引文格式:仝照民,刘耀林,张紫怡,等.骑行流密度聚类下的共享单车源汇空间识别[J].武汉大学学报(信息科学版),2025,50(1):184-196.DOI:10.13203/j.whugis20220467

Citation: TONG Zhaomin, LIU Yaolin, ZHANG Ziyi, et al. Bike-Sharing Source-Sink Space Recognition Based on Riding Flow Density Clustering Method[J]. Geomatics and Information Science of Wuhan University, 2025, 50(1): 184-196. DOI: 10.13203/j.whugis20220467

骑行流密度聚类下的共享单车源汇空间识别

仝照民[1] 刘耀林[1,2,3,4] 张紫怡[5] 安 睿[1] 朱 奕[6]

1 武汉大学资源与环境科学学院,湖北 武汉,430079
2 武汉大学地理信息系统教育部重点实验室,湖北 武汉,430079
3 地球空间信息科学协同创新中心,湖北 武汉,430079
4 自然资源部数字制图与国土信息应用重点实验室,湖北 武汉,430079
5 东华理工大学测绘工程学院,江西 南昌,330032
6 成都市国土规划地籍事务中心,四川 成都,610072

摘 要:共享单车的时空分布不均给运营商和用户造成了不便,有必要识别共享单车的空间发散区(源)及聚集区(汇),对其空间源汇区进行针对性的管理和调度。对共享单车骑行空间社区进行检测后,使用路网约束下共享K最邻近密度聚类方法,从空间流视角识别共享单车的典型源汇区域,探究早晚高峰骑行模式的共性和差异性,结果表明:(1)工作日共享单车分时订单曲线M型特征明显,而休息日曲线更接近梯台型,骑行以交通小区单元内部流动及邻接流动为主。(2)共享单车骑行社区边界受天然屏障的约束显著大于行政区划,早高峰共享单车使用的高值区更为集中。(3)工作日通勤早高峰时段,丛集流以"居住-就业"流向为主,聚散区域呈现"一面-一线-多点"的空间分布特征;晚高峰时段,丛集流以"地铁站-居住区"为主,"共享单车-地铁站-共享单车"的联运特征突出。

关键词:共享单车;流聚类;源汇空间;空间社区检测

中图分类号:P208 **文献标识码**:A **收稿日期**:2023-05-08

DOI:10.13203/j.whugis20220467 **文章编号**:1671-8860(2025)01-0184-13

Bike-Sharing Source-Sink Space Recognition Based on Riding Flow Density Clustering Method

TONG Zhaomin[1] *LIU Yaolin*[1,2,3,4] *ZHANG Ziyi*[5] *AN Rui*[1] *ZHU Yi*[6]

1 School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China
2 Key Laboratory of Geographic Information System, Ministry of Education, Wuhan 430079, China
3 Collaborative Innovation Center for Geospatial Information Science, Wuhan 430079, China
4 Key Laboratory of Digital Cartography and Land Information Application, Ministry of Natural Resources, Wuhan 430079, China
5 Faculty of Geomatics, East China University of Technology, Nanchang 330032, China
6 Chengdu Land Planning and Cadastral Affairs Center, Chengdu 610072, China

Abstract: **Objectives**: The shared bike is an important tool for residents to connect to public transport and travel a short distance, but its uneven space-time distribution has caused inconvenience to operators and users, especially in the morning and evening rush hours of commuting on weekdays. Therefore, it is necessary to identify the spatial divergence area (source) and convergence area (sink) of shared bikes, and carry out targeted management and scheduling of their spatial source-sink area. However, most studies aggregate flow data into polygon units, which lose the precision of flow data and cannot provide accurate spatial location of source and sink areas. We use a method combining spatial thinning with flow density clustering to effectively identify the spatial location of the source and sink area of bike-sharing flow. **Methods**: Spatial community detection algorithm and density-based clustering method for origin-destination riding flow are

基金项目:国家自然科学基金(42230107,42471454);国家重点研发计划(2017YFB0503601)。

第一作者:仝照民,博士,主要研究方向为地理大数据挖掘。tongzm2215@126.com

通信作者:刘耀林,博士,教授,国际欧亚科学院院士。wdzhlyl@163.com

used to identify the typical source and sink regions of shared bikes. The first thing is to clean the original dataset, including deleting duplicate orders, orders with missing attributes and abnormal orders, and performing coordinate matching. After that, using the fast Fourier transform algorithm to test the time periodicity of the dataset. Then the fast unfolding algorithm is used to detect the spatial communities and dilute the order data of shared bikes on this basis. Finally, the spatial distance of bike-sharing flow based on the road network is defined, and the shared K nearest neighbor flow clustering method is used to identify typical source and sink regions. **Results:** (1) The curve of bike-sharing timely orders on weekdays has obvious M-shaped characteristics, while the curve of weekends is more similar to the ladder type. At the scale of traffic analysis zone, the flow of shared bikes is mainly adjacent flow and internal flow. (2) The boundary of bike-sharing community is significantly more constrained by natural barriers than administrative divisions, and the high-value bike-sharing usage areas are more concentrated in the morning peak than evening. (3) During the morning rush hours of commuting on weekdays, the unidirectional flow clusters are mainly "residence to employment", and the convergent and divergent areas show the spatial distribution characteristics of "one area, one line and many points". In the evening peak hours, the flow clusters are mainly the pattern of subway station to residential area, and the intermodal characteristics of "shared bikes-subway station-shared bikes" are prominent. **Conclusions:** Based on the perspective of flow space, we analyze the characteristics of residents' short-distance shared travel. The proposed method effectively identifies the spatial source and sink area of shared bikes, which fills the smallest link in the study of urban internal mobility links, and provides support for the optimal allocation of shared bicycle space.

Key words: shared bikes; flow clustering; source and sink space; spatial community detection

近年来兴起的共享单车为城市居民提供了高效便捷的短途出行服务,促进了城市绿色可持续发展,同时作为公共交通常用的接驳工具,共享单车有效解决了第一/最后一千米问题[1]。然而诸如车辆闲置、乱停乱放、区域"潮汐式"供给不足等问题也日益凸显,由于居民的通勤行为和城市职住功能的分离,此类问题在一天中的固定时段(工作日早晚高峰通勤时间)和特定区域(地铁站附近)频繁发生,这类现象极大影响共享单车系统的运行效率和用户体验,阻碍了轨道交通系统的高效运行。因此,迫切需要定量研究共享单车骑行流的时空模式和规律,加深共享单车使用与城市人流活动耦合关系的理解,为城市规划管理部门进行交通服务设施优化以及共享单车的停放管控提供决策依据,也为共享单车企业制定运营策略提供支撑。

目前,起点-终点(origin-destination,OD)数据为骑行流的时空模式研究提供了数据基础,挖掘地理流中隐含的空间模式已经成为流数据研究的首要任务[2-4]。学者们基于不同的研究单元和尺度开展共享单车使用的时空模式研究,如文献[5]在社区尺度使用多尺度时空聚类方法分析了共享单车的潮汐骑行特征,发现工作日早高峰办公区和商场的流入趋势强,而流出区多为居民区和工厂;文献[6]在格网尺度对接驳地铁站的共享单车所在的源汇格网开展了聚类分析,发现了5类源汇格网及影响不同类型源汇格网形成的主导因子;文献[7]将共享单车骑行路段视为研究单元,利用路径匹配算法和共享单车OD数据提取骑行热点路段和骑行强度空间分布特征,为优化城市非机动车设施提供建议;文献[8]基于共享单车的骑行社区网络结构,识别了在工作日和休息日均保持稳定的骑行热点区,进而指导共享单车的投放。以上研究取得了丰硕成果,但大多局限于将共享单车OD流数据聚合至研究单元,关注骑行网络结构及热点区,而忽视了从骑行流空间视角对共享单车的时空模式进行探讨。这将在一定程度上损失流数据的精度,缺乏对具有任意形状的骑行流簇识别及综合分析,不利于共享单车运营商直接掌握共享单车的时空集散动态,给车辆的调度和管理造成了困难。

而在此基础上,针对个体空间流的分析能够细粒度地把握共享单车使用特征,共享单车作为直接通勤和接驳地铁站的重要工具,其时空动态模式可以抽象为流空间"源"与"汇"的聚集、发散和交互过程[9],如在空间位置相对较近的大型居住区与就业区之间,使用共享单车作为直接通勤工具的空间交互是"源汇区"空间流的丛集现象;而在地铁站附近,共享单车的潮汐使用现象是典型的"源"与"汇"的空间交替,即空间流的聚散现

