相同的土地覆被上承载了不同的人群活动,外形相近的建筑可能承担不同的社会功能。通信大数据为城市空间分析提供了基于全样本的微观数据研究的可能,从人类活动的角度更加全面有效地解读城市空间和土地利用情况^[12]。城市空间可视化主要包括空间热度、空间区域划分、空间功能和空间动态。

(1) 城市空间热度可视化

城市空间热度主要用于体现城市中的人群分布及人群密度^[43]。结合地图和热力

图,可以直观地展示地理空间中不同地点的人群密集情况。Li Z H等人^[13]使用热力图展示人群密度,并通过对比不同时间的热力图发现人群移动规律。Jo A等人^[44]基于手机数据,通过热力图展现不同性别用户在不同地区的活跃程度,从而为城市规划中的性别平等提供实证依据,图6(a)展示了女性人口与男性人口的差值(SMW)的分布热度。与热力图相似,在参考文献[45-46]中,研究者将地图分区,并使用渐变色来对每个分区进行着色,以显示不同区域电信用户的密度情况,图6(b)展示了葡萄牙里斯本(Lisbon)周边与海滨度假圣地卡帕里卡海岸(Costa da Caparica)附近的人口分布渐变情况^[46],其中不同颜色反映了主要假期和工作期的人口密度相对差值(relative difference)。Zhong G等人^[47]基于基站数据在地图上标注点,体现空间内的人群活跃度。

(2) 城市区域划分可视化

城市区域划分的目的是简化城市中的大规模时空数据,将研究对象从数量庞大的具体个体转化为经过划分的聚集区域。同时,根据区域的时空特征自动发现区域的范围和功能配置也是城市分析的研究方向之一。基于通信数据的城市区域划分主

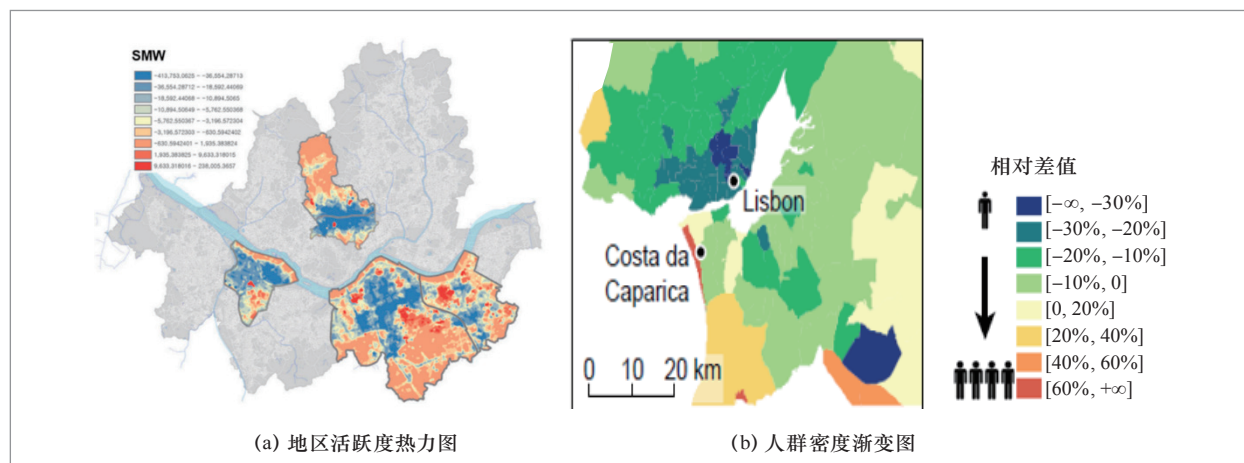


图6 人群分布可视化

要有城市边界发现、城市行政区域发现和城市功能区发现3个方面,相应的可视化方式主要是不同形式的地图分割和按区域差别化着色,也存在采用图论的方式展示城市区域信息的方法。向峰^[12]根据区域内用户的通话特征自动发现了城市的行政分区,并将不规则形状的分区用不同的颜色编码,如图7(a)所示。Wu W C等人^[48]为了展示城市不同移动模式的区域,对地图以六边形为单位进行分割,并用不同的颜色编码不同的分区类型,如图7(b)所示。Chen L B等人^[16]通过分析基站业务模式对区域进行划分,同样使用不同颜色对不同类型的区域进行编码,以方便用户概览城市的分区情况。参考文献[49]根据用户活动将城市区域功能分为吃饭、购物、休闲等类型,然后将城市地图以正方形网格为单位对分区进行着色,如图7(c)所示。向峰^[12]将城市区域抽象为节点,将区域间的联系抽象为边,然后用节点-连接图的方式表示不同分区,以及分区内部和分区之间的联系,如图7(d)所示。

(3) 城市区域动态可视化

区域动态可视化使用户能够从多时间尺度感知城市区域的动态变化。与城市区域动态相关的维度包括时间、空间和人,根据分析目标的不同,可能需要基于不同组合查看城市动态,如固定区域,查看人

群随时间的变化动态;固定两个时间点,查看人群在不同区域之间的移动性;或者同时查看空间、时间和人的变化。

在表示区域随时间的动态变化方面,采用较多的是不同时间的地图快照,或者基于时间的地图动画。Shi L等人^[50]对北京的通话密度的动态变化进行了对比,如图8(a)所示,采用基于时间的6个地图分别展示6个时段的通话密度变化情况。地图左上角的圆形表示钟表,橙色的扇形区域表示对应的时段。可以发现,在早上和深夜,城市区域间的差别不大;但是在下班后的时段,不同区域间却出现了很高的多样性,这可能是因为市民下班后有一些娱乐活动。另一种相对具有更高视觉冲击力的区域动态表达方式是变形地图(cartogram),变形地图按照各区域单元属性值的区际比例调整每个区域单元的几何面积,同时保持各个区域单元的空间邻接关系。Graells-Garrido E等人^[51]采用变形地图的可视化方式展示城市内区域人口总量的动态情况。如图8(b)所示,绿色表示居住区,黄色表示工作区,左右两幅地图分别可视化了早晨和中午的城市人口总量分布,可以看出,早上居住区占据了较大的空间,而中午工作区大幅度扩张,居住区被压缩到很小的空间。

在固定两个时间点,可视化人群在不

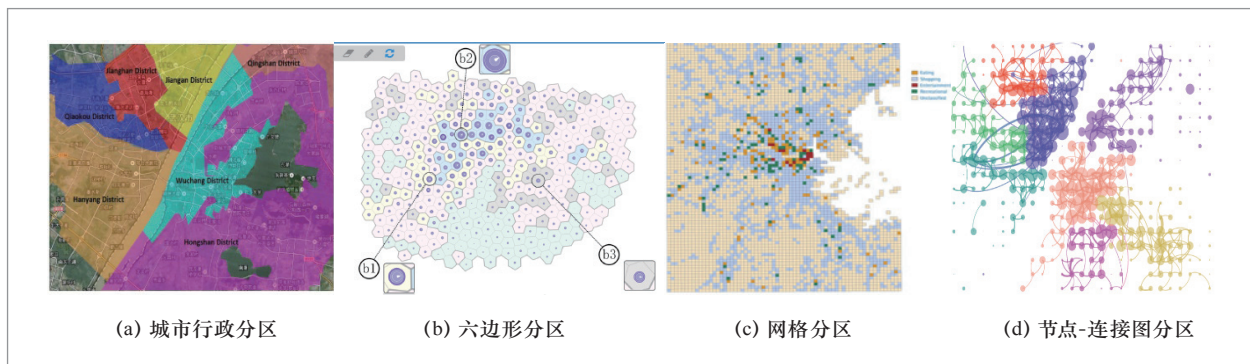


图7 区域划分

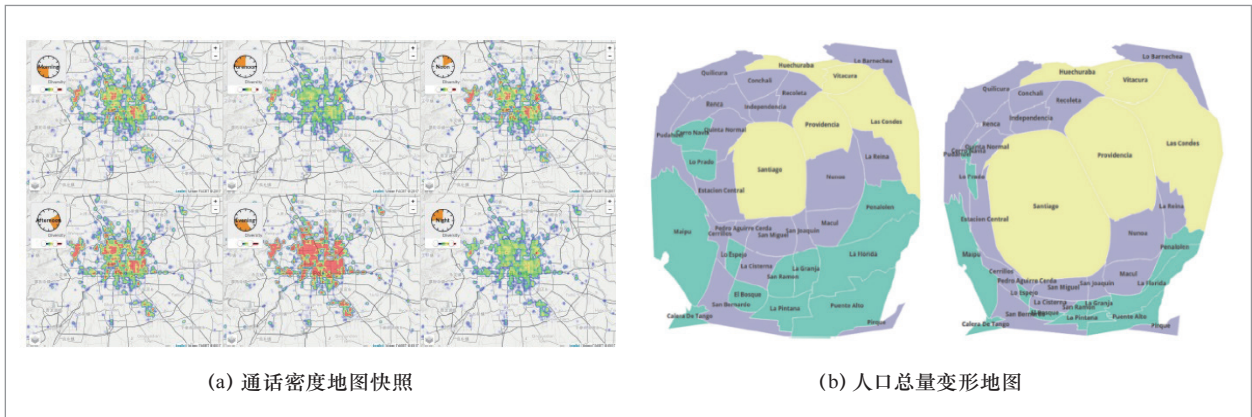


图 8 区域动态可视化

同区域之间的移动性方面, 可采用的方式较为丰富。首先, 可以将人的通话记录看作一个时空序列, 如果两个相邻的时间点之间发生了位置改变, 那么就构成了一条起点到终点的OD轨迹。人口移动性可视化主要是对用户OD轨迹的展示, 主要包括基于地图的直接OD可视化、聚合OD可视化和矢量场OD可视化等。Lorenzo G D等人^[1]

