



北京大学
PEKING UNIVERSITY



社会感知及地理大数据思维

刘瑜 教授
地球与空间科学学院遥感与地理信息系统研究所
2019-7-3 @河海大学



目录

CONTENTS

- 1/ 大数据与社会感知
- 2/ 社会感知方法
- 3/ 社会感知应用
- 4/ 总结



北京大学

大数据时代



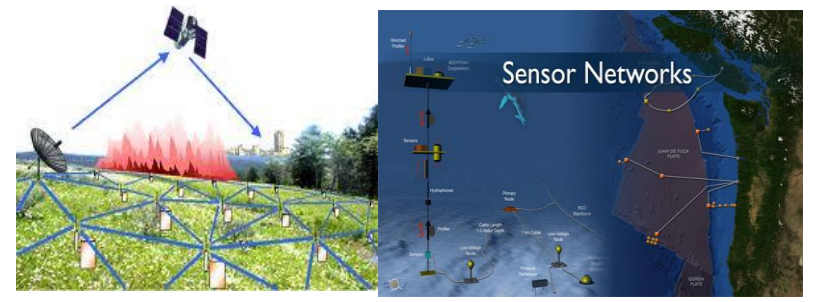
社交媒体与社交网络



科学装置



移动设备



传感网



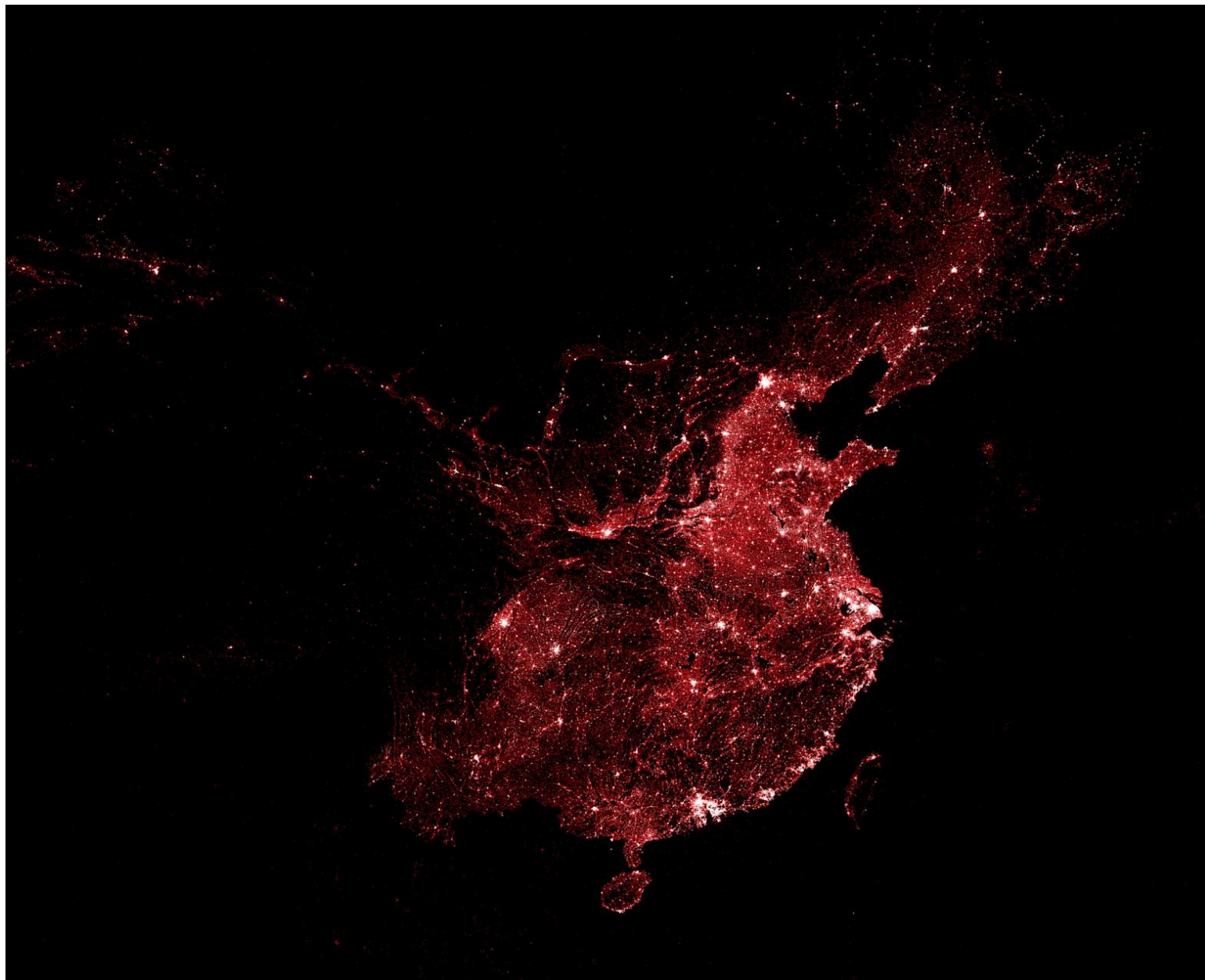
- 地理大数据特点(除了N-Vs)
 - 每条数据往往可以关联到**个体**
 - 具有**时空标记**
- 地理大数据类型
 - 手机数据、社交网络、银行卡数据、车辆轨迹、公交卡数据...

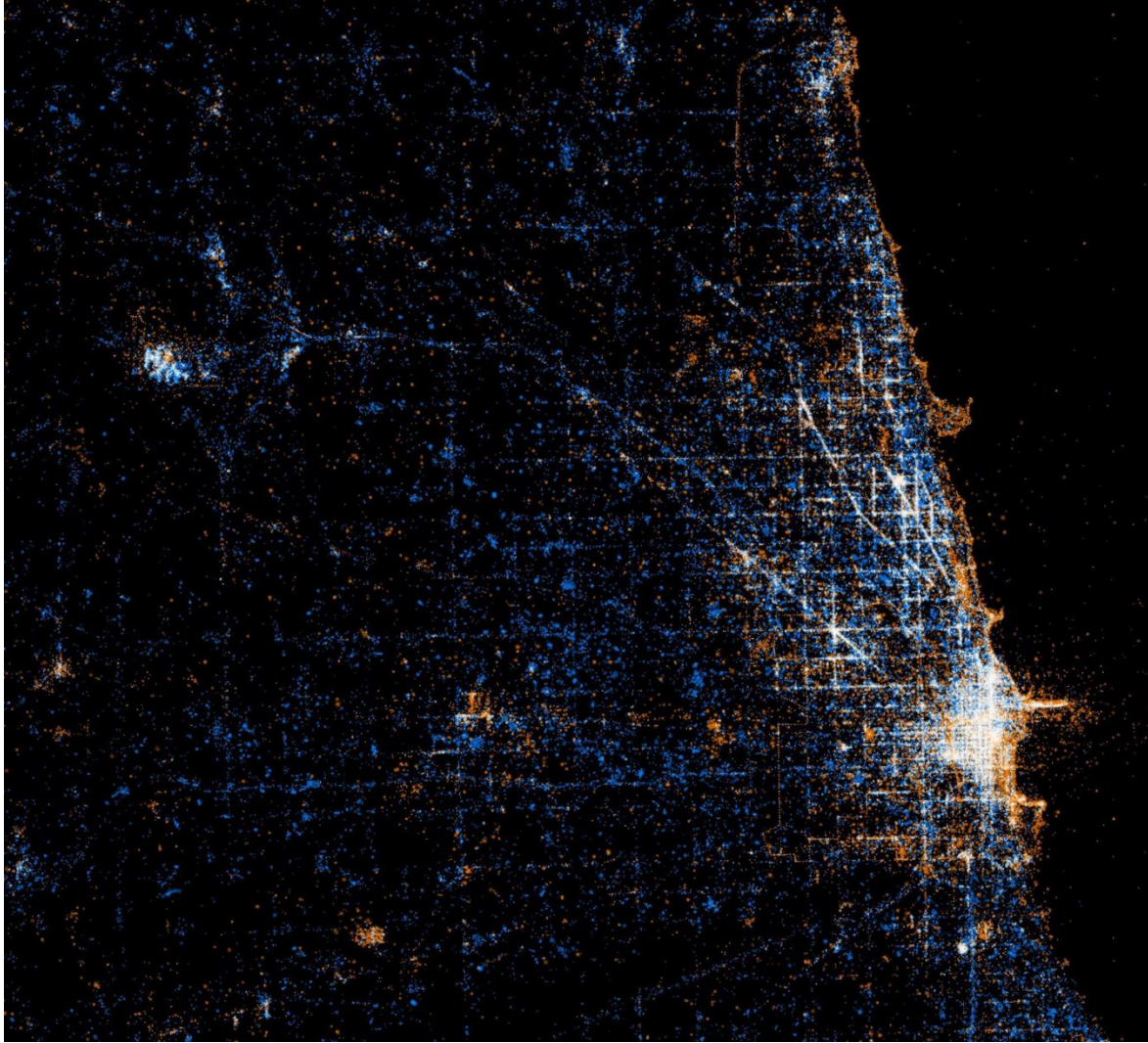


北京大学

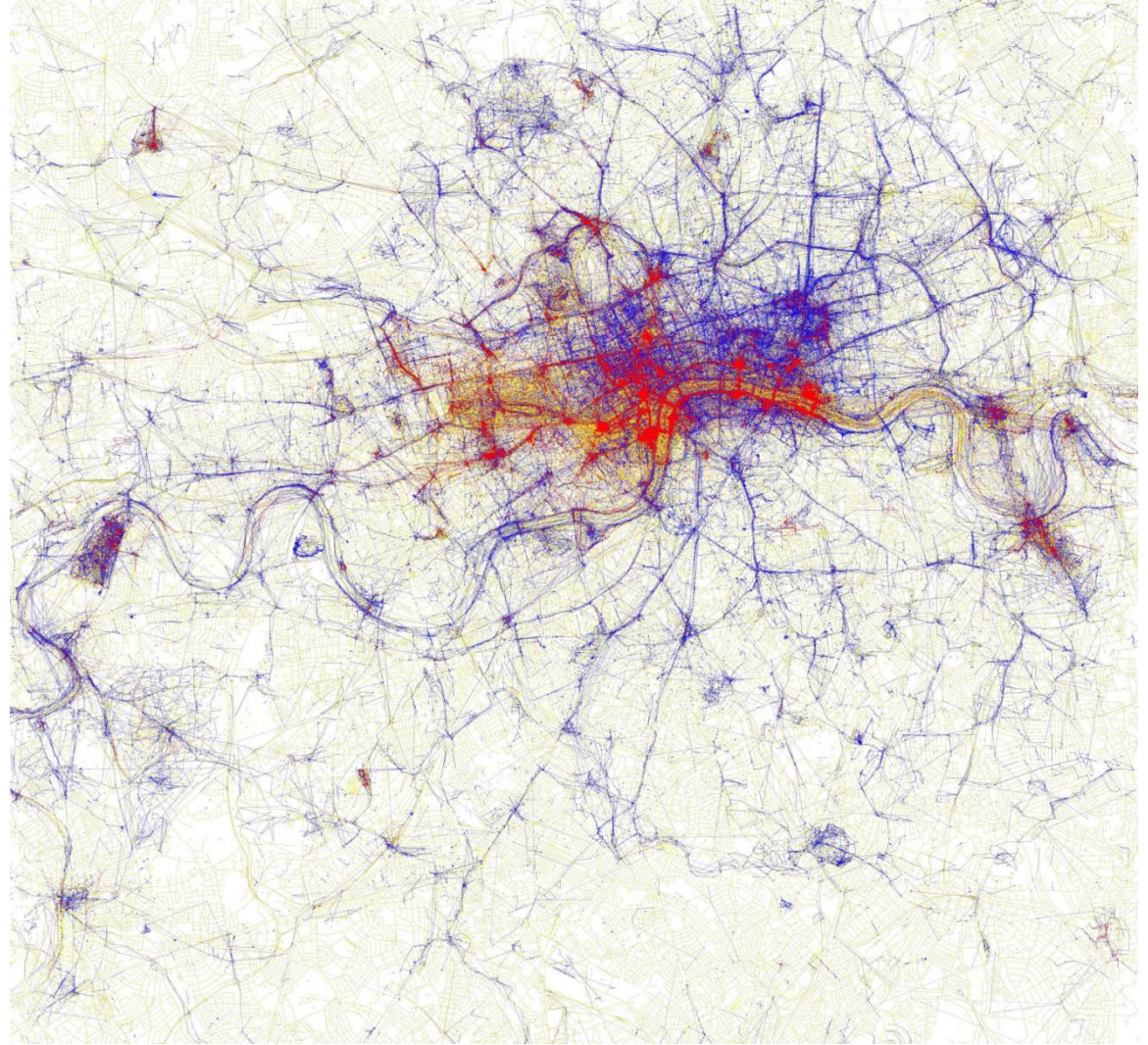
地理大数据实例

全国定位数据





美国芝加哥市，红点是Flickr，蓝点是Twitter，白点是两者都有



伦敦Flickr点的分布，游客所拍（红色），本地居民所拍（蓝色），不确定（黄色）



- 借助于各类地理大数据研究人类时空间行为特征，并进而揭示社会经济现象的时空分布格局、联系以及演化过程的理论和方法
- 相比于传统遥感技术主要用于捕获地理空间中的自然现象，社会感知方法则长于发现社会经济现象，从而有效弥补遥感方法的不足
- 社会感知和传统遥感又有相通和相似之处，但是也有传统遥感所不具有的长处

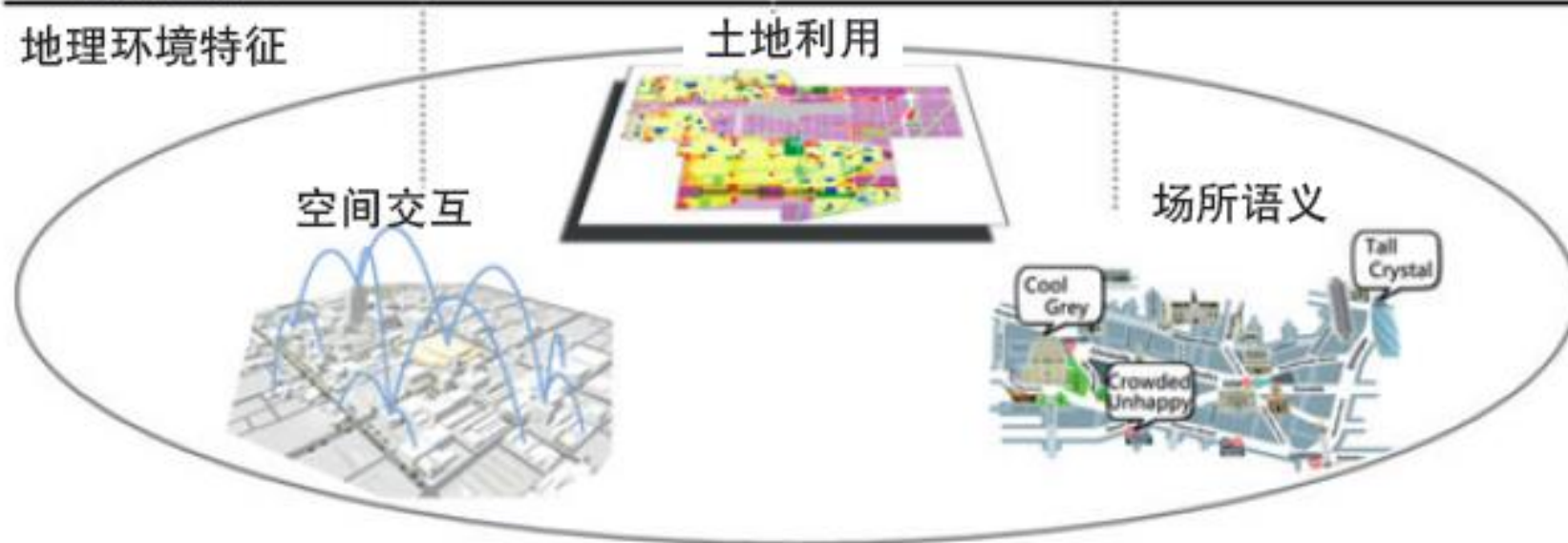


人



地理环境特征

地





- 地理异质性 (Geographical heterogeneity)
 - 地理要素的空间分布不是均质的，这是地理学研究的基础问题之一
- 几个层面的异质性
 - 景观的异质性
 - 人群的异质性
 - 行为的异质性



北京大学

地理异质性



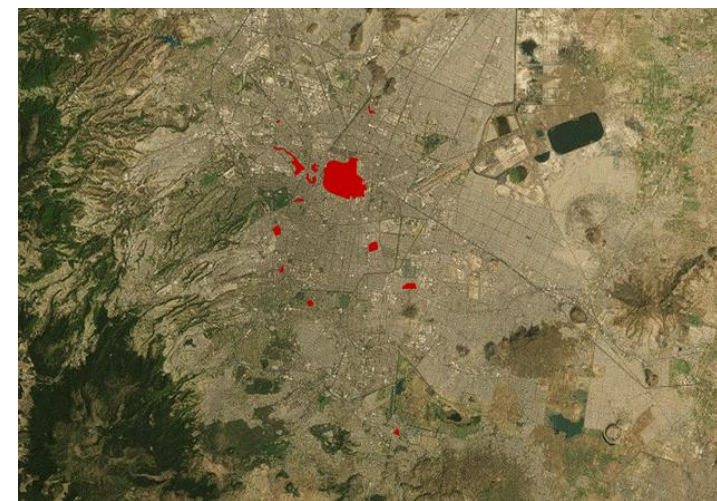
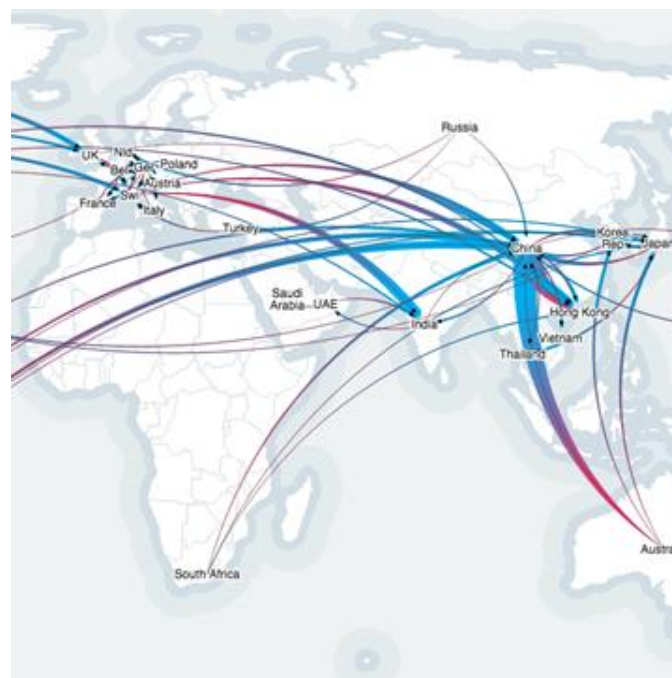


目录

CONTENTS

- 1/ 大数据与社会感知
- 2/ 社会感知方法
- 3/ 社会感知应用
- 4/ 总结

动态演化感知



- **场所**是表达地理分异格局的基础，可以表征为一个一阶场： $y = f(x)$
- **交互**是刻画地理空间结构的关键，可以表征为一个二阶场： $y = f(x_1, x_2)$
- 场所之间的差异是交互的基础，交互消弭场所间差异
- 场所和交互共同帮助我们理解不同尺度地理空间：城市、区域、国家、全球
- 分布一阶场和交互二阶场都随时间演化，表达了不同的地理**过程**，如潮汐通勤、城市扩展、区域发展



- 空间视角的概念
 - 形状
 - 距离
- 场所视角的概念
 - 人们的情感、体验

2. 场所





(Piccadilly Circus)

- 不同place活动不同，不同place的人地交互不同
- 活动依赖于其社会与文化准则，也导致了不同place有不同的活动适宜度



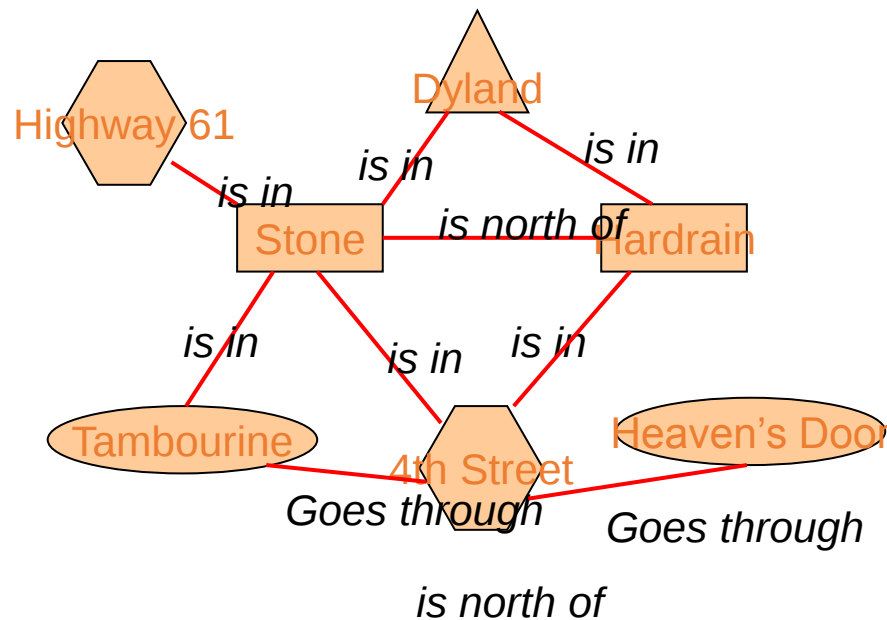
- 尺度无关，从房间小角落到整个星球
- 不依附于行政边界，内在模糊 (vague) 且依赖时间 (time-dependent)
- 不同个体认知下的空间范围不同

