

基于地理流空间与空间交互网络的网约车运营模式分析

摘要：网约车作为当今城市居民出行的重要模式，研究其运营模式能够很好的反映居民的出行特点，地理流空间理论为分析城市居民出行模式提供了新视角。本文以武汉市的出租车轨迹数据作为数据源，利用基于地理流的L函数分析出行流的总体性质，基于流相似性度量挖掘出行流的模式分布；基于复杂网络理论与方法，构建热点区域交互网络，研究不同时间段人群出行特征。研究发现：①网约车出行流在空间上具有异质性。②工作日与节假日的出行模式具有差别。工作日影响力较高的区域为居民区与工作地，节假日影响力较高的区域为商圈与车站。③工作日主要出行流为“居民地—工作地”，节假日主要出行流为“商圈—居民地”或“商圈—商圈”。④节假日相对工作日跨越长江的交互更加频繁。本文发现可为车辆调度优化和交通资源分配提供支持。

关键词：流空间；流聚类；空间交互网络；网约车；社区探测；轨迹数据

1 引言

随着移动定位技术的快速发展，海量、高精度出租车运营数据为研究人群移动性以及城市空间结构的联系^[1-2]提供了数据支撑。目前已有众多学者对出租车轨迹数据进行分析与挖掘，但大多只关注特定的出行特征。本文更多关注城市人群出租汽车的总体出行模式，以地理流空间理论为主，复杂网络理论为辅，全面深入地分析城市人群出行模式，深化对城市人群出行模式的理解。

武汉市作为中国14座特大城市之一，人口众多，出行需求较高，因此交通资源的合理分配十分重要。本文以武汉市出租车轨迹数据为数据源，针对工作日与节假日的出行OD点数据进行分析。基于丛集、汇聚、发散和社区模式四种典型的出行流模式挖掘方法研究居民出行流的模式分布^[9-12]；基于复杂网络的理论和方法构建城市热点区域交互网络，研究城市热点区域的时空分布及交互作用，最后对比分析工作日与节假日的城市居民出行特征差异。

2 研究方法

2.1 流总体性质判断

流模式总体性质分为随机性和异质性^[3]，随机性表示流在流空间中呈现随机分布，不满足随机性则呈现异质性。本文采用Ripley's原始K函数的一种变体 $L(d)$ 来分析流总体性质^[13]：

$$L(d) = \sqrt{\frac{A \sum_{i=1}^n \sum_{j=1, j \neq i}^n k_{ij}}{\pi n(n-1)}} \quad (1)$$

其中 d 表示距离， n 为空间中流的总数目， A 代表流的总面积且 k_{ij} 为权重。当出行流 i 和出行流 j 距离大于 d 时， $k_{ij} = 0$ ，否则 $k_{ij} = 1$ 。对于给定的距离 d ，当 L 观测值大于 L 预期值，表明在该空间尺度的流模式呈现异质性，可进一步分析丛集模式、汇聚模式、发散模式与社区模式。

2.2 丛集模式挖掘

本文采用基于流相似性分析的层次聚类方法进行丛集模式挖掘。

通过DTW算法计算两流之间的归整路径距离分析流相似性。

$$\text{Dist}(W) = \sum_{k=1}^K \text{Dist}(w_{ki}, w_{kj}) \quad (2)$$

其中 w_{ki} 和 w_{kj} 分别为两条流的相接递增点, 满足:

$$w_{1i} = O_i; w_{1j} = O_j; w_{ki} = D_i; w_{kj} = D_j \quad (3)$$

$$I(w_{ki}) \leq I(w_{(k+1)i}) \leq I(w_{ki}) + 1; J(w_{kj}) \leq J(w_{(k+1)j}) \leq J(w_{kj}) + 1 \quad (4)$$

$$(I(w_{(k+1)i}) - I(w_{ki})) + (J(w_{(k+1)j}) - J(w_{kj})) \geq 1 \quad (5)$$

其中 O_i, D_i 为流 I 的起点和终点, O_j, D_j 为流 J 的起点, 终点。 $I()$, $J()$ 两函数分别用于求取路径点在流 I 和流 J 中的序号。

层次聚类算法采用将所有数据点分为不同的类簇并逐次迭代逐层合并的方式进行聚类, 类簇的合并根据类簇间的距离是否小于设定阈值进行, 最终类簇数等于目标类簇数时结束聚类。其时间复杂度为 $O(n^2)$ 。