象。对于个体空间流的分析主要分为流空间自相关统计分析[10]、流聚类分析[11]和个体流的异常探测[12],此外,通过扩展二元及多元空间点分析方法,学者们在二元空间流的聚类模式[13-14]和多元空间流的共现模式[15]中也取得了系列进展。共享单车流的源汇区识别属于个体流聚类分析,常用方法包括统计判别、层次聚类和密度聚类[16-18],但不同于人口迁徙流、行车流、通勤流等空间跨度大且方向特征突出、空间邻接关系明显的长地理流,共享单车流空间位移小、时间延续较短且空间分布不均匀,因此难以通过构建空间拓扑关系识别流簇,更适用于使用密度聚类方法。此外,由于城市中心和城市边缘区路网建设密度的巨大差异,共享单车的空间流簇存在疏密程度的变化,常用的具有噪声的空间密度聚类算法(density-based spatial clustering of applications with noise,DBSCAN)能有效识别空间高密度的集群,但在识别不同密度的流簇时效果较差[19],可能会忽略非中心城区共享单车流的聚散模式;基于点排序识别聚类结构方法(ordering points to identify the clustering structure,OPTICS)能够检测不均匀流簇,且降低了对密度连接半径的敏感度,但识别精度仍被邻域半径和密度阈值设置的主观性所掣肘[20-21]。因此,如何依据局部流疏密情况自适应调整流簇密度阈值,避免聚类参数设置的主观性,是从个体流空间识别共享单车流聚类簇的关键。

综上所述,目前国内外的相关研究主要是将共享单车的OD数据聚合至特定单元,探究其时空特征及骑行热点区,而从个体空间流视角识别共享单车流的源汇区及聚散模式的研究匮乏。因此,本文以中国湖北省武汉市主城区共享单车早晚高峰时段OD流数据为例,在骑行流社区检测后,使用基于道路网络约束的共享K最邻近密度聚类算法,将空间流进行归一化,减小流簇疏密不均的影响;并使用非参数估算方法,避免密度聚类阈值设置的主观性[17];进而从流空间视角识别骑行流的典型源汇区域,分析共享单车的丛集及聚散模式,探索早晚高峰骑行模式的共性和差异性,为深入理解共享单车骑行流时空特征,缓解共享单车的潮汐使用问题,针对性提升高频骑行区域的骑行友好型道路建设提供参考。

1 数据预处理与研究方法

1.1 研究区及数据来源

截至2020年底,武汉市共投入58万辆共享单车,运营企业包括美团、哈罗和青桔3家,市场占有率分别为55.86%、38.11%和6.03%,共享单车累计使用量超3.7亿次,是仅次于公交、地铁的第三大绿色出行方式。以武汉市主城区为研究区(见图1,底图是审图号为武汉市S(2021)033号的武汉市地图,行政界线无修改),共享单车数据为摩拜单车用户2019年11月1日(星期五)到11月7日(星期四)一周的订单数据,原始数据约包含600万条订单数据,工作日平均每日约100万条订单,周末约80万条,该时间段内天气状况良好且无特殊事件,数据集格式如表1所示。武汉市主城区共享单车的使用量占全市总骑行订单量的88%,能够有效代表全市总体情况,为全国同等级城市提供可借鉴的研究案例。

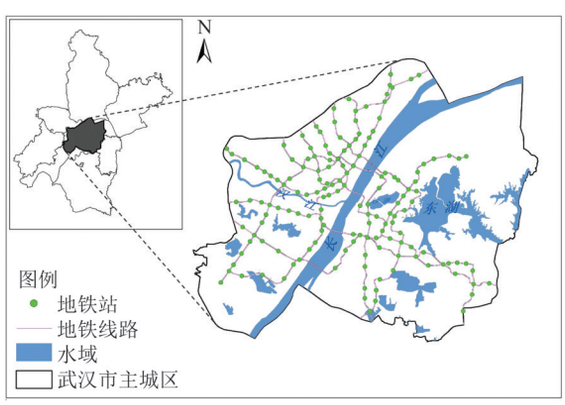


图1 研究区范围

Fig. 1 Overview of the Study Area

表1 共享单车订单数据样例

Table 1 The Samples of Bike-Sharing Orders

用户ID	车辆ID	开始时间	起点经度	起点纬度	结束时间	终点经度	终点纬度
7511553	53264	2019-11-01T07:56	114.445 4°E	30.476 7°N	2019-11-01T08:05	114.433 9°E	30.473 0°N
210412	68412	2019-11-02T11:32	114.133 5°E	30.680 5°N	2019-11-02T11:49	114.147 9°E	30.685 0°N
1025473	35412	2019-11-04T21:19	114.991 0°E	30.668 9°N	2019-11-04T21:25	114.981 9°E	30.662 2°N

1.2 共享单车数据预处理

由于GPS定位误差、坐标漂移、订单重复、订单属性缺失等问题,需要对共享单车原始订单数据进行预处理,包括3类异常数据的清洗:重复订单、属性缺失订单、异常订单(异常订单时长、车速、骑行距离),以及坐标匹配(流程见图2)。

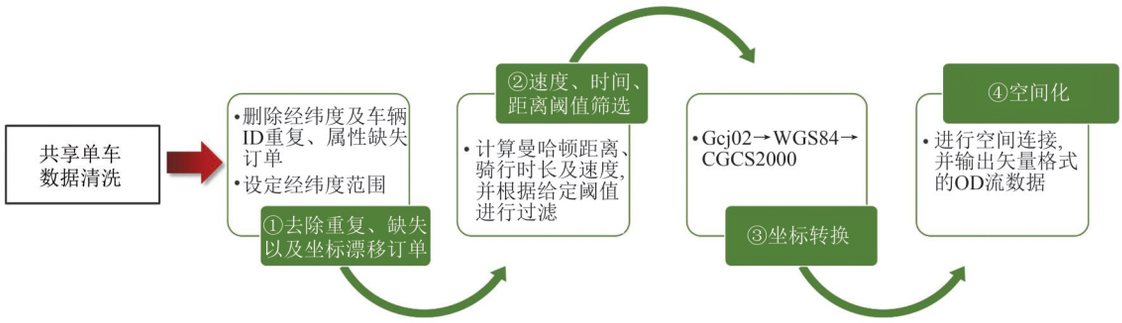


图 2　数据清洗流程

Fig. 2　Process of Data Cleaning

删除属性重复及属性值缺失的订单后,根据主城区的经纬度范围筛选订单起始点和终止点均位于该范围内的记录,根据研究时段筛选开锁时间和关锁时间均位于该时段内的记录。结合实际骑行情况,考虑到未关锁、GPS 信号异常等原因,删去骑行距离及时间过短或过长的订单,设定骑行距离阈值为 0.5～6 km 之间,骑行时间为 1 min～2 h 之间。骑行速度的设置参考《2018 武汉共享单车报告》,结合主城区内交通系统运行情况,设置速度阈值为 5～20 km/h 之间。

已有研究指出,共享单车的使用在工作日内存在明显周期性规律,因此借助快速傅里叶变换算法(fast Fourier transform,FFT)对工作日和周末共享单车分时订单量进行周期性检验。FFT 是信号处理领域检验周期性的常用方法,通过将时间序列的时域转换到频域,以确定时间序列的主周期[22],结果如图 3 所示。工作日呈现出明显的 24 h 周期性,最大峰谱对应的第一、第二主分量周期分别为 24 h 和 48 h,意味着即使是一周的数据也可以代表较长一段时间内共享单车的使用情况;而休息日的周期性并不明显,最大峰谱对应的第一、第二主分量周期分别为 9 h 和 24 h,一方面受限于数据跨度,另一方面周末共享单车出行随机性较强,且使用者大多没有工作通勤目的,因此对工作日的研究更能展现共享单车使用与城市人群活动的耦合关系。考虑到工作日共享单车使用呈现稳定的周期特征,为降低聚类算法的时间复杂度,选取 11 月 6 日(周三)代表工作日的使用特征,对出行早高峰(7 点—9 点)和晚高峰(5 点—7 点)进行分析。