3. 人地关系

- 除活动之外place的语义
- 很多源于personal memories 或 feelings of belonging to 或纪念意义



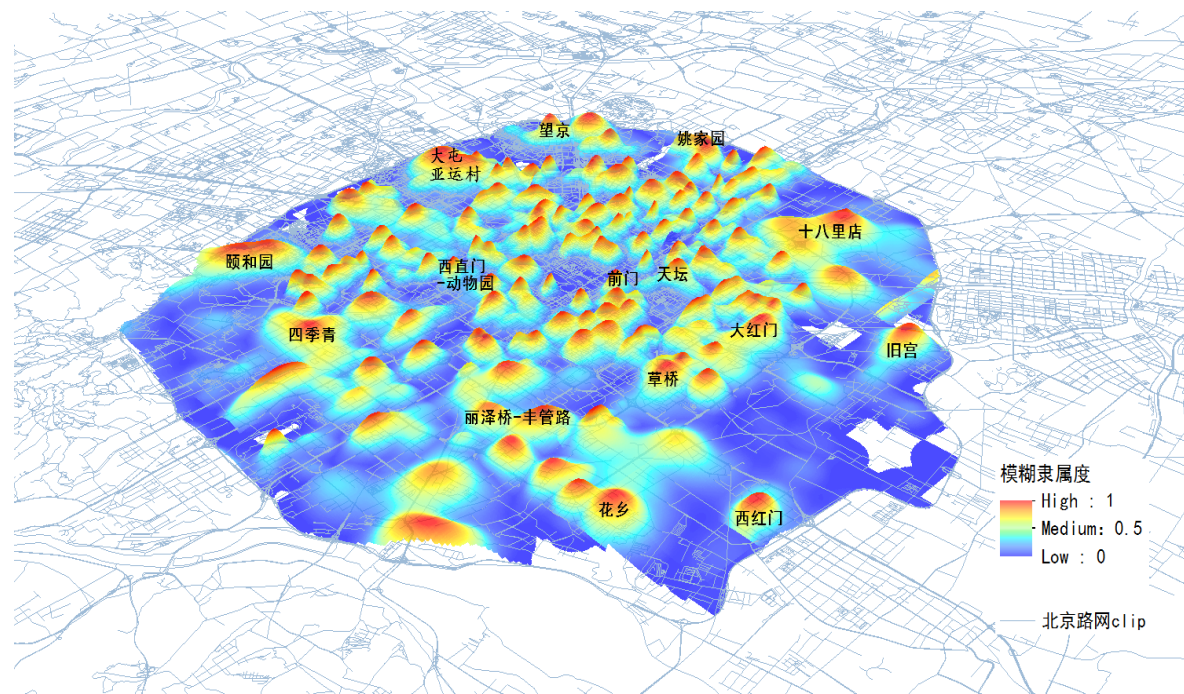
理解空间异质性的单元



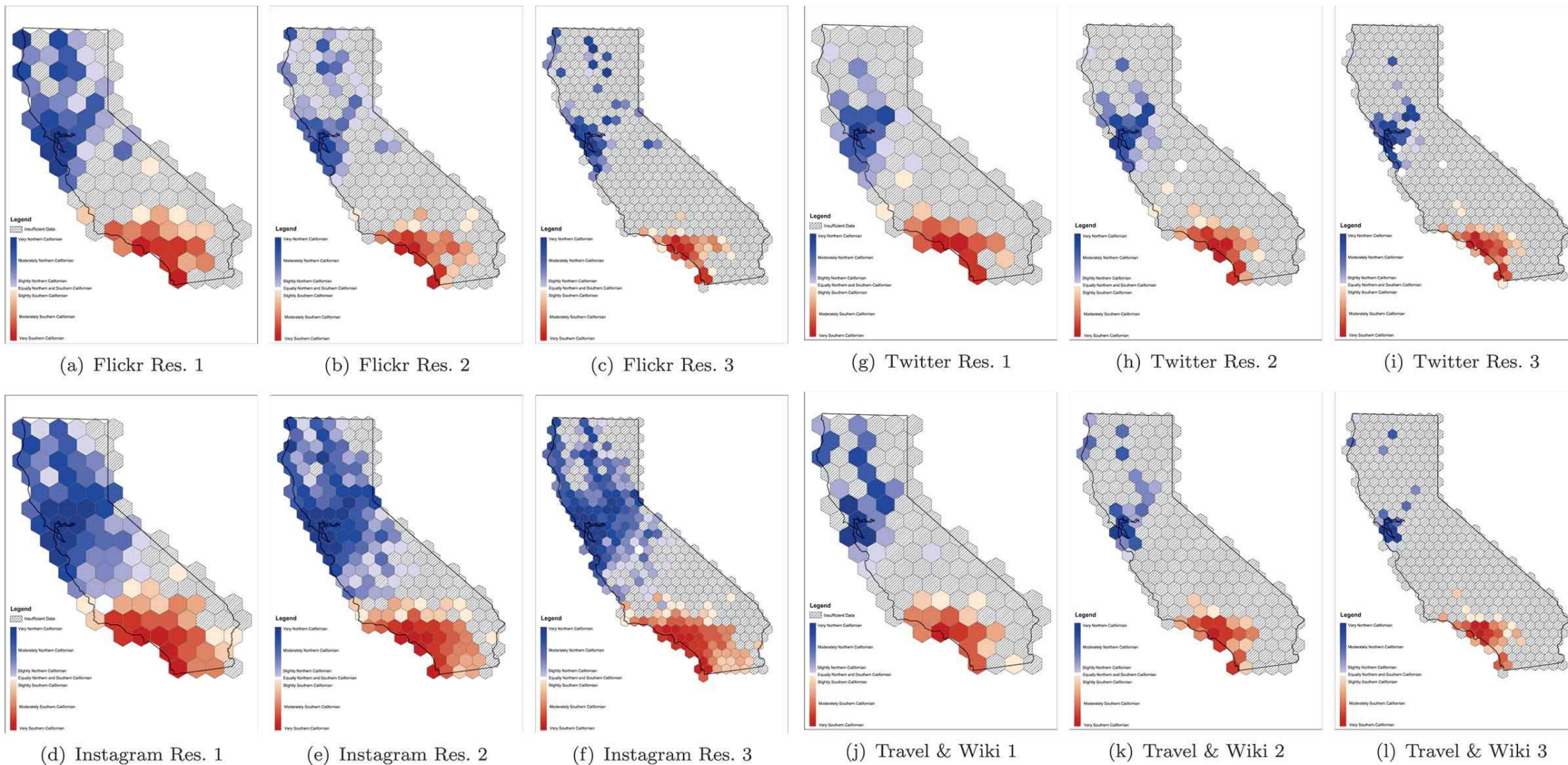
表达地理知识的节点



构建人地耦合的桥梁

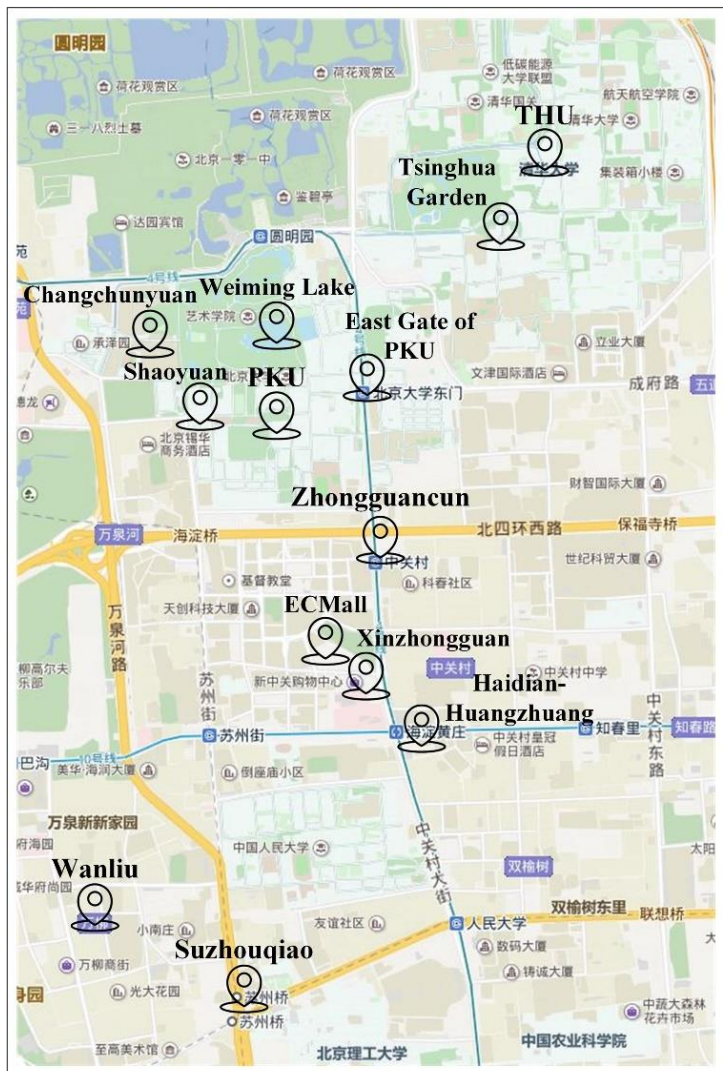


4. 含糊性

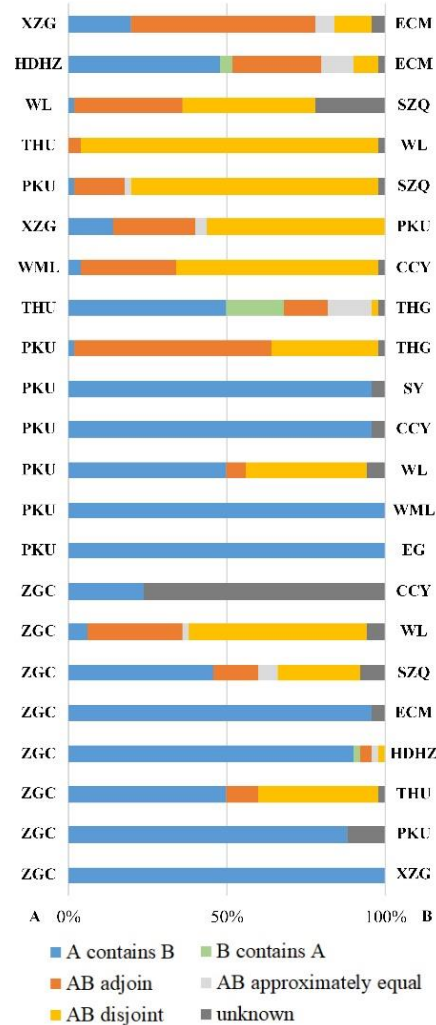


Gao, S., et al. A data-synthesis-driven method for detecting and extracting vague cognitive regions. International Journal of Geographical Information Science, 2017, 31(6), 1245-1271.

场所感知(场所的层次关系)



(a)



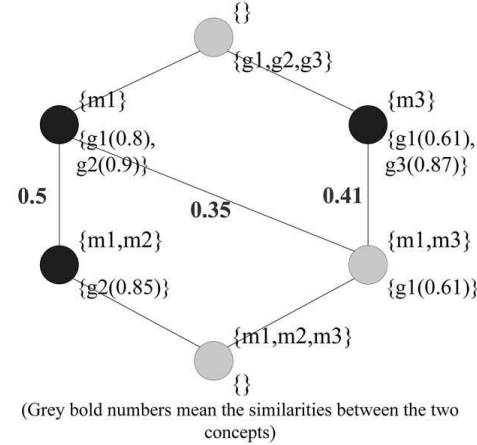
(b)

Fuzzy Formal Context

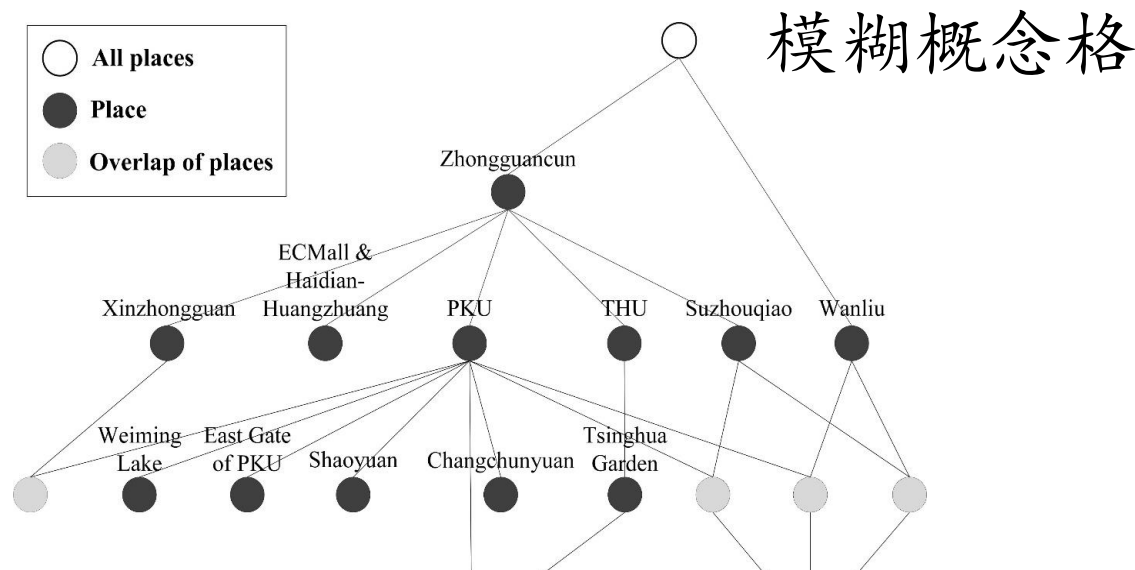
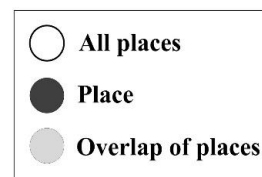
	m1	m2	m3
g1	0.8	0.12	0.61
g2	0.9	0.85	0.13
g3	0.1	0.14	0.87

(a)

Fuzzy Formal Concept Lattice

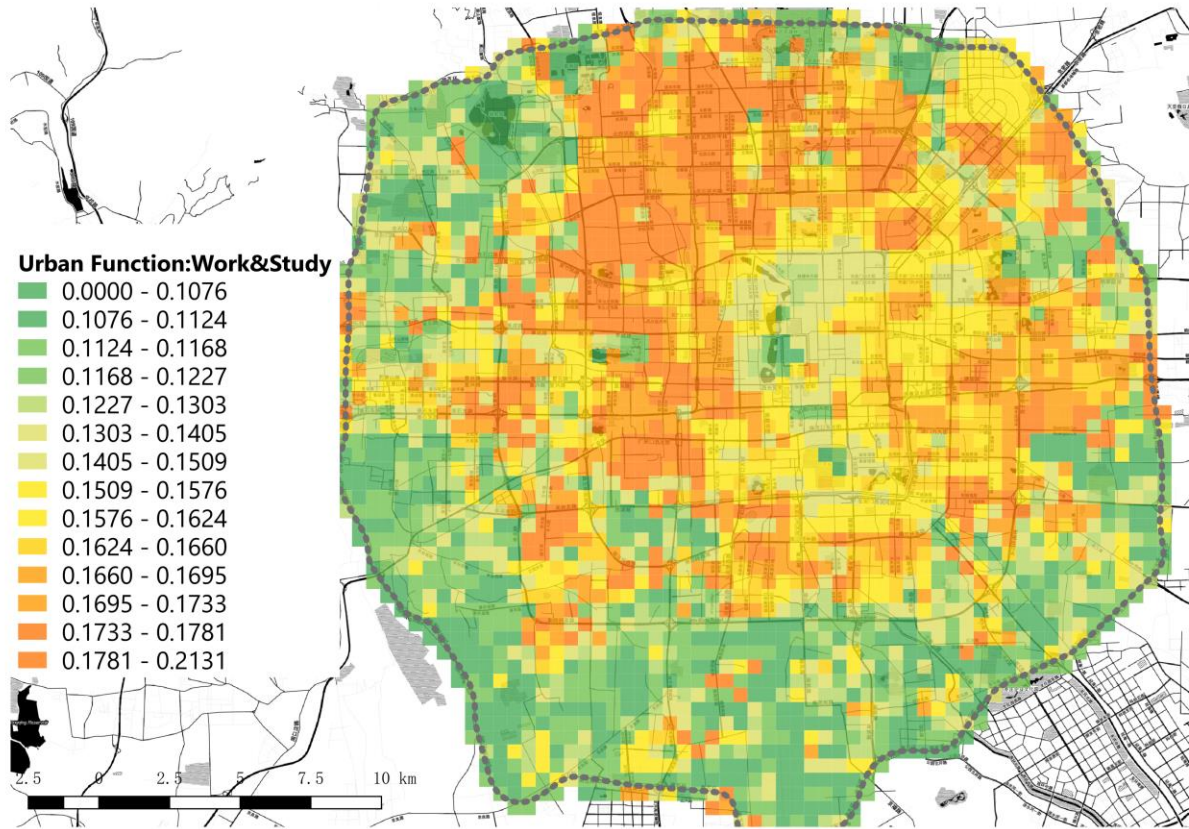


(b)

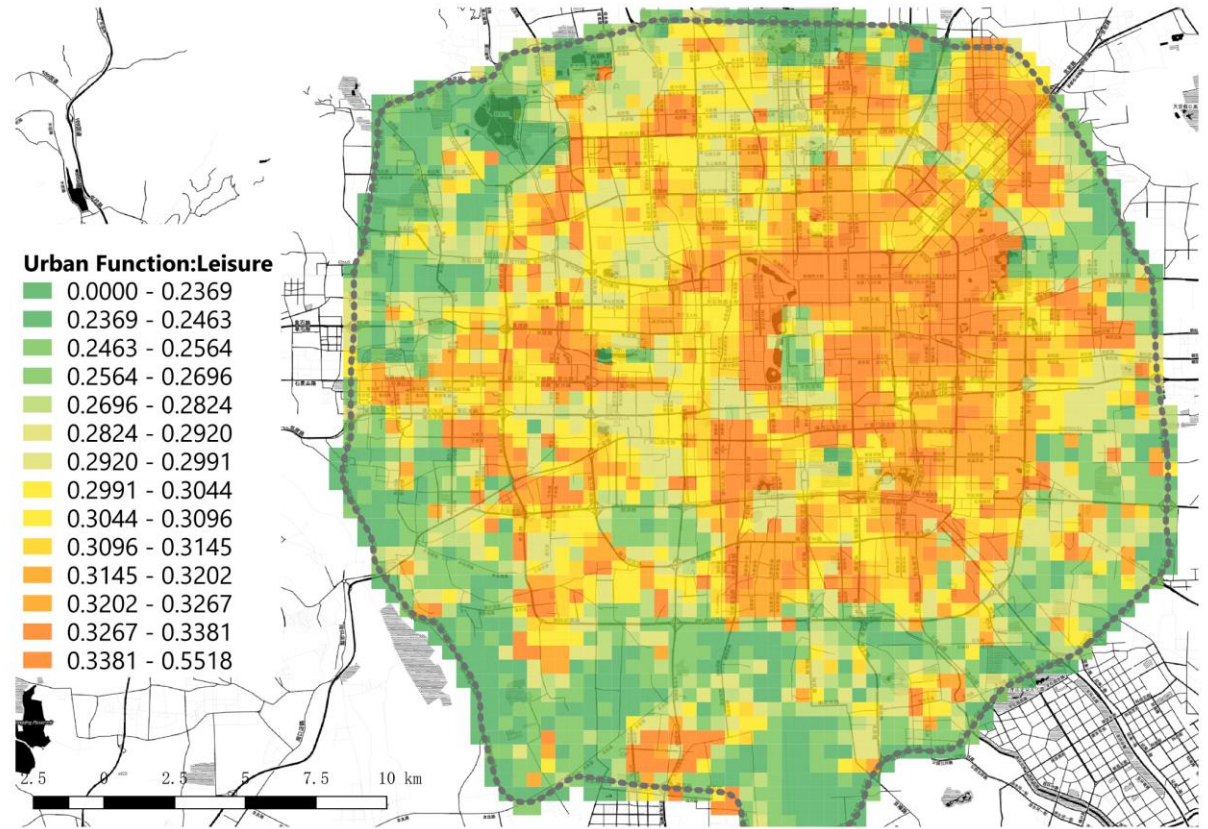


模糊概念格

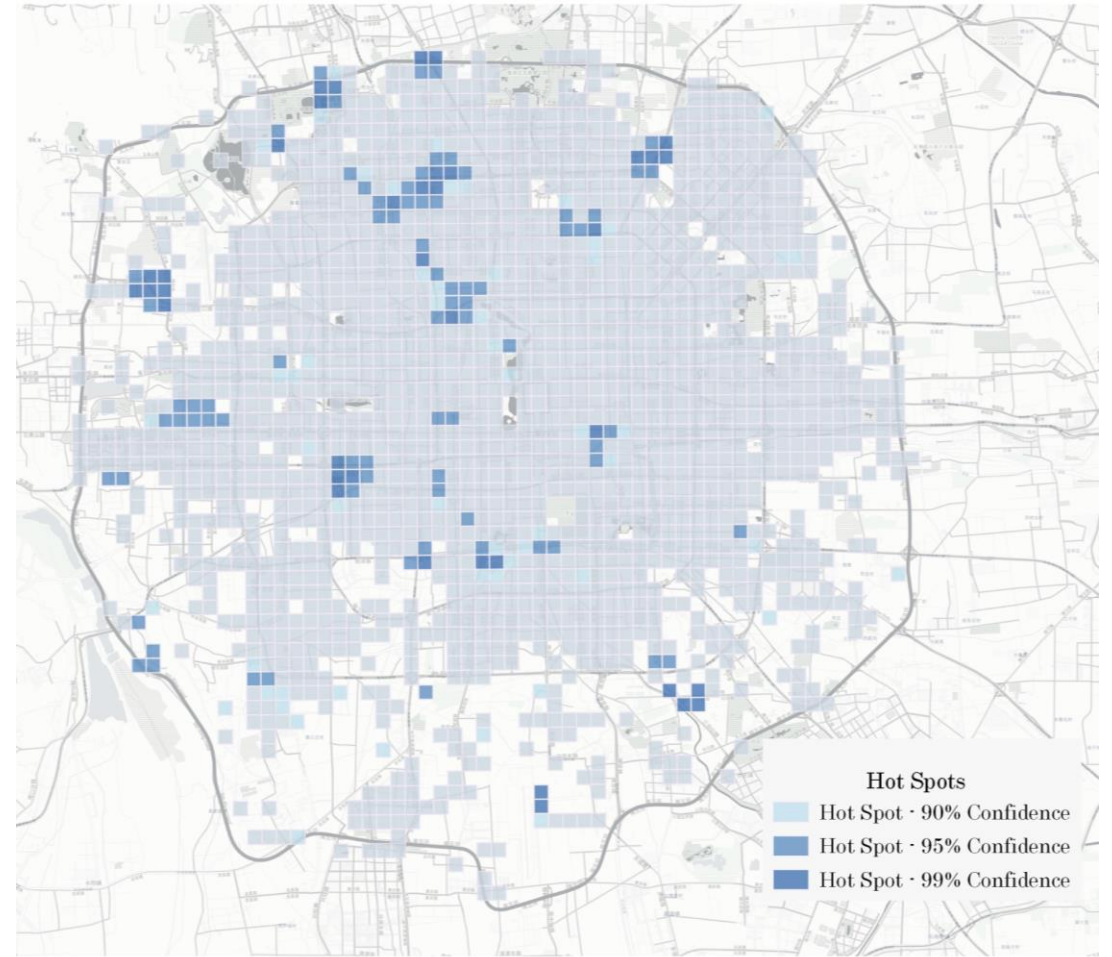
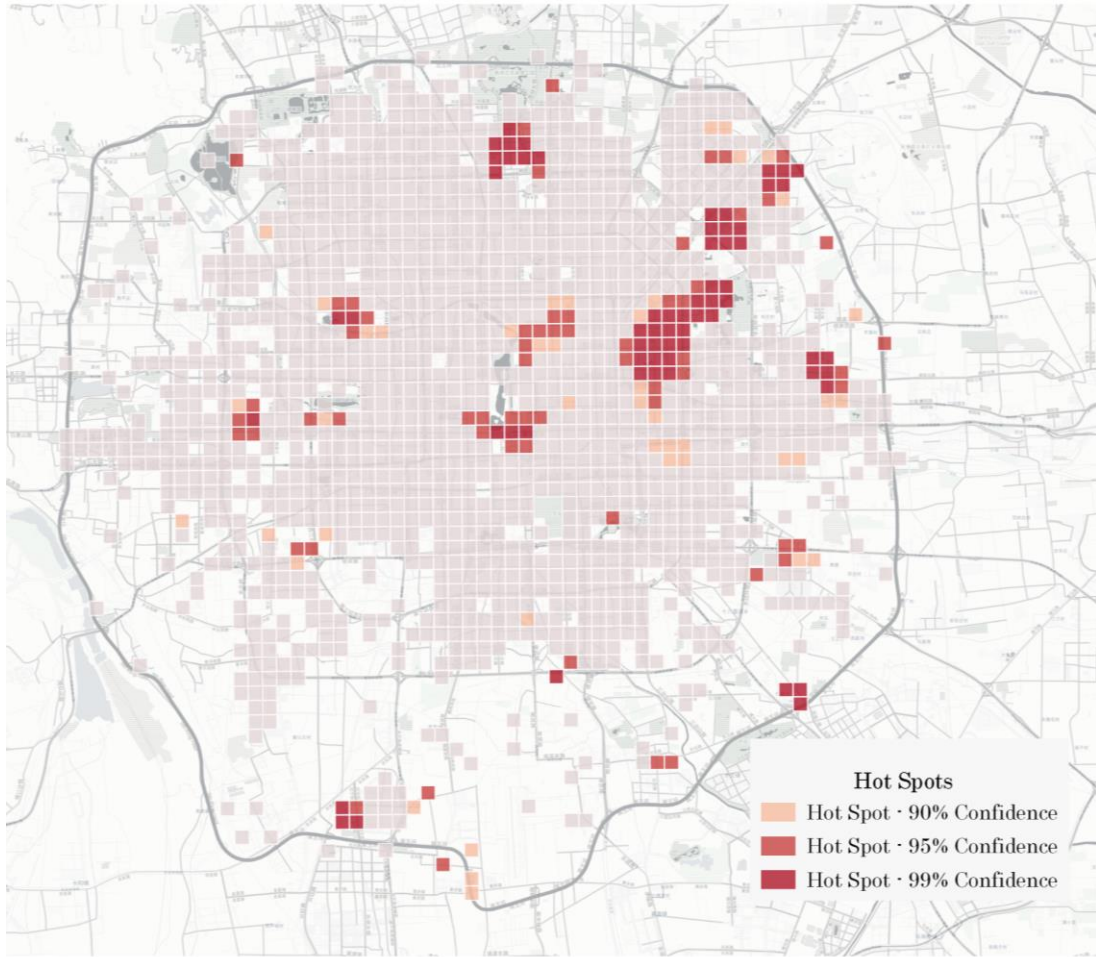
5. 层次性和尺度