不同类簇之间距离的计算方法采用离差平方和, 其表达式为:

$$\text{dist}(C1, C2) = \frac{1}{|C1| \times |C2|} \sum_{P_i \in C1, P_j \in C2} (\text{dist}(P_i, P_j))^2 \quad (6)$$

2.3 汇聚和发散模式挖掘

汇聚模式定义为流在流空间中终点聚集、方向随机或排斥、长度任意; 而发散模式与其差异为流起点聚集^[3]。本文采用基于丛集模式的识别方法, 识别汇聚和发散模式。

首先在 $flow_i$ 的OD点建立以 r 为半径的缓冲区, 如图1所示。

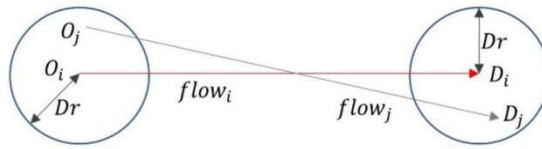


图1 OD出行流空间相似度关系图

其中, O_i, D_i, O_j, D_j 分别为出行流 $flow_i, flow_j$ 的O点和D点, 半径 Dr 是判断 $flow_i$ 与 $flow_j$ 空间相似的距离阈值。

设置距离阈值 Dr 为:

$$Dr = \begin{cases} \text{len}(f)/4 & \text{len}(f) \leq 2km \\ 500m & \text{len}(f) > 2km \end{cases} \quad (7)$$

汇聚模式其时空相似性为:

$$GAT_ST_Sim_{ij} = \left(1 - \frac{\text{func}(\text{ratioD})}{2} * \frac{\text{func}(\text{ratioT})}{2}\right) * 4/3 \quad (8)$$

发散模式时空相似性为:

$$DIV_ST_Sim_{ij} = \left(1 - \frac{\text{func}(\text{ratioO})}{2} * \frac{\text{func}(\text{ratioT})}{2}\right) * 4/3 \quad (9)$$

$$\text{ratioO} = \text{dist}(O_i, O_j) / Dr; \text{ratioD} = \text{dist}(D_i, D_j) / Dr \quad (10)$$

$$\text{ratioT} = \text{timeDiff}(O_i, O_j) / \text{timeSpan} \quad (11)$$

$$\text{func}(\text{ratio}) = \begin{cases} \text{ratio} + 1 & \text{ratio} \leq 1 \\ \infty & \text{ratio} > 1 \end{cases} \quad (12)$$

其中, timeSpan 为时间阈值。计算得到相似性指标^[8], 当 $GAT_ST_Sim_{ij}$ 与 $DIV_ST_Sim_{ij}$ 位于区间 $[0,1]$ 时, 说明 O_i 与 O_j 相似、 D_i 与 D_j 相似。

2.4 社区模式挖掘

地理流社区模式定义为在流分布的研究区域内, 划分多个子区域。在每个子区域内部出行流量较高, 而不同子区域间交互较少。本文采用基于模块度和社区发现算法来识别社区模式。社区挖掘算法采用基于

边介数删除弱连接的 GN 算法^[5]。GN 算法是一种分裂算法，将所有节点看成一个社区，通过删除边将社区划分为更小的社区^[6]。

3 数据及预处理

本文分析的数据为从网约车公司购得经过脱敏处理的武汉市网约车轨迹数据，选用 2018 年 11 月 12 日（星期一）作为工作日数据，2018 年 11 月 18 日（星期日）作为节假日数据。具体的数据项包括：订单编号，定位时间，网约车所处经纬度，订单的起点终点信息等。由于传感器设备故障会造成部分数据丢失，位置点偏移等现象，导致采集到的数据不符合客观逻辑。因此，我们对原始数据进行清洗，主要通过对特定数据项设置约束，路网匹配等方法完成了对数据的预处理，最终得到共计 32800 条数据。清洗后网约车的订单数据字段见表 1。

表 1 清洗后订单数据字段

ID	getonDate	getonLon	getonLat	getoffDate	getoffLon	getoffLat
601**3	1541995**9	114.280**2	30.589**1	1541995**0	114.28**0	30.589**3
601**8	1541995**9	114.260**3	30.628**8	1541995**0	114.29**0	30.630**5
601**4	1541995**1	114.314**3	30.515**5	1541996**0	114.38**4	30.507**2