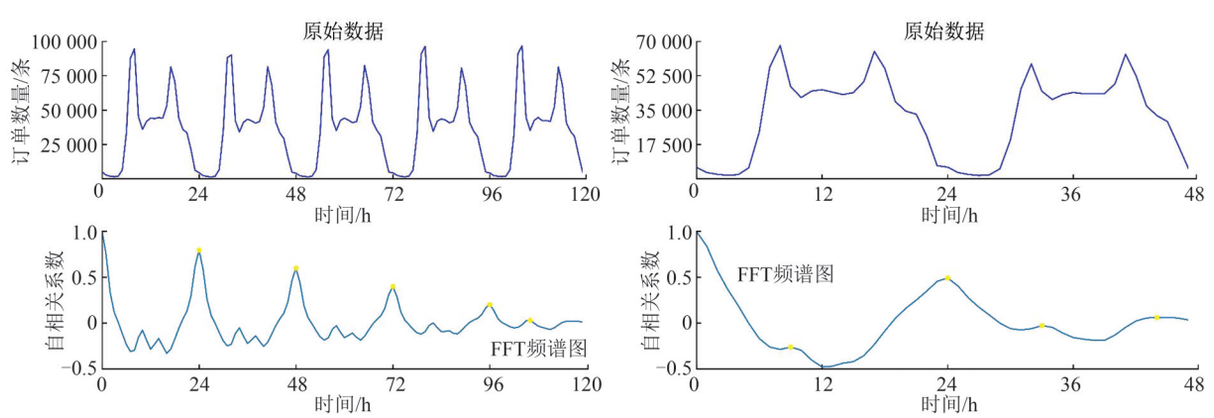


图 3　共享单车订单数量周期性检验

Fig. 3　Periodic Test of Bike-Sharing Orders

1.3　研究方法

1.3.1　共享单车流的空间社区结构发现方法

共享单车作为居民短途出行的重要工具,往往受城市空间结构和建成环境的限制,检测共享单车空间社区结构有助于理解居民骑行的空间集群,辅助自行车道网络体系的规划。本文使用快速展开算法对共享单车骑行流的空间社区结构进行检测。快速展开算法是一种高效的无监督社区检测算法,用于识别复杂网络中的紧密和稀疏关系,将具有强交互作用的部分划分为一个社区,社区内部存在频繁的交互,而社区间的交互较少[23]。在共享单车骑行网络中,社区内的共享单车流关系密切,“逃离”社区的可能性较低,较好地契合了共享单车流动区域性强的特性,可

以有效地识别骑行区的空间结构。

以订单OD数据中出现过的每个栅格作为节点,以节点间的订单量作为边权重,构建空间交互网络。快速展开算法通过最大化模块度进行社区划分,模块度衡量社区内部联系的紧密度,取值范围在−1到1之间,其计算公式为:

$$Q=\frac{1}{2m}\sum_{i,j}\left[A_{i,j}-\frac{k_i k_j}{2m}\right]\delta(c_i,c_j) \quad (1)$$

式中,Q表示模块度;m表示网络中的总订单数;$A_{i,j}$表示节点i与节点j间的边权重;k_i和k_j分别为节点i与节点j边权重的和;$\delta(c_i,c_j)$用于判断节点i与节点j是否在同一个社区,若是,则$\delta(c_i,c_j)=1$,若不是,则$\delta(c_i,c_j)=0$。

快速展开算法通过迭代两个步骤使模块度最大化:(1)模块度优化,将节点划分到其相邻节点所在社区,直到不再增加模块度,在初始化阶段,每个节点都被视为一个社区;(2)社区聚合,将第(1)步发现的同一社区中的节点分组,并构建一个新网络。重复以上两个步骤,直至网络结构不再变化。

1.3.2 共享单车流的空间特征识别方法

本文使用基于路网约束的共享K最邻近OD流聚类算法对共享单车骑行流进行空间流聚类,识别共享单车流的空间汇聚、发散及丛集特征,算法主要包括3个步骤:定义路网约束下共享单车的K最邻近流、基于共享最邻近流密度识别流簇、基于非参数方法估算最邻近K值。算法主要流程如图4所示。

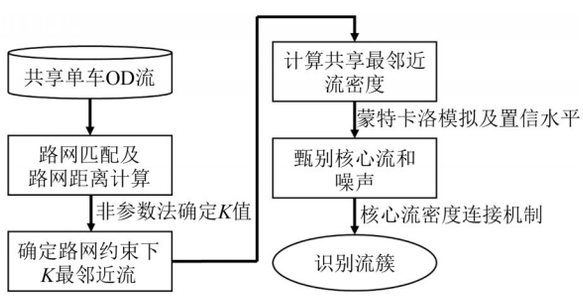


图4 基于路网约束的共享K最邻近OD流聚类算法流程

Fig. 4 Framework of Shared K Nearest Neighbor Flow Clustering Algorithm Based on Road Network

1)定义路网约束下共享单车的K最邻近流。共享单车的原始流具有时间延续性弱且空间位移短的特点,使用欧氏距离对原始流进行聚类难以有效约束聚类簇形状,并且可能高估聚集状态。共享单车流作为典型的城市OD流,受骑行道路网络的强烈约束,因此使用路网匹配算法将共享单车OD点映射到可供骑行的城市道路网(剔除城市快速路、高架等不可骑行道路),从而借助道路网络的约束突出聚类簇和背景噪声的密度差异,并且计算映射节点间网络距离能够有效避免计算冗余[24]。

建立路网约束后,需要重新定义流距离用于聚类,由于共享单车骑行流的空间位移短,流长度差异较小,因此在定义流f_i与流f_j距离时不考虑流的长度。基于路网的流距离计算公式为:

$$d_{f_{ij}}^{\text{net}}=w_1 d_{O_{ij}}^{\text{net}}+w_2 d_{D_{ij}}^{\text{net}}\ (w_1>0,w_2>0) \quad (2)$$

式中,$d_{f_{ij}}^{\text{net}}$为路网约束下流f_i与流f_j的距离;$d_{O_{ij}}^{\text{net}}$为流f_i与流f_j映射在路网上起点的网络距离;$d_{D_{ij}}^{\text{net}}$为流f_i与流f_j映射在路网上终点的网络距离;w_1和w_2分别为起点与终点的权重,在不同的应用背景下起点和终点的关注度不同,本文认为起点与终点是等权的($w_1=w_2=1$)。之后以d为距离阈值,对每个流的起点O和终点D分别构建基于路网的缓冲区,选取路网映射后网络距离小于d的O点和D点作为候选邻近流,通过设定最邻近K值和计算流距离,得到流的K最邻近结果,以流f_i设定K为3的最邻近流如图5所示。

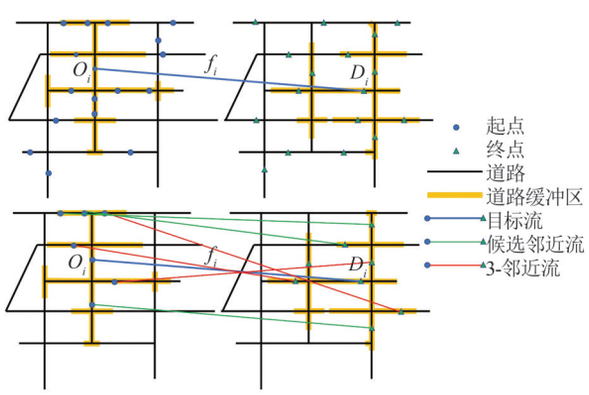


图5 识别距离目标流f_i最邻近的3个流($K=3$)

Fig. 5 3 Flows Nearest to Target Flow f_i ($K=3$)

2)基于共享最邻近流密度识别流簇。由于共享单车的使用密度与可供骑行的道路网络密度关系密切,中心城区的路网密度往往高于市区边缘,使用DBSCAN和OPTICS等聚类算法的密度定义方式可能会忽视市区边缘区共享单车的集聚使用,难以全面反映共享单车流动的空间特征。因此,在确定每个流的K最邻近流后,采用共享最邻近的方法计算流密度,即被相似流包围的程度,通过对全局流密度的归一化,突出流簇与噪声流的密度差异,同时检测高骑行密度的聚类簇和较为稀疏的聚类簇[25]。

在计算共享最邻近流密度前,首先定义流f_i与流f_j的共享最邻近相似性$S(f_{ij})$:

$$S(f_{ij})=\left|N_{f_i}^{k}\cap N_{f_j}^{k}\right| \tag{3}$$

式中,$N_{f_i}^{k}$和$N_{f_j}^{k}$分别为流f_i与流f_j的K最邻近流。若$S(f_{ij})\geqslant K/2$,则认为流f_i与流f_j是直接相连的,而流f_i的共享最邻近流密度$D(f_i)$则是与流f_i直接相连的流的数量。