工作和学习



休闲娱乐

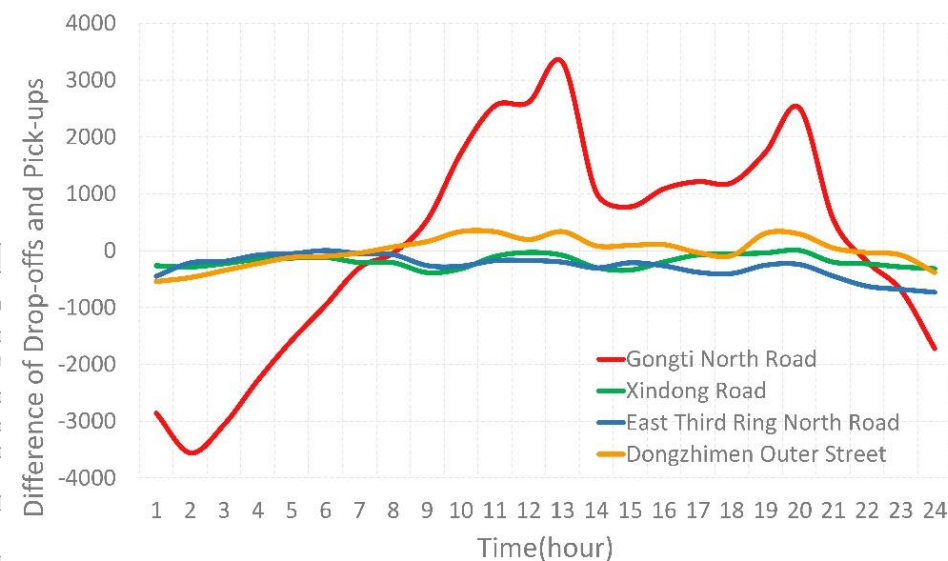
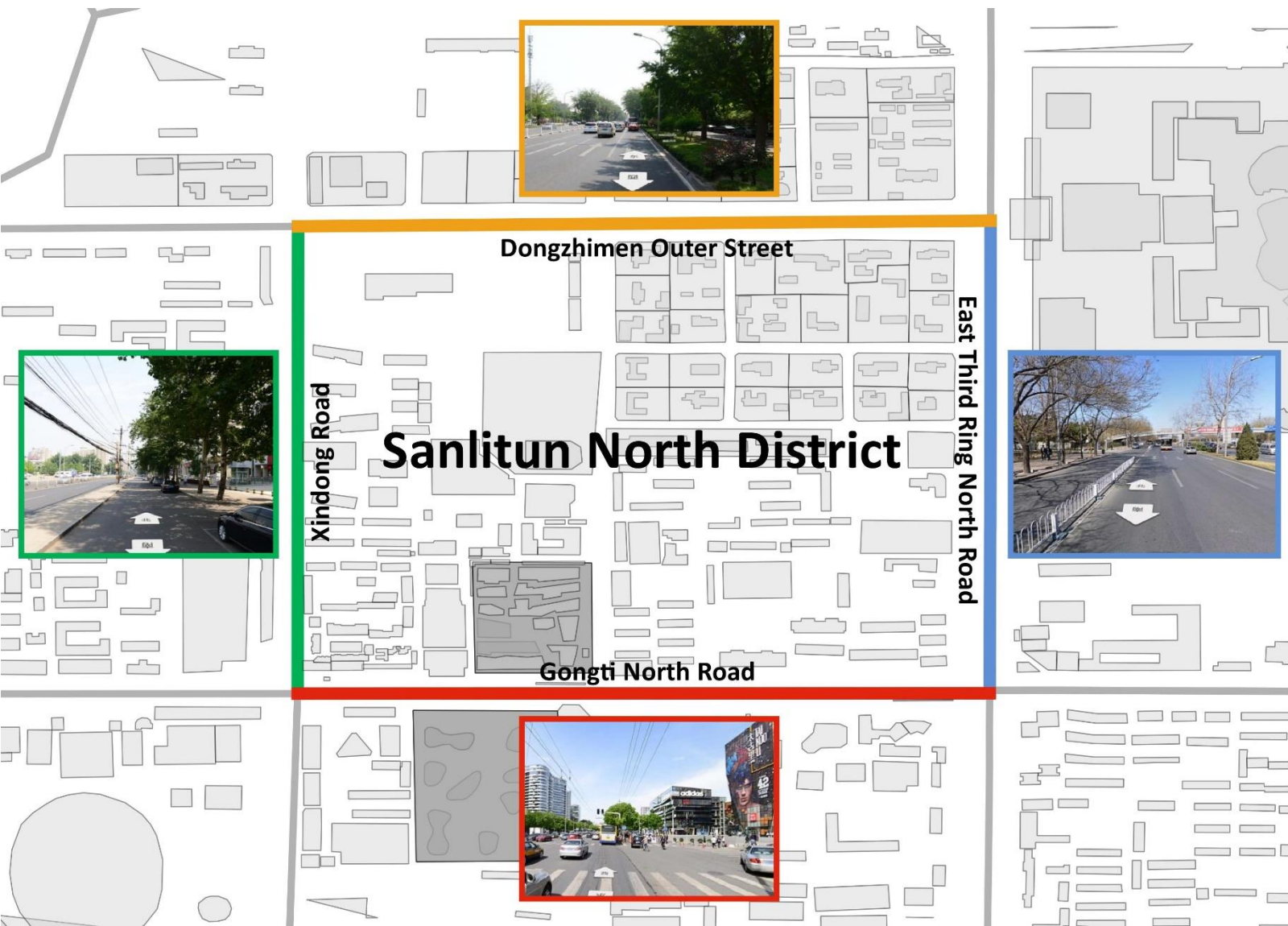


北京市基于微博数据的正面及负面分析热点分布



三里屯北小区 街道视角解析

- 对传统面状研究单元的分解
- 工体北路与小区另外三条街道的明显差异
- 类似的实证分析可以应用于其他地块和商圈 (如北京大学, 火车站等)





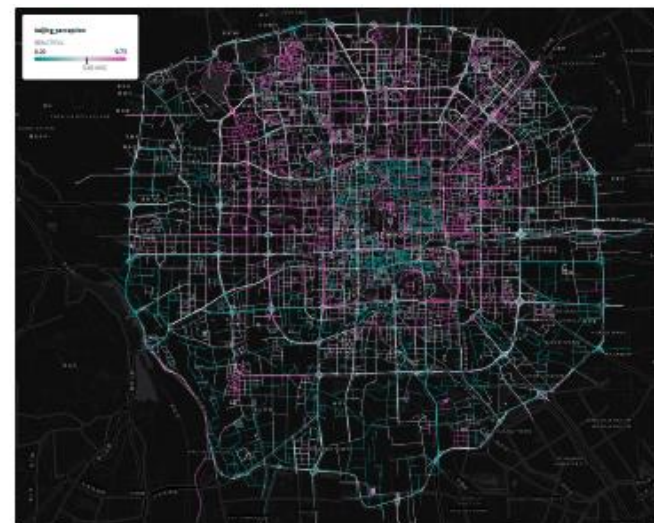
基于照片量化
北京市街道的
不同感知维度



(a) Safe



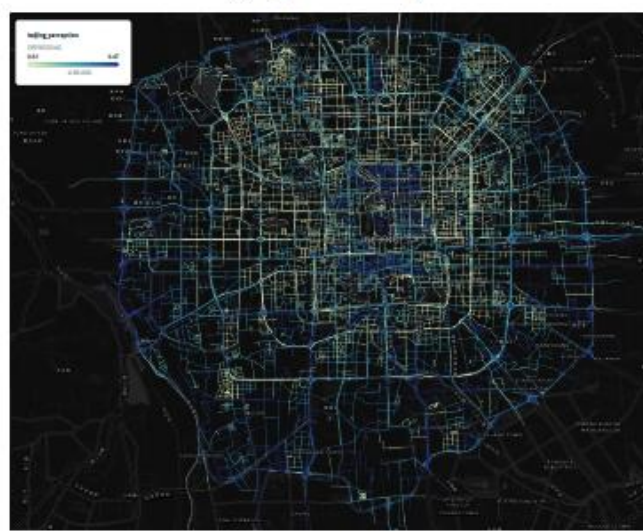
(b) Lively



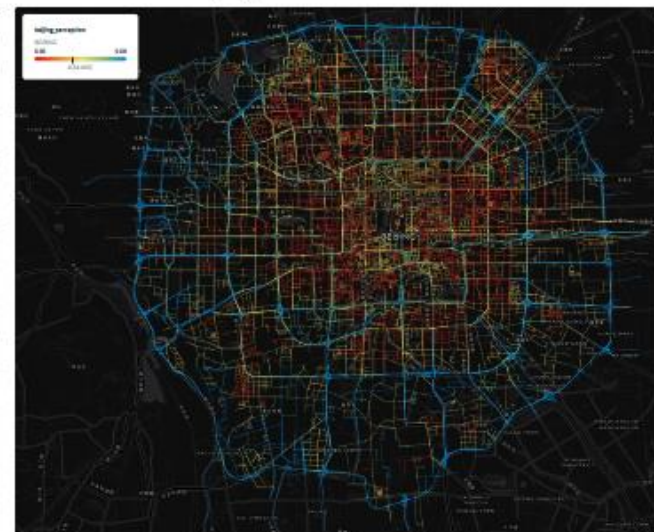
(c) Beautiful



(d) Wealthy



(e) Depressing



(f) Boring



(a)



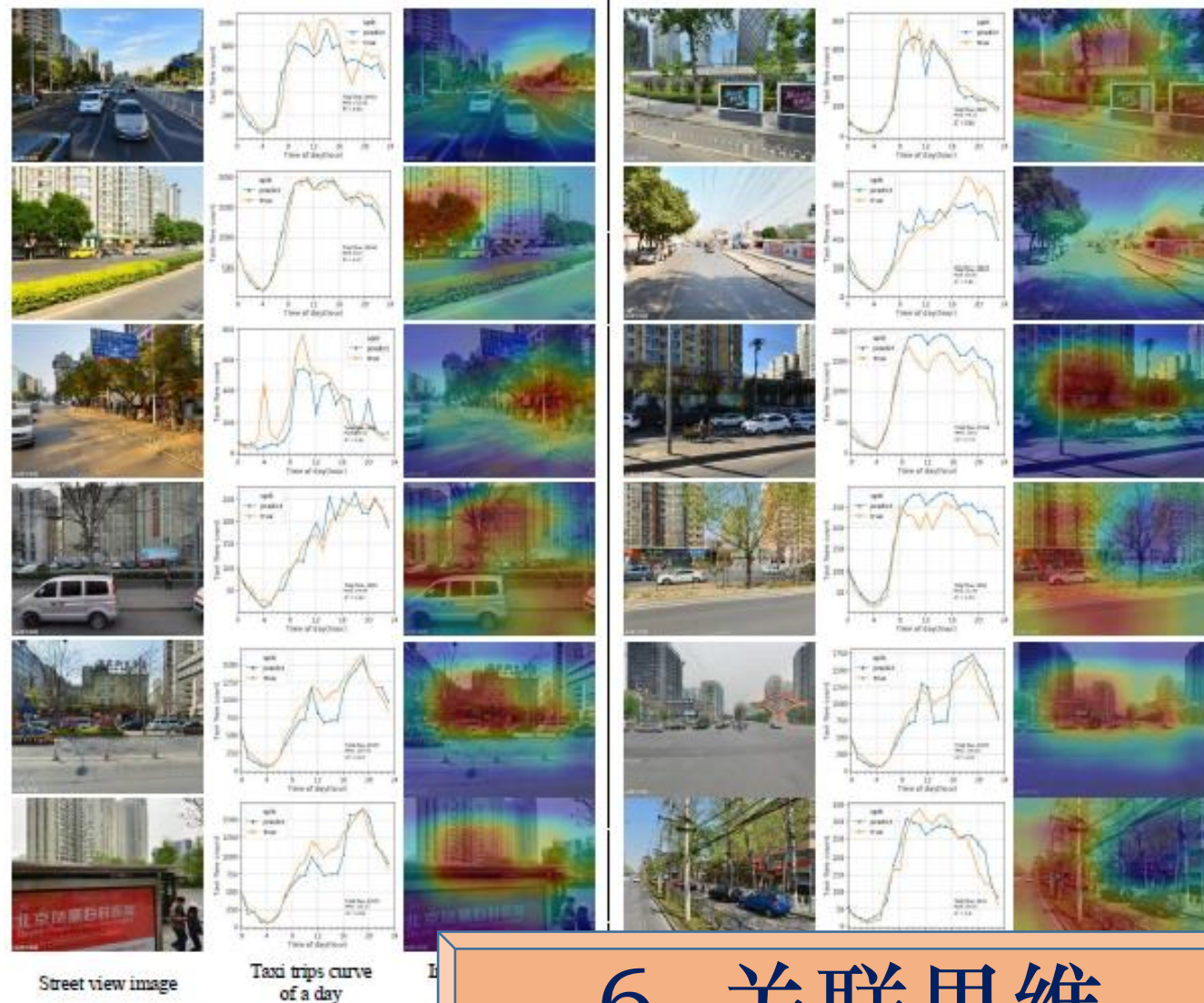
(b)



(c)



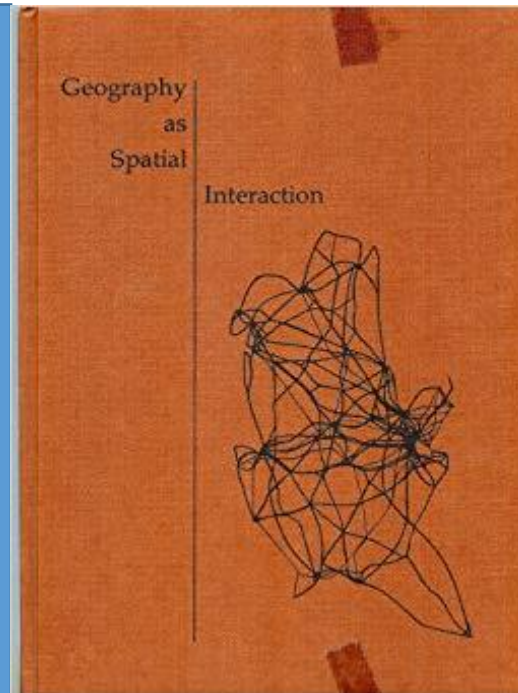
(d)





- 空间交互(Spatial Interaction)
- 空间中任何事物都不是孤立存在的，而是以物质、能量、信息等形式在空间不同位置之间发生着作用和联系

**Geography as
Spatial Interaction.**
--E. L. Ullman (1954)

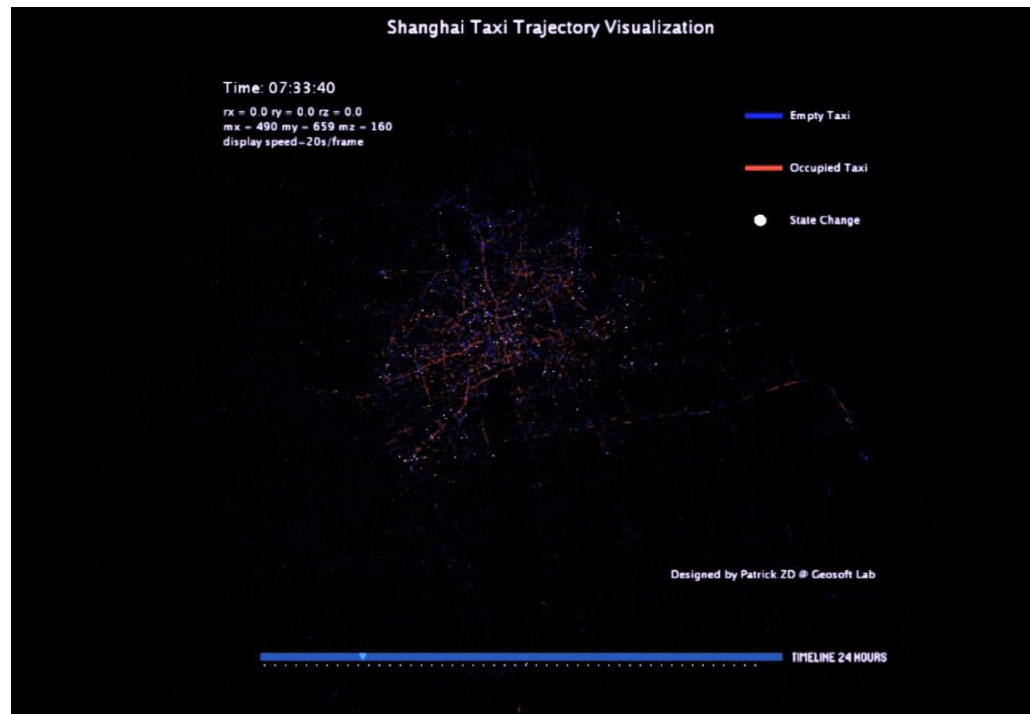




• 空间交互产生的三个基本条件 (Ullman 1956)

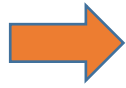
- 互补性：以资源为例，两地发生交互，其中一方可以提供这种资源，而另一方恰好需要这种资源。
- 中介机会：当货物在A和B两地间输送时，A和B两地间介入了另一个能够提供或消费货物的C地，从而产生所谓中介机会，引起货物运输原定起止点的替换。中介机会是改变区域空间相互作用格局的因素
- 可运输性：距离越长，运输成本越高，产生相互作用的阻力越大，即距离衰减效应





- 带有时空标记的、个体粒度的地理大数据为构建空间交互网络，进而为揭示地理空间格局及演化带来了全新的感知手段
- **优势**：从微观和宏观两个视角提供了对空间交互的观察，并为其机理揭示提供了可能性

8. MAUP和生态学谬误



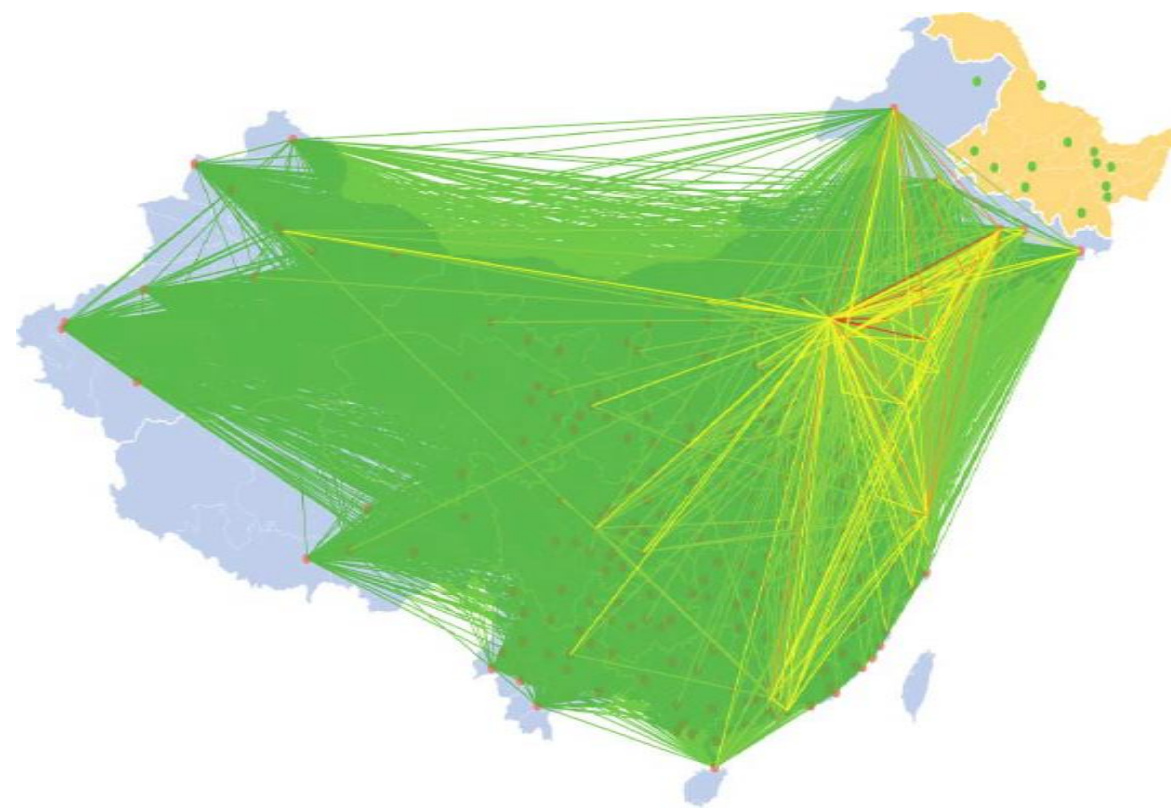
距离衰减、组团结
构、空间依赖、非
线性过程...

要了解空间，就必须了解流和交互；而要了解流和交互，就必须了解网络。

—英国皇家学会院士 M. Batty (2013)



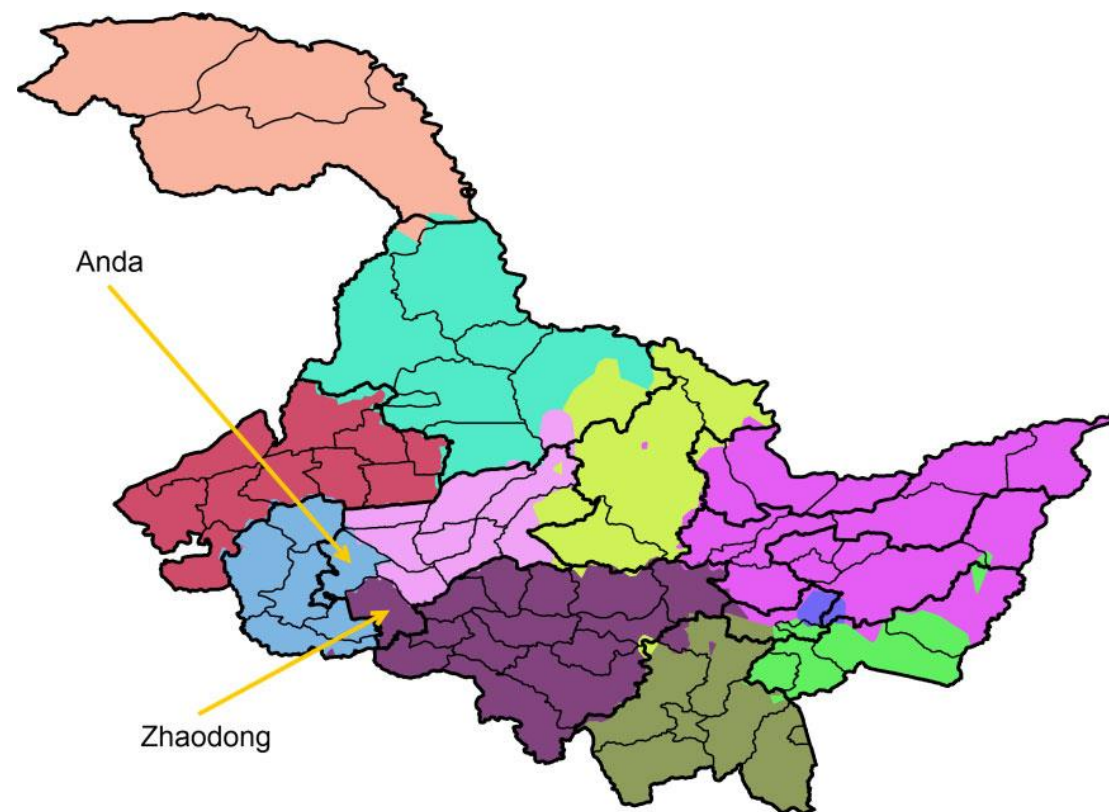
社交媒体数据



手机通讯数据



社交媒体数据

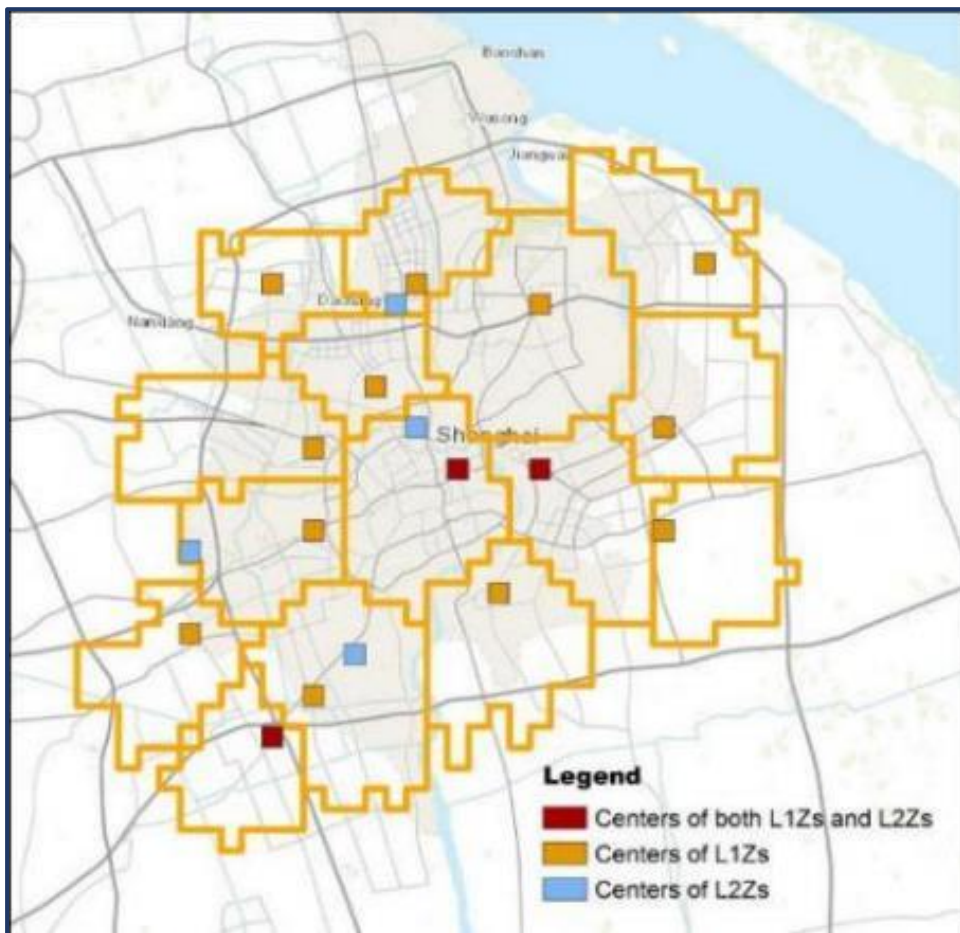


手机通讯数据

识别流场中的“界”