4 研究结果及分析

4.1 流总体性质分析

本文主要分析武汉市工作日与节假日早晚高峰时段网约车的出行流模式差异。基于 K 函数的变体分析出行流总体性质。距离变化次数设置为 50。

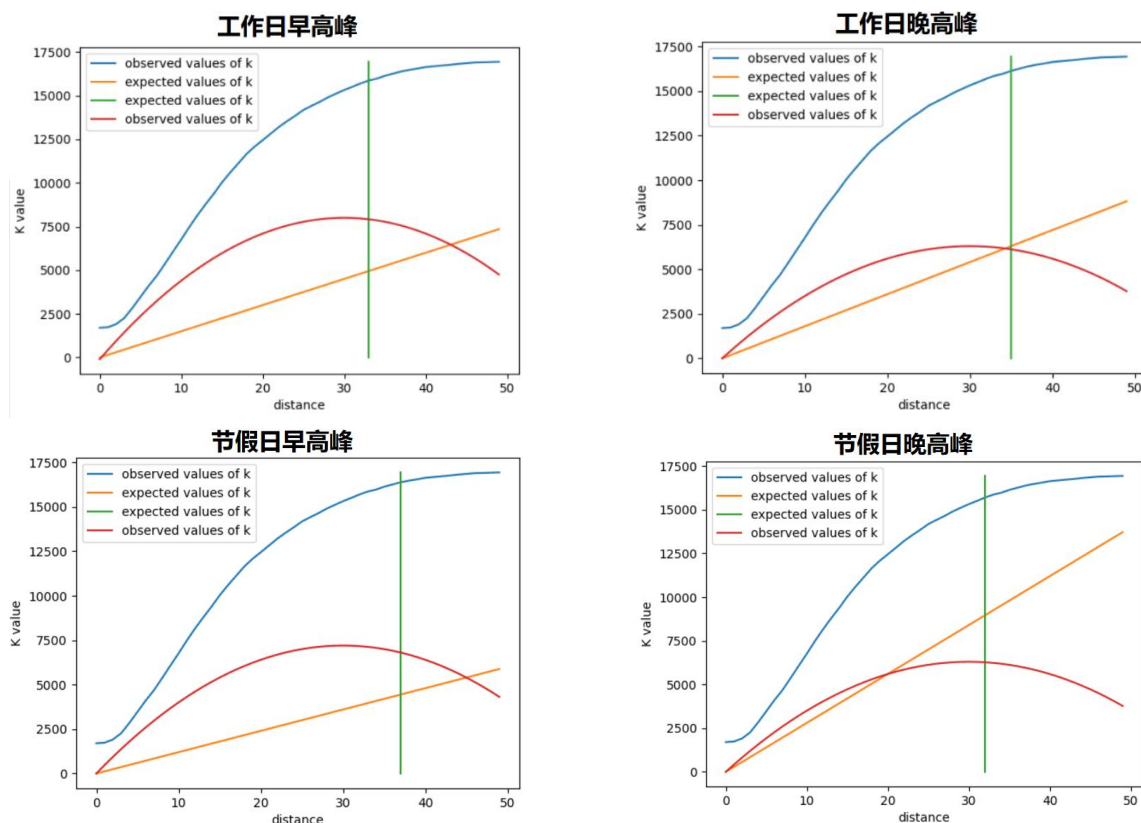


图 2 流总体性质图

图 2 显示网约车在工作日与节假日的早晚高峰出行流 K 函数观测值都大于期望值，说明出行流都呈现异质性，在流空间非随机分布。图中绿线代表此时观测 K 值与预期 K 值的差距最大，聚类层度最高。基于出行流流空间异质性，分别基于丛集模式、汇聚模式、发散模式和社区模式来分析工作日与节假日车辆运营特征差异。

4.2 丛集模式分析

基于时空相似性指标的层次聚类算法分别识别网约车工作日与节假日早晚高峰的出行流丛集模式。

如图 3 所示，工作日早高峰时段出行流丛集模式中，流量较高的五条为同济医学院到井冈山小学，关西小区到金地天悦，越秀星汇云锦到汉江公寓。多条丛集模式流汇聚于武广商圈 ICC 环贸中心和街道口商圈等。晚高峰高流量丛集模式为光谷软件园到中南民族大学，万象新城到列电小区。发现大多数丛集流起点为小区而终点为工作园区、医疗机构、学校等。

节假日早高峰时段出行流丛集模式中，流量较高的五条为 ICC 武汉环贸中心到同济医学院，复兴新村小区到凌云社区，以及柴零社区到景苑小区。大多丛集流起点位于居住地附近，终点为大型商场。晚高峰高流量丛集模式为南湖锦城到中南民族大学，光谷坐标园到东一产业园，巡司河体育公园到柏木岭北公园。大多数丛集流模式起点与终点都位于商圈或景区。