密度阈值是识别高密度聚类簇的常用方法,但在确定密度阈值时往往主观性较强,因此借助统计方法将密度阈值转换为显著性水平,减轻主观参数设定对结果的影响[17]。在路网上随机生成与流数据集数量相同的模拟流,执行R次蒙特卡洛模拟过程,则流f_i的统计检验值$p(f_i)$定义为:

$$p(f_i)=\sum_{j=1}^{R}\frac{l_j\left(D_{\text{random}}(f_i)\geqslant D_{\text{observed}}(f_i)\right)}{1+R} \tag{4}$$

式中,$D_{\text{random}}(f_i)$为模拟流密度;$D_{\text{observed}}(f_i)$为实际计算的流密度。若$D_{\text{random}}(f_i)\geqslant D_{\text{observed}}(f_i)$,则$l_j$为1;否则为0。在显著性水平$a$下,若$p(f_i)\leqslant a$,则说明观测流密度与蒙特卡洛模拟生成的流密度差异并非由偶然性所导致,此时拒绝零假设,确定流f_i为核心流[10]。由于模拟流生成在道路网络密集的区域可能性较大,因此使用统计方法确定核心流能够防止聚类簇过度集中于高路网密度区域,极大程度上保留了全局不同密度层级的聚类簇。

基于密度连接机制,流聚类簇的构建过程与DBSCAN等密度聚类算法思想相同[26],在不考虑流长度和流角度的情况下,基于起点和终点距离的流聚类算法往往见长于识别两种类型的流簇[9]:(1)流向相似的单向集群流,该类流起点和终点都聚集在相近的区域,即共享单车的丛集区域,如图6(a)所示;(2)起点随机或终点随机的多向聚集流,该类流簇中相邻流的流向相似,从而完成密度连接机制,一般可分为终点相近的汇聚型流簇,以及起点相近的发散型流簇,即共享单车的聚散区域,如图6(b)和图6(c)所示。

3)基于非参数方法估算最邻近K值。最邻近K值决定了计算共享最邻近流密度的局部区域大小,即共享单车流簇的地理分析尺度。若K值过小,则距离流簇中心最远的邻近流与噪声的差异不明显,将会导致流簇密度的不稳定;若K值过大,则流簇可能包含噪声流,高估共享最邻近流密度。当K值增加时,距离最远的邻近流和噪声流的方差逐渐趋同,因此,参考已有确定聚类簇数量方法的研究[27],定义非参数指数H衡量K值变化与流簇方差的关系,计算公式为:

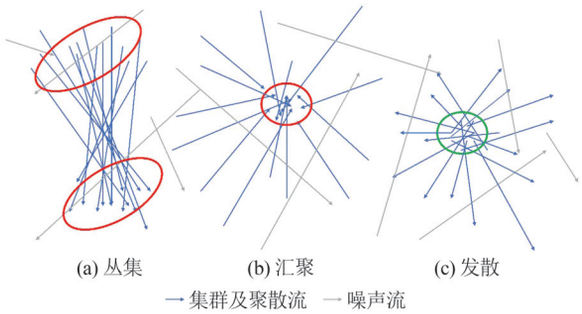


图6　流空间特征分类

Fig. 6　Types of Flow Clusters

$$H=\frac{V_{k+1}^{*}(x)}{V_{k}^{*}(x)}/H_k \tag{5}$$

式中,$V_{k+1}^{*}(x)$、$V_{k}^{*}(x)$分别为第$k+1$和第k邻近流的方差;H_k是一个常量,即均匀流集$V_{k+1}^{*}(x)$与$V_{k}^{*}(x)$的期望比率。H表征了流簇边缘与噪声的相似程度,当K值变大时,其值将降低并趋近期望比率,即趋近1,但降低速率将逐渐减小。因此,肘部法则可以用于估算最适当的K值,即最适当的K值对应了H下降最快且变化逐渐趋缓的区间,此时流簇规模稳定且与噪声流差异明显。使用非参数方法估算最邻近K值能够最大程度上避免主观性,减轻参数设置不当对聚类效果的影响。

2　结果与分析

2.1　共享单车骑行总体特征分析

1)骑行订单数量。根据2019年11月1日—7日一周的分时骑行订单数据统计骑行订单数量,结果如图7所示。可以看到,工作日共享单车订单总量要明显高于休息日,工作日共享单车分时订单曲线M型特征明显,而休息日共享单车分时订单曲线更类似梯台型,日间波动相对较小。工作日7时—9时时段和17时—19时时段出现了明显的共享单车使用高峰,并且早高峰的订单数量要高于晚高峰。早高峰的峰值出现在8时,但从7时开始订单量开始陡增,7时到8时增长速率放缓,曲线变化呈梯台状;而晚高峰的峰值出现在17时,随后便快速下降,曲线变化呈尖峰状,这可能是由于工作日通勤时早间上班时间不统一(互联网及金融企业早上上班时间较晚),而晚上下

班时间集中在17时。而休息日仍存在一定的早晚高峰特征,尤其是周六早高峰时段相较于周日更为明显,这可能是一些高新技术企业在周六加班所致。此外,在非早晚高峰时段,工作日和休息日的订单数量差距较小,休息日甚至略高于工作日,这也说明了共享单车的使用更多是作为辅助通勤的手段,在非通勤高峰时段,共享单车的使用多以休闲娱乐为主。

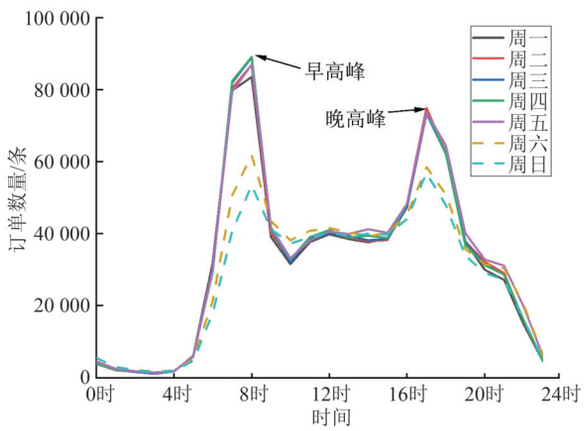
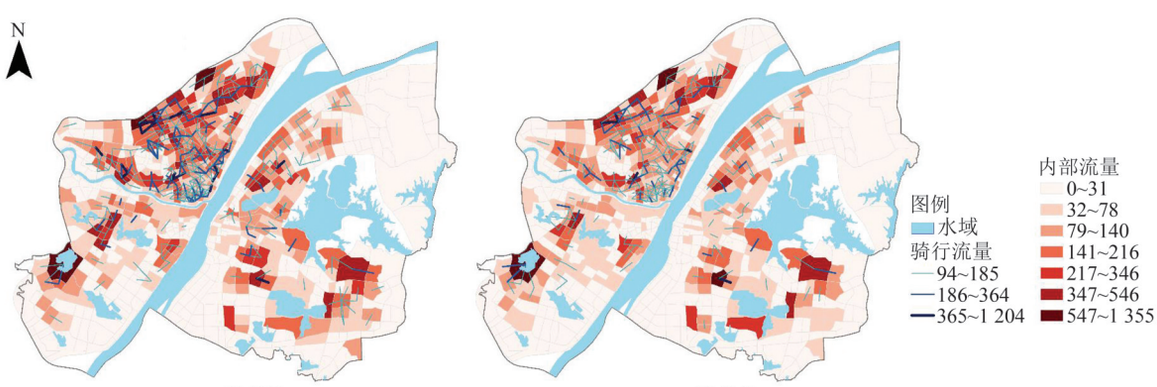


图7 一周骑行订单数量分时统计图

Fig. 7 Daily Trend of Bike-Sharing Usage

2)骑行空间特征。为了分析共享单车流的总体空间特征,将分析单元定为交通小区(traffic analysis zone,TAZ)。相较于组团、街道、社区等研究单元,TAZ是由主干路围成的,内部具有相似交通特征和关联的单元,常用于分析居民通勤、流动等交通特征[28]。以武汉市主干道为主要分界线,考虑河流、湖泊等空间分布,结合土地利用特征对TAZ边界进行调整,最后得到武汉市主城区586个TAZ,平均面积为0.96 km^2,标准差为1.3 km^2,中心区域单元面积较小,而外围区域单元面积逐渐增大。