采用直接可视化的方式展示了8:00到9:00的6万条OD流, 如图9(a)所示。对于像通信这样大规模的数据来说, 直接OD可视化存在视觉遮挡、混乱, 且绘制效率较低的问题。这时需要采取进一步的方法优化轨迹可视化设计, 如根据线路的方向和距离的聚合线路的边绑定算法、基于拓扑重构的聚合方法。参考文献[52]根据轨迹的

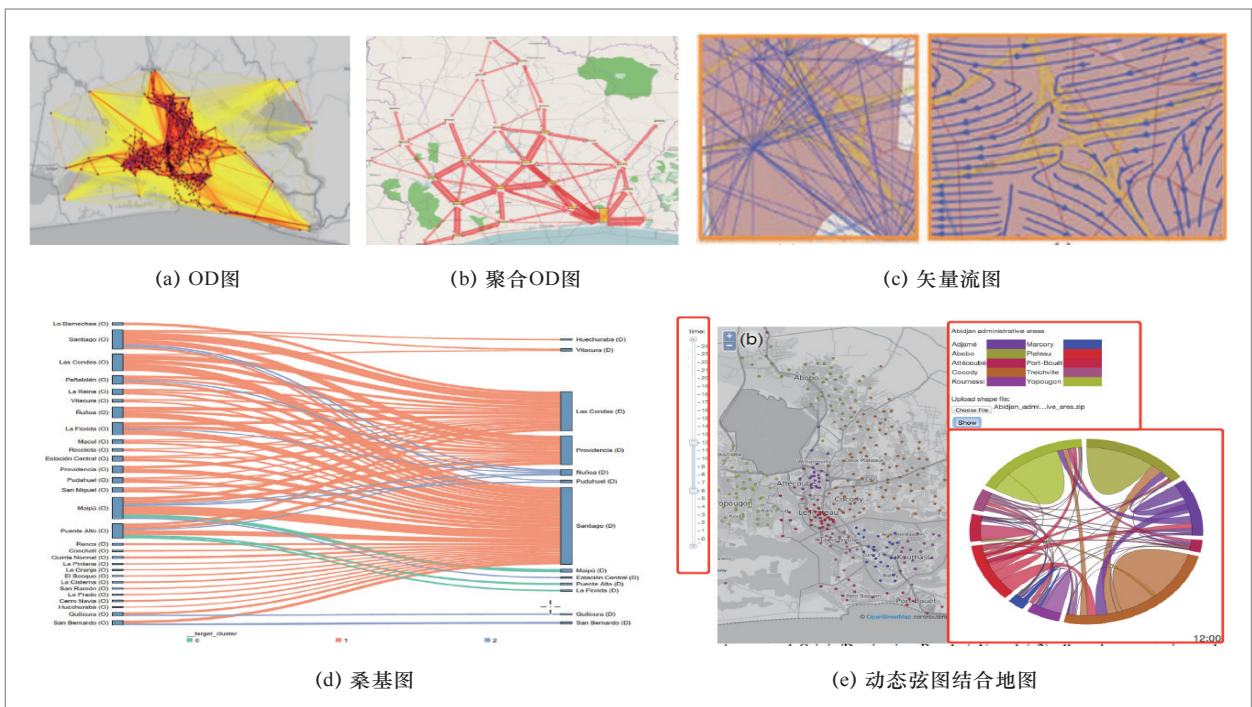


图 9 人群移动性可视化

距离对相邻的轨迹进行聚类,以提高大规模轨迹的可读性。Andrienko G等人^[8]采用了基于区域聚合的方式,如图9(b)所示,用户的轨迹被汇总为28个聚合区域之间的迁移线,区域间的迁移强度用线的宽度表示,区域内的用户数用区域中心的黄色条形表示。另外,也有一些基于矢量流的移动性表示方法。Wang F等人^[14]将城市通信中的人群移动表示为矢量场,并提出了一种根据相邻区域的主要矢量重构新的人造矢量的方式,使用该方式优化轨迹呈现,以矢量流图的形式进行可视化,使得大规模人群的移动直观易读,而且减少了视觉混乱和遮挡。如图9(c)所示,左侧是一个立交桥原始OD的可视化,右侧为优化后的矢量流图可视化,矢量流图将细节表达得更清晰。此外,除了基于地图的OD可视化,如果更关心区域间的流量转移情况,也可以采用流图、桑基图或者弦图的方式可视化区域之间的移动性。Graells-Garrido E等人^[51]采用桑基图(如图9(d)所示)来展示区域之间的人流情况,突出展示了流量分

布和相互关系,但是缺失区域的地理坐标。Lorenzo G D等人^[1]采用动态弦图结合地图的方式展示人群的移动,如图9(e)所示。通过在弦图和地图中采用同样的颜色编码,读者既可以从弦图中观察区域间的迁移分布情况,又可以在地图中根据颜色直观地查看区域的地理位置。

城市动态研究有时需要同时观察时间和空间两个方面的变化,这时需要采用相对复杂的时空可视化方法来展示数据,时空动态可视化的主流方法是时空立方体。Kang C G等人^[53]采用3D时空立方体的可视化方法描述个人移动模式,如图10(a)所示,水平面为地图平面,竖轴表示周一到周日的的时间范围,中间的轨迹线描述了一个人一周的活动路径。可以看出,在工作日,用户行为在时间和空间上都具有很高的规律性,而在周末具有明显的不同。Senaratne H等人^[54]对时空立方体可视化进行了扩展,以处理因天线信号等技术问题导致的部分移动位置缺失引起的用户路径不确定性,他们将时空棱镜(如图10(b)所示)作

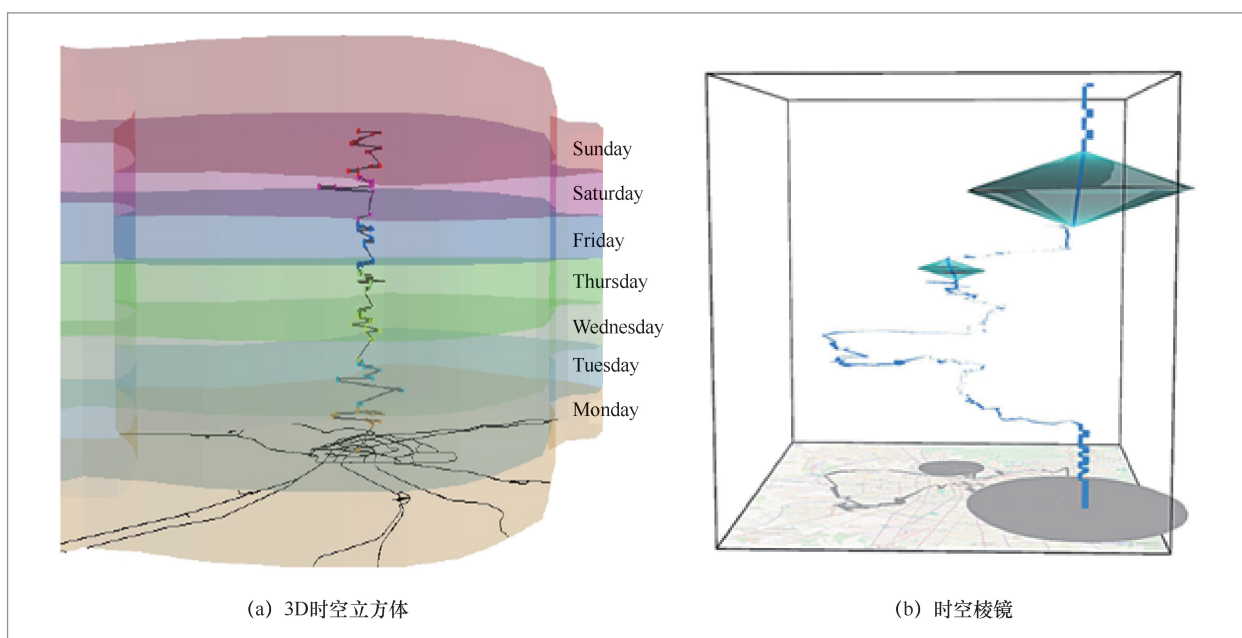


图10 时空立方体

为位置不确定性的标识图符,可视化了某个选定用户一周的行动路径,两个时空棱镜表示两个不确定路段的可能区域范围。

4 通信数据的城市可视分析

通信数据具有时间、空间和用户关系等多方面的信息维度,在时序性上,蕴含用户行为、事件序列等信息;在空间上,具有地理空间数据信息。许多学者基于通信数据分析城市态势,Calabrese F等人^[55]总结了城市感知的常见任务,包括评估人口分布、评估城市不同地区的活动类型、评估人群移动模式、分析本地事件以及分析社交网络的地理位置。本节将从城市语义感知、城市预测与异常分析、城市资源优化3个角度对这些研究进行梳理,并介绍其中的代表性工作,见表3。

