基于线图视角的城市社区发现可视分析方法

林正轩[1,2)], 王超[1)*], 邵彬[3)], 邓紫坤[4)], 王永恒[3)], 巫英才[2)]

[1)] (之江实验室数据枢纽与安全研究中心 杭州 311121)
[2)] (浙江大学 CAD&CG 国家重点实验室 杭州 310058)
[3)] (之江实验室天文计算研究中心 杭州 311121)
[4)] (华南理工大学软件学院 广州 510006)
(wangchaogis@zju.edu.cn)

摘 要: 现有的社区发现方法通常以城市区域作为研究对象, 探究热点区域间的空间交互关系, 然而在对OD(origin-destination)对之间关联关系的表达上存在一定局限性, 难以开展面向 OD 对的城市社区结构发现. 为解决上述问题, 提出一种融合线图和轨迹表征的城市社区发现方法. 首先基于出行轨迹数据构建传统的空间交互网络; 然后交换空间交互网络的节点和边, 建立以 OD 对特征为节点、以 OD 流的起点和终点为边的线图视角下的新型轨迹交互拓扑图; 最后基于轨迹交互拓扑图, 使用顾及节点属性和拓扑结构的社区发现算法开展城市社区发现. 进一步地, 设计并实现了基于线图视角的城市社区发现可视分析系统, 帮助用户探究能够反映人们真实活动空间的城市子区域, 为城市管理提供参考. 通过 2 个案例和用户评估实验, 证明了所提方法及系统的实用性与有效性.

关键词: 社区发现; 空间交互; 可视分析; OD 流
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023-00622

A Visual Analytic Approach for Discovering Urban Community Based on Line Graph

Lin Zhengxuan[1,2)], Wang Chao[1)*], Shao Bin[3)], Deng Zikun[4)], Wang Yongheng[3)], and Wu Yingcai[2)]

[1)] (*Research Center for Data Hub and Security, Zhejiang Lab, Hangzhou* 311121)
[2)] (*State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou* 310058)
[3)] (*Research Center for Astronomical Computing, Zhejiang Lab, Hangzhou* 311121)
[4)] (*College of Software, South China University of Technology, Guangzhou* 510006)

Abstract: The existing community discovery methods usually focus on urban areas as research objects, exploring the spatial interaction relationships between hotspot areas. However, there are certain limitations in expressing the correlation between OD (origin destination) pairs, making it challenging to conduct urban community structure discovery for OD pairs. Therefore, this paper proposes an urban community discovery method that integrates trajectory representations based on the line graph. First, a traditional spatial interaction network is constructed based on travel trajectory data. Then, the nodes and edges of the traditional topological graphs are exchanged to establish a new trajectory interaction topology graph from the perspective of a line graph with OD pair features as nodes and starting points and end points of the OD flow as edges. Finally, this paper uses a community discovery algorithm that considers node attributes and network topology to discover urban communities based on the trajectory interactive topological graph. Furthermore, a visual analytics system is designed for urban community

收稿日期: 2023-09-20; 修回日期: 2024-03-27. **基金项目**: 国家重点研发计划(2022YFE0137800); 国家自然科学基金青年科学基金(42201510); 国家自然科学基金区域联合基金(U22A2032); 浙江省自然科学基金(LTGG24D010003). **林正轩**(1998—), 男, 硕士研究生, 主要研究方向为可视化; **王超**(1993—), 女, 博士, 副研究员, CCF 会员, 论文通信作者, 主要研究方向为大数据分析; **邵彬**(1993—), 男, 硕士, CCF 会员, 主要研究方向为大数据挖掘; **邓紫坤**(1996—), 男, 博士, 副教授, 博士生导师, CCF 会员, 主要研究方向为可视化; **王永恒**(1973—), 男, 博士, 副研究员, CCF 会员, 主要研究方向为大数据分析; **巫英才**(1983—), 男, 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为可视化、可视分析、大数据分析.

discovery from the line graph perspective, which helps users explore urban sub-regions that can reflect people's real activity space, and provides reference for urban management. The practicality and effectiveness of the proposed method and system have been demonstrated through two case studies and user evaluation experiments.

Key words: community discovery; spatial interaction; visual analytics; origin-destination flow

开展合理的城市分区管理一直是城市国土空间规划面临的关键问题, 对于统筹空间利用、完善路网建设、提升城市数字化水平具有重要意义[1]. 传统的城市分区管理往往依赖自上而下的行政区域划分, 难以细粒度地揭示居民出行活动的行为规律和移动模式[2]. 近年来, 随着物联网、云计算以及人工智能技术的快速发展, 城市出行轨迹数据已经成为微观尺度下刻画人类社会活动不可或缺的重要支撑, 基于轨迹数据开展自下而上的城市空间结构分布研究成为一种新的趋势[3].

基于交通网络的社区发现是最广泛使用的城市分区方法之一[4]. 这类方法将道路网络中的热点区域视为节点, 节点间的 OD(origin-destination)对视为边, 从而构造出空间交互网络; 然后在该网络运行社区检测算法, 得到城市分区(社区)结果. 每个分区可作为城市分析单元运用到后续的城市大数据分析应用中, 如交通模式分析、OD 需求预测、公交线路优化等. 然而, 这类方法存在以下局限性:

(1) 城市内部道路网络结构复杂, 热点区域众多, 构建的空间交互网络具有节点数目多、密度不均等特点, 城市规划领域的专家难以快速地从这种复杂且密集网络中获取有价值的信息;

(2) 常规建立的空间交互网络只能揭示热点区域节点间的可达性, 对 OD 对间潜在的时空关联关系的揭示存在局限性;

(3) 对复杂拓扑网络中轨迹边特征的建模存在不足, 不易体现 OD 对内路线选择的多样性, 难以对轨迹之间的关联关系进行挖掘.

为了解决上述问题, 本文提出一种融合线图和轨迹表征的城市社区发现方法, 突破常规城市空间交互网络的构建模式, 交换复杂拓扑图的节点和边, 建立以 OD 对特征为节点、以 OD 流的起、终点为边的新型交互网络拓扑图, 在此基础上开展融合轨迹特征表达的细粒度城市社区发现. 面对如何将复杂拓扑图、城市空间上下文、轨迹特征等相互异构的动态时空信息互相关联起来的挑战[5], 本文提出一种基于线图视角的可视分析方法, 并实现了相应的可视分析系统, 帮助专家开展面向交通 OD 流的城市分区, 并支持从 OD 对和轨迹 2 个层面进行分析, 以理解交通模式和规律. 最后, 基于真实的杭州市出租车轨迹数据, 挖掘兴趣区域内 OD 对间潜在关联关系, 并通过用户评估实验证实了该可视分析方法和系统的有效性.