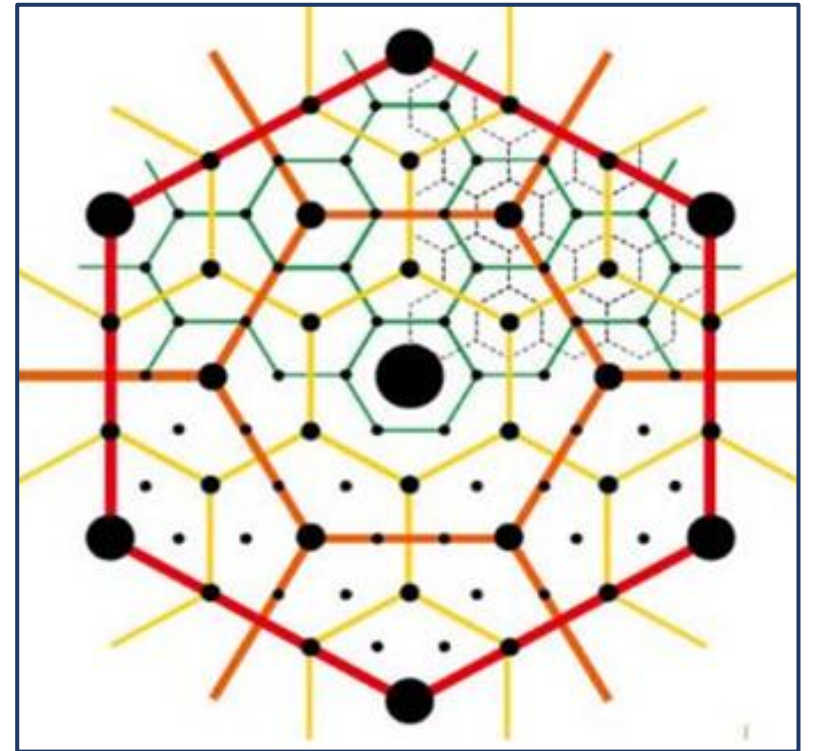


提出顾及地理空间特性的**空间交互网络分割**方法，识别空间组团结构，并支持**交通网络优化**

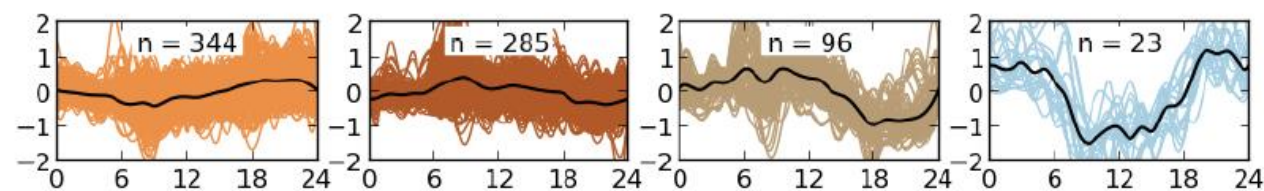
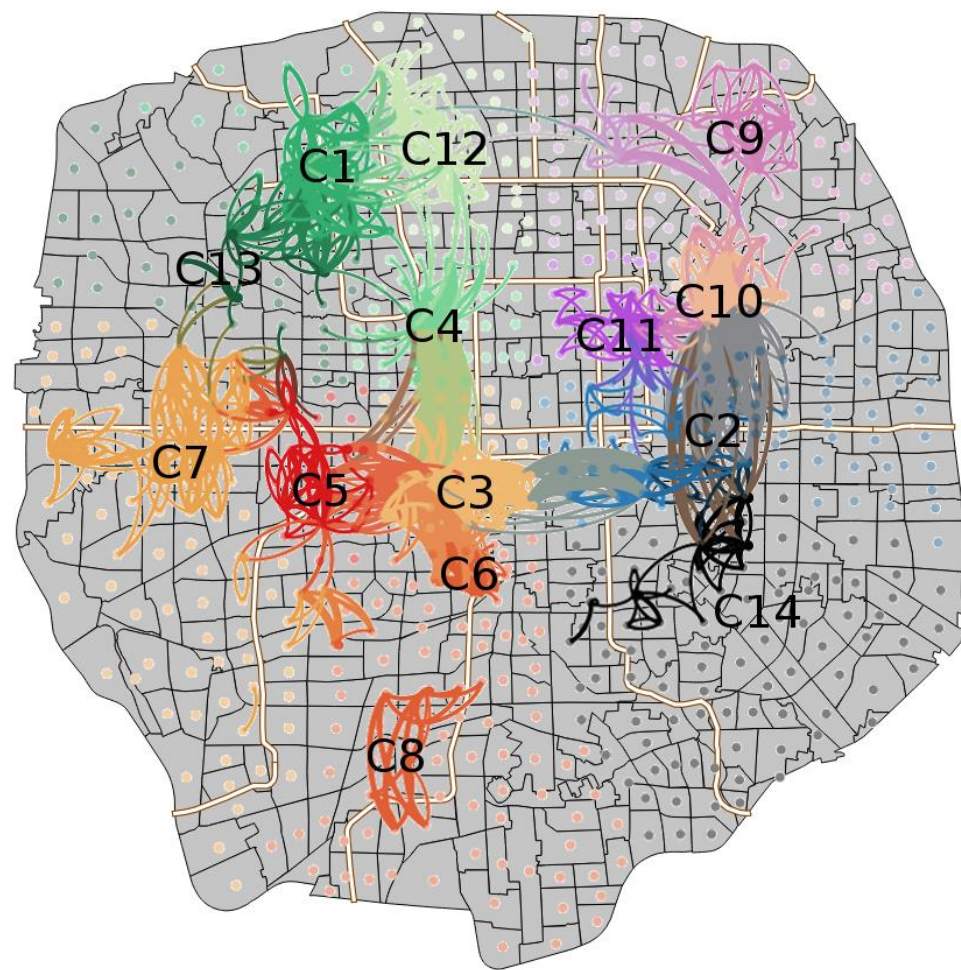


$$Q = \frac{1}{M} \sum_{i,j} \left(A_{ij} - \frac{w(i)w(j)}{M} \right) \delta(c(i), c(j))$$

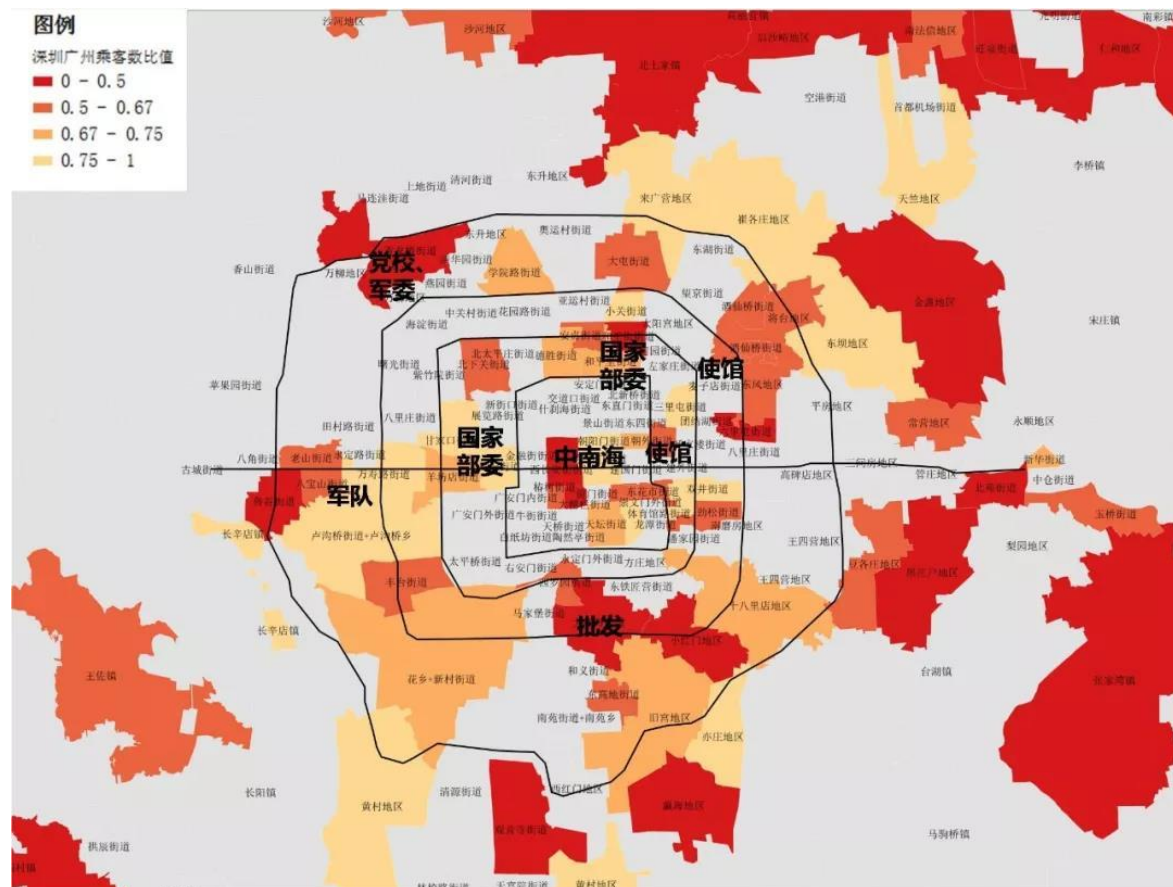
$$A_{ik} = \frac{C_{ij}}{\hat{C}_{ij}} = \frac{C_{ij} D_{ij}^{\beta}}{\hat{P}_i \hat{P}_j}$$



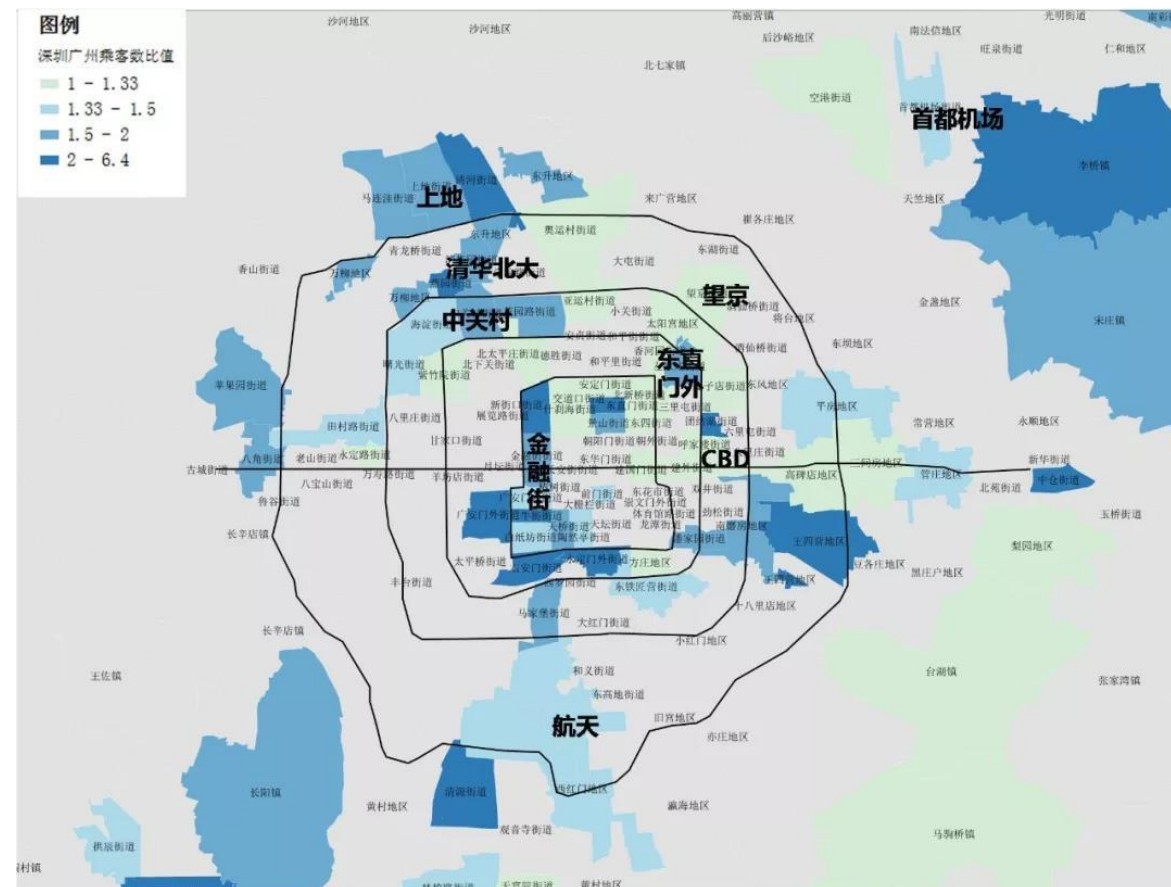
揭示上海市**多中心、同心环状**结构，对克里斯塔勒中心地理论进行了实证



(a) Weekday



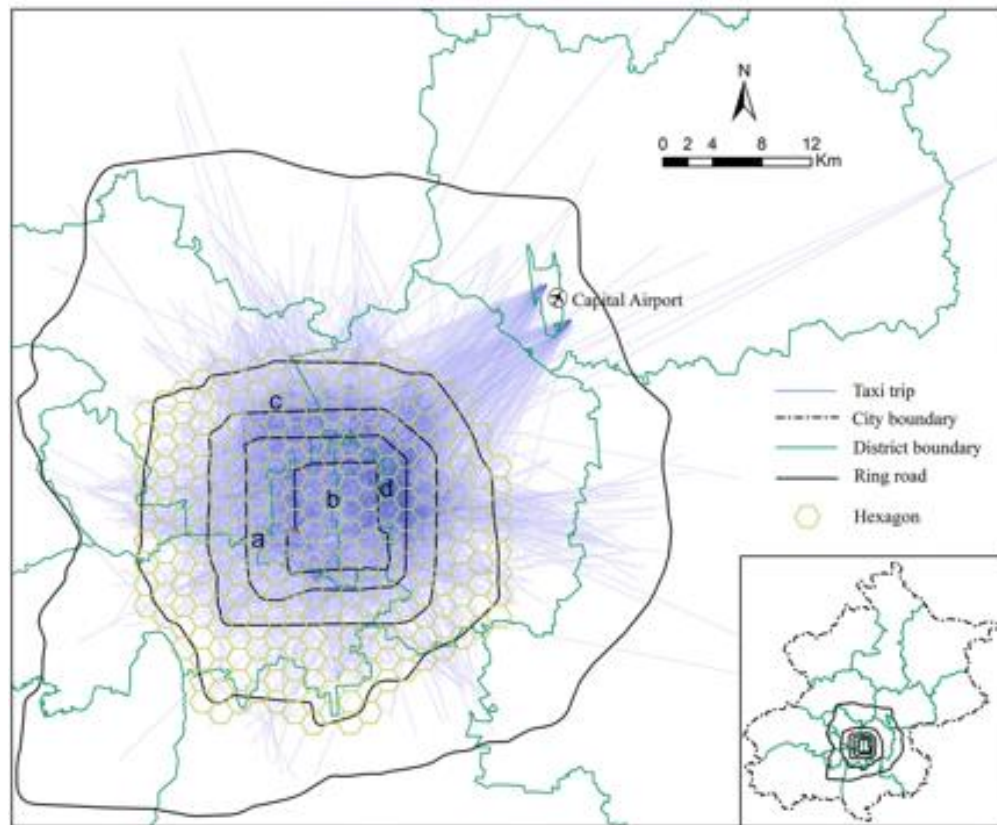
飞广州大于飞深圳旅客数量的乡镇街道



飞深圳大于飞广州旅客数量的乡镇街道

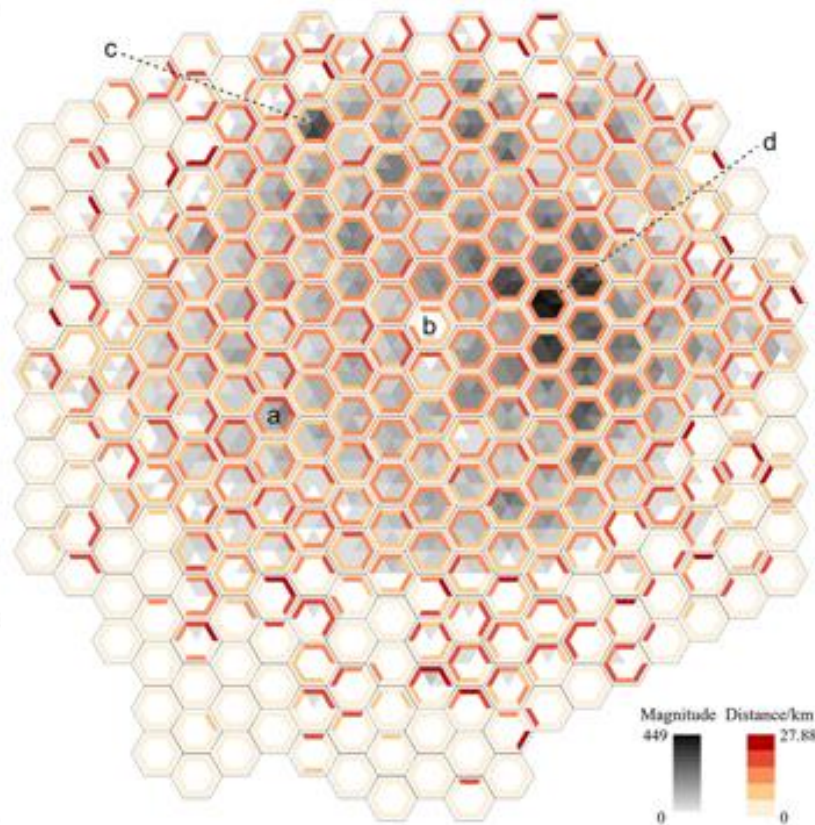
Spatial interaction data

Beijing taxi trips

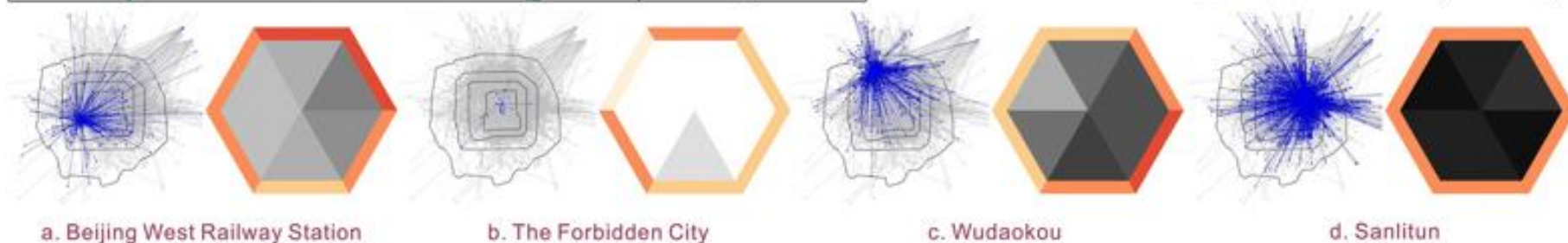


Direction-based pattern map

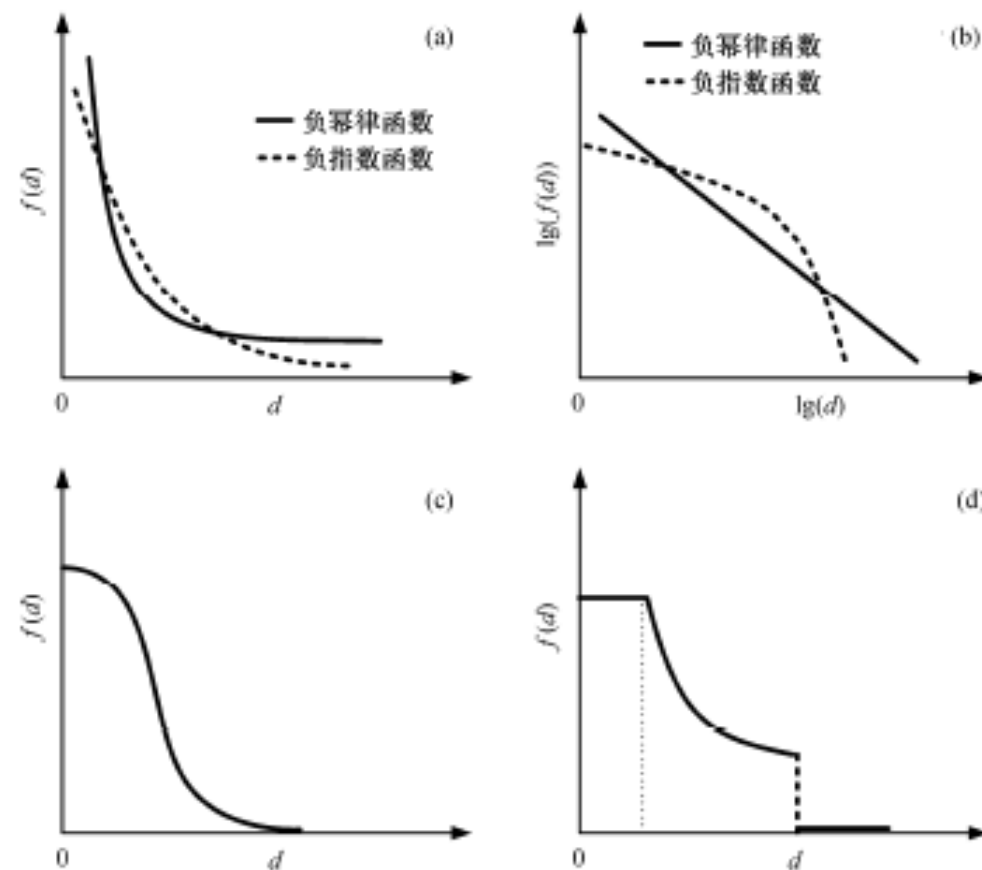
1:00 - 5:00 a.m.



地理空间单元交互特征的可视化



模型	距离	评价
重力模型	幂律函数 $f(d) = d^{-\beta}$	简单易用，但是缺乏理论基础 参数 P 和 β 的取值对模型效果影响很大
Wilson最大熵模型	指数函数 $f(d) = e^{-\beta d}$	有理论基础 参数计算过程复杂
中介机会模型	相对距离 机会数目	绕开了争议较大的 β 取值问题，中介机会思想对后期个体移动模型的产生具有重大作用



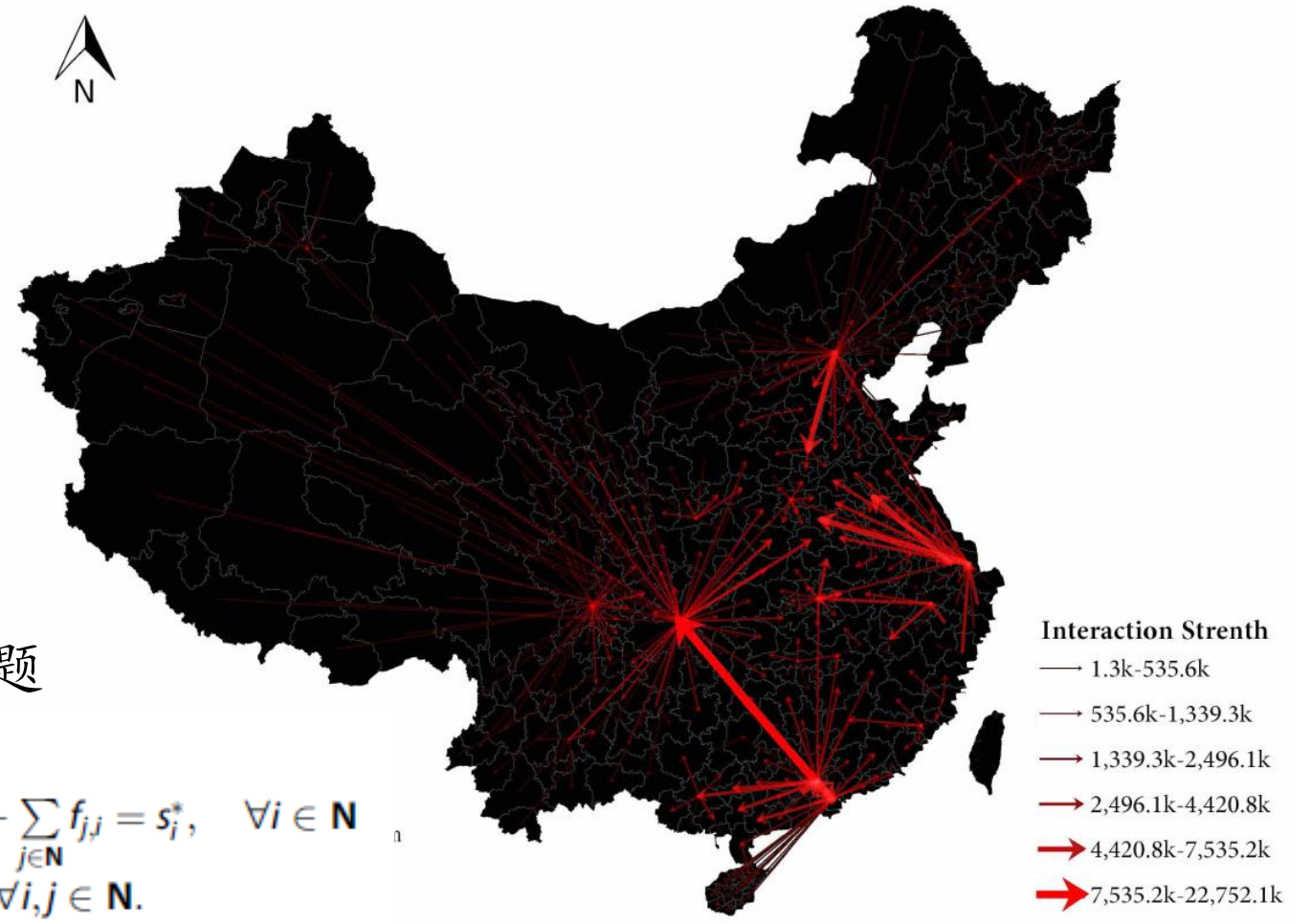
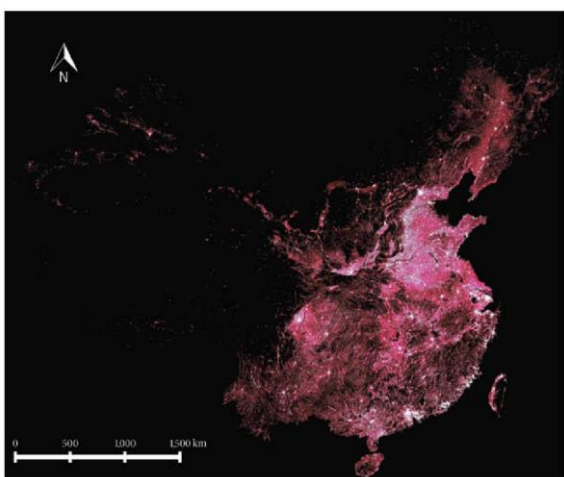
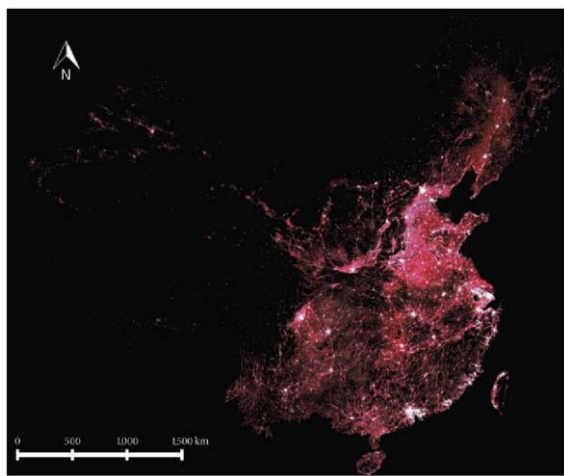
几种不同的距离衰减曲线 (Liu et al., 2014)