4.1 城市语义感知

城市是由人类行为决定和演化的生命系统。移动通信网络的深度普及和大规模通信数据的产生提供了一种从人类真实行为角度感知和探索城市语义的通道。移动

数据特有的基于位置的服务和实时动态反馈也给城市动态感知带来了前所未有的时间尺度细节。

(1) 城市人的语义

人是城市生活的主体。城市语义感知的一大主题是关于人类语义的感知,人类语义包括人的社会角色发现、人群画像、行为语义、用户关系感知等。

- 用户角色和关系语义: Jiang H Y等人^[28]根据用户的时间和空间分布特征推测用户的角色,分析出学生、上班族等用户角色,然后结合用户相似度和地理空间距离推测用户的关系,发现家庭关系和同事关系。张海旭等人^[30]提取了用户通话在日常移动模式和社交生活方面的特征,采用K-means算法聚类,然后勾勒用户画像。向峰^[12]将用户的通话联系抽象为无向图,将用户间的亲密度映射为边的权重,然后基于用户关系网络拓扑和地理重合度来分析和预测用户之间关系。Zhang H Q等人^[35]将用户通话看作随机过程,采用亲密度度量用户之间的相似性,可以发现社会团体,以及短期、中期和长期的朋友关系。Fan X等人^[56]根据智能手机的使用记录分析用户个人生活信息,根据事件种类和详细的事件序列的多视图协调交互分析和层次

表3 通信数据城市可视分析的典型工作

城市可视分析类别	可视分析任务	分析方法	代表性工作
城市语义感知	城市人的语义感知	统计方法、多视图协同、相似性分析、相关性分析、群体状态立方体	[18][28][42][46][50][57]
	城市空间语义感知	非负矩阵分解(nonnegative matrix factorization, NMF)、词嵌入、相似性计算、统计方法、聚类分析	[13][19][48]
	城市动态语义感知	时空相似性计算、POI向量、t-SNE降维、统计方法、蓝噪声采样、词嵌入、3D时空路径建模、哈希、聚类	[14][17][27][53][58]
城市预测与异常分析	通信设备异常检测	统计方法、聚类、假设检验	[6][42][59-60]
	社会群体异常检测	社区发现算法、空间概率密度、分类算法、ego网络	[28][34][61]
	社会事件异常检测	聚类、相似性计算、预测算法、纯可视分析	[5][7][62]
城市资源优化	基站和通信设施优化	MuLSTM预测、聚类、统计方法	[6][16][63]
	城市交通设施优化	K-means、DBSCAN、OD流聚合、统计方法	[1][9][64]
	城市配套资源优化	统计方法、机器学习SVM、决策树	[65-67]

关联,可以发现用户的不同风格和爱好。Slingsby A等人^[18]基于移动通话的log,从时间和空间上探索用户社会关系网络,用户界面如图11(a)所示,采用可缩放的地图A和时间线B查看用户概览,然后通过筛选和缩放选定感兴趣的区域和人, D区域采用矩阵图展示用户之间的通话频度。

- 人类行为语义: Phithakkitnukoon S等人^[49]以人类活动模式感知分析为目标,提出了活动感知地图(activity-aware map)。通过100万条手机记录数据提取并捕捉个人日常活动模式,结合城市POI数据进行设计,基于活动感知地图描述空间区域最相关的活动,分析不同人或人群和区域之间的相关性,对海量的人类活动赋予吃饭、购物、娱乐等语义。Lu M M等人^[68]基于用户手机App的使用记录,发掘用户日常生活中的个人习惯、兴趣和行为模式。Alexander L等人^[69]首先根据手机用户频繁访问地点的到达时间和停留时间等推测停留地点的语义,如家、工作地或者其他,然后根据停留地点语义推测人的活动语义。

- 人类群体和宏观语义: Deville P等人^[46]基于移动通信数据进行国家和不同行政

级别区域内的动态人口研究,推测区域内一天不同时间、一周不同日期,以及季节性的人口动态变化。Shen Z Q等人^[25]在异构网络中呈现社会和空间信息的理念,使用“行为环”表示个人或群体的行为模式。Wu W C等人^[70]采用双簇可视化等新颖的可视化设计,帮助专家深入探索城市用户移动的并发性。Zheng Y等人^[57]通过用户移动模式感知人群移动模式,提出了一个全面的可视分析系统TelcoFlow。系统采用基于状态的行为模型等高级定量分析和嵌入状态符号的扩展流视图等可视化设计,能够支持对群体行为的高效和深入分析。系统用户界面如图11(b)所示,其中视图A是多侧面过滤视图,流视图B提供了基于状态模型的群体行为概览,径向分布图C和径向距离图D展示了一个状态簇的细节并支持比较分析,序列图E可视化了提取的代表不同群体行为的状态簇序列,分析师可在此视图查找感兴趣的目标,并在流图F中进一步查看。

(2) 城市空间语义感知

根据其所承载的功能和配置的不同,城市空间具有不同的语义。单独从城市土地覆盖和建筑外形难以准确推测其社会功能。利

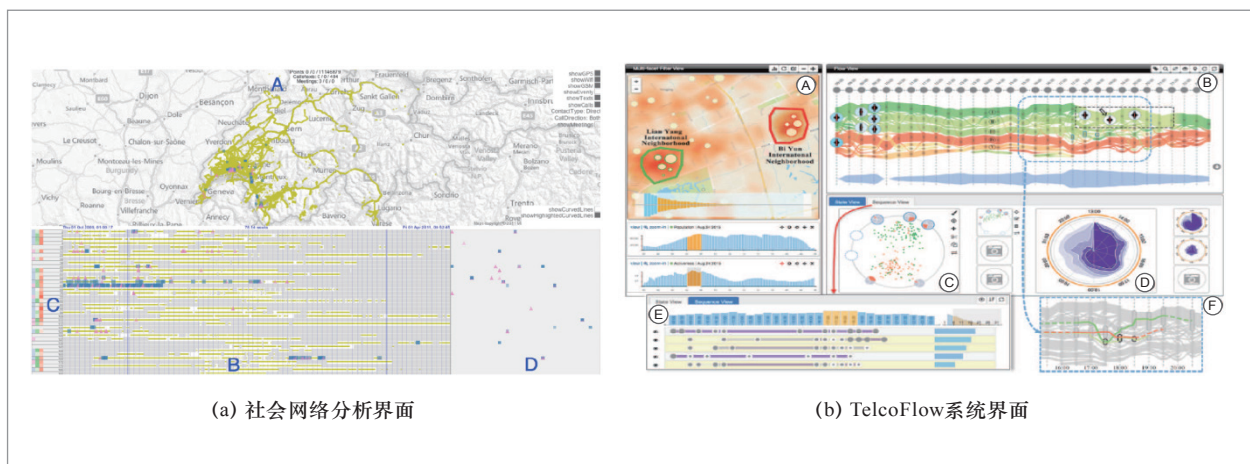


图 11 城市人的语义可视化

用通信大数据提取人类活动分布特征,能从人群活动的角度更为全面有效地解读城市土地的利用情况,感知城市空间语义。

Wu W C等人^[48]提出了一种交互式视觉分析系统MobiSeg,通过将城市区域划分为共享相似活动模式的区域,对人们的活动进行探索。联合通信、出租车和地铁射频识别(radio frequency identification, RFID)3种数据,通过非负矩阵分解和有监督聚类分析人们的移动活动模式,将市区划分为不同的活动区域。系统的用户接口如图12(a)所示,全局地图视图A提供整体移动概览,并支持空间和时间过滤;观察视图B展示了分区结果及区域内活动视图;细节视图C可视化了移动特征向量的时间序列,以进行深入的分析 and 对比;活动模式视图D便于对提取的潜在活动模式进行解读。

Li Z H等人^[13]设计了Trajectory2Vec可视分析系统,系统界面如图12(b)所示。根据词嵌入模型,将基站的时空信息映射为向量,通过计算基站间的高层语义的相似规律来分析地理区域的功能性信息;再将带有时空变化信息的手机用户移动轨迹映射至向量空间,使基站地理

坐标与轨迹结合,从而获取更加丰富的语义信息。

Liu C F等人^[19]通过运动模式分析获得指定市区的功能。采用基于深度学习的移动电话通信数据,使用聚类分析来探索城市的结构。从图像处理中获得灵感,并构建通信快照图以表示每个区域。使用深度学习方法提取特征后,使用无监督学习查找城市的相似区域。

(3) 城市动态语义感知

交通、流行病学调查和灾害防备等需要在更细粒度上实时感知城市区域的动态变化规律,人口普查数据仅能反映城市行政区划粒度的静态分布数据。手机是感知用户移动轨迹的可靠设备^[71],基于人对通信网络的使用情况建立城市动态感知模型,可支撑运营商和城市管理者多侧面探索城市区域的时间动态、区域间人类的移动性和城市的时空动态。