以TAZ为单元,汇总共享单车流形成586×(586−1)个骑行流对(除去TAZ内部流动情况),分析工作日和休息日及早晚高峰的差异,绘制各TAZ内部流动及TAZ间流动空间图,颜色的深浅和连线的粗细反映流动强度,如图8和图9所示(骑行流量小于93未在图中表示)。总体来看,区域间的骑行流动联系以邻接流动为主,跨多个区域的骑行流动联系很少。江汉大学、塔子湖、东湖附近内部骑行流较高,而江汉、江岸、硚口区区域间流动较多。工作日早晚高峰TAZ内部流动情况相对一致,但早高峰的区域间流动更为频繁,汉兴街、长青街、宗关街、唐家墩街早高峰区域间流动更多。休息日由于订单数量的减少,区域间流动较少,但内部流动相比工作日变化不大。



(a) 早高峰 (b) 晚高峰

图8 工作日骑行流空间特征

Fig. 8 Spatial Characteristics of Weekday Riding Flow

由于共享单车骑行的短距短时特点,在汇总至统计单元后,无法对其空间特征进行细粒度的描述,难以有效识别空间源汇区,因此从流尺度对骑行订单进行分析,能够更深入地识别共享单车流动的空间特征。

2.2 早晚高峰共享单车骑行社区检测

对共享单车骑行流的社区检测能够发现空间上骑行联系紧密的区域,理解骑行流的空间分布格局。使用快速展开算法,在剔除结果中规模较小的社区和孤立的社区后,早高峰时期共产生12个社区,晚高峰期间共产生13个社区,模块度分别为0.80和0.79,当模块度大于0.5时,说明网络社区结构明显,每个社区代表一种典型的流动模式,而社区间联系较弱。

如图10所示,骑行社区与行政区的边界区别明显,说明共享单车的使用受行政区划约束较小。由于武汉大江大湖的自然条件,骑行社区被长江和汉江划分为3个片区,虽然武汉市2019年底已建成8座跨江大桥、多个跨江通道,但很少有居民使用共享单车作为跨江通行的工具。此外,

诸如沙湖、东湖等主要湖泊也成为了分割骑行社区的自然屏障。相较早高峰,晚高峰时段武昌区划分为了南北两个骑行社区,武昌区南部多为高校和科研机构,北部则以居住区为主导。从骑行高值上看,早高峰共享单车的使用更加集中于汉口中央活动区,而晚高峰共享单车使用高值区向武昌区、塔子湖、南湖等组团扩散,高值范围更广。

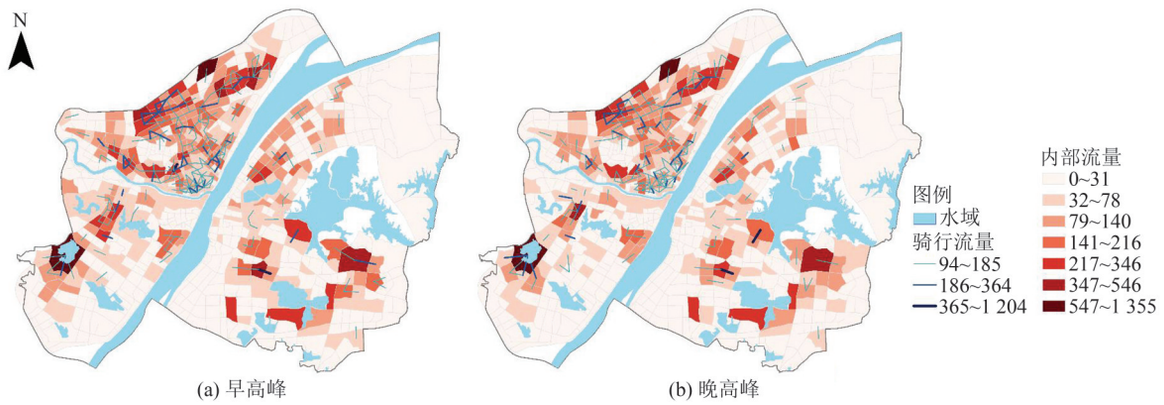


图 9　休息日骑行流空间特征

Fig. 9　Spatial Characteristics of Weekend Riding Flow

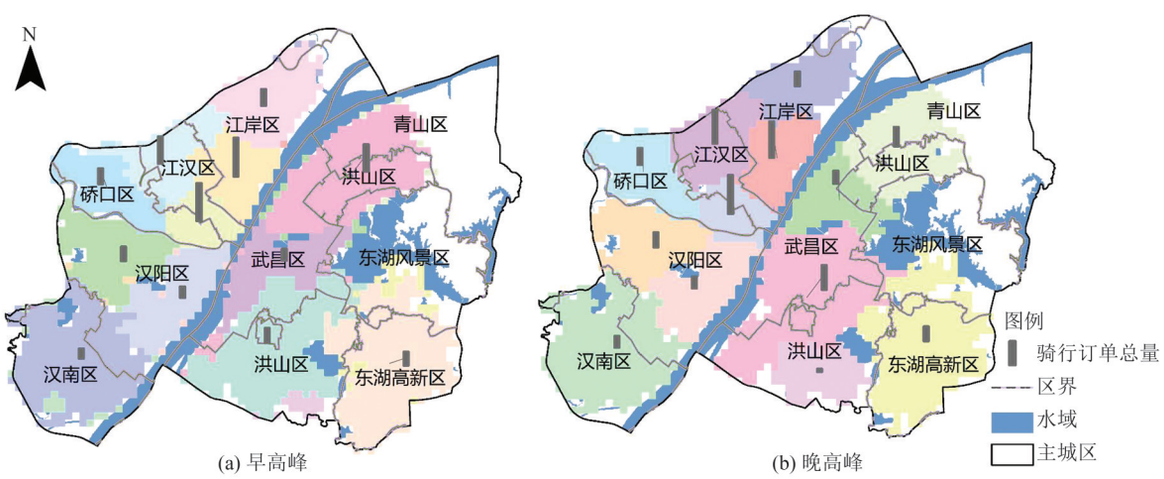


图 10　工作日骑行流社区检测结果

Fig. 10　Community Detection Results of Weekday Riding Flow

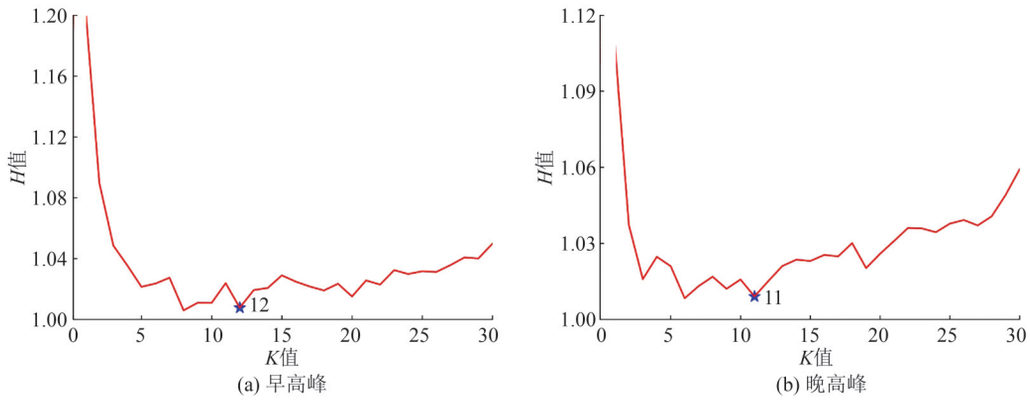
在确定共享单车出行社区后,以出行社区为空间分层抽样单元对共享单车订单进行抽稀,以突显聚类结果的社区内联系,避免流簇包含过多噪声流,同时并降低算法的时间成本。

2.3　共享单车骑行流空间聚类

从OD骑行流中识别空间集聚和扩散现象,对研究共享单车的移动特征和位置之间的空间交互至关重要,共享单车的流动在城市系统中,有助于精细化研究城市区域间的联系。考虑到工作日早晚高峰共享单车的使用强度较高,且存在明显差异,因此分别对工作日早晚高峰时段共享单车的OD流进行聚类,解释共享单车骑行流的空间集聚、扩散和丛集特征。通过肘部法则确定最适当的K值,同时考虑到早晚高峰聚类簇大小的相对一致性,在早高峰设为12,晚高峰设为11,如图11所示。武汉市主城区内有90%的共享单车骑行距离在0.8~3.5 km之间,因此将距离阈值d设为3.5 km,显著性水平a设为0.05,蒙特卡洛模拟次数设为99。