数据	系数	论文
上海出租车数据	1.08	Liu et al. 2012, Journal of Geographical Systems
手机数据所提取的轨迹	1.3-1.8	Kang et al. 2012, Physica A
微博数据所提取的轨迹	0.8	Liu et al. 2014, PLOS ONE
航空网络数据	0.87 (2001), 0.45 (2008)	Xiao et al. 2013, Professional Geographer
手机通话数据	0.5	Kang et al. 2013, IJGIS
Web页面地名共现	0.2	Liu et al. 2014, TGIS



提出了一种基于连续空间分布反演空间交互模式的方法，在推算中国春节期间城际返乡流分析中取得了理想结果



空间优化问题

$$\begin{aligned} & \text{minimize} \\ & \text{subject to} \\ & C^T \times F \\ & s_i - \sum_{j \in N} f_{ij} + \sum_{j \in N} f_{ji} = s_i^*, \quad \forall i \in N \\ & f_{ij} \in \mathbb{Z}^n, \quad \forall i, j \in N. \end{aligned}$$

• 流的空间自相关 (*Liu et al., 2015*)

全局空间自相关指数

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (u_i u_j + v_i v_j)}{\sum_i (u_i^2 + v_i^2)}$$

局部空间自相关指数

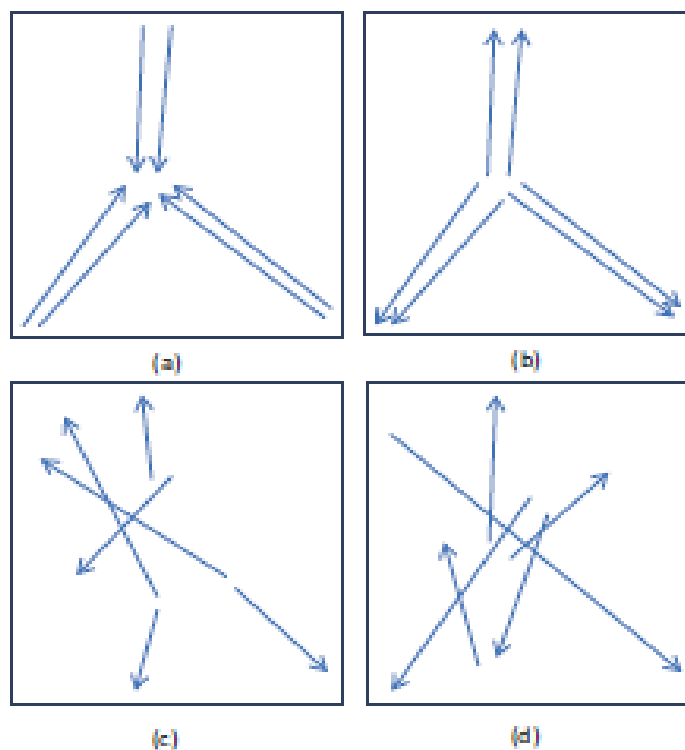
$$I_i = \frac{u_i}{m_2} \sum_j w_{ij} u_j + \frac{v_i}{m_2} \sum_j w_{ij} v_j$$

式中：

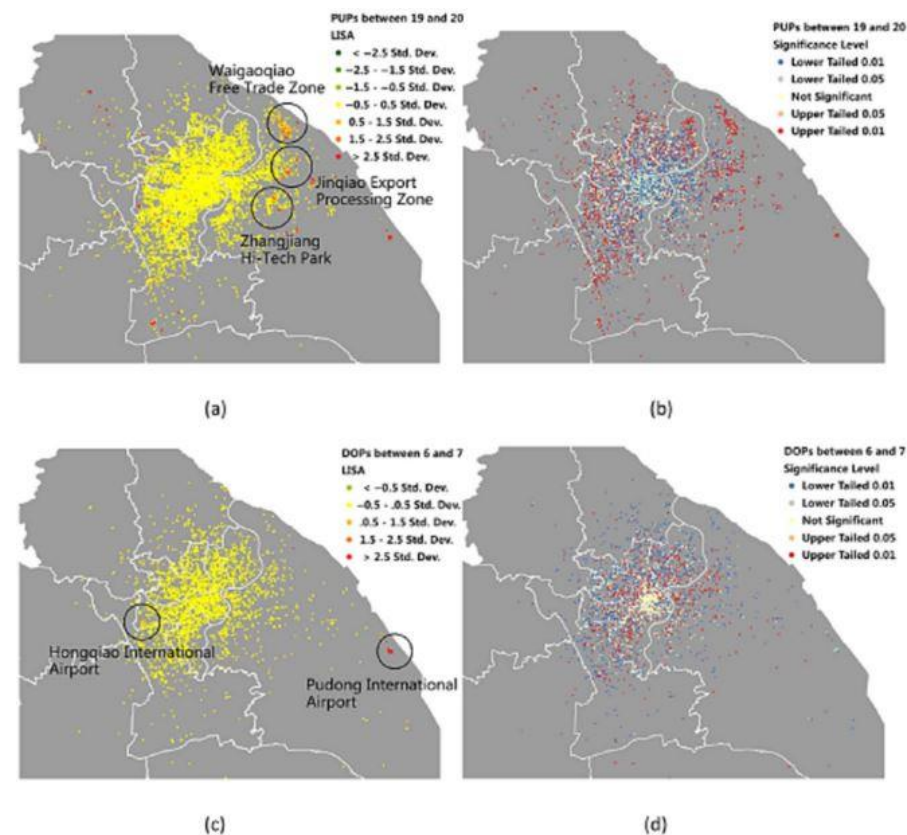
$$v_i = (y_i^D - y_i^O) - (\bar{y}^D - \bar{y}^O)$$

$$u_i = (x_i^D - x_i^O) - (\bar{x}^D - \bar{x}^O)$$

$$m_2 = \frac{\sum_i (u_i^2 + v_i^2)}{n}$$



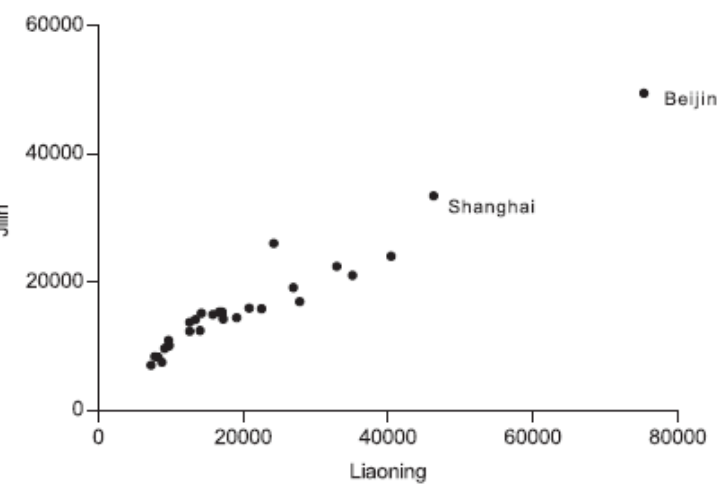
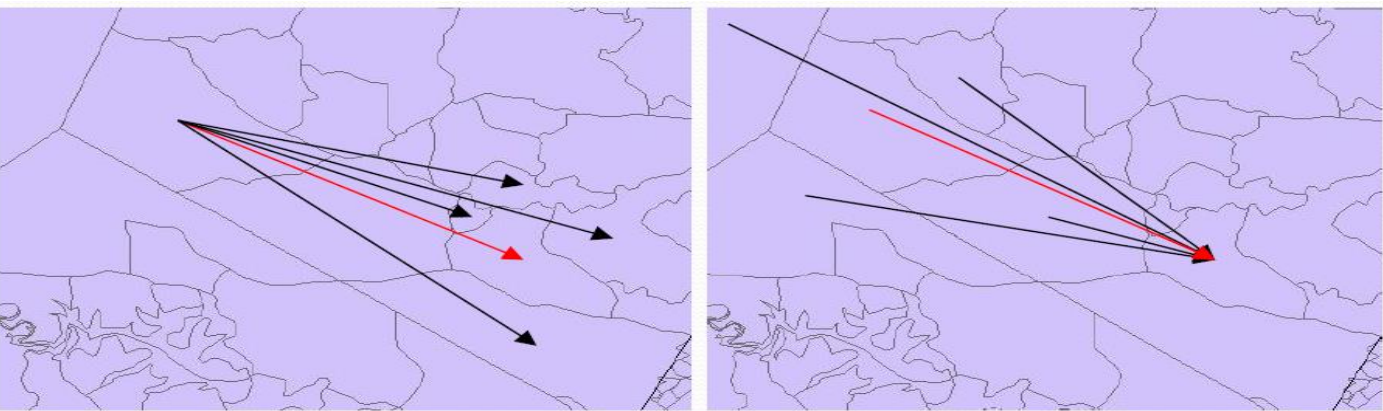
向量的空间自相关



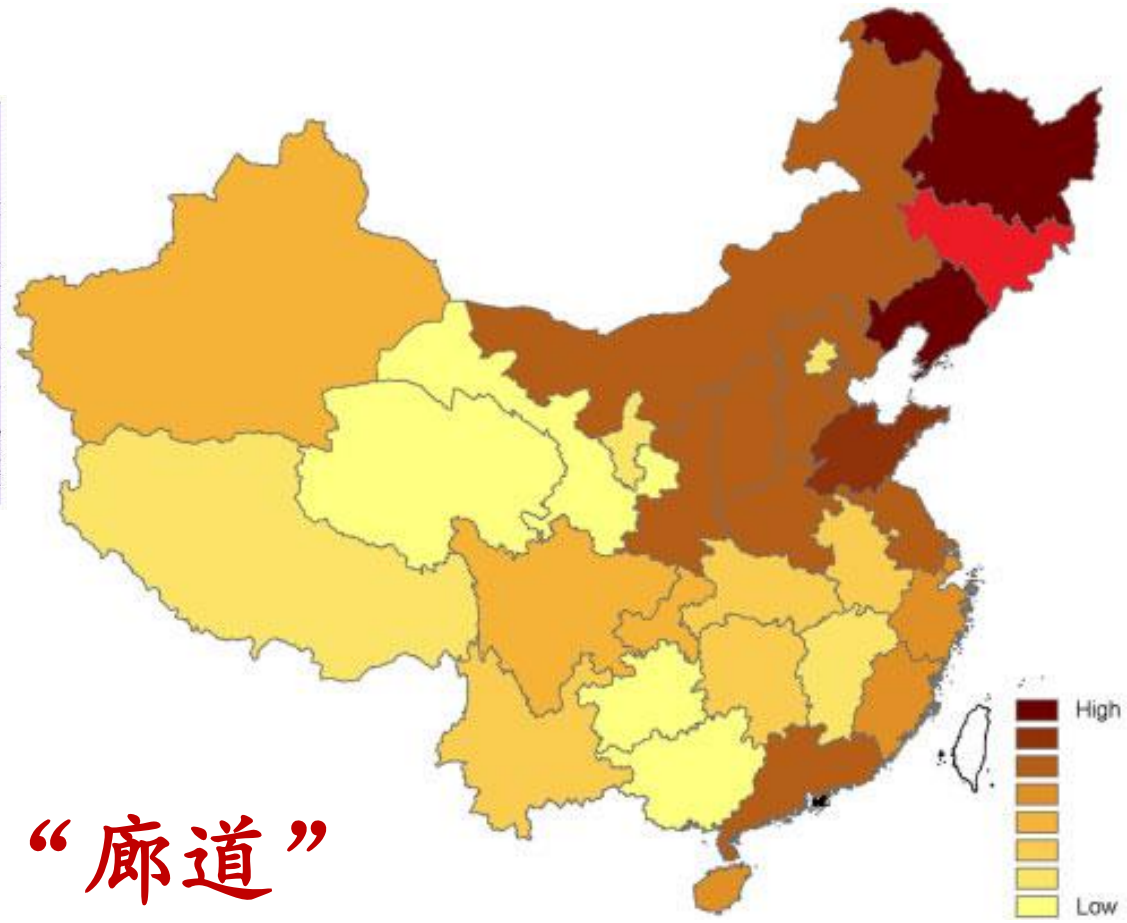
上海市出租车数据OD局部空间自相关指数及显著性水平



相近OD的交互总量相近→正的空间自相关



识别流场中的“廊道”



吉林省和其它省份交互向量的相似性



• 基于流地图的分析方法

空间聚集测度 - 流数据的K函数 ([Tao et al, 2016](#))

$LocK_i = E(\# \text{ of flow events within Flow distance } r \text{ of flow } i)$

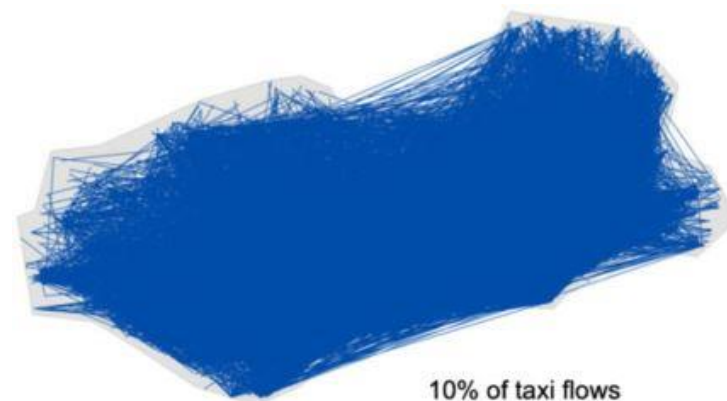
式中流距离：

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (u_i - u_j)^2 + (v_i - v_j)^2}$$

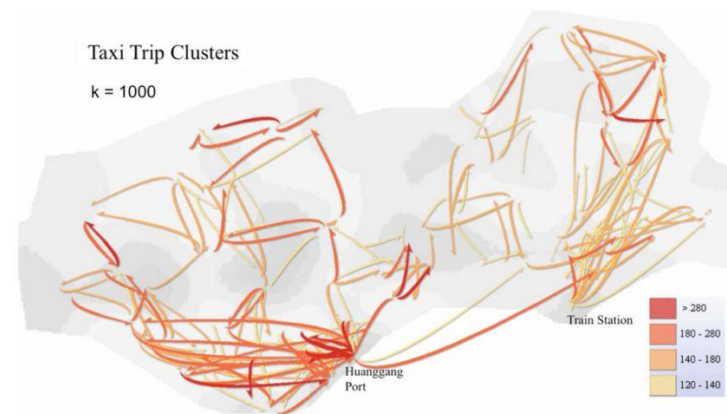


北卡罗莱纳州夏洛特市
交通“盗窃-找回”数据的聚集

流聚类 ([Zhu et al, 2014](#))



10% of taxi flows



深圳市原始出租车数据层次聚类

• 非负矩阵分解 Non-negative Matrix Factorization

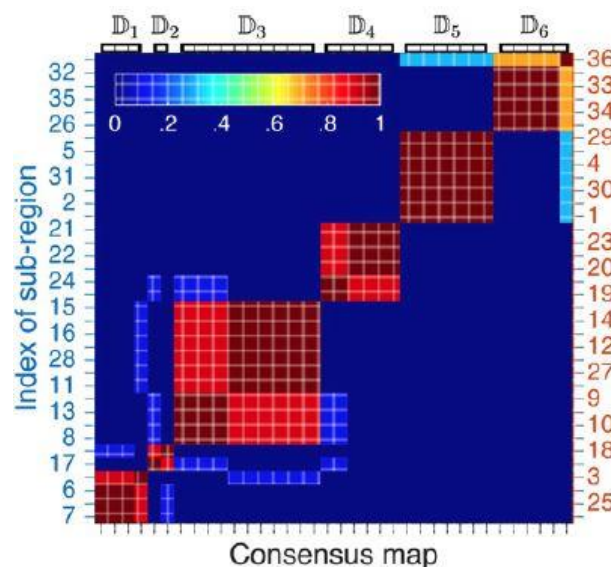
空间交互的OD位置来自不同的空间过程，并一起构建了全局性交互分布，矩阵分解能分离出这些空间过程，从而发现全局模式中的子结构

$$A_{mn} = W_{mk} \times H_{kn}$$

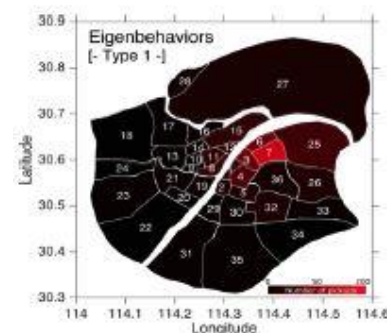
$$\begin{bmatrix} a_{11} & \dots & a_{1m} \\ a_{21} & \dots & a_{2m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{bmatrix} = \begin{bmatrix} w_{11} & \dots & w_{1k} \\ w_{21} & \dots & w_{2k} \\ \vdots & \ddots & \vdots \\ w_{n1} & \dots & w_{nk} \end{bmatrix} \begin{bmatrix} h_{11} & \dots & h_{1n} \\ h_{21} & \dots & h_{2n} \\ \vdots & \ddots & \vdots \\ h_{k1} & \dots & h_{kn} \end{bmatrix}$$

约束：

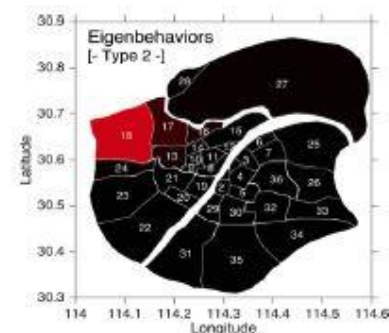
$$\begin{cases} a_{ij} \geq 0 \\ w_{ij} \geq 0 \\ h_{ij} \geq 0 \end{cases}$$



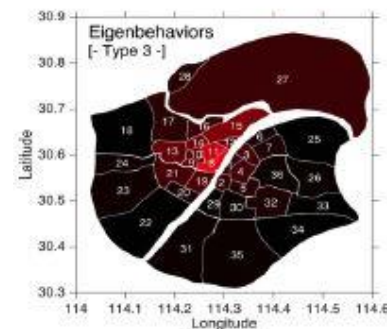
武汉市36个区域出租车交互矩阵分解得到的6个区块组团及其空间分布
(Kang and Qin, 2016)



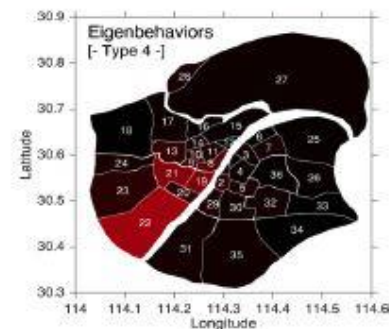
(a) BB-1 service scheme



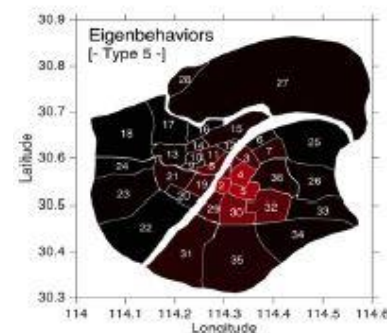
(b) BB-2 service scheme



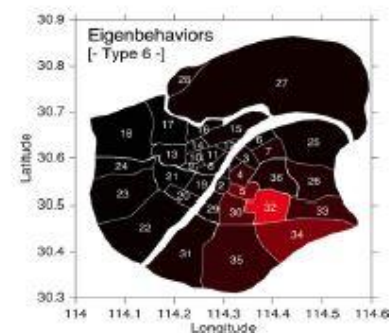
(c) BB-3 service scheme



(d) BB-4 service scheme

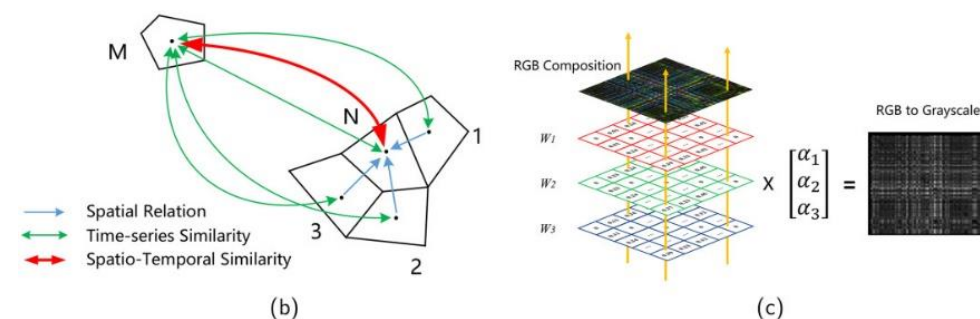
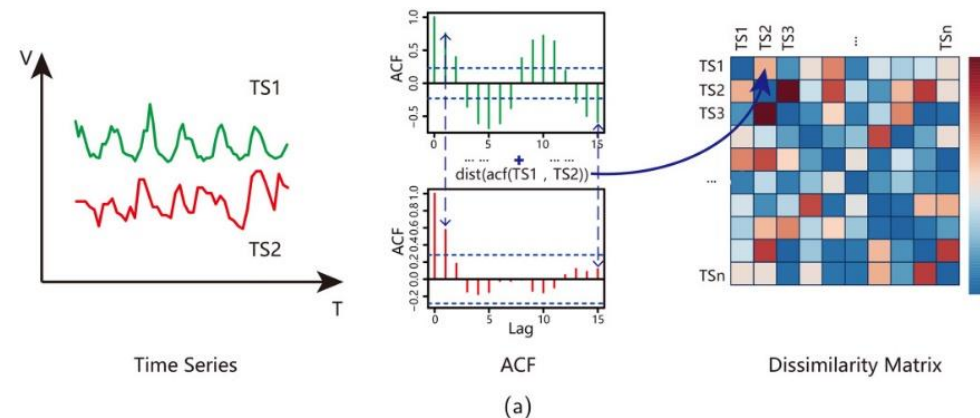
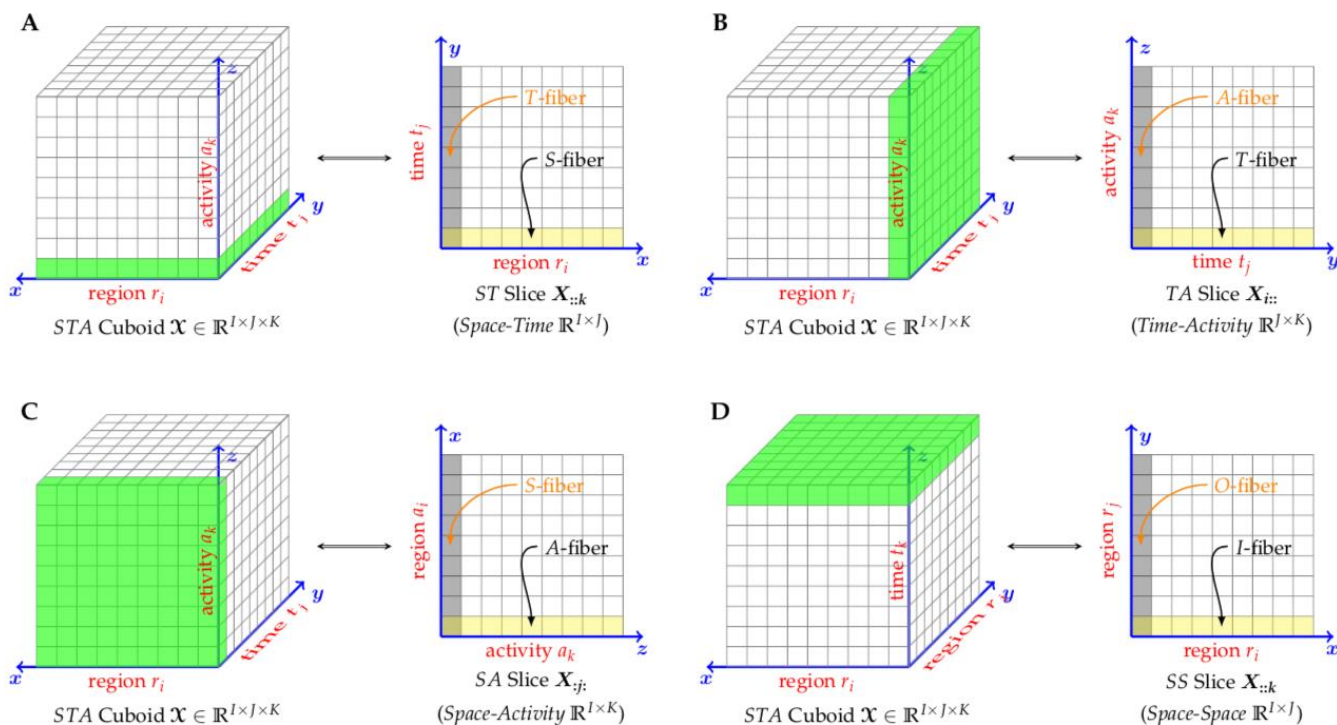