Shi L等人^[50]设计了一套多侧面的城市动态可视概览分析系统UrbanFACET,用户界面如图13(a)所示。基于城市居民的手机App签到数据,可发现并可视化多个城市及其居民的动态情况。他们通过提出一套新的基于信息理论的度量方法来表征

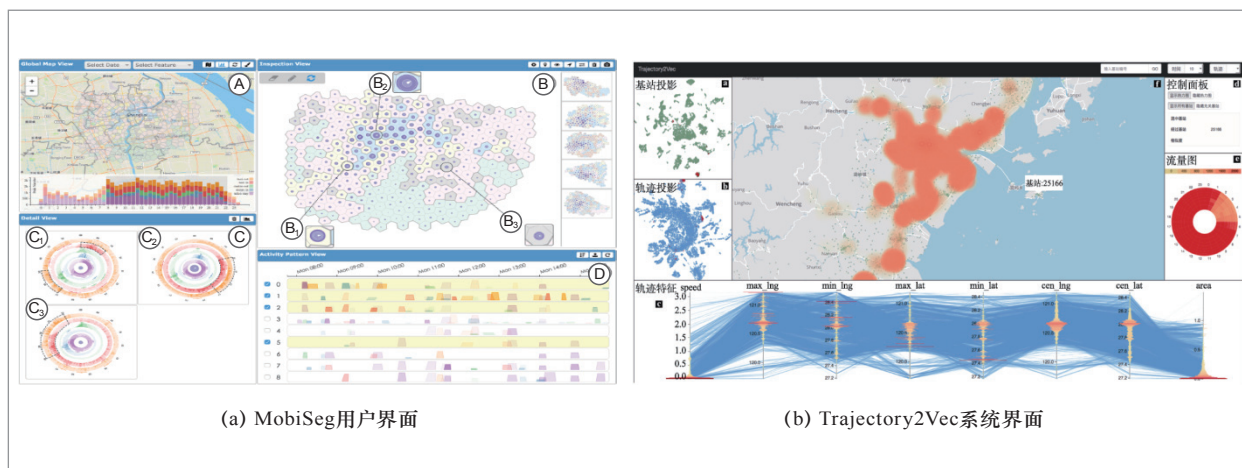


图12 城市空间语义感知可视化

城市区域和居民群体的流动模式,从而分类和显示城市的隐藏功能,发现城市居民和城市功能间的关系和影响。度量方法包含的指标包括流动性、活力、交换、多样性和密度等,并能进一步可视分析和比较不同地区城市指标的异同。

Kang C G等人^[53]通过使用中国大城市的数百万条原始手机通话记录,计算一天中不同时间和一周中不同日期的手机使用情况的统计特征,从而得出数百万手机用户个人的汇总移动性模式和不同群体的流动模式。Senaratne H等人^[54]提出了一种用于全球移动通信系统(global system for mobile communications, GSM)移动网络数据探索和搜索的可视化分析方法。该方法定义了交互导航数据的地理空间和矩阵表示,将数据可视化与合适的数据分析算法集成在一起,允许在空间和时间上比较手机的使用情况,识别规律,以及发现跨越区域和用户组的用户移动模式的异常。Zhou Z G等人^[58]基于OD流与自然语言处理(natural language processing, NPL)项的类比,建立了OD流的特征,设计了一种迭代的多元目标采样方案,可在向量化的表示空间中选择OD流,并设计了一组

有意义的可视编码来表示OD流的交互。

Wang F等人^[14]设计了一套自适应的时空表示方法来描述城市人口移动,系统用户界面如图13(b)所示,包括时间组件视图A、流视图B、控制面板和交互工具C、堆栈时序图D。用户可以在不同的时间尺度上交互动地检索和探索感兴趣区域中人口流动的轨迹。他们将人口移动表示为矢量场,以流图的形式呈现,使得大规模人群的移动直观易读,而且减少了视觉混乱和遮挡。

4.2 城市预测与异常分析

移动电话产生的大规模数据提供了一个群体移动性的显微镜,用来观测人类群体的隐藏模式,发现群体异常行为和城市生活中的异常事件^[7]。基于通信数据的城市异常发现为建设更安全的城市提供了强有力的参考。基于通信数据的城市异常分析主要包括通信基站异常检测、社群异常检测和社会事件异常检测。

(1) 通信基站异常检测

通信基站异常检测的一方面是发现基站不正常的工作模式。Lee W H等人^[60]结合时空数据挖掘和地理信息系统(geographic

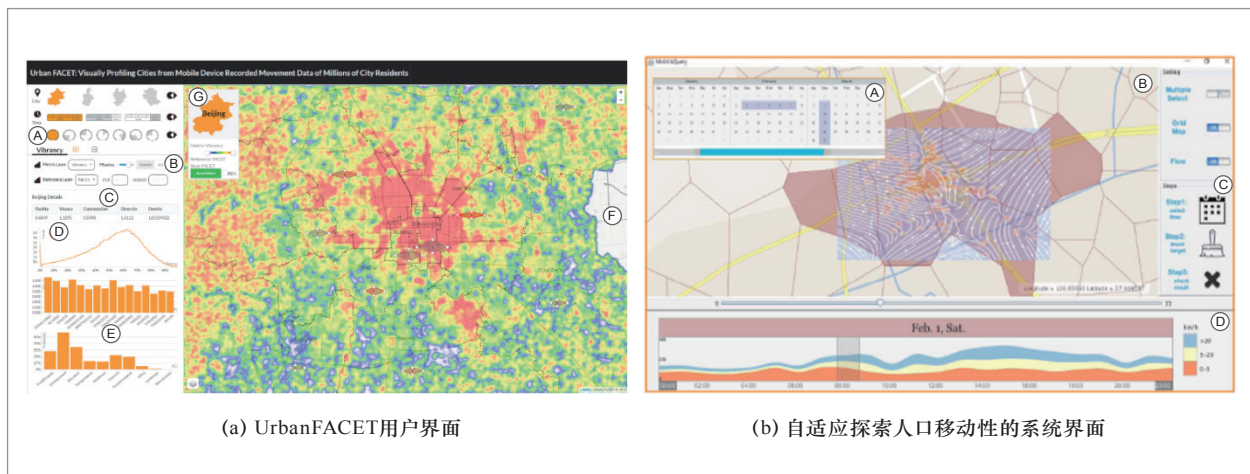


图13 城市动态语义可视化

information system, GIS) 技术, 整合GIS道路网络和基站数据库, 提出了一种新颖的基站信号覆盖发现模型SCHADIM。该模型能够在移动网络中发现信号覆盖漏洞区和信号覆盖弱区。Jiang H Y等人^[6]根据基站流量的时间分布模型, 可以发现发生了宕机的基站。

通信基站异常检测的另一个方面是发现伪基站。伪基站又称假基站, 是非法份子为牟取利益而非法改装的基站设备。伪基站启动后会干扰和屏蔽一定范围内的运营商信号, 之后则会搜索附近的手机号, 并将短信发送到这些手机号上。伪基站既影响用户的正常通信, 又对用户财产和信息安全构成了威胁。然而由于伪基站一般被装载在车辆上, 具有一定的流动性, 伪基站的检测和定位具有一定的难度。蒲誉文等人^[42]提出了基于多用户垃圾短信数据进行伪基站活动轨迹可视分析的方法。通过用户上传垃圾短信的时间、相对位置、内容等信息, 追溯伪基站的近似活动轨迹, 并通过设计多种可视化视图, 实现了一个多视图组合的交互式可视分析系统。徐江阳等人^[59]提出了基于恶意短信数据进行伪基站行为可视分析的方法, 通过设计基于地图的热力视图、散点视图、日历热力视图、时间堆

栈视图, 开发可视分析系统FBSVA, 从而探求伪基站的行为规律。

(2) 社交群体异常检测

通信网络中具有异常行为的用户群体是社会的潜在威胁。从用户社交的角度出发, 识别异常的社交关系和社交群体, 并通过可视化技术提供上下文和交互分析支持, 有助于安全部门高效识别、对比和探索潜在的危险用户。

Traag V A等人^[61]构建了一个贝叶斯位置推理框架, 能够从大规模通信数据中发现异常的大规模聚集事件, 以及进一步推断出谁参与了事件。Jiang H Y等人^[28]从通信网络用户社交和时空特征的角度出发, 提出了一种社团和用户行为可视分析模型, 以支持安全部门识别异常用户事件和异常群体活动。该系统能够识别大规模社交群体, 并识别出其中的关键成员。用户界面如图14(a)所示, 左上为社团的活动地图和时间分布趋势, 右侧是社团弦图, 左下是展示了社团Top-N用户和他们之间的相互联系的甘特图。

Han M等人^[34]基于图模型, 从ego网络的角度交互式地探索移动用户的通信行为, 提出了一个异常通信行为分析系统egoStellar。egoStellar用于全面地探索移动