(1) 提出融合线图和轨迹表征的城市社区发现方法, 考虑了 OD 对中多样化路线的轨迹特征, 以期为探究不同时间段下挖掘 OD 对之间的关联关系, 以及城市网络空间结构的模式发现提供研究思路上的启发.

(2) 开发了线图视角下的城市社区发现可视分析系统, 使用户可通过数据视图、GIS 视图以及力导向视图分析特定时间和空间范围内的 OD 对连通关系, 识别细粒度的城市社区.

(3) 通过 2 个案例的研究和用户实验探究, 证实了该方法及系统的实用性与有效性.

1 相关工作

1.1 基于交通数据的社区发现

任一复杂系统可以视作由内部节点和节点之间的连接边构成的复杂网络, 网络中的节点构成的具有一定意义的内聚子图被称作社区[6]. 社区发现算法[7]是一类揭示复杂网络中社区结构的技术, 可用于社交关系挖掘[8]、通勤社区发现[9]以及疾病检测[10]等研究. 在城市规划领域, 研究人员可以利用基于位置的服务技术产生的海量数据, 采用社区发现算法测度区域之间的关联紧密程度, 并将挖掘的社区结构与现有城市规划方案进行对比评价, 对城市规划管理提供相应的参考[11].

常规的交通网络图构造方法通常以热点区域为节点, 以车辆在热点区域间的上下车关系为边, 将上下车次数作为网络边的权值[12]. Chen 等[13]使用核密度估计、密度聚类等方法探究不同区域的交通便捷性, 并借鉴图社区发现的思想研究网络内部的聚类特征和节点间的交互异质性. 但是, 现有方法只顾及热点区域节点间的连通性, 缺乏对节点间轨迹多样性特征的考虑.

出租车 GPS 轨迹数据是一种记录人类行为活动的出行大数据，基于出租车轨迹数据开展社区发现，可以帮助人类从复杂网络视角挖掘城市中更精细的区域结构. Zhao 等[14]使用社区发现方法检测路网和出租车出行 OD 网络的子结构，评价城市空间的相互作用并追踪其动态演化过程；Liu 等[15]采用上海出租车出行数据，基于社区发现视角探究子区域结构；周博等[16]基于复杂网络分析方法，运用空间可视化技术分析深圳热点区域之间的交互性. 但是，现有方法仅考虑了车辆的流动特征和各区域之间的空间交互关系，缺乏对周边兴趣点(point of interest, POI)信息和轨迹信息的综合建模.

1.2　OD 流可视化

城市中不同区域并非孤立的，而是通过各种方式连接在一起. 在交通研究中，居民出行数据揭示了居民在不同区域间的移动规律，可以很好地刻画区域间的连通关系. 其中，出行轨迹数据的起点和终点组成 OD 对，大量的 OD 对数据汇聚成 OD 流. 通过对 OD 流数据的可视化，可以呈现数据中的隐藏信息.

现阶段，OD 流数据的可视化方法有 OD 矩阵、OD 嵌套地图和 OD 流图 3 种. OD 矩阵可以看作是图的矩阵表示，行表示起始点，列表示终点，矩阵元素的颜色编码表示 OD 对数量[17]. OD 嵌套地图利用嵌套方式解决 OD 矩阵空间信息表达不够直观的问题，如 Yang 等[18]提出的 MapTrix 方法通过对矩阵旋转以及连线重排算法，清晰地呈现起终点间的流量关系. OD 流图以起点和终点作为节点，以节点间的有向边表示 OD 对的方向. 为了考察区域间的联通关系及联通频次，Zhou 等[19]在图的基础上，基于节点大小、边的宽度等方式展示 OD 流频次信息，并通过不同时段拓扑图中节点和边形态的变化，宏观地反映区域间车辆流动模式. 在传统流图中，流之间的交叉重叠问题会严重影响数据的可视化效果，von Landesberger 等[20]基于 DBSCAN 算法对图中的节点进行合并，简化有向图后再进行可视化展示；Yang 等[21]提出的 MapsLink 方法实现了三维地球和流图的协同可视化，以解决视觉混淆问题. 但是，现有方法仅考虑区域连通性和频次的可视化表达形式，缺乏对节点属性和联通路径多样性信息的呈现.

此外，OD 流隐含了时空和语义等信息[22]，使用可视分析系统对出行轨迹数据进行多维度的探索成为一种研究趋势. 例如，Al-Dohuki 等[23]将轨迹采样点映射为街道名称和 POI，从语义、行驶速度等方面为轨迹建立索引，通过一个可视分析系统完成对轨迹数据的管理和分析. 罗月童等[24]将轨迹序列与城市道路进行匹配，学习道路的特征表示，设计了一个可视分析系统，从道路特征相似度、道路地理位置和 POI 信息等视角探索各时间段的道路交通情况. 上述研究工作都为本文建设基于轨迹数据的可视分析系统奠定了坚实的基础.

2　研究背景

2.1　需　求

为了深入理解城市规划师在交通社区发现中的具体需求，与 3 位来自该领域的专家开展了密切的合作，将 3 位专家记为 E1，E2 和 E3. 其中，E1 为有 8 年地理科学研究经验的资深研究人员，E2 为有 8 年交通数据分析经验的资深研究人员，E3 为有 20 年可视分析经验的资深研究专家. 通过每周与这 3 位专家进行一次讨论，最终明确了该领域从业人员在探测细粒度城市子区域过程中的需求和挑战，并归纳出以下设计需求.

R1. 研究区域的筛选. 面对海量的出租车轨迹数据，用户需要了解数据在时间维度的大致分布情况并进行快速定位. 系统需要以地图为载体，为用户提供交互式的兴趣区域选择方式.

R2. 线图视角下的 OD 对可视化. 线图不仅需要表示研究区域内 OD 对的拓扑结构，还需要编码 OD 对的行程长度、轨迹数量、空间相对距离和 POI 分布等信息. 用户在与线图交互时，地图部分同时发生联动，以帮助用户查看线图结构的对应关系.

R3. 社区发现与结果展示. 为了满足用户开展线图视角下融合轨迹特征的社区发现需求，系统需要提供一种社区发现可视化方法，并在结果的可视分析过程中忽略一些低参考价值的子区域.

2.2　问题定义

线图是由 Whitney[25]提出的一个图论中的概念. 给定一个空间交互网络拓扑图，本文需要将其转换成代表 OD 流拓扑结构的有向线图，在线图上执行融合了轨迹特征的社区发现算法，可以发现由若干属性和拓扑结构相似度较高的节点组成的区域子图.