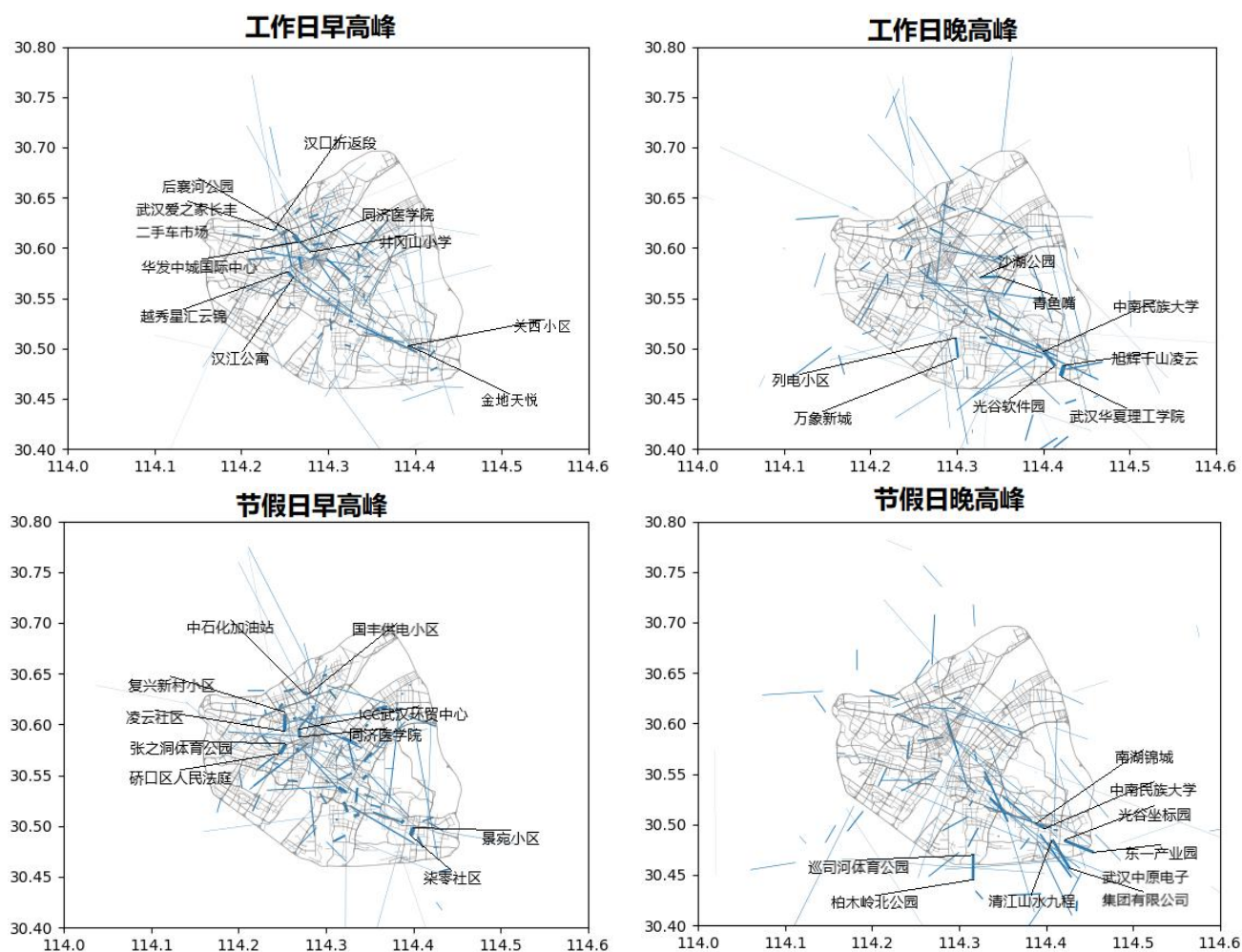


图 3 丛集模式

4.3 汇聚和发散模式分析

图 4、5 分别为网约车在工作日和节假日的汇聚和发散模式，分别呈现出不同的时空模式。较多知名地标既是聚集点也是发散点，如车站、商圈广场等。

图 4 中工作日聚集区域主要为商圈（街道口商圈、光谷广场商圈）、工作区（武汉市青少年宫周边）、交通设施（汉口火车站）等。节假日网约车聚集区域分布在汉口区（ICC 武汉环贸中心、循礼门周边）、雍豪府、同馨花园居民区、武汉摩尔城。