(e) BB-5 service scheme



(f) BB-6 service scheme

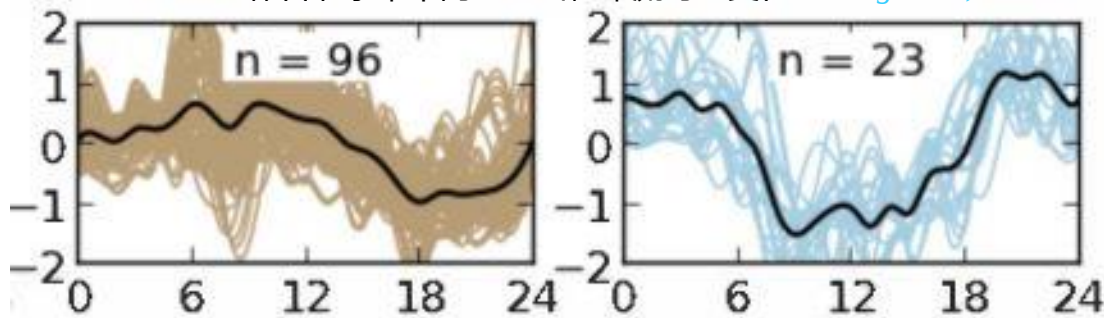
提出了基于多源地理大数据的城市空间解析模型，为理解城市出行多样性及其相互作用提供了方法支撑



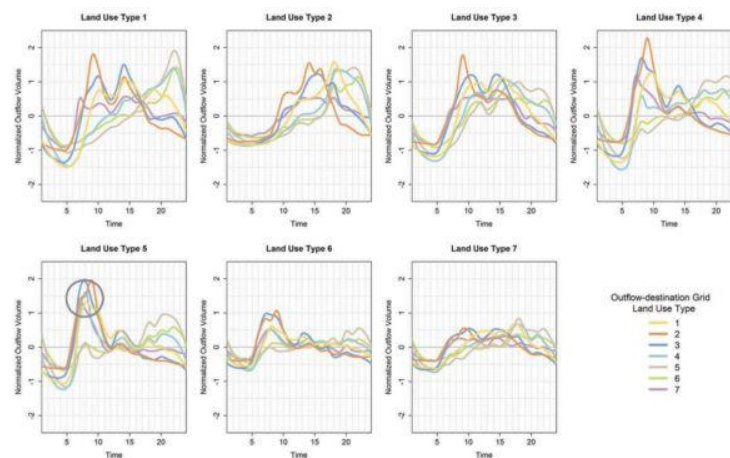
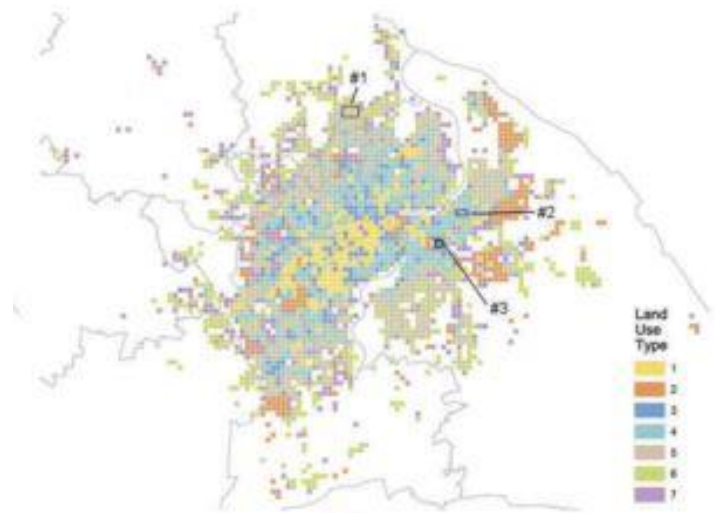
10. 数据有偏性和集成思维



工作日北京市不同地区出租车流时态变化 (Kang et al., 2016)



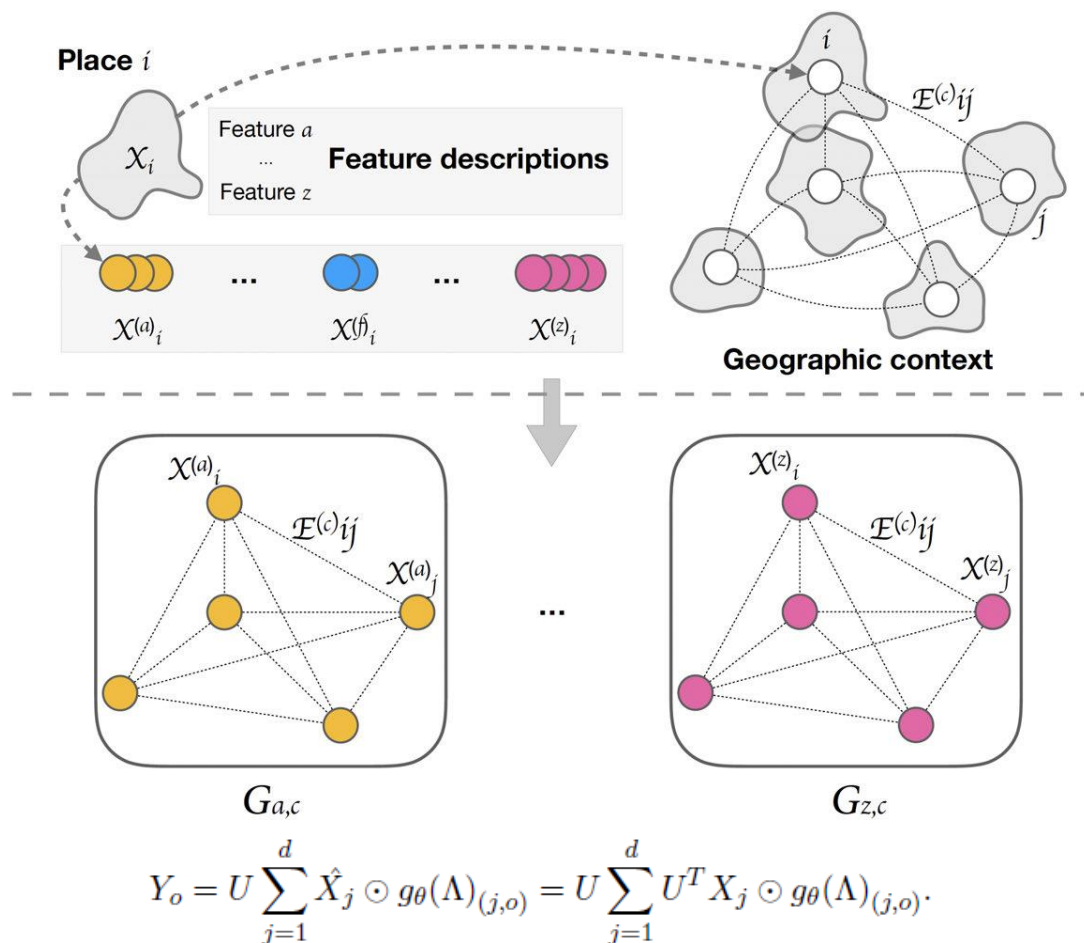
$$a, b \in (T_1) \text{ and } c, d \in (T_2) \rightarrow \text{sim}(\text{inter}(a, c), \text{inter}(b, d))$$



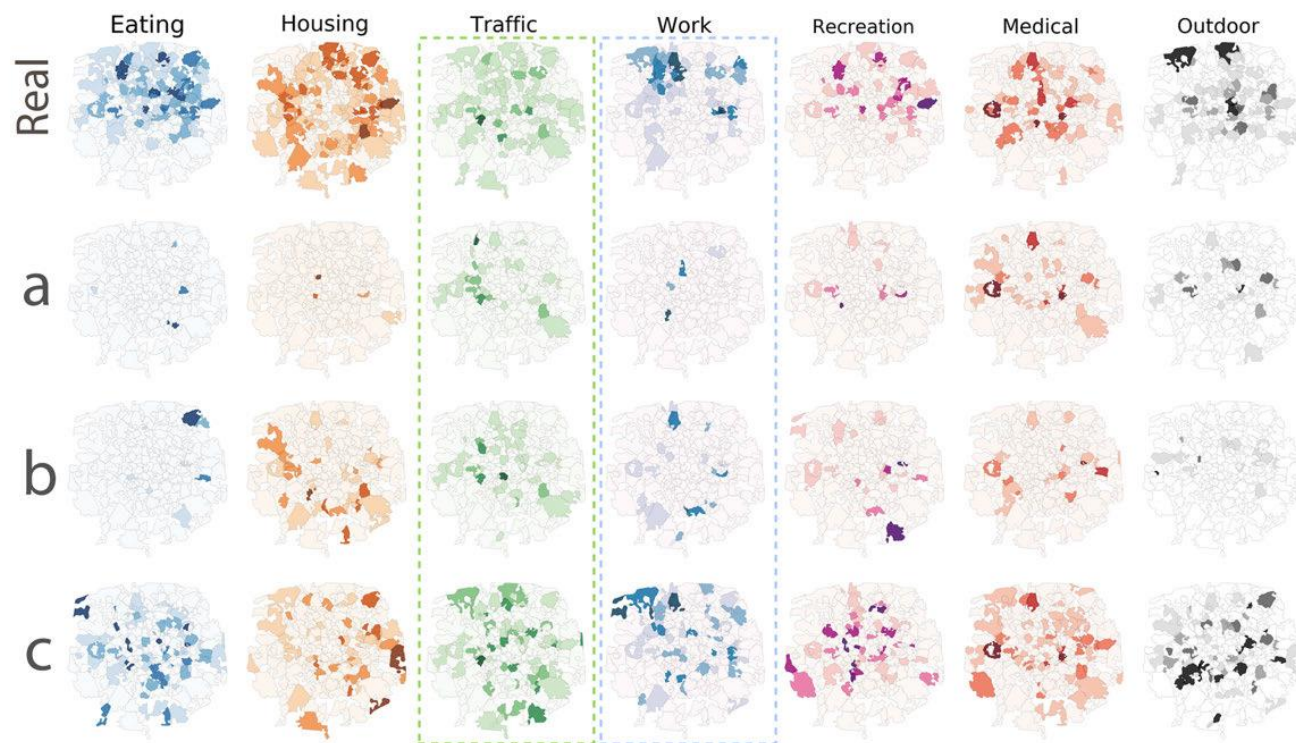
↑ 基于分组交互特征
(grouped interaction
signature)
的上海市用地功能分类

←
不同类别的交互时态变化

(Liu et al., 2016)



图卷积网络构建



基于图卷积网络集成交互特征

• 介数中心性 (Betweenness centrality) 分析

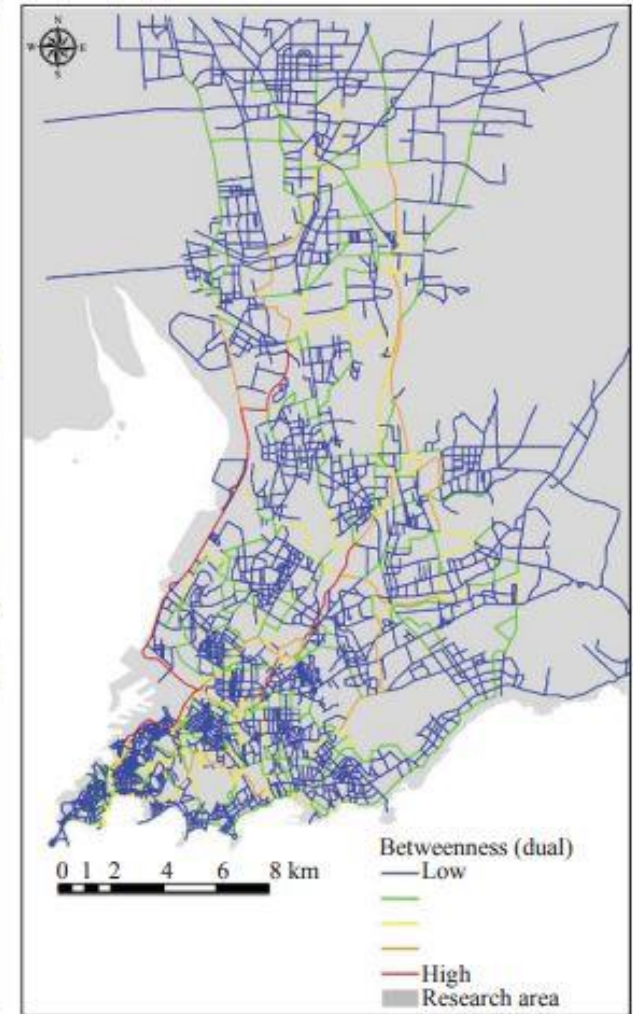
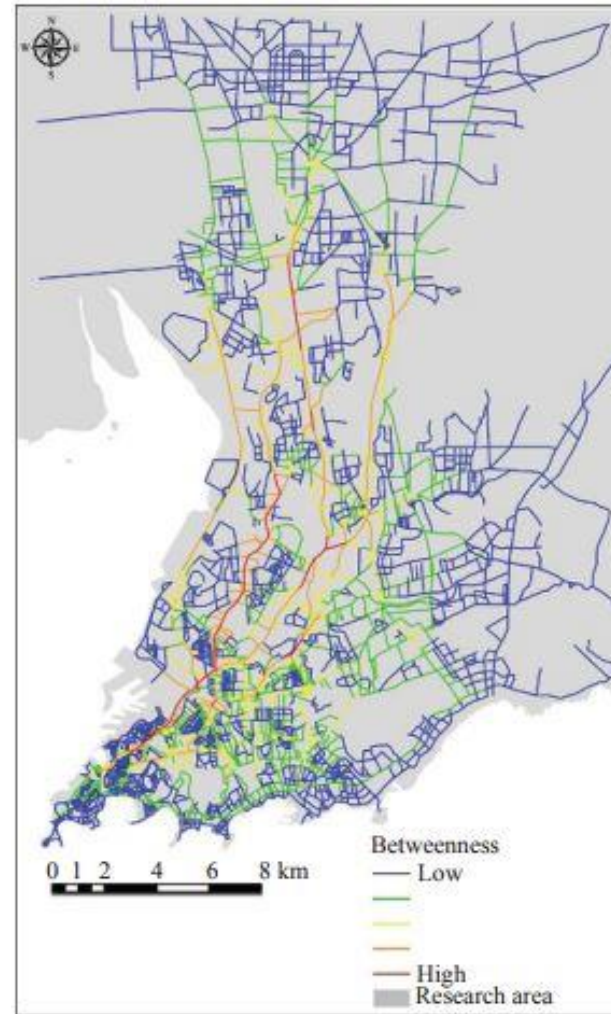
给定包含N个节点的网络，任一节点的介数中心性定义为：

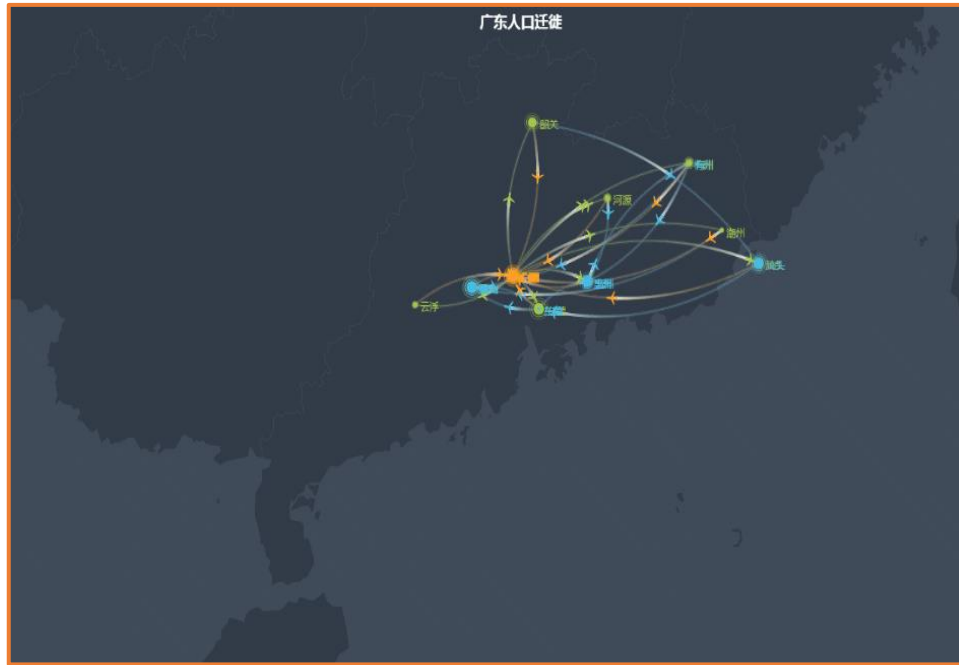
$$C_i^b = \sum_{j=1; k=1; i \neq j \neq k}^N \frac{n_{jk}(i)}{n_{jk}}$$