对于一个原图 $G=(V,E)$，其对应的线图为 $L(G)=(E,\tilde{E})$，且 $(e_1,e_2)\in\tilde{E}\Leftrightarrow e_1\cap e_2\neq\varnothing$. 其中，$\tilde{E}$ 表示线图的边集. 假设 e_1 和 e_2 是线图中的任意 2 个节点，它们之间存在一条边当且仅当它们在原图 G 中对应的边相邻. 由于本文需要基于有方向

的 OD 对组成的图建立新的线图，因此要在定义线图时加入有向边约束，以构建有向线图．为了弥补传统社区发现过程中未考虑 OD 对中路径选择多样性的缺陷，提取 OD 对在不同时间段的轨迹特征，作为线图节点的属性．

2.3 数 据

本文收集了杭州市出租车 GPS 轨迹数据和 POI 数据．

出租车GPS轨迹数据蕴含大量的居民出行信息，不仅具有更新频率快、数据量大、覆盖率广等优点，更是出租车司机依据经验择优选择的结果，能较好地反映真实的居民空间活动规律，有利于发现空间形态多样且具有跨行政区特性的交通分析单元．本文采用由浙江省杭州市交通运输局提供的 2020-01—2020-05 的真实出租车 GPS 轨迹数据作为实验数据集．原始的轨迹数据包含车辆在每一时刻的经纬度坐标、时间戳、载客状态以及速度等信息，本文提取载客状态下的轨迹数据并进行清洗，载客段轨迹 $\boldsymbol{T}_j$ 表示为 $\{(p_1,p_2,\cdots,p_i,\cdots,p_n),d_j\}$．其中，$p_i=(x_i,y_i,t_i)$ 表示轨迹 $\boldsymbol{T}_j$ 的第 i 个轨迹点；x_i，y_i 和 t_i 分别表示轨迹点 p_i 的经度、纬度和时间戳；d_j 表示当前轨迹发生的日期．

为了结合地理语义对居民的出行模式进行分析，本文通过高德 API 接口获取了杭州市的 POI 数据，包含实体名称、具体地址、经纬度坐标等信息，涵盖餐饮、住宅、医疗等 14 种城市功能类型．

3 本文方法

本文基于图论中线图的概念，提出一种融合线图和轨迹表征的城市社区发现方法，包含空间交互网络构建、轨迹交互拓扑图构建、轨迹表征建模以及融合线图和轨迹表征的城市社区发现 4 个核心模块，其流程如图 1 所示．

空间交互网络构建模块通过三角剖分聚类方法获取 OD 点簇作为节点，以 OD 点簇之间的出入关系作为边，构建网络的拓扑结构．轨迹交互拓扑图构建模块将空间交互网络转换为线图．轨迹表征建模模块提取轨迹数据的特征向量．融合线图和轨迹表征的城市社区发现模块执行同时考虑节点属性和拓扑结构的区域子图社区发现．

3.1 空间交互网络的构建

OD 对保留了轨迹的起终点信息，是一种对轨迹的粗粒度表达，用于较简洁地探究居民出行活

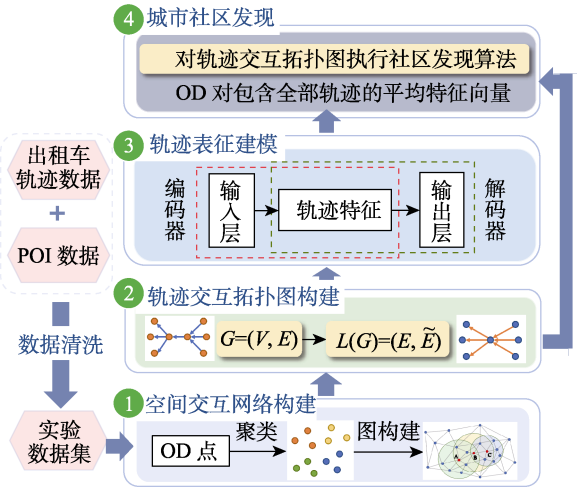


图 1 本文方法流程

动规律．研究人员常基于出行轨迹数据选择居民活动较频繁的空间位置作为热点区域，开展热点区域间的 OD 对关联关系研究．传统的热点区域划分方法包括按照网格划分、对轨迹点进行聚类或依据行政区域划分．为了让空间划分的尺度足够灵活，使得划分结果既能够根据轨迹点的密度反映周边地理实体的聚集情况，又能够保证热点区域的空间大小在可控范围内，本文采用三角剖分聚类算法[26]对轨迹端点进行划分．首先从轨迹数据中提取起点和终点坐标，构成 OD 点集合；然后对集合内的所有空间节点构建三角剖分图；最后计算三角剖分图中所有边的长度．假设任意边 e 具有 A 和 B 这 2 个端点，2 点间的距离计算公式为

$$D(A,B)=1-\frac{W\left(N(A,k)\cap N(B,k)\right)}{W\left(N(A,k)\cup N(B,k)\right)}.$$

其中，$N(A,k)$ 表示包括 A 及其 k 个最近邻居共同组成的点集合；$W\left(N(A,k)\right)$ 表示集合中点的个数．如图 2 所示，$D(A,B)=1-5/11=0.45$，$D(B,C)=1-4/12=0.67$，因此，B 到 A 的距离小于 B 到 C 的距离．

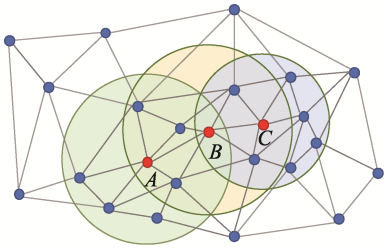


图 2 三角剖分聚类的距离计算示意图

对所有边的长度进行升序排列，并将每个节点都看作一个簇．遍历边集合，对于每条边 e 的端

点 u 和 v，查询它们所在的集合 C_u 和 C_v．若 $C_u \neq C_v$ 且 $W(C_u)<\theta$ 或 $W(C_v)<\theta$，则合并这 2 个集合，其中，$W(\cdot)$ 表示查询簇大小的操作函数；θ 表示约束每个簇最小范围的预设参数．聚类后的每个簇是一个 OD 点簇，可以被看作空间交互网络的节点．通过分析 OD 点簇之间的出入关系，在空间交互网络中添加有向边，构建网络的拓扑结构．