图 5 中工作日发散区域主要为居民区，如硃口公园周边居民区，武汉船舶学院周边居民区，由人群上班通勤、就医等方式驱动，其中大多数都在工作区域附近，如环贸中心、万象城、光谷广场等。节假日发散区域集中于交通设施（汉口火车站）、商场（华发中城国际中心）、景区（解放公园）。相比于工作日，节假日发散区域大多位于长江两岸。

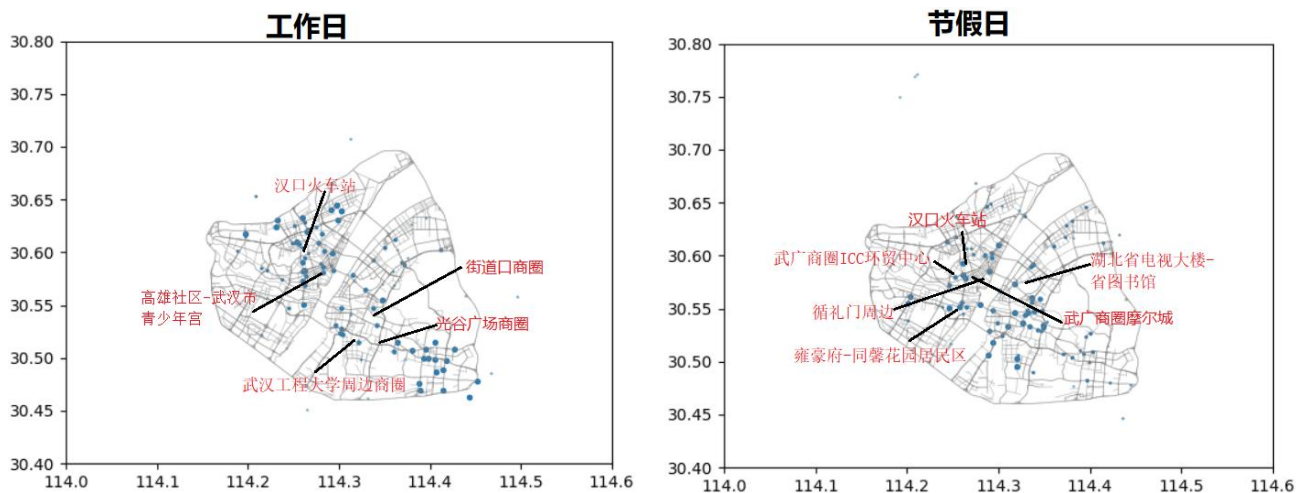


图 4 汇聚模式

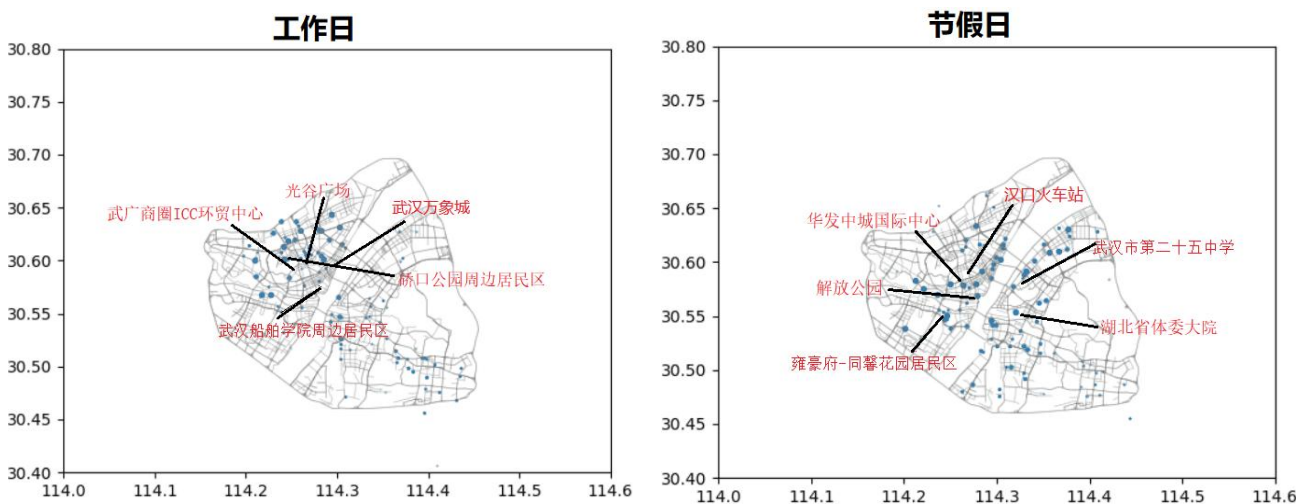


图 5 发散模式

4.4 热点区域交互网络分析

对工作日与节假日的 OD 点数据进行 DBSCAN 聚类，利用上下车点热点数据构建热点空间交互网络。并对出行热点交互网络进行指标测度计算。结果如表 2 所示。

由图 6 可以得出，工作日和节假日的热点交互网络基本都以中间两个节点为中心。但相对来说工作日交互网络中节点间联系更加紧密，节假日交互网络中节点分布更加均匀分散。说明在工作日，由于上班人们的出行范围固定为某一小区域。而节假日人们的娱乐方式多样，相对来说出行的范围也更大，不同热点区域间交互更加平均。