2.3.1　共享单车骑行流源汇空间识别

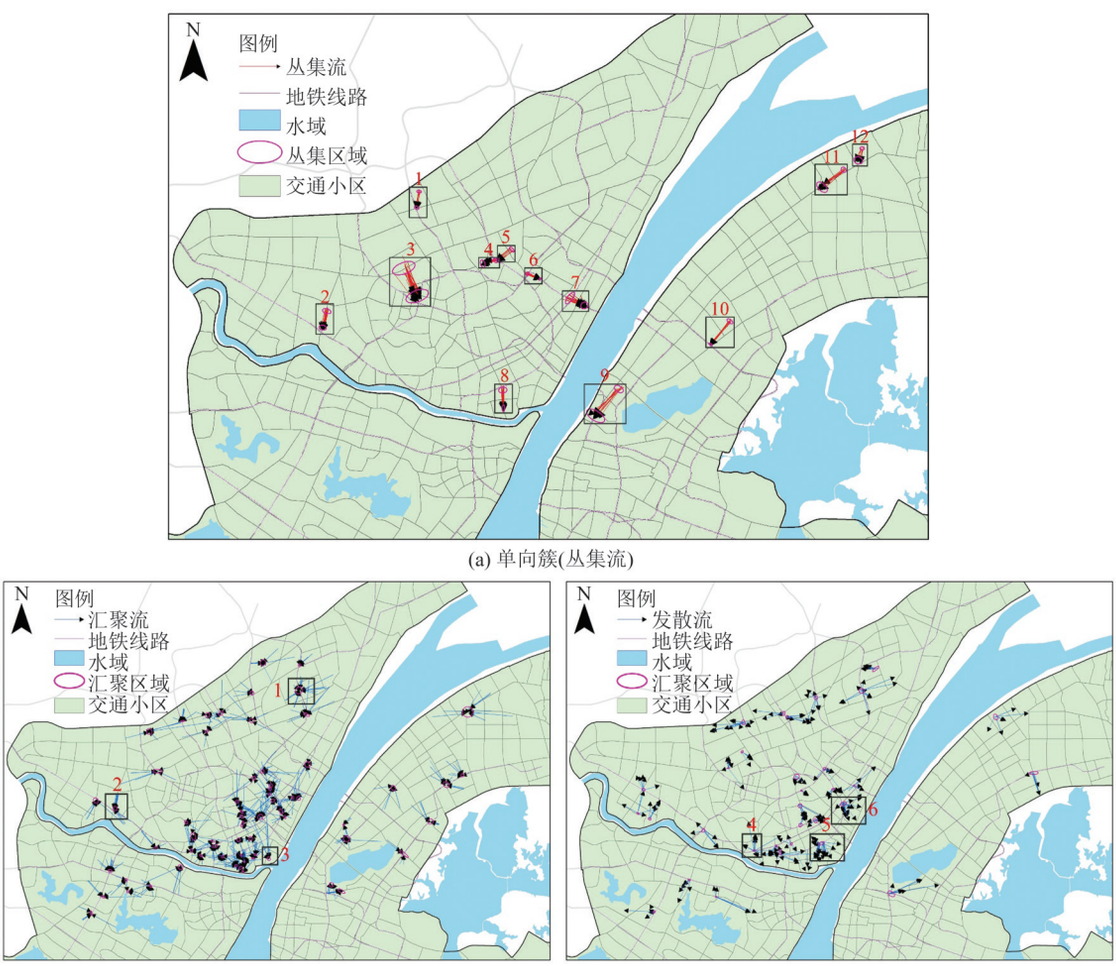
1)早高峰共享单车聚散特征。将聚类结果划分为丛集、汇聚和发散3种模式,并可视化核心流,如图12所示,源汇区域主要位于江汉、江岸和硚口区,武昌区和青山区的沿长江地带汇散特征也较为明显。江北地区作为武汉市的经济和商业中心,拥有武汉国际广场、武汉世界贸易中心等国际会展中心,江汉路、汉正街等商业街,以及中山公园等休闲游憩场地,在商业活动、通勤、游憩等目的下,共享单车的使用量较大。东湖、南湖、武汉理工大学附近共享单车的使用强度也较



(a) 早高峰 (b) 晚高峰

图11 不同K值对应的H值

Fig. 11 H Values Corresponding to Different K Values



(a) 单向簇(从集流)

(b) 典型收敛簇及发散簇(汇聚区及发散区)

图12 工作日早高峰共享单车流聚类结果

Fig. 12 Clustering Results of Bike-Sharing Flow in Weekday Morning Peak

高,但景点附近共享单车的使用随机性较强,空间聚散特征不明显,在工作日早晚高峰时段未能识别到显著的源汇区。

早高峰时段,从集流主要位于江北区域的武汉商务区、三阳路、菱角湖、三眼桥、长港路,武昌区的积玉桥,以及青山区的青山居住组团,以“居住-就业”流向为主。其中武汉商务区的从集流较为明显,武汉商务区属于中央活动区,是武汉市规划的集金融、保险、商贸等业务为一体的现代服务业中心,有很强的就业吸引力,可以看到从集流的起点位于武汉商务区北部的居住组团,集中了蓝天花园、复兴新村、中城悦城等多个居住小区,由于地理位置毗邻,较多人选择共享单车前往武汉商务区,两个区域间呈现明显的有向从

集效应。沿长江地带的汉正街、三阳路、积玉桥是武汉市规划的滨江活动区,同样属于“居住-商业”复合功能区,区域内布设有多家大型写字楼、政府机关和颇具规模的商业街,同时也有与就业匹配的居住社区,在便捷的道路交通网络支撑下,共享单车能较好地连接地铁站与就业区,满足早高峰通勤需求。在青山区和平大道沿线,青山公园及和平公园附近也出现了丛集现象,2019年底武汉市5号线青山区段尚未开通,以共享单车为出行工具的“红钢城住宅区-青山印象城商业区”“武汉工业科技学校-武汉英赛商务中心”的集中出行流特征明显。

早高峰时段,共享单车的聚散区域呈现“一面-一线-多点”的空间分布特征,“一面”为汉口中央组团的汉江-长江沿带,该区域有着越秀星汇维港、武广商圈、江汉路商圈、汉正街商圈等,“一线”为“江兴路-常青五路-后湖大道”这一东北-西南轴线,多点包括螃蟹岬、湖北大学、铁机路、金地广场、青山印象城等。可以发现,多数共享单车的聚集区在地铁站、商业活动中心、工业园区、政府机关附近,这说明早高峰时段,共享单车有向经济、商业、行政办公密度高的区域聚集使用趋势;而共享单车的发散区同样集中在地铁站出入口附近,如江汉路地铁站、大智路地铁站、取水楼地铁站等,也有诸如湖北省交通厅运管局宿舍、碧园社区、丰竹园等居住区,说明早高峰时段共享单车的发散区多集中在地铁站点、居住区附近,而诸如长港路地铁站、硚口路地铁站、香港路地铁站、王家湾地铁站等地铁站附近,共享单车的汇聚和发散区有重叠,说明“共享单车-地铁-共享单车”的通勤接驳特征显著。由于共享单车的聚散现象,早高峰时段大型就业区及城市边缘的居住区附近的地铁站,共享单车可能存在堆积情况;而大型就业区附近的地铁站及城市边缘的居住区,共享单车可能存在供应不足。

2)晚高峰共享单车聚散特征。如图13所示,晚高峰共享单车的空间特征与早高峰出现了较大差异,相较于早上上班时间的明确约束,晚高峰下班时间较灵活,共享单车在晚高峰的空间丛集和聚散区域都有所减少。此外,居民选择共享单车出现的目的不再局限于通勤,也可能开展购物、休憩、娱乐等活动,此类活动的骑行流空间趋向不集中,因此难以捕捉流空间特征。晚高峰共享单车的聚散区同样聚集在江北地区的长江沿带,且更向长江沿岸集中,可能沿江活动区在晚间对居民有较强的汇聚引力。