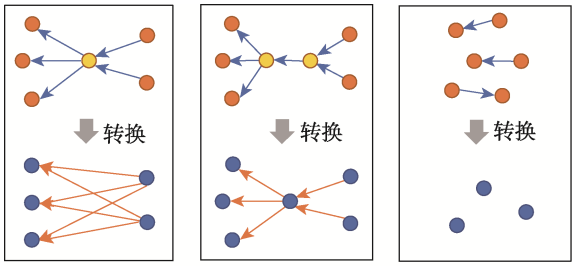
3.2　轨迹交互拓扑图的构建

在第 3.1 节构建的空间交互网络拓扑图(记为 G)的基础上，将其节点和边进行转换，生成新的有向线图(记为 L)，本文称为轨迹交互拓扑图．

Step1. 初始化 $L=\varnothing$，G 中的 OD 对集合记作原图 G 的有向边集 E，在线图 L 中为每条边 $e\in E$ 初始化一个顶点 v，v 中保留了 e 包含的 OD 对数据．

Step2. 遍历 G 中的任意边 e_1 和 e_2，若 e_1 和 e_2 共享了公共顶点 u，且 u 既是 e_1 的终点也是 e_2 的起点，则在 L 中添加一条从 v_1 到 v_2 的有向边，同时在这条边中保留 u 包含的 OD 点簇信息．

在构建的轨迹交互拓扑图中，存在 3 类与原图具有对应关系的典型结构，其中蕴含了具有一定现实意义的交通特征，如图 3 所示．



a. 全连接结构　　b. 桥形结构　　c. 散点结构

图 3　轨迹交互拓扑图中的典型结构

将图 3 中的 3 种典型结构的对应关系分别记为模式①、模式②和模式③，每个分图的上图代表原图，下图代表线图视角下的新图．其中，新图中的点代表 OD 对，边代表 OD 点簇．模式①中，黄色节点代表的 OD 点簇被频繁经过，可能是潜在的热点区域；模式②表明，2 个黄色节点代表的 OD 对两端的交通流量大，如果该 OD 对内的轨迹数较多，则其作为 2 个热点区域的连接可能存在交通瓶颈；模式③表明，存在一些零散的 OD 对，它们彼此之间没有相互关联，通常代表低频的 OD 对．在线图视角下，轨迹交互拓扑图的结构由若干相连或分离的相同模式或不同模式组合而成，不同结构之间分界明显．这 3 类典型结构可以帮助用户在线图视角下快速地判断出交通流特征．

为了在轨迹交互拓扑图中保留原图 OD 对间的空间距离关系，本文定义 OD 对之间的距离计算公式为

$$d=\min\left(F(o_1,o_2)+F(d_1,d_2),F(o_1,d_2)+F(o_2,d_1)\right) \quad (1)$$

其中，d 表示 2 个 OD 对之间的空间相对距离；$F(\cdot)$ 表示 2 个坐标点间欧几里得距离的计算函数；o_1 和 d_1 以及 o_2 和 d_2 分别表示 2 个 OD 对的起点和终点．由此可以在可视化展示中，将 OD 对之间的空间距离相对关系融入轨迹交互拓扑图中，揭示 OD 对的空间结构．

3.3　轨迹表征建模

本文基于序列到序列(sequence-to-sequence, Seq2Seq)[27]框架，采用门控循环单元(gated recurrent unit, GRU)[28]学习轨迹数据的有效特征表示．首先将研究区域划分为均等大小的空间网格，将训练数据中所有轨迹点的经纬度坐标映射至网格，视为自然语言处理领域中的离散标记(token)，筛选命中次数超过 k 的网格作为词汇表 V，每条轨迹可表示为 token 序列[29]；然后取轨迹数据集中的任一条轨迹为 x，通过对 x 进行一定程度的失真处理，如添加噪声、随机丢弃部分采样点等，生成低采样率的轨迹 x'．模型的训练目标为最大化条件概率 $P(x|x')$，即让 x' 经过编码和解码后尽可能地接近原始轨迹 x，损失函数[30]为 $L=-\ln P(x|x')$．

3.4　融合线图和轨迹表征的城市社区发现

为了在社区发现过程中同时考虑图中的节点属性和拓扑结构，基于 SA-Cluster 算法[31]开展融合线图和轨迹表征的城市社区发现研究．SA-Cluster 算法首先基于欧几里得距离计算节点属性向量之间的相似度，然后通过随机游走距离度量拓扑结构的相似度，最终以二者加权平均的结果作为距离度量．

本文将线图形式的轨迹交互拓扑图作为社区发现算法的输入，图中的每个节点表示一个 OD 对，同一 OD 对可能在不同时间段具有不同的热点路线，且不同的路线有着不同的特征向量，蕴含着差异化的城市潜在路线信息．将一个 OD 对的特征向量表示为若干个轨迹特征向量的聚合形式．对于一个 OD 对，即线图中的节点 $v_i\in V$，将其属性定义为 $\boldsymbol{F}_{v_i}=\mathrm{avg}\left(\boldsymbol{F}_{t_j}\right), t_j\in\boldsymbol{T}_{v_i}, v_i\in V$．其中，$\boldsymbol{T}_{v_i}$ 表示节点 v_i 代表的 OD 对中的所有轨迹，对 $\boldsymbol{T}_{v_i}$ 中所有轨迹的特征 $\boldsymbol{F}_{t_j}$ 计算均值得到 $\boldsymbol{F}_{v_i}$，作为节点属性．由此，本文在 SA-Cluster 算法中融入轨迹特征表示，以挖掘城市在不同时间段的空间结构．该算

法最终将一个带有属性的图分成多个密集连接子图，子图中的节点具有相似的属性.

4 本文系统

4.1 系统介绍

根据设计的需求，本文设计并实现了一个线图视角下的城市社区发现可视分析系统，包括 5 个可视化视图，其可视界面如图 4 所示. 其中，数据视图包含轨迹数据选择和概览功能；GIS 视图提供 OD 点数据的聚类及展示功能；力导向视图提供线图视角下的轨迹交互拓扑图以及社区划分结果的展示功能，且用户与该视图的交互操作会同步在 GIS 视图中产生联动，帮助用户发现 OD 对之间的关联关系；轨迹详情视图根据用户在力导向视图中选择的节点，展示节点包含轨迹的详细信息和路况概览；轨迹特征视图展示了轨迹交互拓扑图所含的轨迹的特征在二维空间的分布，并按社区发现结果进行颜色区分，帮助用户了解轨迹特征和社区划分的相关关系.