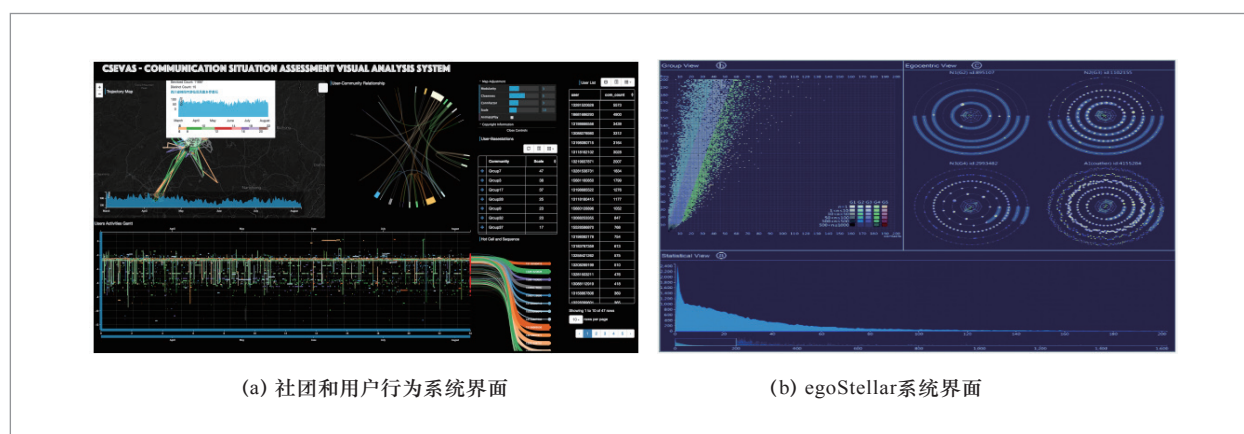


图 14 社交群体异常检测

用户的行为分布情况,系统界面如图14(b)所示,左侧用户组视图用于对用户进行分类并提取特征,以进行异常检测和比较;右侧ego网络视图展示了以自我为中心的视图,用于显示详细的自我与接触者之间的交互分布。

(3) 社会事件异常检测

通信群体的“异常现象”可以提供对城市环境功能配置的更多见解,从而提高城市管理者对城市事件识别的时空意识。

Dong Y X等人^[7]提供了一套能够从低时空分辨率的CDR数据中检测异常事件的通用框架和原型系统。该框架通过定义柱状簇(cylindrical cluster)解决CDR数据时空稀疏性的问题,通过定义近距离人群的概念刻画群体聚集性,通过柱状簇、近距离人群和用户移动性定义异常事件发现

算法,最终形成了一个通用的异常发现框架和原型系统。该系统允许城市管理人员等分析人群的形成和演变,并研究不同参数对所得结果的影响,根据分析目标启发性地提出可能的变化,以获得更有意义的结果。系统的用户接口如图15(a)所示,主要由两部分组成:上部视图a为城市的地图概况,下部视图b为用户、人群和事件的统计数据。该系统可以基于CDR数据高效地检测出多种异常社会事件。

与采用统计或者机器学习方法相区别,Sagl G等人^[5]刻意仅采用直接可视映射的方式发现通信数据中人类群体移动的时空特征和异常时空模式,并建立了异常模式与真实世界的关联性。作者发现群体的整体移动模式具有令人惊奇的相似性和对称性,他们提出的模型和原型系统能鉴

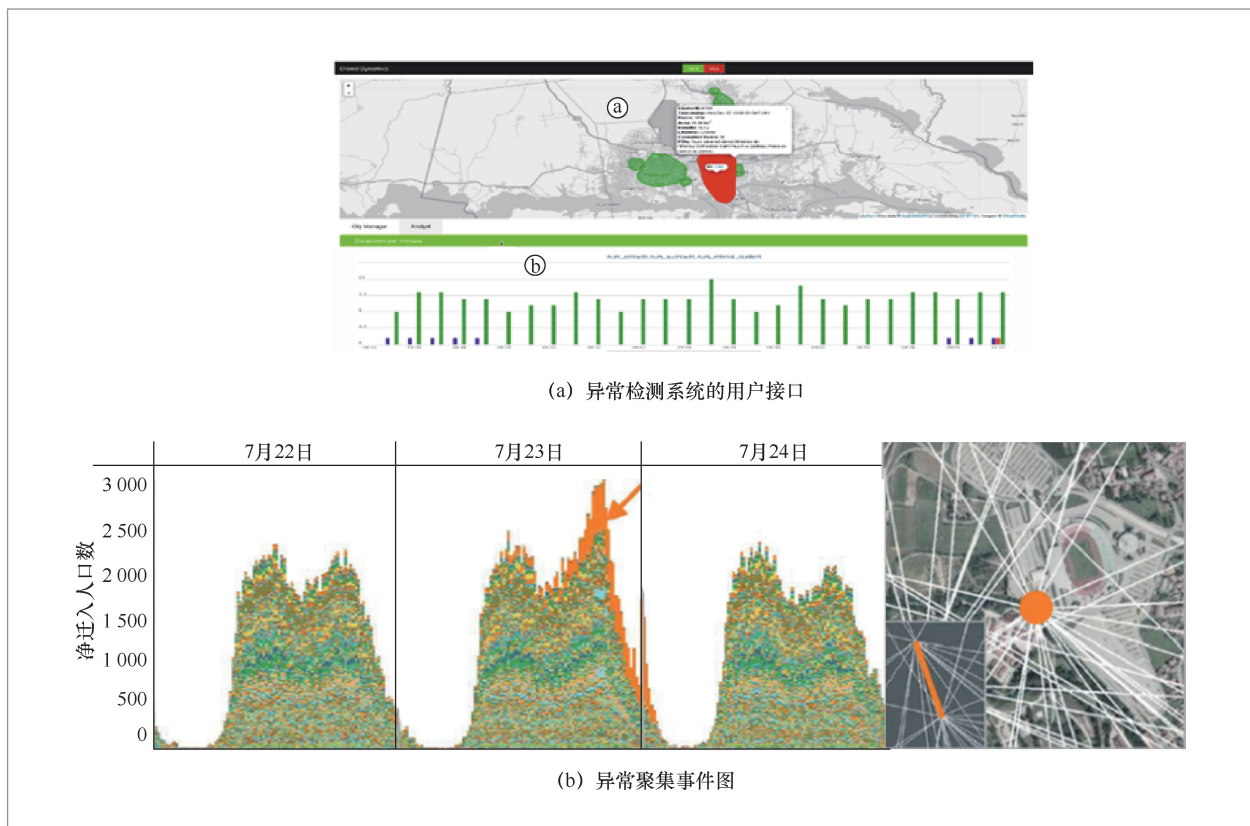


图 15 城市异常事件检测

别出多种城市异常事件,如演唱会、球赛和游行等。如图15(b)所示,左侧3个图显示了选定区域7月22日—24日3天的净迁入人口随时间变化的堆叠柱状图,右侧为一个关联地区的航拍图,其中7月23日出现了一个显著的橙色峰值,进一步分析得知该时段举办了一场演唱会。

Zhu Q等人^[62]基于CDR数据设计了一种时间序列分析与机器学习相结合的异常提取方法。这种方法可以将具有相似流量模式的区域单元进行聚类,并将一个城市划分为不同的群体。然后,在分组区域中,作者使用聚类技术检测蜂窝网络的异常行为,并使用从在线来源收集到的地面真实信息来验证结果的准确性。结果表明,在特定的地点和时间,用户行为异常与活动量的突然增加有关。

4.3 城市资源优化

城市的快速发展使城市结构分析变得复杂。使用访谈和调查等传统方法很难学习城市生态系统。大规模通信数据提供了一个从多尺度时空粒度观测人与城市基础设施互动情况的窗口,这有助于分析人员更深入地理解城市基础设施的利用情况,进而为城市设计与规划提供有效的参考。

(1) 基站和通信设施优化

随着移动电话用户的增加,电信供应商需要建设和调整基站,以达到更高的地理区域覆盖率和更好的通话信号质量,因此对基站规划的合理优化变得迫切。

Chen L B等人^[16]提出了一种新颖的将在时间上流量互补的基站聚合到同一个基带处理单元(baseband unit, BBU)的聚合方案,以优化无线接入网的容量利用率和部署成本。作者利用多元长短时记忆(multivariate long short-term memory, MuLSTM)模型学习基站流量

模式的时间依赖性和空间关联性,并对未来一段时间内的流量做出准确的预测。根据基站的业务量,建立一个加权图模拟基站间的互补性,把住宅区、商业区和交通枢纽等在时间上互补的基站聚集到一起,并连接到同一个BBU,从而实现带宽共享,提高BBU的利用率。

Ghahramani M等人^[63]提出了一种探索性的空间数据分析算法,对不同区域进行优先排序,快速准确地检测城市热点。研究结果可为城市规划与发展及电信基础设施升级提供参考。Jiang H Y等人^[6]设计了名为Aureole的可视化分析系统来感知基站的时空分布和利用率。系统采用环形构图理论,让用户在不丢失上下文信息的情况下专注于感兴趣的区域。Aureole的用户界面如图16所示,视图A是选定基站簇的全局信息,视图B是基站在一周的每天,以及一天内的通话时间分布,视图C通过一个扇区模型展示了感兴趣基站的聚合时间动态。用户通过该系统对基站进行逐层深入的探索分析,高效定位过载基站、空闲基站及其时间上的详细流量模式。