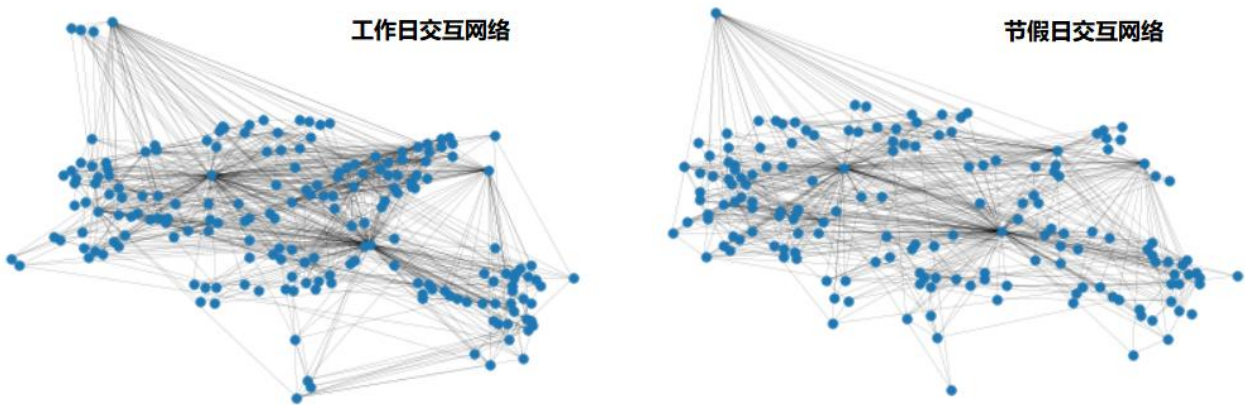


图 6 热点交互网络

表 2 网约车工作日、节假日出行网络参数

	工作日	节假日
节点数	193	166
边数	978	738
平均度	10.13	8.89
网络密度	0.053	0.054
平均聚类系数	0.613	0.615
路径平均长度	2.07	2.02
度同配性	-0.29	-0.30

区域网络中的度中心性体现该热点区域与其他热点区域间的交互作用^[7]。度中心性越高，表明该热点区域在网络中的重要性越高。图 7 为热点区域网络中度中心性排名前 20 的节点。表 3 为其前 5 个节点的具体位置。从表中可知，在工作日由于居民的工作需求，工作地如环贸中心，居民区如湖北大学周边居民区与其他热点区域联系跟为密切，度中心性较高的为工作地和居民区。而节假日由于娱乐需求，商圈如街道口商圈更加活跃，度中心性较高的多为商圈和车站。此外工作日和节假日的出行网络均呈现出小部分节点值很高，大部分节点值很低，说明出行网络具有“无标度性”。