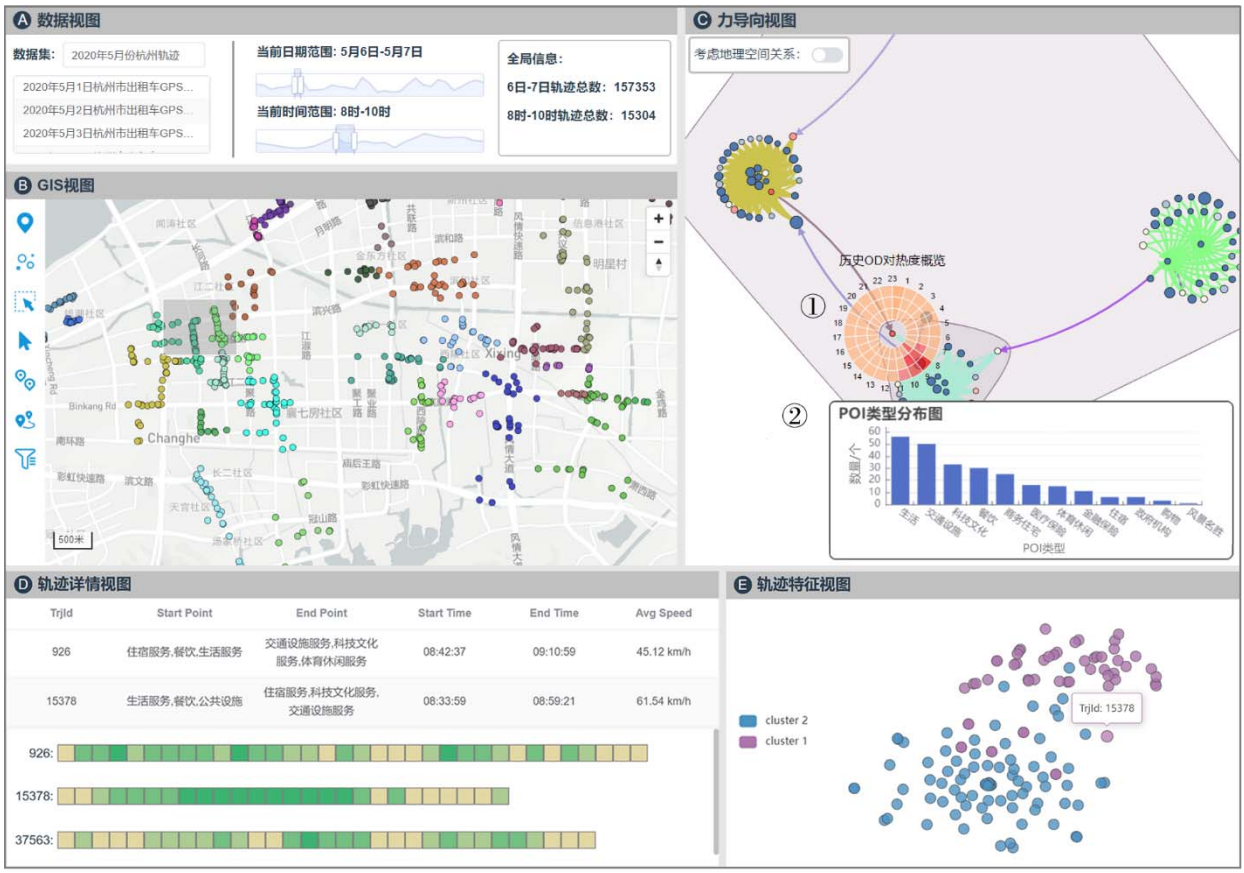


图 4 本文系统可视界面

4.2 数据视图

数据视图包含数据上传、时间范围选择和全局信息展示 3 个部分. 用户上传数据后，视图会展示数据集涵盖的各时间段的轨迹数量概览，并提供日期、时间范围选择等功能，支持用户发现和聚焦感兴趣的数据.

4.3 GIS 视图

GIS 视图以中国地理信息图层作为底图，展示当前所选轨迹数据的所有 OD 点. 用户设置聚类参数可以获取 OD 点簇的聚类结果，并通过不同颜色进行展示；也可以点击菜单切换至刷选模式，通过在地图中使用矩形框绘制工具刷选要研究的区域，系统会高亮展示与所选区域内的簇存在可达关系的簇，并隐藏其余不相关的点簇；此外，还可以通过移动和缩放矩形框以调整区域范围. 接着，系统将刷选的所有簇作为节点，簇间的可达关系作为边构建空间交互网络，并运行线图转换算法得到轨迹交互网络，由右侧的力导向视图进行展示.

4.4 力导向视图

构建轨迹交互网络时，系统按照第 3.4 节中的方法融合该网络的节点属性和拓扑结构进行社区发现，并通过凸包形式展现社区划分结果，整个网

络的拓扑结构通过力导向布局在视图中展示. 若用户需要为节点加入空间相对距离的约束, 则可以点击视图上方的按钮切换图布局模式, 系统将按式(1)重新计算图节点间的距离和位置关系. 线图视角下的轨迹交互网络包含原图的所有信息, 其中, 节点的半径代表 OD 对的起、终点间的空间距离; 节点的颜色反映 OD 对内的轨迹数量, 从低到高分别用由蓝色、白色和红色组成的渐变色表示. 线图中, 颜色相同的边对应 GIS 视图中同样颜色的 OD 点簇.

进一步, 点击线图中的节点, 系统会通过极坐标热力图呈现该 OD 对近 3 天内的轨迹数量概况, 并由外到内依次展示每天的轨迹数量, 颜色越深则表明轨迹越多, 如图 4 中 C 区域的组件①所示. 同时, GIS 视图会高亮显示对应的 OD 对, 以及起、终点所在的 OD 点簇.

当鼠标悬浮或点击线图中的边时, 与该边属于相同簇的边高亮显示, 其他边被隐藏, 此时 GIS 视图中高亮显示的是与该 OD 点簇关联的所有 OD 对. 如图 4 中 C 区域的组件②所示, 视图右下角会同步出现 POI 详情窗口, 对 OD 点簇范围内所有 POI 类别按数量降序排列, 便于用户分析该区域周边的城市功能性分布情况.

4.5 轨迹详情视图

用户在力导向视图中点击线图节点后, 轨迹详情视图同步展示该节点对应的 OD 对在当前时间段下所有轨迹的详细信息, 包括起、终点周围的主要 POI, 行程起止时间以及平均行驶速度. 为了通过轨迹的速度变化反映相应路线的路况信息, 该视图下方用热力图展示每条轨迹的行驶速度概况, 其中, 每个单元格的颜色表示当前轨迹在 2 个 GPS 采样点间的平均速度, 从低到高由黄色到绿色的渐变色表示, 便于比较相同 OD 对内不同路线的差异性.

4.6 轨迹特征视图

当力导向视图完成社区发现算法时, 系统使用 *t*-SNE 算法[32]对轨迹交互拓扑图中所有节点包含的轨迹特征进行降维可视化, 并在轨迹特征视图中展示. 其中, 每个坐标点代表一条轨迹在二维特征空间内的位置, 坐标点间的距离代表对应轨迹的相似度. 系统根据轨迹所在的社区将对应轨迹的坐标点进行染色, 使得用户可以在轨迹特征的视角上查看社区划分的结果, 并根据不同社区内坐标点的离散程度发现社区内轨迹的相似模式.