晚高峰时段,共享单车丛集流主要位于江岸区的后湖居住组团,从后湖大道地铁站出发到幸福人家居住小区;江岸区的塔子湖组团,从杨汊湖地铁站出发到跃进嘉园等居住小区;汉阳区的汉江沿岸,从五里墩地铁站到月湖琴声等居住小区;洪山区的杨园组团,从铁机路地铁站到保利城等居住小区;洪山区的“洪山-青山”组团,从仁和路地铁站到沙湖港湾等居住小区。晚高峰时段的丛集流均为地铁站-居住区的流动模式,在这几个典型区域,居民大多选择共享单车作为接驳地铁站的交通工具,可见共享单车在解决最后一千米问题上作用显著。

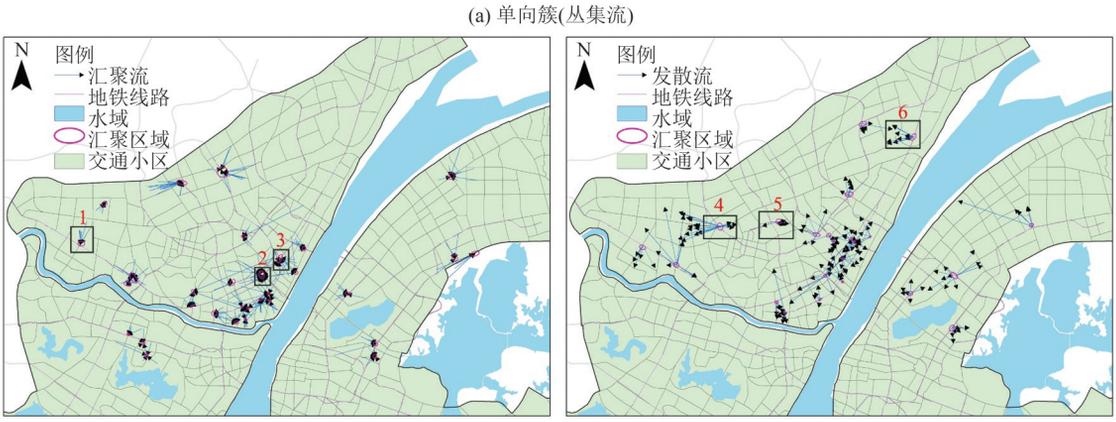
晚高峰时段,共享单车的聚散区域明显减少,汇聚区包括循礼门地铁站、大智路地铁站、江汉路地铁站、汉正街地铁站、硚口路地铁站、武商城市奥莱、欢乐星城、华侨城等区域,共享单车的汇聚区除了汉口中央活动组团就业密集区附近的地铁站外,还在洪山区与东湖风景区邻接的华侨城、硚口区的团结佳兴园等居住小区出现了聚集。共享单车的发散区更多分布在汉口中央活动区边缘的地铁站附近,如常码头地铁站、范湖地铁站、赵家条地铁站;也集中在后湖组团、青山组团等居住组团的地铁站附近,如丹水池地铁站、后湖大道地铁站、仁和路地铁站、古田三路地铁站等。这说明晚高峰时段,“共享单车-地铁站-共享单车”的联运特征同样突出,共享单车的聚集区主要在武汉市中心区域就业组团的地铁站附近及城市边缘的居住区附近,居民下班后骑行前往地铁站,或在一些偏远居住小区,居民在附近地铁站下车后,骑行回到住所;发散区主要分布在中央活动区边缘的地铁站及居住区附近的地铁站。武胜路地铁站、崇仁路地铁站、江汉路地铁站附近,共享单车的汇聚和发散区有重叠,说明该区域用地功能混合度较高,职住发展平衡。可以看出,绝大部分共享单车的聚散区域都围绕着地铁站分布,共享单车骑行流的源汇区域与地铁站分布呈现空间一致性,进一步证实了共享单车在居民日常接驳轨道交通中的重要作用。此外,共享单车的聚散区在早晚高峰呈现显著不同,说明共享单车的使用在时间维度上呈现潮汐特征。

2.3.2　与OPTICS密度聚类方法的比较

OPTICS聚类算法是识别不均匀密度簇的代表性方法,以工作日早高峰共享单车流为对象进



(a) 单向簇(从集流)



(b) 典型收敛簇及发散簇(汇聚区及发散区)

图13 工作日晚高峰共享单车流聚类结果

Fig. 13 Clustering Results of Bike-Sharing Flow in Weekday Evening Peak

行对比。算法参数设置上,由于OPTICS聚类算法对半径敏感度低,为了确保邻域足够大,将邻域半径参数设置为共享单车的平均骑行距离1.3 km,将密度阈值设为12(与K最邻近值相同)。

如图14所示,OPTICS算法能够识别主城区边缘密度较稀疏的聚类簇,但高估了江汉、江岸城市中心共享单车流的聚集,此外,由于长江南岸共享单车的使用聚集性不强,未能识别到明显的聚类簇。OPTICS算法虽然能够识别聚类簇的稀疏,但由于共享单车流动区域性强、分布依赖路网,在流未经归一化处理时,算法将会过多关注城市中心的高密度聚类簇,且参数设置的主观性仍会对算法效果造成较大影响。

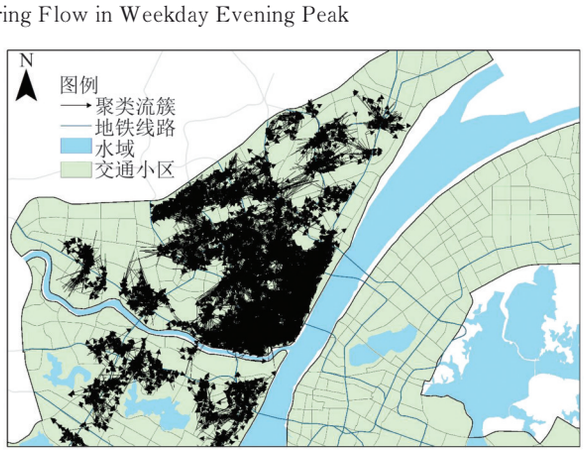


图14 工作日早高峰共享单车流聚类结果(OPTICS算法)

Fig. 14 Clustering Results of Bike-Sharing Flow in Weekday Morning Peak (OPTICS Algorithm)

3 结 语

本文在对武汉市共享单车骑行的总体特征进行分析的基础上,通过社区结构检测算法识别了共享单车骑行的空间集群;使用道路网络约束下的共享K最邻近流聚类算法,分析了共享单车早晚高峰时段骑行的源汇区域,为揭示城市居民短途流动模式,优化共享单车投放和调配,促进骑行友好型城市建设提供了参考。

1)共享单车分时订单曲线表明,工作日早晚高峰潮汐峰值更明显,共享单车通勤目的更强,休息日共享单车分时订单曲线更平缓,共享单车

更多服务于休闲娱乐目的;空间流动联系上,共享单车的内部流动和TAZ间邻接流动更多,跨多个TAZ的骑行订单较少。

2)共享单车骑行社区受自然屏障约束更明显,骑行社区边界与行政区边界空间一致性较弱;早高峰共享单车骑行高值区集中于汉口中央活动区,而晚高峰骑行高值区向武昌区、塔子湖、南湖等组团扩散。

3)骑行空间社区和道路网络的双重限制有效约束了流簇的形状,全局流密度的"归一化"保留了不同密度层级的聚类簇,识别出的丛集、汇聚和发散区从流尺度刻画了共享单车的空间流动特征。早高峰时段,共享单车的丛集流以"居住-就业"流向为主,聚集区在地铁站、商业活动中心、政府机关附近,发散区多集中在地铁站、居住区附近;晚高峰时段,共享单车的丛集流以地铁站-居住区的流动模式为主,而聚散区域明显减少,共享单车与地铁的联运特征突出。

共享单车聚散区域有助于识别骑行热点区及地铁站站前空间的使用负荷,从而针对性地进行改善。空间上毗邻的聚散区,可以通过短距离的互相调度实现彼此的动态平衡。在时间尺度上出现潮汐骑行现象的地铁站,如后湖、杨汊湖等城市边缘的大型居住区及其周边地铁站,运营商应注意早高峰共享单车的停放疏导,并及时将地铁站范围内积压的共享单车转移至附近的居住社区,晚高峰时应保证地铁站附近共享单车的供给,避免供不应求的问题发生;而武广、江汉路等商圈及就业区易发生相反的潮汐现象,运营商应在早高峰前完成共享单车的重点投放。丛集流发生的区域,共享单车更多作为直接通勤工具,规划者应着重关注区域内非机动车道规划、骑行护栏、绿道等慢行友好型设施建设。

本文仍存在一些局限:方法上,本文所使用的仍是空间聚类模型,虽然将时间分为早晚高峰进行分析,但时间信息仍存在缺失,未来可引入流之间的时空相似性度量方法,从数据中发现显著的时空趋势。数据上,本文使用的共享单车订单数据时间跨度仅为一周,虽然共享单车的使用呈现明显的周期性,但共享单车使用模式可能存在季节性或与天气相关的差异,共享单车使用模式的年度变化可以反映城市发展建设对居民出行行为的影响,这些规律需要从更长时间跨度的数据中分析得到。

参 考 文 献

[1] 罗建平,裘凝远,尹杰丽,等. 基于大数据分析的共享单车与地铁接驳方案优化研究[C]//第十九届中国智能交通年会,浙江杭州,2024.
LUO Jianping, QIU Ningyuan, YIN Jieli, et al. Intelligent Transportation Systems and Applications Conference[C]//The 19th Annual Conference of China Intelligent Transportation Systems, Hangzhou, China, 2024.