(2) 城市交通设施优化

交通设施和市民日常生活息息相关,合理的交通设施规划和优化对市民的日常工作起着非常重要的作用。通信和交通具有一定的相关性,可以通过分析城市用户的通话数据来感知市民的移动模式和动态,进而揭示交通设施的可用情况。

Lorenzo G D等人^[1]提出了基于手机数据的智能分析系统AllAboard,帮助城市规划者在视觉上探索城市流动性和优化公共交通。作者将通信数据和公交网络关联,允许公交运营商直观地探索城市用户在空间和时间上的出行需求,以非常精细的尺度评估公交网络的服务质量,并对优化方案提供对比支持。AllAboard用户界面如图17



图 16 Aureole 用户界面

所示, 视图A是OD过滤组件和交互组件控制面板; 视图B基于地图展示OD流和公交线路图, 流量的颜色表示流量的大小; 视图C是OD流的时空比较图, 用时间顺序的圆圈表示每个移动性时间序列。

为了挖掘广州市的交通特征, Mai W M等人^[9]从出行时间估计和城区重点线路分析两个方面进行研究。他们提出了一种基于交通语义属性划分交通区域的城市出行时间估计方法, 利用CDR数据提取用户轨

迹, 利用DBSCAN算法描述广州市的关键路线。为了研究城市内的旅客出行行为, 在划分交通分区的过程中引入了交通语义属性, 对城市中任意两个位置之间的出行时间进行了估计, 使分区之间和分区内部的特征更符合实际交通状况。该方法为低成本估算城市出行时间提供了思路。此外, 利用从CDR数据中提取的单个轨迹, 采用DBSCAN轨迹聚类方法对广州市重点线路进行聚类分析。这些路线可以通过城市

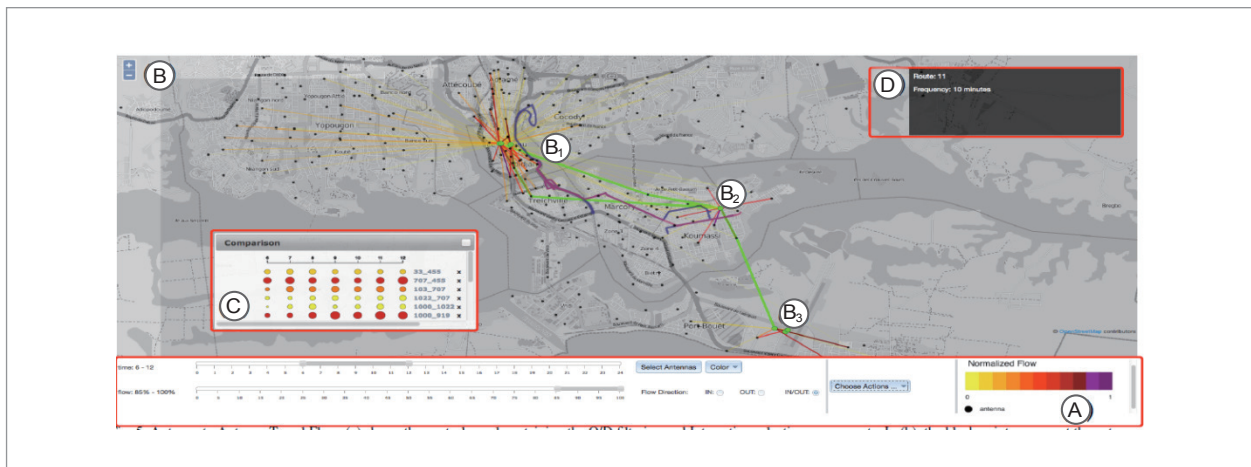


图 17 AllAboard 用户界面图

的实际情况进行验证,为城市道路规划和城市交通的进一步研究提供一些新的发现和直观的参考。

Zhong G等人^[64]提出了一个新颖的系统性方法,用于从通信数据中获取城市用户的行程信息。作者以一个大型的交通枢纽为研究中心,将通信用户分为市内行程和市间行程,设计了用于提取行程出发/到达时间的方法,分析了行程的时间分布、空间热度和相关性,能够为城市规划者和交通研究者提供有效的参考。

(3) 城市服务机构优化

人口因素是城市服务规划的首要参考因素,传统的人口调查是一个以行政边界为空间粒度的静态人数统计。而在现实世界中,人口是动态流动的,而且即使分布在很小的区域内也可能有明显的变化。移动电话基于位置服务的特性为城市规划者提供了一个更真实和更细的时空粒度的人口观察窗口,基于移动通话的人流分析为城市服务机构选址和更好的现场管理提供了较为可靠的依据。

翟书颖等人^[65]根据CDR数据推断基站周边的人流量特征,结合交通和房价等多源数据,建立了针对连锁企业选址的位置推荐模型和规模推荐模型。邓轲等人^[66]根据CDR数据确定地图格栅范围内的基站信息,提取区域内居住人数、工作人数和流动人数,然后结合交通状况、用户价值和竞争网点的数量,建立银行网点选址推荐模型。Krisp J M等人^[67]以提高消防和城市救援服务效率为目标研究通信数据。作者研究随着时间的推移人口分布与人口密度热点之间的关系,以有效地选择消防、救援和其他城市保护机构的位置。同时,移动电话数据也可以提供事故地点的真实人口密度,从而帮助救援工作人员更好地做应急准备。

5 结束语

采用可视分析技术解读移动通信数据蕴含的大规模人类关系拓扑和时空动态分布,对数据驱动的多尺度城市感知具有重要意义。本文总结了通信数据的主要来源、特点和处理方法,归纳了通信数据城市可视分析中的人、通信设备、城市空间3种内在对象相应的可视化方法,并对近年来基于通信数据的城市可视分析的代表性工作进行了梳理。尽管基于通信数据的城市可视分析已经取得了较为丰硕的成果,然而未来面对日益增长的城市智能分析需求,依旧存在诸多的挑战和机遇,具体如下。

一是如何有效保护用户的隐私。通信数据涉及详细的个人活动轨迹和社交关系等高度敏感的信息。研究采用的数据大多进行了匿名化处理,但是依旧存在被反匿名化的可能。有研究将用户的真实地理位置进行偏移处理,也有研究将用户呼叫记录中的空间、时间信息作为输入,创造一种合成的数据来保护真实话单数据中的隐私。但是,这方面的研究相对稀缺,未来如何有效保护用户隐私是一个重要的研究方向。

二是如何融合多源数据进行相互补充和印证。移动电话数据虽然具有较好的用户和时空特性,但是也存在轨迹记录稀疏、基于基站定位的空间精度不高以及缺乏明确语义信息等问题。如果能够结合高精度的GPS数据,或者具有语义的微博、Twitter社交网络数据等数据源,各数据互相补充印证,能够获得更加立体的城市感知。未来,大规模多时空数据的融合、关联和纠错技术将是一个研究难点。

三是如何高效处理大规模通信数据。

当前绝大部分基于移动通信数据的可视分析是基于离线的统计或者聚合进行处理的,而城市灾备等高时效性的服务需要更快的数据处理能力。有研究引入了Hadoop分布式架构、并行处理算法或者优化的可视交互手段^[15],但是此领域的研究总量相对较少,分析效率提升是未来的一个研究方向。

四是如何设计领域特定的可视化方案,以降低用户门槛。由于城市分析目标和通信数据的复杂性,城市可视分析的用户门槛相对较高,如何沉淀已有的研究成果,结合人工智能和领域知识,设计一套更加自动化的面向城市分析的可视分析系统也具有较强的应用价值和意义。

参考文献:

- [1] LORENZO G D, SBODIO M, CALABRESE F, et al. AllAboard: visual exploration of cellphone mobility data to optimise public transport[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2016, 22(2): 1036–1050.
- [2] International Telecommunication Union. Measuring digital development: facts and figures 2020[Z]. 2020.
- [3] ZHENG Y X, WU W C, CHEN Y Z, et al. Visual analytics in urban computing: an overview[J]. *IEEE Transactions on Big Data*, 2016, 2(3): 276–296.
- [4] 张平文, 鄂维南, 袁晓如, 等. 大数据分析与应用技术创新平台[J]. *大数据*, 2018, 4(4): 86–93.
ZHANG P W, E W N, YUAN X R, et al. Big data analysis and application technology innovation platform[J]. *Big Data Research*, 2018, 4(4): 86–93.
- [5] SAGL G, LOIDL M, BEINAT E, et al. A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic[J]. *International Journal of Geo-Information*, 2012, 1(3): 256–271.
- [6] JIANG H Y, TANG K, ZHAO W X, et al. Aureole: a multi-perspective visual analytics approach for green cellular networks[J]. *Journal of Visualization*, 2018, 21(3): 485–494.
- [7] DONG Y X, PINELLI F, GKOUFAS Y, et al. Inferring unusual crowd events from mobile phone call detail records[C]// *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Heidelberg: Springer, 2015: 474–492.
- [8] ANDRIENKO G, ANDRIENKO N, FUCHS G. Multi-perspective analysis of mobile phone call data records: a visual analytics approach[Z]. 2015.
- [9] MAI W M, XIE S H, CHEN X. Travel time estimation and urban key routes analysis based on call detail records data: a case study of Guangzhou City[C]// *International Conference on Machine Learning and Intelligent Communications*. Heidelberg: Springer, 2019: 659–676.
- [10] 李艳妮, 敖成凤, 吴亚东, 等. 基于移动通讯数据的用户行为可视分析[J]. *西南科技大学学报*, 2017, 32(4): 60–67.
LI Y N, AO C F, WU Y D, et al. Visual analysis of user behavior based on the mobile phone data[J]. *Journal of Southwest University of Science and Technology*, 2017, 32(4): 60–67.
- [11] 黄文彬, 徐山川, 马龙, 等. 利用通信数据的移动用户行为分析[J]. *现代图书情报技术*, 2015(5): 80–87.
HUANG W B, XU S C, MA L, et al. Analysis of mobile user behaviors with telecommunication data[J]. *New Technology of Library and Information Service*, 2015(5): 80–87.
- [12] 向峰. 基于移动网络数据的用户行为与城市感知研究[D]. 武汉: 华中科技大学, 2014.
XIANG F. Research on behavior analysis and urban sensing based on mobile network data[D]. Wuhan: Huazhong University of Science and Technology, 2014.