5 案例评估

本文邀请了 E1 和 E2 这 2 位专家使用本文系统进行案例研究. 2 位专家在案例中获得了针对城市空间中 OD 对的空间分布与交互情况的重要启示; 案例研究后, 对他们进行一对一的访谈, 收集他们的反馈意见.

5.1 案例 1

考虑互联网等高新技术产业通常以科技园的形式在某地聚集, 且具有与传统行业不同的工作节奏, 人员流动可能具有特殊的模式. 因此, 本节以杭州市滨江区某科技园为例, 分析其在工作日早晚高峰时段 OD 流的空间分布特征, 如图 4 所示.

专家选择杭州市 2020-05 的出租车轨迹数据集, 从图 4 中 A 区域可以观察到该月轨迹数量的大致分布, 并选择 2020-05-06T08:00—10:00 早高峰的数据; 然后, 专家在 B 区域中执行聚类算法获得 OD 点簇的划分结果, 再定位并刷选出位于滨江区的若干高新科技园区, C 区域会联动展示线图视角下的轨迹交互网络和社区发现结果, E 区域则联动展示出该网络中所有轨迹的特征降维表示.

如图 4 中 C 区域所示, 4 个不同颜色的连通子图共同构成一个大社区, 且子图 1 同时自成一个小社区. 这 4 个子图都与第 3 节的模式①相匹配, 子图间通过模式②的稀疏边连接. 专家通过 B 区域和 C 区域间的联动分析, 进一步探究这 3 种模式隐含的出行规律. 具有不同颜色边的子图代表不同被频繁经过的 OD 点簇, 边的数目体现了簇的客流吞吐量. C 区域中还有一些独立于社区外的散点对应模式③, 代表研究区域内的部分子区域, 这些子区域包含从研究区域外的地方到达的 OD 对.

为了进一步了解子图的内部结构, 专家检查子图各节点对应的 OD 流发现, 在早高峰时间不仅有大量流入科技园的轨迹, 还有部分流出的轨迹, 且这些方向相反的 OD 流通常都由子图间的稀疏边连接. 专家从图 4 中 C 区域的 POI 类型分布图发现, 生活区内民生、交通和住宅类的兴趣点占比很大, 这是因为研究区域内存在一些居民区, 一些员工在早晨会打车前往上班地点, 也有一些员工从公司前往居民区, 后者可能是因为值夜班结束的员工打车回家, 所以存在频繁地往返于科技园和生活区的轨迹数据, 即早高峰时间的居民区是出租车司机载客的热门区域和中转站. 专家经过对非高峰期的轨迹数据分析发现, 只有零星的 OD 对流向科技园, 且几乎没有流出. 基于上述分析, 专

家想从轨迹层面分析早晚高峰的交通状况，因此在力导向视图中点击节点以查看包含的轨迹详情，如图 4 中 D 区域所示. 通过表格中的起、终点 POI 信息，起止时间与 GIS 视图中高亮的 OD 对，专家判断这些轨迹是属于一个从生活区流向科技园的早高峰通勤 OD 对；随后从图 4 中 D 区域下方的热力图中发现，许多这类轨迹行程末尾的网格都呈淡黄色，行驶速度处于较低的状态，推测是高峰时期园区附近有较多车辆到达导致的拥堵.

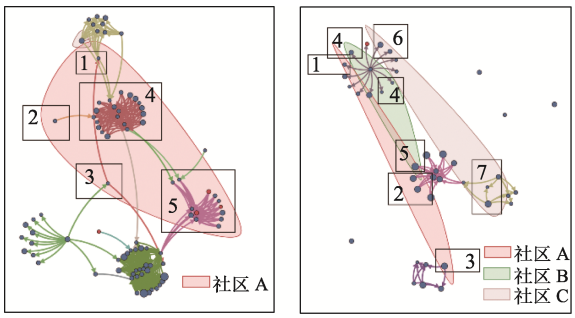
进一步，专家回到力导向视图分析下班通勤流，以每 2 h 为间隔，分析研究区域从 18:00—次日 2:00 的 OD 对，发现了大量由若干模式①构成的线图结构，持续存在大量流出科技园的 OD 对，也有少量 OD 对从生活区流入. 通过在图 4 中 C 区域中对近 3 天 OD 对历史热度图的分析发现. 科技园至生活区的热度较大且较为集中地分布在晚高峰时段，而生活区至科技园的热度较小且规律不明显，这可能与科技园员工的工作习惯有关，部分员工需要在晚高峰甚至更晚的时间下班. 根据线图节点的颜色专家发现，晚高峰的返程轨迹在 20:00 后的数量开始增多，且从轨迹详情视图中发现，部分轨迹也出现了行程开头行驶速度较低的情况，于是推测该时间段可能是科技公司的晚班交通补贴生效时间，因此造成了小范围的拥堵. 但由于下班时间较为灵活，人流量相对较少，轨迹末尾呈淡黄色的网格数量少于早高峰的数量，因此拥堵程度较低.

专家使用本文系统完成了对兴趣区域早晚高峰 OD 流的分布分析发现，高新产业园区的员工倾向于选择园区附近的居民区作为居住地，且下班晚高峰区间较长，从 21:00 至凌晨仍有许多的 OD 返程流量. 此外，由于部分企业有值夜班要求，也存在频繁地往返于科技园和生活区的轨迹数据.

5.2 案例 2

为了探究城市子区域随时间的变化，分析某区域在 2020-05-08T19:00—20:00 和 2020-05-08T 20:00—21:00 的社区分布情况，线图视角下的社区发现结果如图 5 所示，其中，该区域包含某党校和某企业的一个园区.

如图 5a 所示，社区 A 中代表 OD 对的节点与研究区域内的企业园区相关，即 OD 对的起点或终点在该园区附近. 专家通过同步分析地图、线图与轨迹特征视图发现，社区 A 中节点所代表的 OD 对方向大多一致，均为长距离东西方向，且轨迹特征之间的距离较近，具有一定的相似性. 其中，力导向视图中的区域 1~区域 3 位于企业园区北门附近，



a. 19:00—20:00 b. 20:00—21:00

图 5 杭州市某区域晚高峰社区发现结果

OD 对连通至附近的住宅区，区域 4 和区域 5 分别位于企业园区的访客接待区和南门附近，部分 OD 对连通至杭州东站、杭州站和杭州北站等交通枢纽.