[2] 刘瑜,姚欣,龚咏喜,等. 大数据时代的空间交互分析方法和应用再论[J]. 地理学报, 2020, 75(7): 1523-1538.
LIU Yu, YAO Xin, GONG Yongxi, et al. Analytical Methods and Applications of Spatial Interactions in the Era of Big Data[J]. *Acta Geographica Sinica*, 2020, 75(7): 1523-1538.

[3] 刘瑜,詹朝晖,朱递,等. 集成多源地理大数据感知城市空间分异格局[J]. 武汉大学学报(信息科学版),2018,43(3): 327-335.
LIU Yu, ZHAN Chaohui, ZHU Di, et al. Incorporating Multi-source Big Geo-data to Sense Spatial Heterogeneity Patterns in an Urban Space[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(3): 327-335.

[4] 涂伟,曹劲舟,高琦丽,等. 融合多源时空大数据感知城市动态[J]. 武汉大学学报(信息科学版), 2020, 45(12): 1875-1883.
TU Wei, CAO Jingzhou, GAO Qili, et al. Sensing Urban Dynamic by Fusing Multi-sourced Spatiotemporal Big Data[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(12): 1875-1883.

[5] 汪晓龙,石岩,王达,等. 基于自适应密度聚类的单车出行源汇时空区域探测方法[J]. 武汉大学学报(信息科学版), 2024, DOI: 10.13203/j. whugis20230362.
WANG Xiaolong, SHI Yan, WANG Da, et al. Spatiotemporal Extent Detection method of Bike Travel Source-Sink Based on Adaptive Density Clustering[J]. *Geomatics and Information Science of Wuhan University*, 2024, DOI: 10.13203/j. whugis20230362.

[6] 高楹,宋辞,郭思慧,等. 接驳地铁站的共享单车源汇时空特征及其影响因素[J]. 地球信息科学学报, 2021,23(1): 155-170.
GAO Ying, SONG Ci, GUO Sihui, et al. Spatial-Temporal Characteristics and Influencing Factors of Source and Sink of Dockless Sharing Bicycles Connected to Subway Stations[J]. *Journal of Geo-infor-*

mation Science,2021,23(1):155-170.

[7] 杨永崇,柳莹,李梁. 利用共享单车大数据的城市骑行热点范围提取[J]. 测绘通报, 2018(8): 68-73.
YANG Yongchong,LIU Ying,LI Liang. Urban Cycling Hot Spot Extraction Based on Sharing-Bikes' Big Data[J]. *Bulletin of Surveying and Mapping*, 2018(8): 68-73.

[8] SONG J, ZHANG L, QIN Z, et al. A Spatiotemporal Dynamic Analyses Approach for Dockless Bike-Share System [J]. *Computers, Environment and Urban Systems*, 2021, 85: 101566.

[9] 裴韬. 地理流空间分析[M]. 北京:科学出版社, 2024.
PEI Tao. Spatial Analysis of Geographical Flow [M]. Beijing:Scinece Press,2024.

[10] TAO R, THILL J. Spatial Cluster Detection in Spatial Flow Data[J]. *Geographical Analysis*,2016, 48(4): 355-372.

[11] GAO X,LIU Y,YI D,et al. A Spatial Flow Clustering Method Based on the Constraint of Origin-Destination Points' Location [J]. *IEEE Access*, 2020, 8: 216069-216082.

[12] CAI J, KWAN M. Detecting Spatial Flow Outliers in the Presence of Spatial Autocorrelation[J]. *Computers, Environment and Urban Systems*, 2022, 96: 101833.

[13] LIU Q, YANG J, DENG M, et al. BiFlowAMOEBA for the Identification of Arbitrarily Shaped Clusters in Bivariate Flow Data[J]. *International Journal of Geographical Information Science*, 2022, 36(9): 1784-1808.

[14] SHU H, PEI T, SONG C, et al. Density-Based Clustering for Bivariate-Flow Data[J]. *International Journal of Geographical Information Science*, 2022, 36(9): 1809-1829.

[15] CAI J, KWAN M. Discovering Co-location Patterns in Multivariate Spatial Flow Data[J]. *International Journal of Geographical Information Science*, 2022, 36(4): 720-748.

[16] 陆川伟,孙群,季晓林,等. 一种基于核距离的车辆轨迹点聚类方法[J]. 武汉大学学报(信息科学版), 2020, 45(7): 1082-1088.
LU Chuanwei, SUN Qun, JI Xiaolin, et al. A Method of Vehicle Trajectory Points Clustering Based on Kernel Distance[J]. *Geomatics and Information Science of Wuhan University*, 2020,45(7): 1082-1088.

[17] LIU Q, YANG J, DENG M, et al. SNN_Flow: A Shared Nearest-Neighbor-Based Clustering Method for Inhomogeneous Origin-Destination Flows [J]. *International Journal of Geographical Information Science*,2022, 36(2): 253-279.

[18] YAO X, ZHU D, GAO Y, et al. A Stepwise Spatio-Temporal Flow Clustering Method for Discovering Mobility Trends[J]. *IEEE Access*, 2018, 6: 44666-44675.

[19] 孙悦朋,郭仁拥,于涛. 基于密度聚类的多向行人流群集区域分布比较[J]. 山东科学, 2021, 34(5): 64-74.
SUN Yuepeng, GUO Renyong, YU Tao. Comparison Among the Collection Region Distributions of Multi-directional Pedestrian Flows Based on Density Clustering[J]. *Shandong Science*, 2021, 34(5): 64-74.

[20] 何少元. 不确定性数据聚类算法及其并行化研究[D]. 重庆:重庆邮电大学,2019.
HE Shaoyuan. Research on Uncertain Data Clustering Algorithm and Its Parallelization [D]. Chongqing: Chongqing University of Posts and Telecommunications, 2019.

[21] SONG C, PEI T, SHU H. Identifying Flow Clusters Based on Density Domain Decomposition[J]. *IEEE Access*, 2020, 8: 5236-5243.

[22] LIN H, YE Y. Reviews of Bearing Vibration Measurement Using Fast Fourier Transform and Enhanced Fast Fourier Transform Algorithms[J]. *Advances in Mechanical Engineering*, 2019, 11(1): 2072050763.

[23] 王婷,吴必虎. 基于关键词共现和社会网络分析的北京城市歌曲中地方意象特征研究[J]. 人文地理, 2020, 35(6): 57-65.
WANG Ting, WU Bihu. Research on Characteristics of Place Image in Beijing City Songs Based on Keywords Co-occurrence Analysis and Social Network Analysis[J]. *Human Geography*, 2020, 35(6): 57-65.

[24] LIU Q L, WU Z H, DENG M, et al. Network-Constrained Bivariate Clustering Method for Detecting Urban Black Holes and Volcanoes[J]. *International Journal of Geographical Information Science*, 2020, 34(10): 1903-1929.

[25] LIU Q, DENG M, BI J, et al. A Novel Method for Discovering Spatio-Temporal Clusters of Different Sizes, Shapes, and Densities in the Presence of Noise[J]. *International Journal of Digital Earth*, 2014, 7(2): 138-157.

[26] ESTER M, KRIEGEL H P, SANDER J, et al. A Density-Based Algorithm for Discovering Clusters

(下转第206页)