式中：

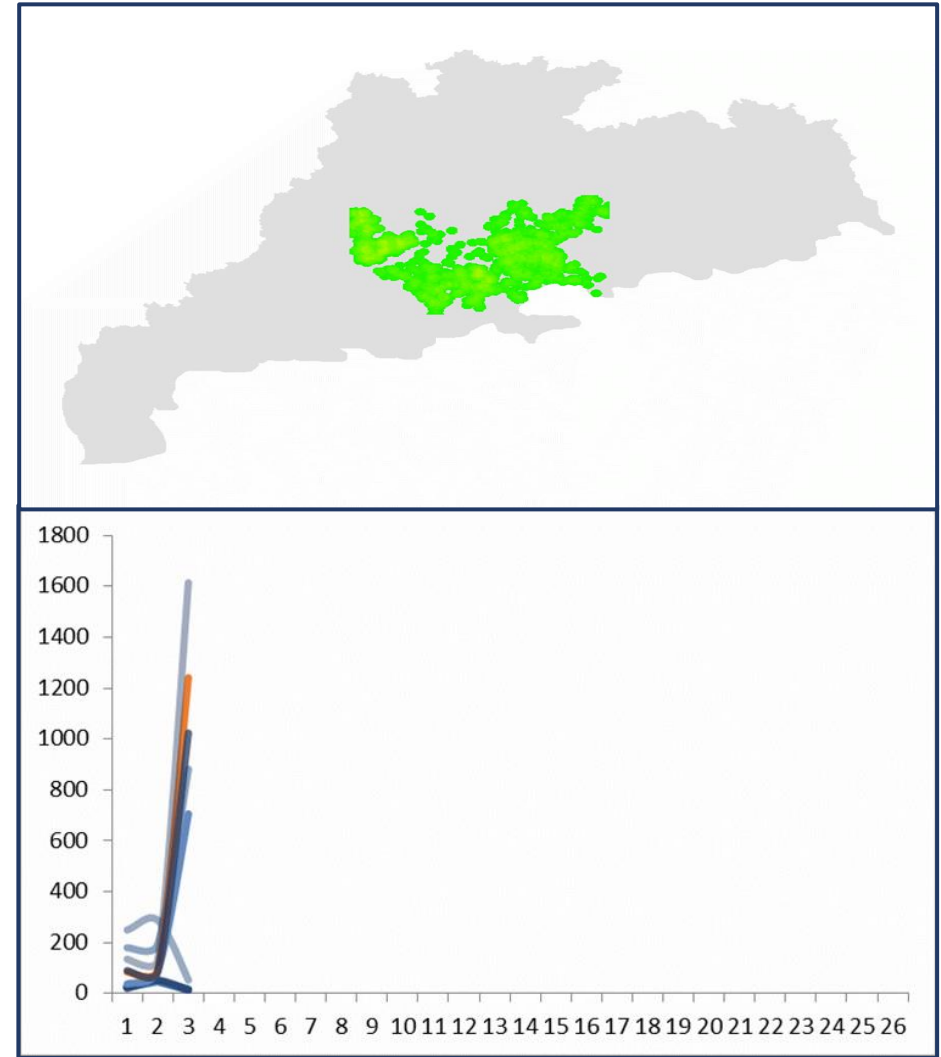
n_{jk} 为节点 j 与 k 之间的最短路径条数

$n_{jk}(i)$ 为节点 j 与 k 之间的最短路径通过节点 i 的条数





基于大数据观察到的空间交互流

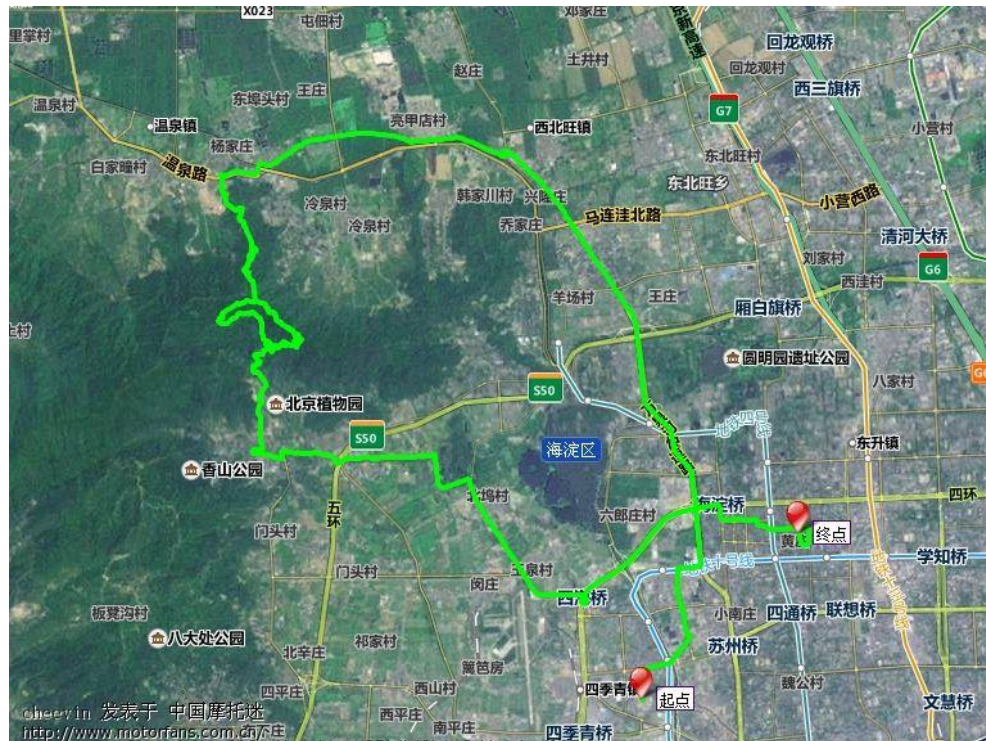


传染病蔓延时空过程

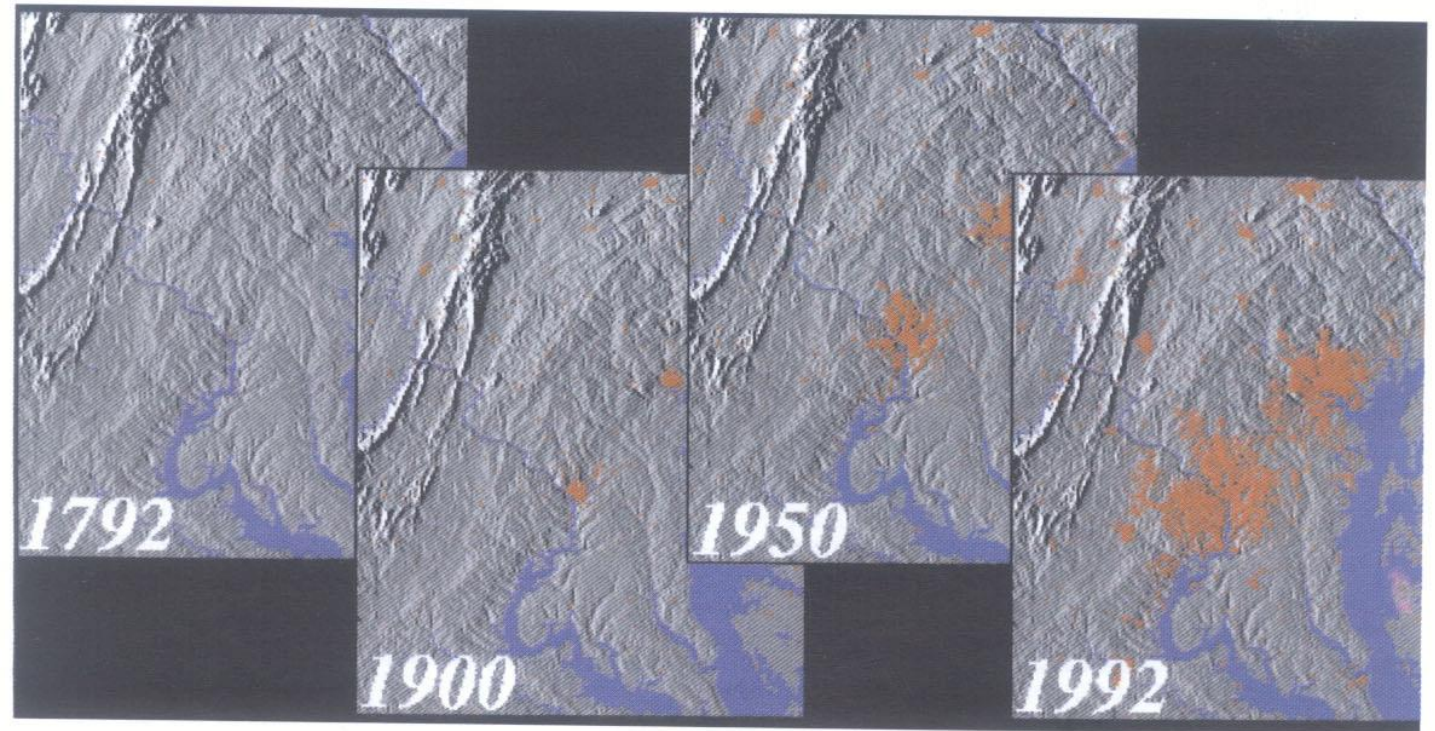


3、演化过程感知

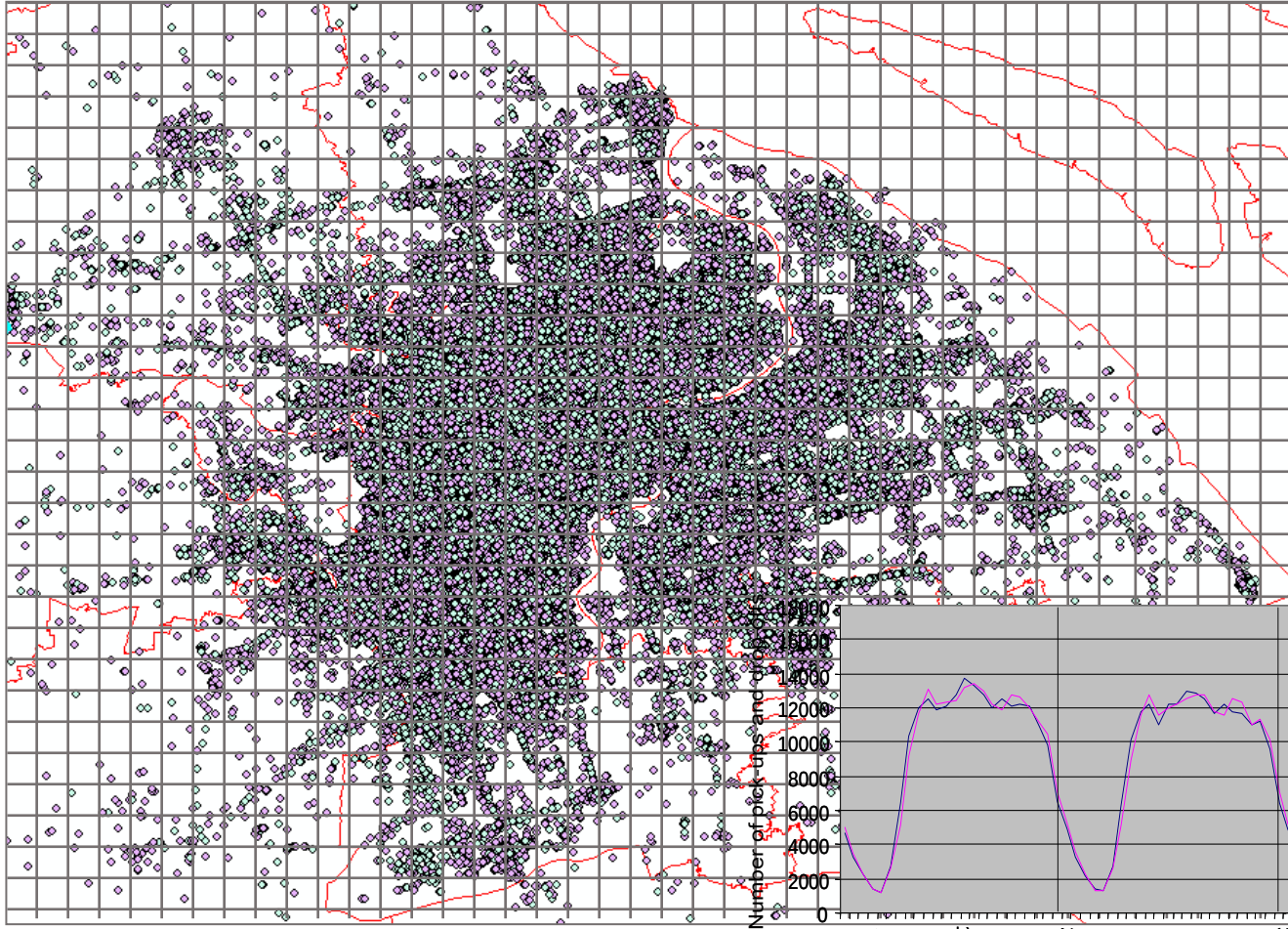
地理世界处在持续不断的变化过程中，大数据所具有的时间标签，为感知“人”和“地”两个方面的动态演化提供了支持。



短时间尺度动态

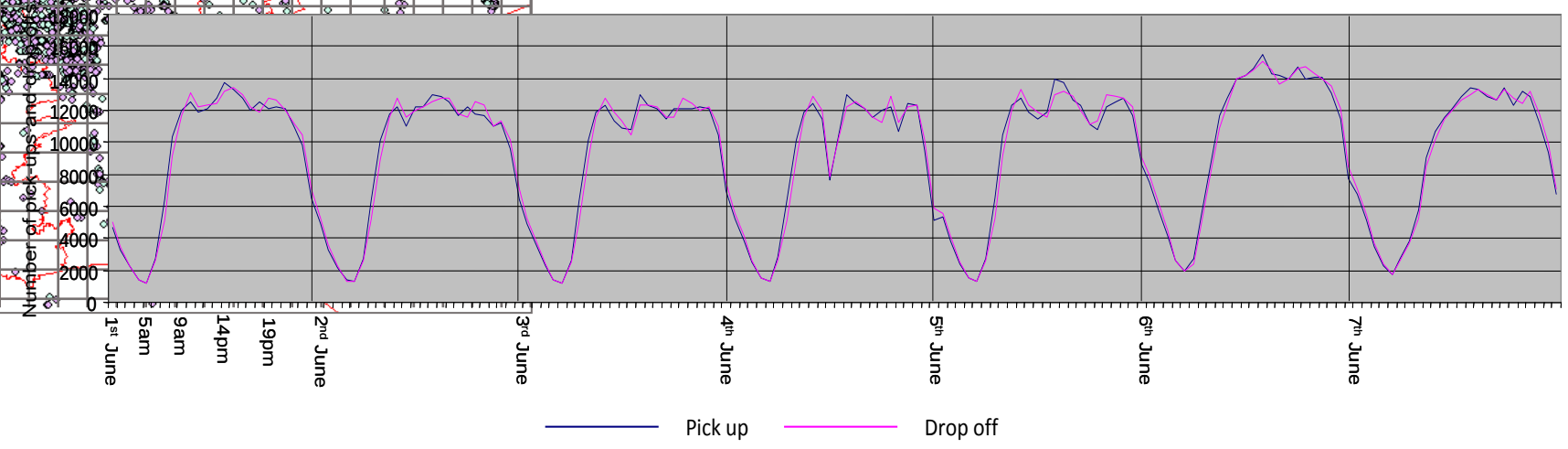


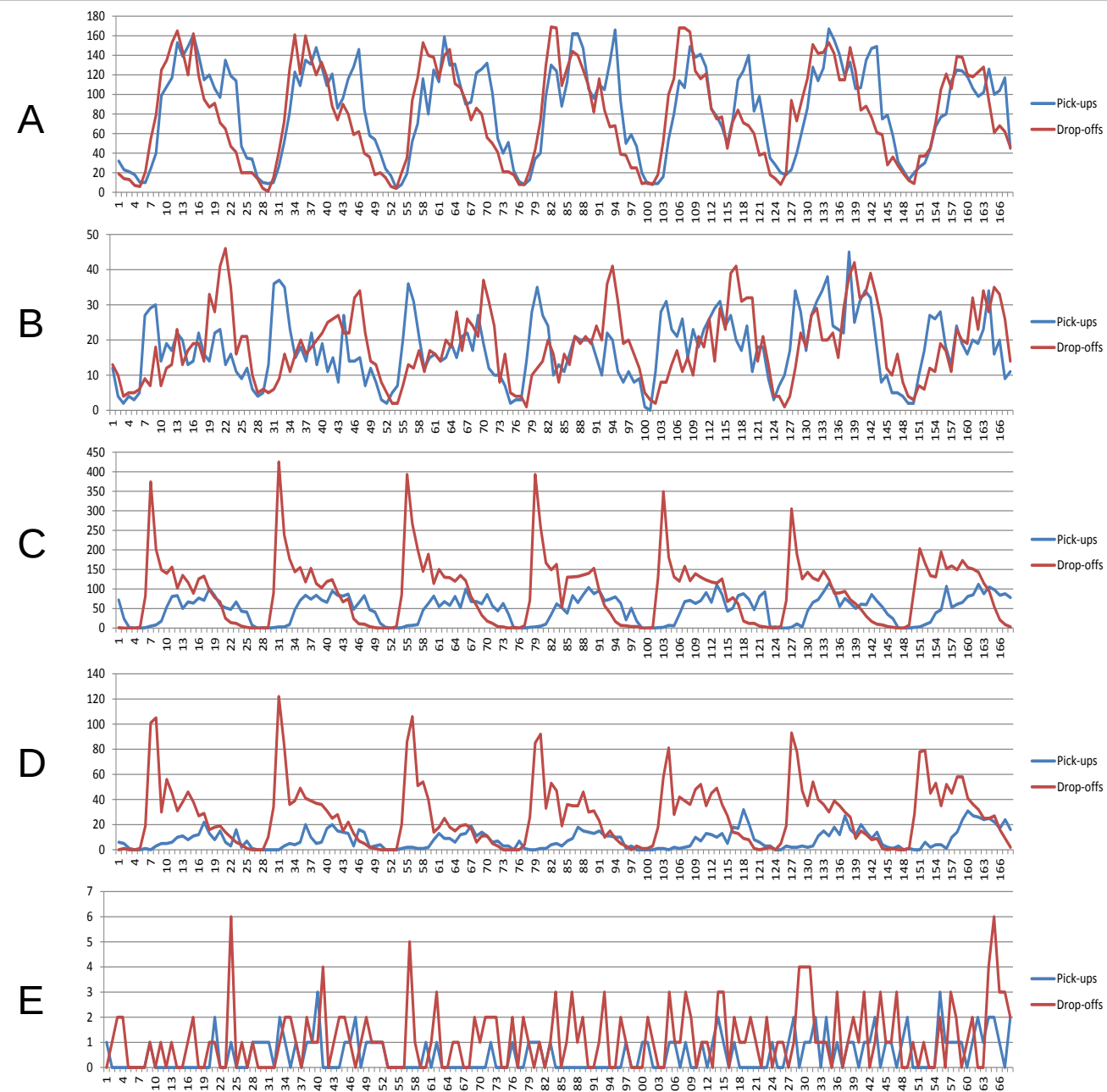
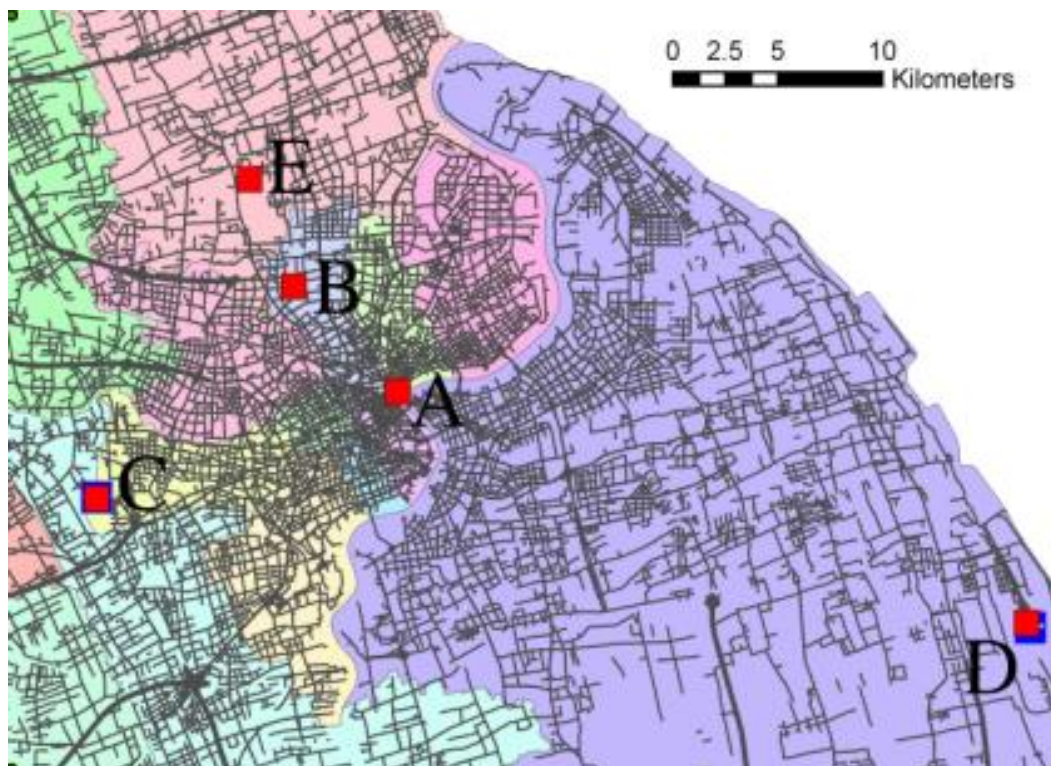
长时间尺度动态

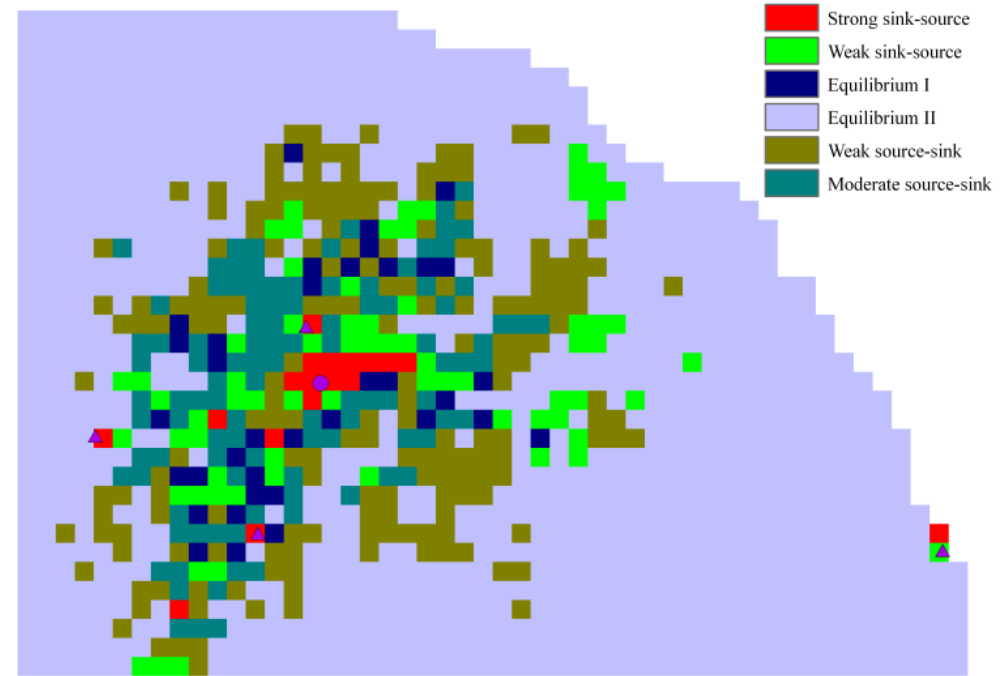
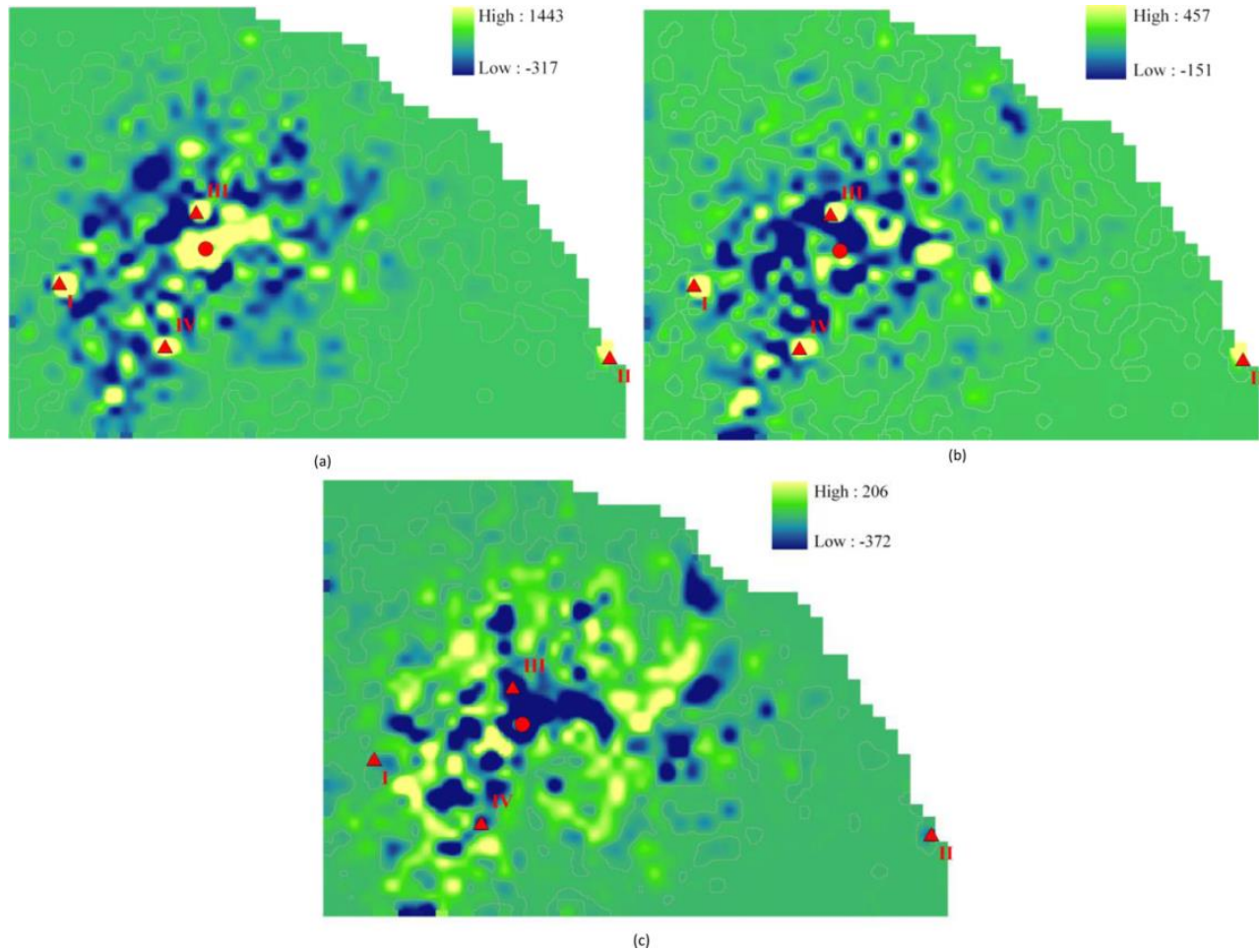