如图 5b 所示，研究区域在 20:00—21:00 发现了 3 个社区. 社区 A 包含区域 1~区域 3，社区 B 包含区域 4~区域 5，社区 C 包含区域 6~区域 7，除了区域 7 中的部分节点与党校关联，其余节点基本都位于企业园区附近. 专家通过对比图 5a 和图 5b 发现，党校周边在 20:00—21:00 打车行为的活跃度较低，而企业园区的打车行为活跃度依然较高. 专家通过查看线图中的每个连通子图探究其在现实地图中的位置关系，经分析，力导向图中区域 3 所在的连通子图由从企业园区出发的东西向距离较近的 OD 对构成，区域 4 所在的连通子图则是由从企业出发但没有统一方向的 OD 对构成，区域 5 所在的连通子图由到达企业园区的东西向 OD 对构成.

进一步，专家分析每个社区内部的 OD 对相关关系. 社区 A 包含东西向往返企业园区的 OD 对及企业园区往返火车站的 OD 对；社区 B 同样包含往返企业园区的 OD 对，但空间分布相对发散，没有统一流向，且距离较近；社区 C 中，区域 6 的节点为从企业靠近党校方向出口出发的 OD 对，区域 7 中节点为到达党校附近的远距离 OD 对. 接着在轨迹特征视图中分别查看每个社区内轨迹特征点的分布，发现社区 A、社区 C 内的特征点之间距离较近，而社区 B 中的特征点离散程度较高，与上述发现吻合. 此外，这些社区中都存在一些与其他轨迹特征点较远的点，通过与线图和地图的联动，可以确认这些特征点对应的轨迹为长距离轨迹，因此与同社区内的其他轨迹相关性较低. 根据对社区及内部 OD 对和轨迹的观察专家发现，社区 C 在 21:00 前有较多出租车轨迹都以党校和某企业附近区域作为上下客点，在空间拓扑和轨迹特征上都具有较强的关联关系，因此认为该社区在当前时段可以视作一个打车热点街道，由此得出“出租车

可以在该街道附近快速获得接单机会”.

从社区发现的结果可以看出, 通过本文方法融入轨迹特征后, 轨迹交互拓扑图中没有直接连通关系的 OD 对间, 依然可能因为其包含的轨迹相似度较高而存在潜在的关联关系, 同一社区内的轨迹也会因为时空特征呈现一定的分布态势, 从而能够揭示不同时段内出行需求差异导致的城市子区域变化情况, 也能从侧面反映出现实城市空间的潜在关联关系.

5.3　专家反馈

在案例分析后, 本文收集了专家的反馈意见, 总结为 4 点.

(1) 可视化设计. 专家们均表示, 可视化设计易于理解. 本文方法采用常用的地图和节点链接图分别展示 OD 流的地理上下文信息和轨迹等抽象信息, 专家们对这 2 种可视化方法均非常熟悉. E2 还特别表扬了线图, 称“线图采用了传统的节点链接图的形式, 却以全新的角度展示了交通 OD 数据, 能够揭示很多有趣的模式” “基于力导向的图布局可以清晰地展示社区”.

(2) 系统功能. 专家们表示, 本文系统功能完备, 能够很好地支持对 OD 对的交互式探索. 系统支持灵活的时空过滤功能, 专家得以根据他们的需求和兴趣探索大尺度的时空数据集. 此外, 用户不仅可以通过地图在地理上下文中理解 OD 流信息, 还可以通过抽象空间中的线图对大量的 OD 对进行分析; 而轨迹详情和轨迹特征的可视化则提供了更细粒度的数据视角, 尤其是从轨迹特征散点图的角度展示社区划分的结果, 能够将融合轨迹特征的社区发现算法发现的潜在信息体现出来, 也为社区发现的结果提供了一些可解释性.

(3) 与现有工作的对比. 请求专家将本文工作与现有的一些方法进行比较, 以评估本文方法可能存在的优势. 通过与 MapTrix 方法[18]的对比, 专家认为, 本文方法与该方法在面对原始地图和密集的 OD 图如何可视化上做出了不同的尝试, MapTrix 方法通过矩阵展示 OD 图, 并通过连接线与原始地图产生映射; 而本文方法则是新颖地用线图表达 OD 图, 避免了原图转换为节点连接图的杂乱性, 通过归纳出的 3 种典型模式辅助分析, 并且与原始地图通过鼠标交互自然地进行关联映射, 能够进行更灵活的探索. 与 MobilityGraphs 方法[20]对比时专家指出, MobilityGraphs 方法使用节点连接图实现了 OD 流时空变化的可视化表示, 并很好地通过聚合简化了节点连接图; 但本文方法还提供了轨迹层面的信息, 加强了对 OD 流变化的可解释性.

(4) 改进. 专家们还提出一些具有建设性的建议. 其中, E1 提到, “如果能够将原始的轨迹也按需地展示出来就更好了, 获取更加原始的信息能够辅助进行可靠的决策”; E2 则提出, “社区发现结果在大规模线图中存在一定程度的重叠和混淆, 可以优化可视化交互, 如对单个社区进行聚合缩放以虚化用户不关注的部分图结构”.

6　结　语

本文提出一种融合线图和轨迹表征的城市社区发现方法, 建立了以 OD 对特征为节点、以 OD 对的起点和终点为边的轨迹交互拓扑图, 可基于出行轨迹数据自下而上地开展子区域发现; 在此基础上, 设计并实现了基于线图视角的城市社区发现可视分析系统. 通过案例探究与用户实验, 证实了本文方法与系统可以挖掘 OD 对之间的关联关系. 未来, 将针对以下 3 点进行优化: (1) 优化用户交互体验, 如改进研究区域选择方式, 增加单选、多选和多边形选择方式, 以适应复杂的地理区域形状, 且在 GIS 视图和力导向视图中减少图形元素的重叠和混淆, 减少用户的心智负担; (2) 让 OD 对包含的轨迹在 GIS 视图中也支持进行可视化和详情分析, 并优化各个视图间轨迹数据联动的交互, 帮助用户更方便地从细粒度的角度对 OD 对关联关系进行分析; (3) 拓展系统的功能, 如接入实时数据和 OD 需求预测等算法, 使得系统更具通用性.