- [13] LI Z H, ZHU M F, HUANG Z S, et al. Trajectory2Vec: a visual analytics approach for urban mobility patterns based on mobile phone data[J]. *Journal of Computer-Aided Design and Computer Graphics*, 2018, 30(1): 68.
- [14] WANG F, CHEN W, ZHAO Y, et al. Adaptively exploring population mobility patterns in flow visualization[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2017, 18(8): 2250–2259.
- [15] ZHANG S H, YIN D D, ZHANG Y Q, et al. Computing on base station behavior using Erlang measurement and call detail record[J]. *IEEE Transactions on Emerging Topics in Computing*, 2017, 3(3): 444–453.
- [16] CHEN L B, YANG D Q, ZHANG D Q, et al. Deep mobile traffic forecast and complementary base station clustering for C-RAN optimization[J]. *Journal of Network and Computer Applications*, 2018, 121(11): 59–69.
- [17] YADAV K, KUMAR A, BHARATI A, et al. Characterizing mobility patterns of people in developing countries using their mobile phone data[C]// *The 6th International Conference on Communication Systems and Networks*. Piscataway: IEEE Press, 2014: 1–8.
- [18] SLINGSBY A, BEECHAM R, WOOD J. Visual analysis of social networks in space and time using smartphone logs[M]. Amsterdam: Elsevier Science Publishers B. V., 2013.
- [19] LIU C F, LIU C, LIU F Q, et al. Clustering analysis of urban fabric detection based on mobile traffic data[J]. *Journal of Physics: Conference Series*, 2020, 1453(1): 012158.
- [20] 牛国庄. 基于用户行为的无线网络功能扩展方法研究[D]. 保定: 华北电力大学, 2014.
- NIU G Z. Research on wireless network communication function extension methods based on the analysis of user behavior[D]. Baoding: North China Electric Power University, 2014.
- [21] 黄诗瑶. 聚类分析在移动通信用户行为分析中的研究与应用[D]. 广州: 广东工业大学, 2013.
- HUANG S Y. The research and application of clustering analysis in mobile user behavior analysis[D]. Guangzhou: Guangdong University of Technology, 2013.
- [22] 彭燕妮, 樊晓平, 赵颖, 等. 时间事件序列数据可视化综述[J]. *计算机辅助设计与图形学学报*, 2019(10): 1698–1710.
- PENG Y N, FAN X P, ZHAO Y, et al. A survey of visualization of temporal event sequence data[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2019(10): 1698–1710.
- [23] 周志光, 石晨, 史林松, 等. 地理空间数据可视分析综述[J]. *计算机辅助设计与图形学学报*, 2018, 30(5): 747–763.
- ZHOU Z G, SHI C, SHI L S, et al. A survey on the visual analytics of geospatial data[J]. *Computer-Aided Design & Computer Graphics*, 2018, 30(5): 747–763.
- [24] 王祖超, 袁晓如. 轨迹数据可视分析研究[J]. *计算机辅助设计与图形学学报*, 2015(1): 9–25.
- WANG Z C, YUAN X R. Visual analysis of trajectory data[J]. *Computer-Aided Design & Computer Graphics*, 2015(1): 9–25.
- [25] SHEN Z Q, MA K L. MobiVis: a visualization system for exploring mobile data[C]// *2008 IEEE Pacific Visualization Symposium*. Piscataway: IEEE Press, 2008.
- [26] 蒋宏宇, 吴亚东, 赵韦鑫, 等. 面向用户行为理解的移动通讯数据可视分析[J]. *浙江大学学报(理学版)*, 2018, 45(1): 40–46.
- JIANG H Y, WU Y D, ZHAO W X, et al. Mobile data visual analysis for human activity understanding[J]. *Journal of Zhejiang University(Science Edition)*, 2018, 45(1): 40–46.
- [27] PU J S, XU P P, QU H M, et al. Visual analysis of people's mobility pattern from mobile phone data[C]// *The 2011 International Symposium on Visual Information Communication*. New York:

- ACM Press, 2011: 1–10.
- [28] JIANG H Y, WU Y D, ZHANG Y, et al. From social community to spatio-temporal information: a new method for mobile data exploration[J]. *Journal of Visual Languages & Computing*, 2017, 41(8): 1–9.
- [29] 王峰, 钟宝荣, 余伟, 等. 基于用户角色与城市地域结构的互推断模型[J]. *计算机科学与探索*, 2016, 10(12): 1662–1672.
- WANG F, ZHONG B R, YU W, et al. Mutual inference model based on user roles and urban functional zones[J]. *Journal of Frontiers of Computer Science & Technology*, 2016, 10(12): 1662–1672.
- [30] 张海旭, 胡访宇, 赵家辉. 基于话单数据的移动通信用户画像研究[J]. *计算机系统应用*, 2018, 27(11): 271–277.
- ZHANG H X, HU F Y, ZHAO J H. Mobile communication user profiling based on call detail records[J]. *Computer Systems & Applications*, 2018, 27(11): 271–277.
- [31] WANG Q, PU J S, GUO Y F, et al. egoPortray: visual exploration of mobile communication signature from egocentric network perspective[C]// *International Conference on Multimedia Modeling*. Heidelberg: Springer, 2017: 649–661.
- [32] RIEGLER V, WANG L N, DOPPLER-HAIDER J, et al. Who talks to whom: an evaluation of a call log visualization[J]. *Journal of Visualization*, 2021, 24: 133–156.
- [33] GOLBAN O, MARCU M, FUICU S, et al. Solution for analyzing interpersonal relationships based on mobile communication history[C]// *2015 IEEE 10th Jubilee International Symposium on Applied Computational Intelligence and Informatics*. Piscataway: IEEE Press, 2015.
- [34] HAN M, WANG Q, WEI L R, et al. egoStellar: Visual Analysis of Anomalous Communication Behaviors from Egocentric Perspective[C]// *International Computer Symposium*. Heidelberg: Springer, 2019: 280–290.
- [35] ZHANG H Q, DANTU R. Discovery of social groups using call detail records[C]// *The OTM Confederated International Workshops and Posters on the Move to Meaningful Internet Systems*. Heidelberg: Springer, 2008: 489–498.
- [36] 赵润乾, 吴渝, 陈昕. 大规模社交网络社区发现及可视化算法[J]. *计算机辅助设计与图形学学报*, 2017, 29(2): 328–336.
- ZHAO R Q, WU Y, CHEN X. An algorithm for large-scale social network community detection and visualization[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2017, 29(2): 328–336.
- [37] 张政, 江南, 张俊, 等. 空间位置耦合的地理社交网络可视化布局算法[J]. *计算机辅助设计与图形学学报*, 2020, 32(6): 865–873.
- ZHANG Z, JIANG N, ZHANG J, et al. Geographical social network visualization layout algorithm based on spatial location coupling[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2020, 32(6): 865–873.
- [38] 任淑霞, 吴涛, 张书博. 嵌入社区半径的力引导与径向树混合布局算法[J]. *四川大学学报: 自然科学版*, 2020, 57(1): 73–81.
- REN S X, WU T, ZHANG S B. Community radius embedded in force-directed and radial tree hybrid layout algorithm[J]. *Journal of Sichuan University(Natural Science Edition)*, 2020, 57(1): 73–81.
- [39] 吴玲达, 张喜涛, 孟祥利. 基于社团结构节点重要性的网络可视化压缩布局[J]. *北京航空航天大学学报*, 2019, 45(12): 2423–2430.
- WU L D, ZHANG X T, MENG X L. Compression layout for network visualization based on node importance for community structure[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2019, 45(12): 2423–2430.
- [40] 张喜涛, 吴玲达, 于少波, 等. 面向多层网络可视化的多力引导节点自动布局算法[J]. *计算机辅助设计与图形学学报*, 2019, 31(4): 129–136.
- ZHANG X T, WU L D, YU S B, et al. A multi-force directed layout algorithm for multilayer networks visualization[J].