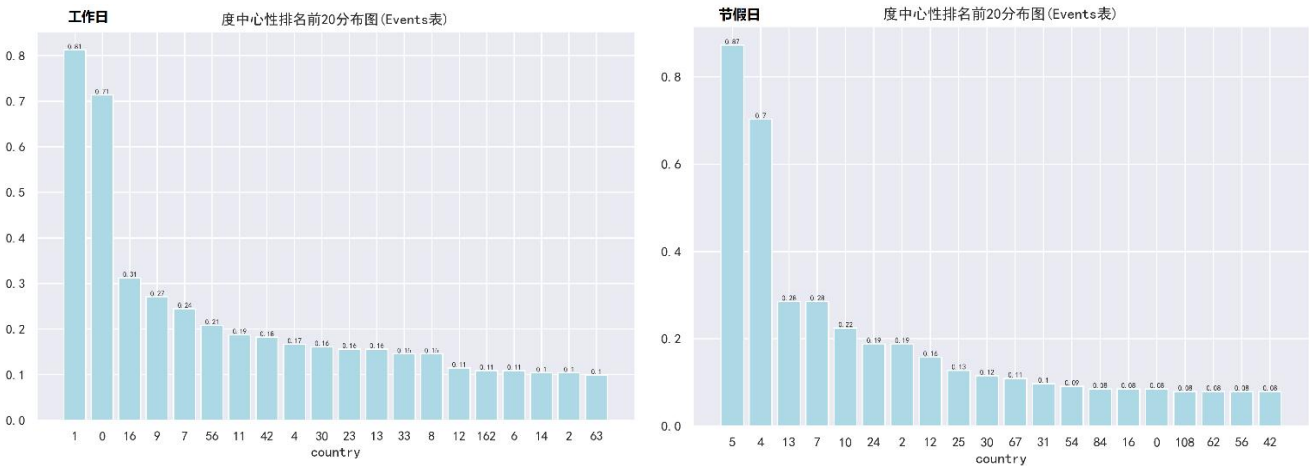


图 7 节点度中心排名前 20

表 3 度中心性 Top5 统计表

排序	工作日	节假日
1	武汉环贸中心，武广商圈	宝通寺-街道口商圈
2	街道口商圈	武汉环贸中心，武广商圈
3	湖北大学及周边居民区	钢都花园及附近商圈
4	武广商圈(创世纪广场)	汉阳区法院-名流公馆居民区
5	南湖-建安街居民区	武汉火车站

区域网络中边权重越大说明对应的热点区域间交互较多。对交互网络中所有连边进行权重排序，其中边权值最大的 4 条边对应的区域如表 4 所示。可以发现，在工作日，居民出行模式多为“工作地—居住地”；而在节假日，则多为“商圈—居住地”，或“商圈—商圈”出行模式。

表 4 边权值 Top4 统计表

排序	工作日	节假日
1	武汉环贸中心—街道口商圈	武汉环贸中心—街道口商圈
2	武汉环贸中心—创世纪广场	街道口商圈—钢都花园及附近商圈
3	街道口商圈—湖北大学及周 边居民区	武汉环贸中心—武汉第八医院居民区
4	街道口商圈—南湖-建安街居 民区	街道口商圈—武汉火车站

4.5 社区模式分析

基于 GN 算法和模块度划分标准对武汉市工作日、节假日的网约车出行网络进行社区划分。

从图 8 可见，工作日出行网络被划分为 4 类，除天河机场，武汉站外，所有热点区域以长江为界被划分为两类。节假日出行网络被划分为 8 类，除天河机场外，其他几类相互渗透，在地域上并没有呈现明显的界限。结果表明，工作日出行热点区域之间交互半径较短，人们更倾向于在居住地不远处工作。而节假日出行热点区域之间交互在地域上更广，跨江出行更加频繁。