参考文献(References):

[1] Min Zhongrong, Yang Xianfang. Optimization of urban spatial structure and control division of urban space: a case study of Nanchang city[J]. Modern Urban Research, 2011, 26(3): 43-47(in Chinese)
(闵忠荣, 杨贤房. 城市空间结构优化与城市空间管制区划——以南昌市为例[J]. 现代城市研究, 2011, 26(3): 43-47)

[2] Wu Shanmei, Wang Yandong, Qiao Mengling, *et al*. Detecting fine-grained urban sub-region using social media data[J]. Geospatial Information, 2020, 18(9): 122-125(in Chinese)
(武善梅, 王艳东, 乔梦玲, 等. 利用社交媒体数据探测细粒度城市子区域[J]. 地理空间信息, 2020, 18(9): 122-125)

[3] Yang Xiping, Fang Zhixiang. Recent progress in studying human mobility and urban spatial structure based on mobile location big data[J]. Progress in Geography, 2018, 37(7): 880-889(in Chinese)
(杨喜平, 方志祥. 移动定位大数据视角下的人群移动模式及城市空间结构研究进展[J]. 地理科学进展, 2018, 37(7):

880-889)

[4] Guo D S, Jin H, Gao P, *et al*. Detecting spatial community structure in movements[J]. International Journal of Geographical Information Science, 2018, 32(7): 1326-1347

[5] Deng Z K, Weng D, Liu S H, *et al*. A survey of urban visual analytics: Advances and future directions[J]. Computational Visual Media, 2023, 9(1): 3-39

[6] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99(12): 7821-7826

[7] Newman M E J. Detecting community structure in networks[J]. The European Physical Journal B, 2004, 38(2): 321-330

[8] Jokar E, Mosleh M, Kheyrandish M. GWBM: an algorithm based on grey wolf optimization and balanced modularity for community discovery in social networks[J]. The Journal of Supercomputing, 2022, 78(5): 7354-7377

[9] Xie Linqi. Research on the characteristics of Shenzhen metro commute based on community discovery[D]. Shenzhen: Harbin Institute of Technology, 2021(in Chinese)
(谢林岐. 基于社区发现的深圳市地铁通勤特征研究[D]. 深圳: 哈尔滨工业大学, 2021)

[10] Shirazi S, Baziyad H, Ahmadi N, *et al*. A new application of Louvain algorithm for identifying disease fields using big data techniques[J]. Journal of Biostatistics and Epidemiology, 2020, 5(3): 183-193

[11] Deng Z, Chen S, Xie X, *et al*. Multilevel visual analysis of aggregate geo-networks[J/OL]. IEEE Transactions on Visualization and Computer Graphics: 1-16[2023-09-20]. https://doi.org/10.1109/TVCG.2022.3229953

[12] Yao X, Gao Y, Zhu D, *et al*. Spatial origin-destination flow imputation using graph convolutional networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(12): 7474-7484

[13] Chen Y, Xu J, Xu M Z. Finding community structure in spatially constrained complex networks[J]. International Journal of Geographical Information Science, 2015, 29(6): 889-911

[14] Zhao K, Chinnasamy M P, Tarkoma S. Automatic city region analysis for urban routing[C] //Proceedings of the IEEE International Conference on Data Mining Workshop. Los Alamitos: IEEE Computer Society Press, 2015: 1136-1142

[15] Liu X, Gong L, Gong Y X, *et al*. Revealing travel patterns and city structure with taxi trip data[J]. Journal of Transport Geography, 2015, 43: 78-90

[16] Zhou Bo, Ma Linbing. Spatial interaction analysis of urban hotspot areas from the perspective of complex networks[J]. Geospatial Information, 2021, 19(6): 115-119(in Chinese)
(周博, 马林兵. 复杂网络视角下的城市热点区域空间交互分析[J]. 地理空间信息, 2021, 19(6): 115-119)

[17] Guo D. Visual analytics of spatial interaction patterns for pandemic decision support[J]. International Journal of Geographical Information Science, 2007, 21(8): 859-877

[18] Yang Y L, Dwyer T, Goodwin S, *et al*. Many-to-many geographically-embedded flow visualisation: an evaluation[J]. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(1): 411-420

[19] Zhou Y, Fang Z X, Thill J C, *et al*. Functionally critical locations in an urban transportation network: identification and space-time analysis using taxi trajectories[J]. Computers, Environment and Urban Systems, 2015, 52: 34-47

[20] von Landesberger T, Brodkorb F, Roskosch P, *et al*. MobilityGraphs: visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering[J]. IEEE Transactions on Visualization and Computer graphics, 2015, 22(1): 11-20

[21] Yang Y L, Dwyer T, Jenny B, *et al*. Origin-destination flow maps in immersive environments[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, 25(1): 693-703

[22] Wang Zuchao, Yuan Xiaoru. Visual analysis of trajectory data[J]. Journal of Computer-Aided Design & Computer Graphics, 2015, 27(1): 9-25(in Chinese)
(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, 27(1): 9-25)

[23] Al-Dohuki S, Wu Y Y, Kamw F, *et al*. Semantictraj: a new approach to interacting with massive taxi trajectories[J]. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(1): 11-20

[24] Luo Yuetong, Ding Tiecheng, Zhu Minfeng. Road2vec: a visual analysis method of urban road with taxi trajectory data[J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(12): 2256-2264(in Chinese)
(罗月童, 丁铁成, 朱闽峰. Road2vec: 一种基于出租车轨迹数据的城市道路可视分析方法[J]. 计算机辅助设计与图形学学报, 2017, 29(12): 2256-2264)

[25] Whitney H. Congruent graphs and the connectivity of graphs[OL]. [2023-09-20]. https://doi.org/10.1007/978-1-4612-2972-8_4

[26] Guo D S, Zhu X, Jin H, *et al*. Discovering spatial patterns in origin-destination mobility data[J]. Transactions in GIS, 2012, 16(3): 411-429

[27] Park S H, Kim B D, Kang C M, *et al*. Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder-decoder architecture[C] //Proceedings of the IEEE Intelligent Vehicles Symposium. Los Alamitos: IEEE Computer Society Press, 2018: 1672-1678

[28] Dey R, Salem F M. Gate-variants of gated recurrent unit (GRU) neural networks[C] //Proceedings of the IEEE 60th International Midwest Symposium on Circuits and Systems. Los Alamitos: IEEE Computer Society Press, 2017: 1597-1600

[29] Wang C, Huang J H, Wang Y H, *et al*. A deep spatiotemporal trajectory representation learning framework for clustering[J/OL]. IEEE Transactions on Intelligent Transportation Systems: 1-14[2023-09-20]. https://doi.org/10.1109/TITS.2024.3350339

[30] Li X C, Zhao K Q, Cong G, *et al*. Deep representation learning for trajectory similarity computation[C] //Proceedings of the IEEE 34th International Conference on Data Engineering. Los Alamitos: IEEE Computer Society Press, 2018: 617-628

[31] Zhou Y, Cheng H, Yu J X. Graph clustering based on structural/attribute similarities[J]. Proceedings of the VLDB Endowment, 2009, 2(1): 718-729

[32] van der Maaten L, Hinton G. Visualizing data using *t*-SNE[J]. Journal of Machine Learning Research, 2008, 9(86): 2579-2605