- Journal of Computer-Aided Design & Computer Graphics, 2019, 31(4): 129–136.
- [41] 唐楷, 赵韦鑫, 蒋宏宇, 等. 基于可视分析的伪基站活动特征分析方法[J]. 西南科技大学学报, 2018, 33(2): 72–78.
TANG K, ZHAO W X, JIANG H Y, et al. A visual analytic method of activity characteristics exploration for pseudo base station[J]. Journal of Southwest University of Science and Technology, 2018, 33(2): 72–78.
- [42] 蒲誉文, 胡海波, 何凌君. 基于多用户垃圾短信数据的伪基站活动轨迹可视分析方法[J]. 计算机应用, 2018, 38(4): 1207–1212.
PU Y W, HU H B, HE L J. Visual analytics on trajectory of pseudo base-stations based on SMS spam collected from mobilephone users[J]. Journal of Computer Applications, 2018, 38(4): 1207–1212.
- [43] CHEN W, HUANG Z, WU F, et al. VAUD: a visual analysis approach for exploring spatio-temporal urban data[J]. IEEE Transactions on Visualization & Computer Graphics, 2018(99): 2636–2648.
- [44] JO A, LEE S, KIM J. Gender gaps in the use of urban space in seoul: analyzing spatial patterns of temporary populations using mobile phone data[J]. Sustainability, 2020, 12(16): 6481.
- [45] CALABRESE F, DI LORENZO G, LIU L, et al. Estimating origin-destination flows using mobile phone location data[J]. IEEE Pervasive Computing, 2011, 10(4): 36–44.
- [46] DEVILLE P, LINAR C, MARTIN S, et al. Dynamic population mapping using mobile phone data[J]. Proceedings of the National Academy of Sciences. 2014, 111(45): 15888–15893.
- [47] ZHONG G, WAN X, ZHANG J, et al. Characterizing passenger flow for a transportation hub based on mobile phone data[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(6): 1507–1518.
- [48] WU W C, ZHENG Y X, CAO N, et al. MobiSeg: interactive region segmentation using heterogeneous mobility data[C]// 2017 IEEE Pacific Visualization Symposium. Piscataway: IEEE Press, 2017.
- [49] PHITHAKKITNUKON S, HORANONT T, LORENZO G D, et al. Activity-aware map: identifying human daily activity pattern using mobile phone data[C]// The 1st International Conference on Human Behavior Understanding. Heidelberg: Springer, 2010.
- [50] SHI L, JIANG T, ZHAO Y, et al. UrbanFACET: visually profiling cities from mobile device recorded movement data of millions of city residents[J]. arXiv preprint, 2017, arXiv:1707.04210.
- [51] GRAELLS-GARRIDO E, GARCÍA J. Visual exploration of urban dynamics using mobile data[C]// International Conference on Ubiquitous Computing and Ambient Intelligence. Heidelberg: Springer, 2015: 480–491.
- [52] 李海生, 黄媛洁, 宋璇, 等. 手机基站定位数据可视分析[J]. 大数据, 2017, 3(1): 61–71.
LI H S, HUANG Y J, SONG X, et al. Visual analyses of mobile phone base station location data[J]. Big Data Research, 2017, 3(1): 61–71.
- [53] KANG C G, GAO S, LIN X, et al. Analyzing and geo-visualizing individual human mobility patterns using mobile call records[C]// 2010 18th International Conference on Geoinformatics. Piscataway: IEEE Press, 2010.
- [54] SENARATNE H, MUELLER M, BEHRISCH M, et al. Urban mobility analysis with mobile network data: a visual analytics approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(99): 1537–1546.
- [55] CALABRESE F, FERRARI L, BLONDEL V D. Urban sensing using mobile phone network data: a survey of research[J]. ACM Computing Surveys, 2014, 47(2): 1–20.
- [56] FAN X, PENG Y, ZHAO Y, et al. A

- personal visual analytics on smartphone usage data[J]. *Journal of Visual Languages & Computing*, 2017, 41(8): 111–120.
- [57] ZHENG Y, WU W, ZENG H, et al. TelcoFlow: visual exploration of collective behaviors based on Telco data[C]// 2016 IEEE International Conference on Big Data. Piscataway: IEEE Press, 2016: 843–852.
- [58] ZHOU Z G, MENG L H, TANG C, et al. Visual abstraction of large scale geospatial origin–destination movement data[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2018, 25(1): 43–53.
- [59] 徐江阳, 张慧军, 张小龙. 基于恶意短信的伪基站行为可视分析[J]. *计算机工程与应用*, 2019, 55(18): 95–102.
- XU J Y, ZHANG H J, ZHANG X L. Visual analysis of pseudo base station behavior based on malicious SMS[J]. *Computer Engineering and Applications*, 2019, 55(18): 95–102.
- [60] LEE W H, TSENG S S, WANG C H, et al. Discovering the radio signal coverage hole and weak coverage area in mobile network by spatiotemporal data mining on location–based services[J]. *Fundamenta Informaticae*, 2010, 98(1): 33–47.
- [61] TRAAG V A, BROWET A, CALABRESE F, et al. Social event detection in massive mobile phone data using probabilistic location inference[C]// IEEE 3rd International Conference on Privacy, Security, Risk and Trust. Piscataway: IEEE Press, 2012.
- [62] ZHU Q, SUN L. Big data driven anomaly detection for cellular networks[J]. *IEEE Access*, 2020, 8: 31398–31408.
- [63] GHAHRAMANI M, ZHOU M C, HON C T. Mobile phone data analysis: a spatial exploration toward hotspot detection[J]. *IEEE Transactions on Automation Science and Engineering*, 2019, 16(1): 351–362.
- [64] ZHONG G, YIN T, ZHANG J, et al. Characteristics analysis for travel behavior of transportation hub passengers using mobile phone data[J]. *Transportation*, 2019, 46(5): 1713–1736.
- [65] 翟书颖, 李茹, 李婧, 等. 一种融合多源数据的连锁企业选址推荐方法[J]. *现代电子技术*, 2019, 42(22): 145–149.
- QU S Y, LI R, LI J, et al. A method for chain enterprise site selection recommendation fusing multi–source data[J]. *Modern Electronics Technique*, 2019, 42(22): 145–149.
- [66] 邓轲, 冯辉宗, 许国良, 等. 基于多源大数据融合的银行网点选址方法[J]. *重庆邮电大学学报(自然科学版)*, 2020, 32(4): 664–672.
- DENG K, FENG H Z, XU G L, et al. Site selection method of banking facility location based on multi–source big data fusion[J]. *Journal of Chongqing University of Posts and Telecommunications(Natural Science Edition)*, 2020, 32(4): 664–672.
- [67] KRISP J M. Planning fire and rescue services by visualizing mobile phone density[J]. *Journal of Urban Technology*, 2010, 17(1): 61–69.
- [68] LU M M, MENG D, PENG Y N, et al. A visual analytics approach for exploring individual behaviors in smartphone usage data[C]// 2016 IEEE Pacific Visualization Symposium. Piscataway: IEEE Press, 2016.
- [69] ALEXANDER L, JIANG S, MURGA M, et al. Origin–destination trips by purpose and time of day inferred from mobile phone data[J]. *Transportation Research Part C Emerging Technologies*, 2015, 58(9): 240–250.
- [70] WU W C, XU J Y, ZENG H P, et al. TelCoVis: visual exploration of co–occurrence in urban human mobility based on Telco data[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2016, 22(1): 935–944.
- [71] BLONDEL V D, DECUYPER A, KRINGS G. A survey of results on mobile phone datasets analysis[J]. *EPJ Data Science*, 2015, 4(1): 1–55.

作者简介



王桂娟 (1981-), 女, 西南科技大学信息工程学院博士生, 西南科技大学计算机科学与技术学院教师, 主要研究方向为可视化与可视化分析、自动可视化。



周锐 (1997-), 男, 西南科技大学计算机科学与技术学院硕士生, 主要研究方向为网络可视化。



蔡梦杰 (1994-), 男, 西南科技大学计算机科学与技术学院硕士生, 主要研究方向为城市计算可视分析。



汤勇 (1988-), 男, 西南科技大学信息工程学院博士生, 四川轻化工大学自动化与信息工程学院教师, 主要研究方向为可视化与可视分析。



李芊芊 (1996-), 女, 西南科技大学计算机科学与技术学院硕士生, 主要研究方向为城市计算可视分析。



陈华容 (1977-), 女, 西南科技大学信息工程学院博士生, 西南科技大学计算机科学与技术学院讲师, 主要研究方向为可视化与可视化分析。



吴亚东 (1979-), 男, 博士, 四川轻化工大学计算机科学与工程学院教授、博士生导师, 主要研究方向为可视化、可视分析、人机交互、虚拟现实。

收稿日期: 2021-01-11

通信作者: 吴亚东, wyd028@163.com

基金项目: 国家自然科学基金资助项目 (No.61872304); 四川轻化工大学人才引进项目 (No.2020RC20)

Foundation Items: The National Natural Science Foundation of China(No.61872304), Talent Introduction Project of Sichuan University of Science and Engineering(No.2020RC20)