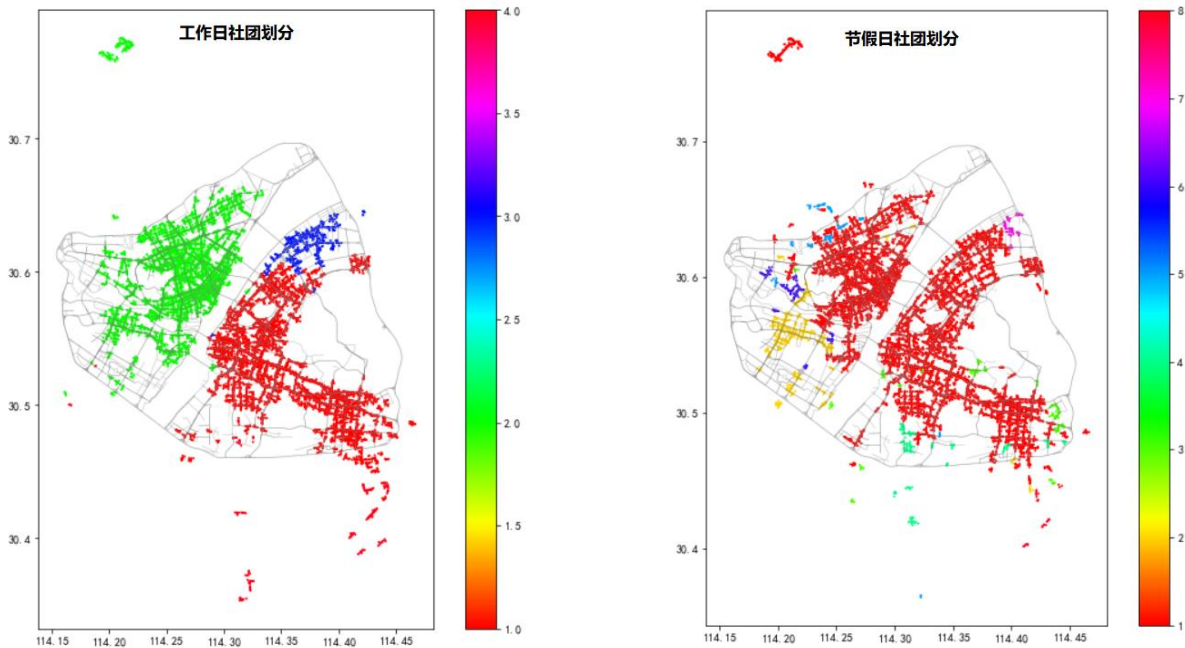


图 8 社区模式

5 讨论与结论

本文基于武汉市出租车轨迹数据提取 OD 点数据并构造出行流。通过定义流相似性指标,基于层次聚类识别流丛集模式;通过定义模块度指标识别流社区模式;通过聚类 OD 点数据识别流汇聚与发散模式;通过研究工作日、节假日城市热点区域的分布模式,为制定分时段的交通管理政策提供参考。

本研究仅采用了工作日与节假日的订单数据来研究网约车人群出行流模式,后续研究中可进一步分析出行模式在不同时间尺度下的差异;可通过分析新冠疫情后的数据与本文成果进行对比,研究疫情对居民出行模式的影响。其次,也可考虑地理环境因素等对不同出行流模式的影响,例如结合土地利用、路网、POI 数据等进行分析,更全面的指导车辆资源调度与交通管理。

参考文献:

- [1] Liu X, Gong L, Gong Y, et al. Revealing travel patterns and city structure with taxi trip data[J]. Journal of Transport Geography, 2015,43:78-90.
- [2] Zhong C, Schlöpfer M, Müller Arisona S, et al. Revealing centrality in the spatial structure of cities from human activity patterns[J]. Urban Studies, 2017,54(2):437-455.
- [3] 裴韬,舒华,郭思慧,等.地理流的空间模式:概念与分类[J].地球信息科学学报,2020,22(1):30-40. [Pei T, Shu H, Guo S H, et al. The concept and classification of spatial patterns of geographical flow[J]. Journal of Geo-information Science, 2020,22(1):30-40.]
- [4] 刘入嘉.基于出租车 GPS 数据的城市出行空间结构研究[D].东南大学,2018.
- [5] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proceedings of the national academy of sciences, 2002,99(12):7821-7826.
- [6] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical review E, 2004,69(2):026113.
- [7] 周勃,秦昆,陈一祥,李志鑫.基于数据场的出租车轨迹热点区域探测方法[J].地理与地理信息科学,2016,32(6):51-56.
- [8] 项秋亮,邬群勇,张良盼.一种逐级合并 OD 流向时空联合聚类算法[J].地球信息科学学报,2020,22(6):1394-1405. [Xiang Q L, Wu Q Y, Zhang L P. An OD flow spatio-temporal joint clustering algorithm based on step-by-step merge strategy[J]. Journal of Geo-information Science, 2020,22(6):1394-1405.]
- [9] 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.
- [10] Chen Y, Xu J, Xu M. Finding community structure in spatially constrained complex networks[J]. International Journal of Geographical Information Science, 2015,29(6):889-911.
- [11] Guo D, Jin H, Gao P, et al. Detecting spatial community structure in movements[J]. International Journal of Geographical Information Science, 2018,32(7):1326-1347.
- [12] 杨喜平,方志祥,赵志远,等.城市人群聚集消散时空模式探索分析——以深圳市为例[J].地球信息科学学报,2016,18(4):486-492. [Yang X P, Fang Z X, Zhao Z Y, et al. Exploring urban human spatio-temporal convergence-dispersion patterns: a case study of Shenzhen city[J]. Journal of Geo-information Science, 2016,18(4):486-492.]
- [13] Shu H, Pei T, Song C, et al. L-function of geographical flows[J]. International Journal of Geographical Information Science, 2021,35(4):689-716.