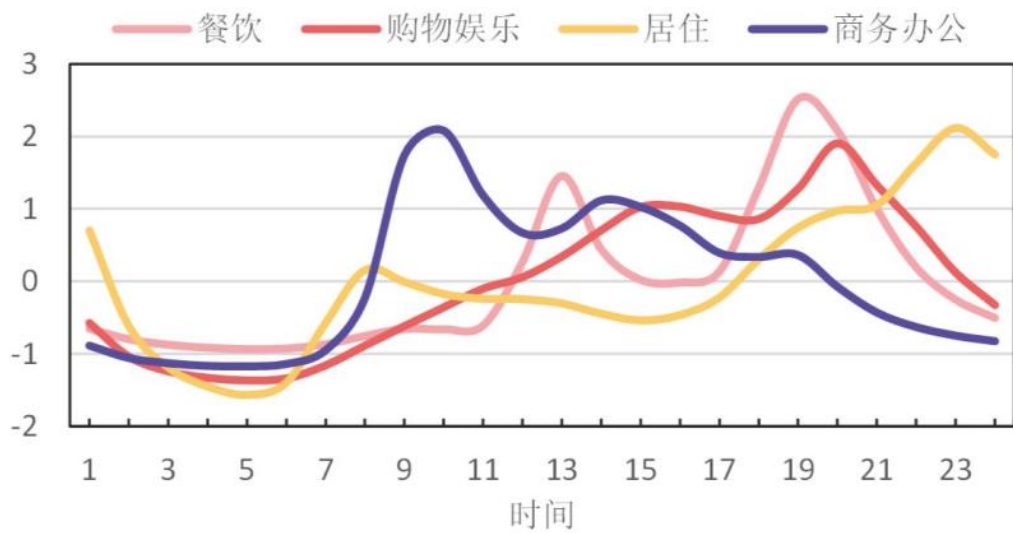
上海市出租车数据，20万+次出行/天

活动的周期性

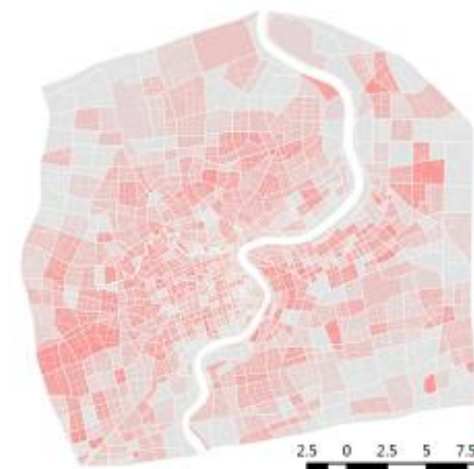








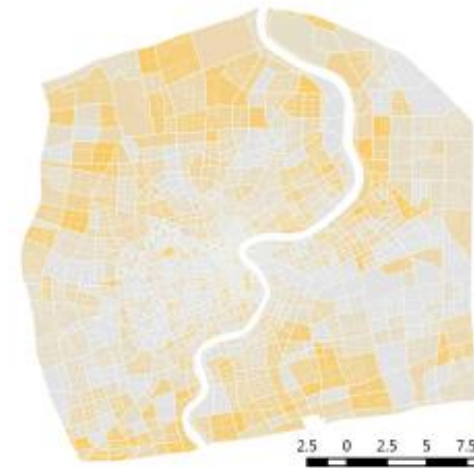
量化城市用地混合度



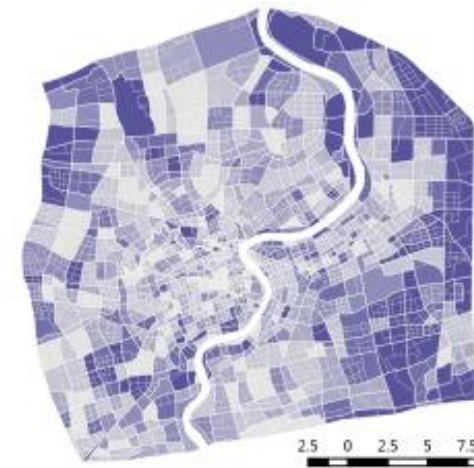
(a) 餐饮基活动



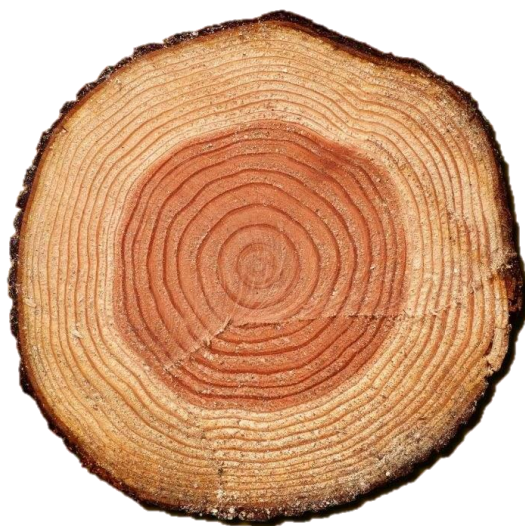
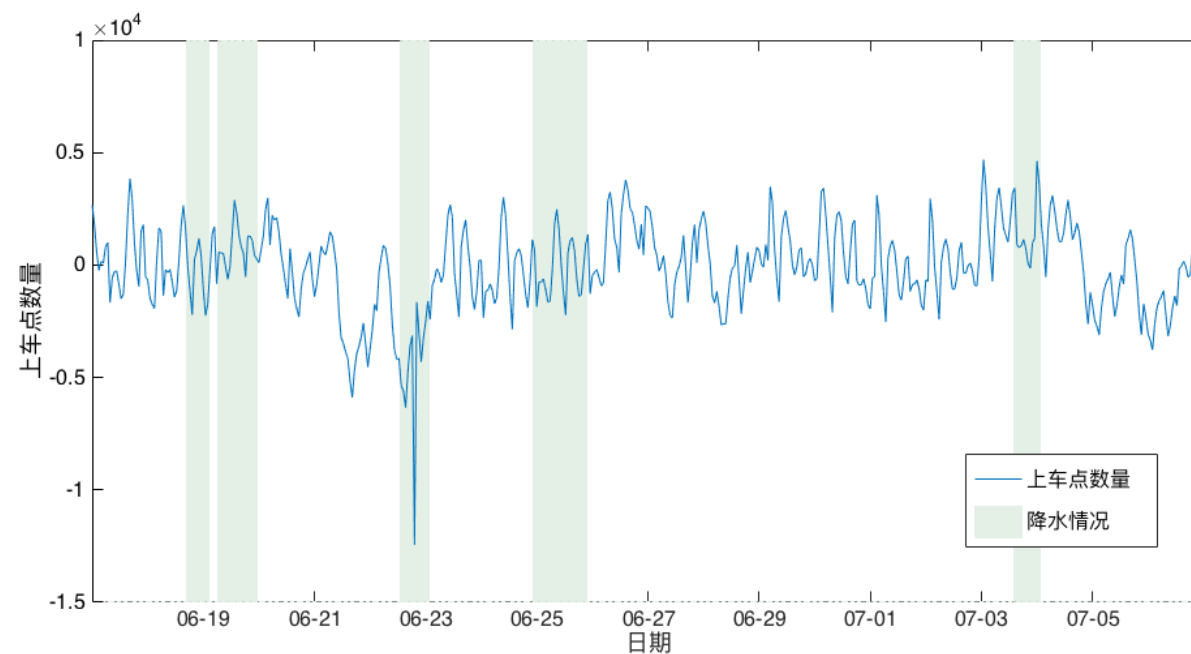
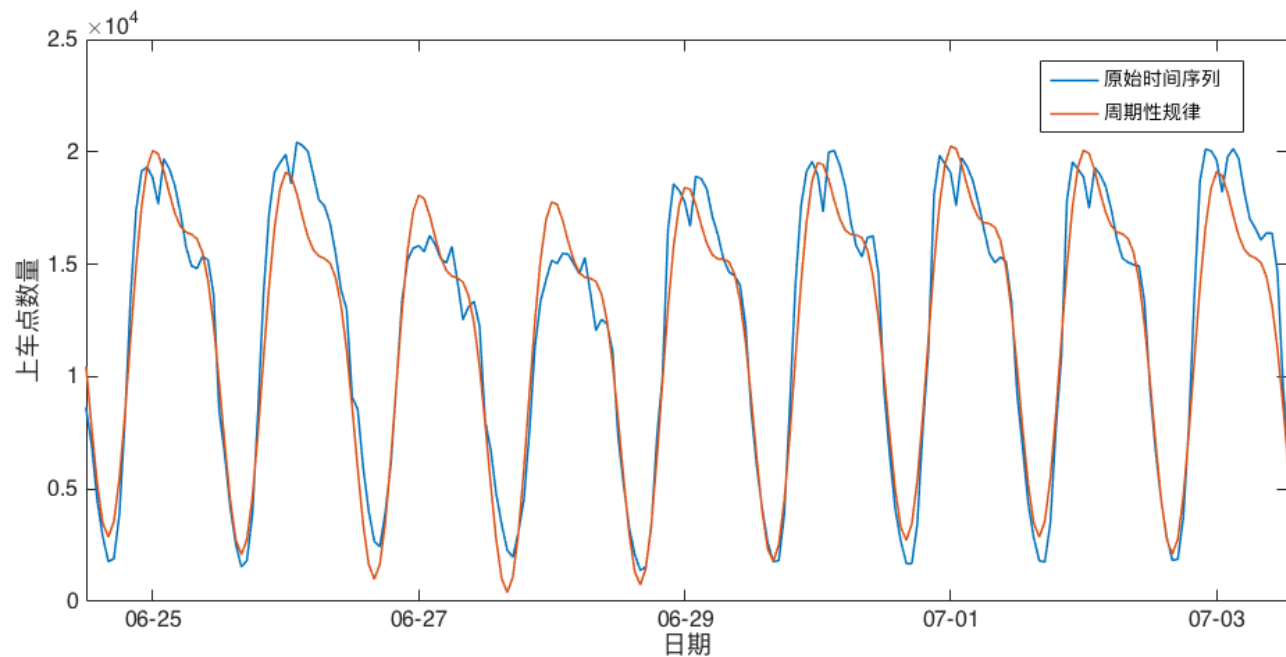
(b) 购物娱乐基活动



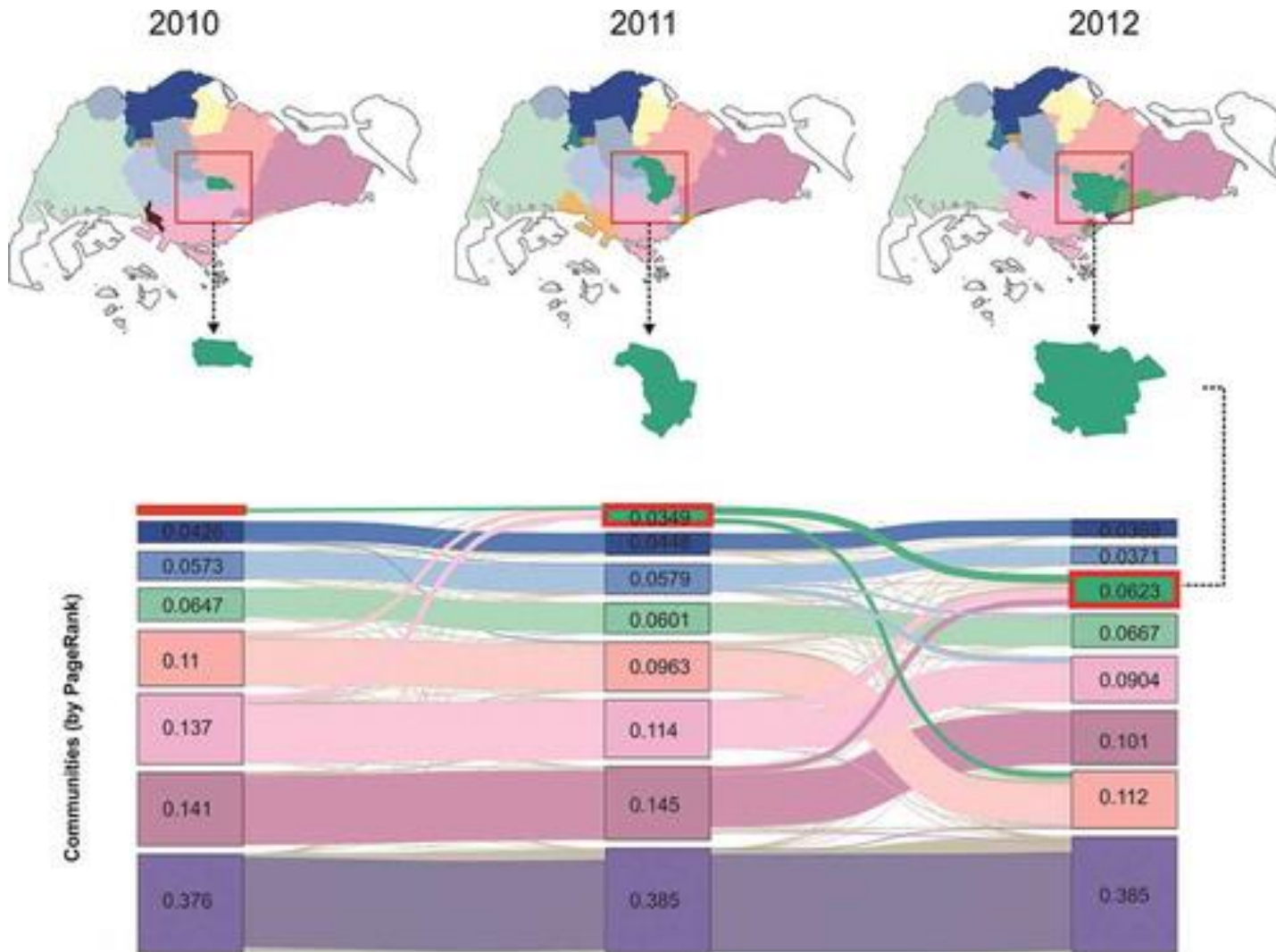
(c) 居住基活动



(d) 商务办公基活动



活动量变化曲线
如同城市的“年
轮”

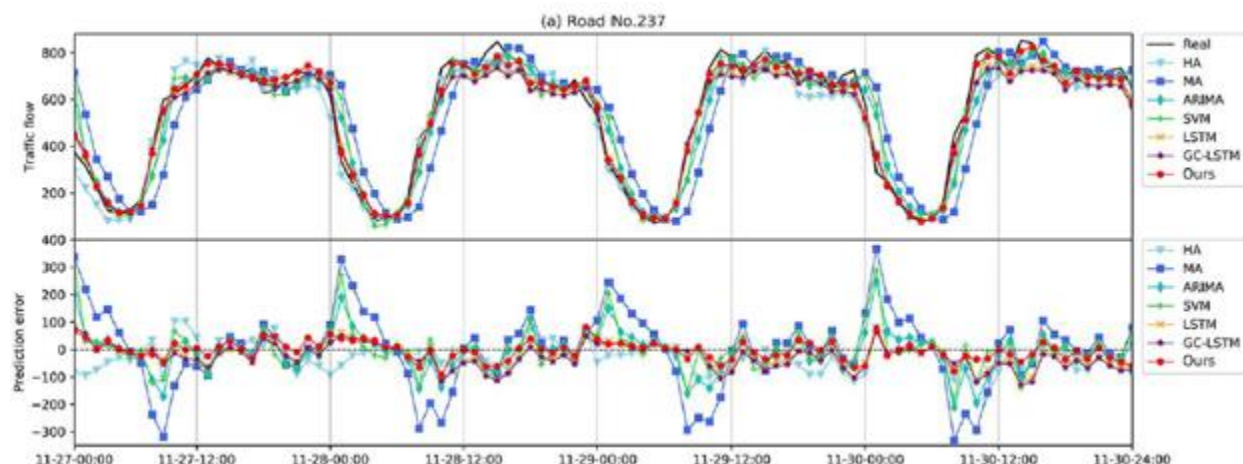
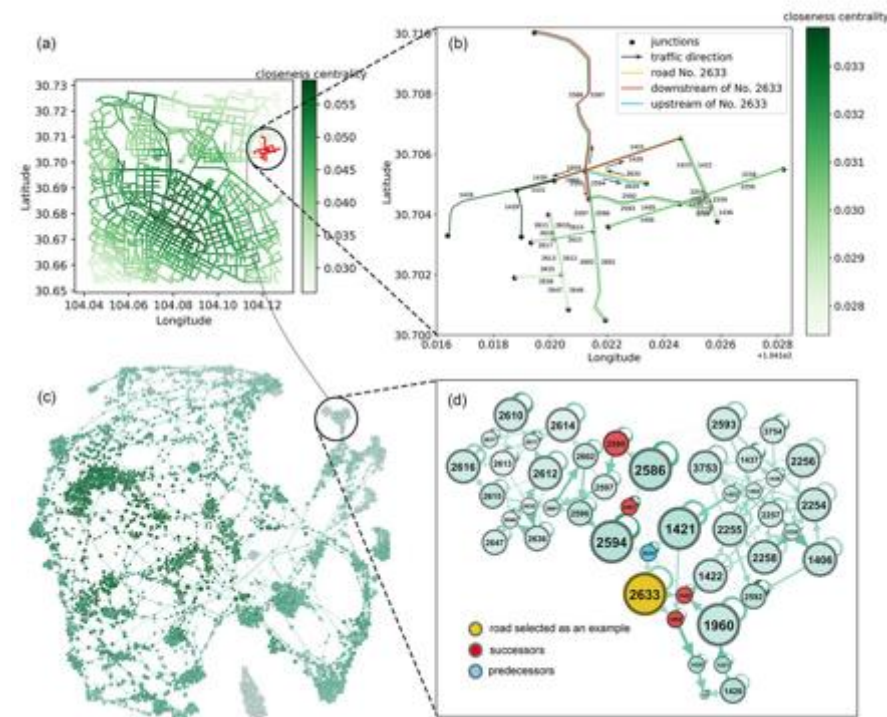
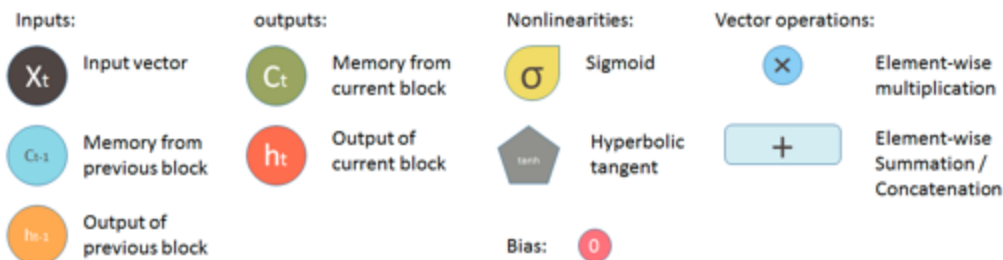
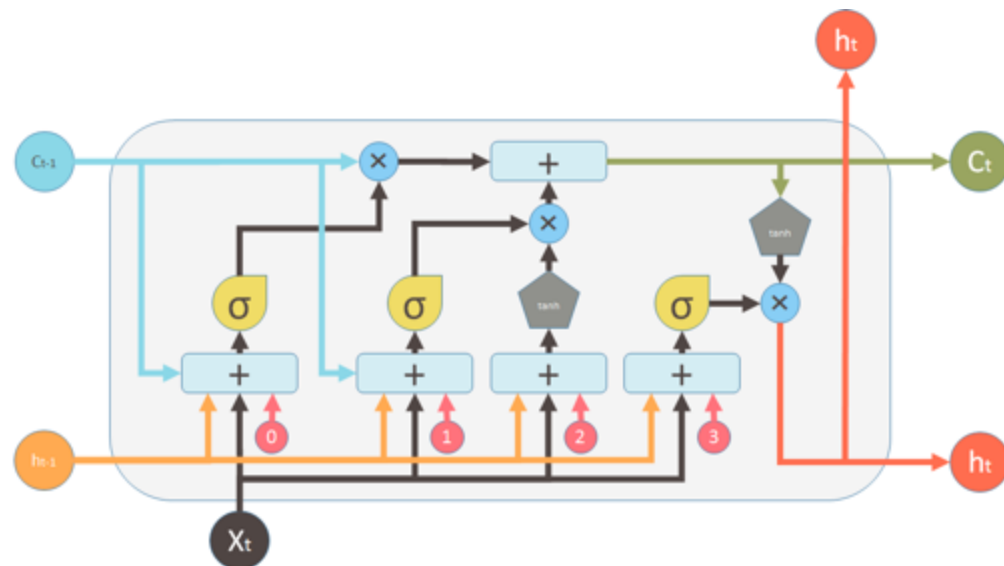


Chen Zhong, Stefan Müller Arisona, Xianfeng Huang, Michael Batty & Gerhard Schmitt (2014) Detecting the dynamics of urban structure through spatial network analysis, *International Journal of Geographical Information Science*, 28:11, 2178–2199



长短周期记忆与流量预测

- 长短周期记忆模型+GCN图卷积
- HA、MA、ARIMA等统计方法

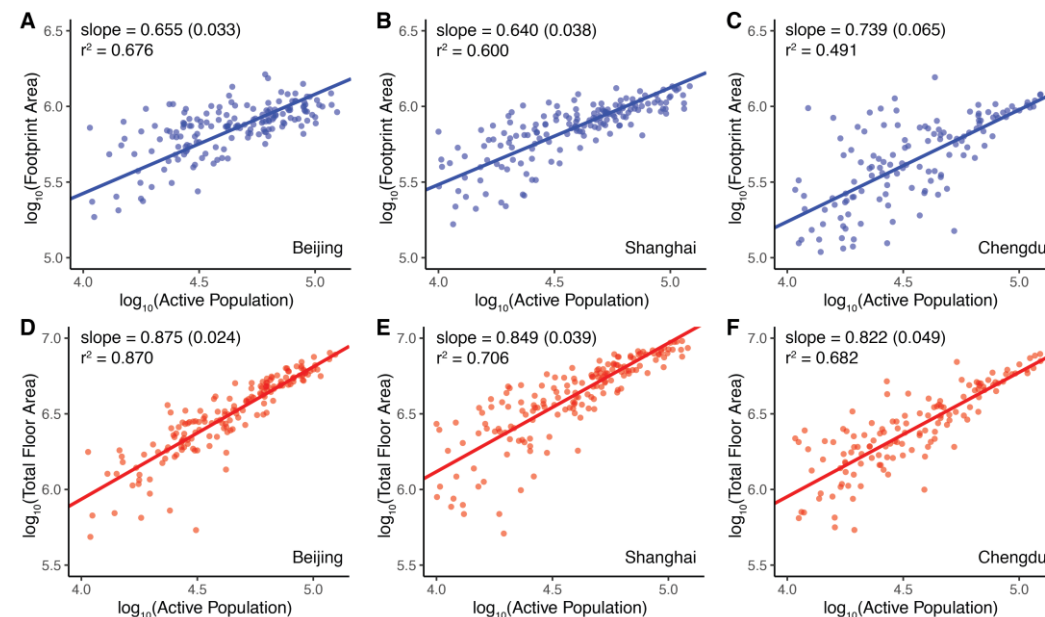
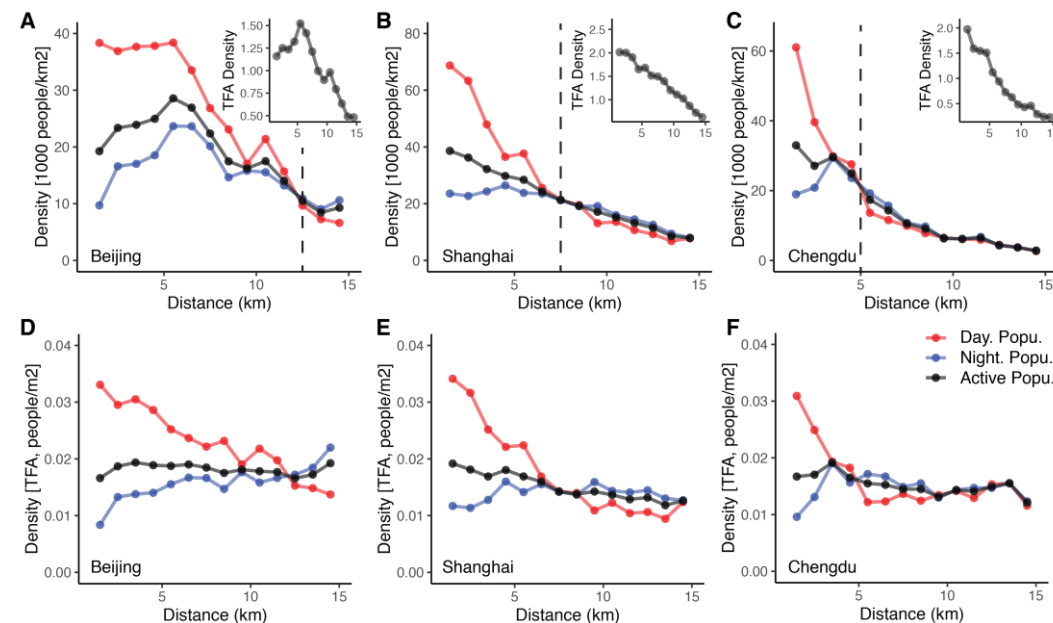
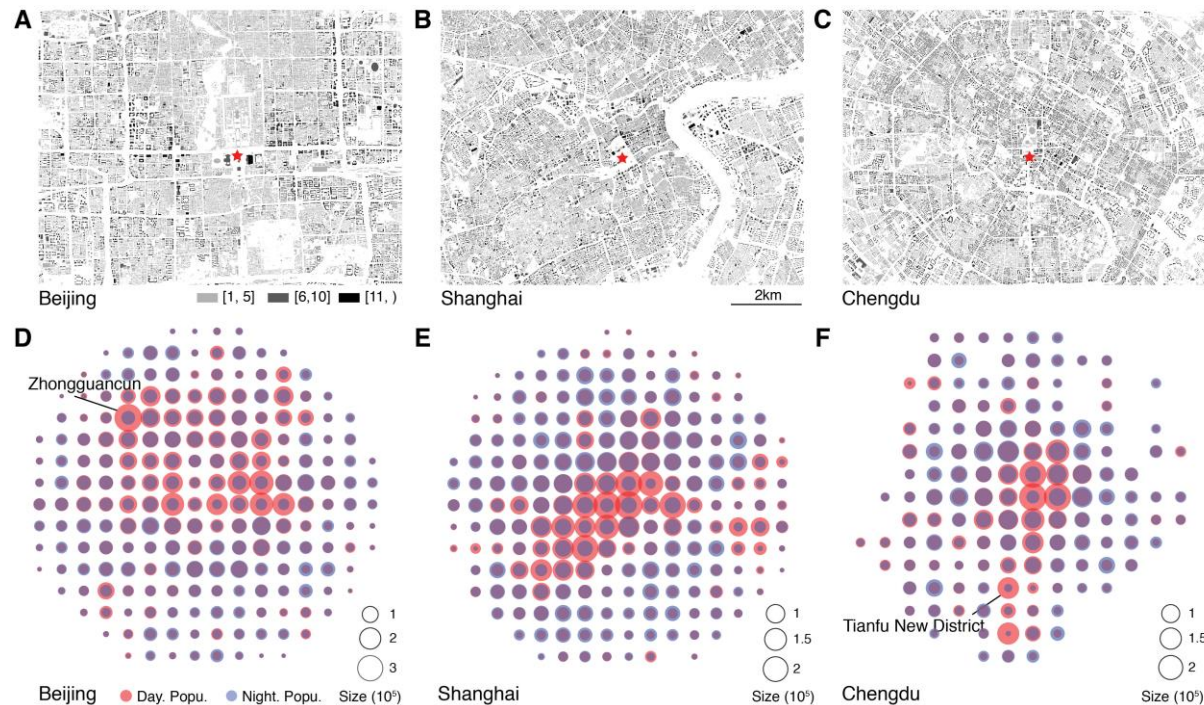




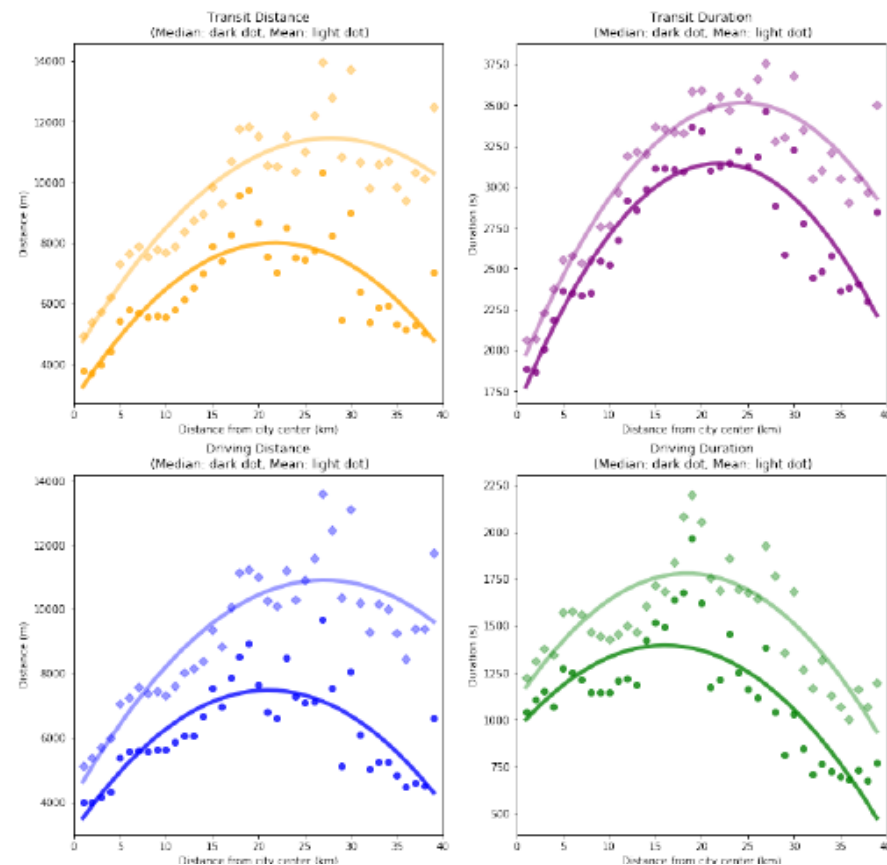
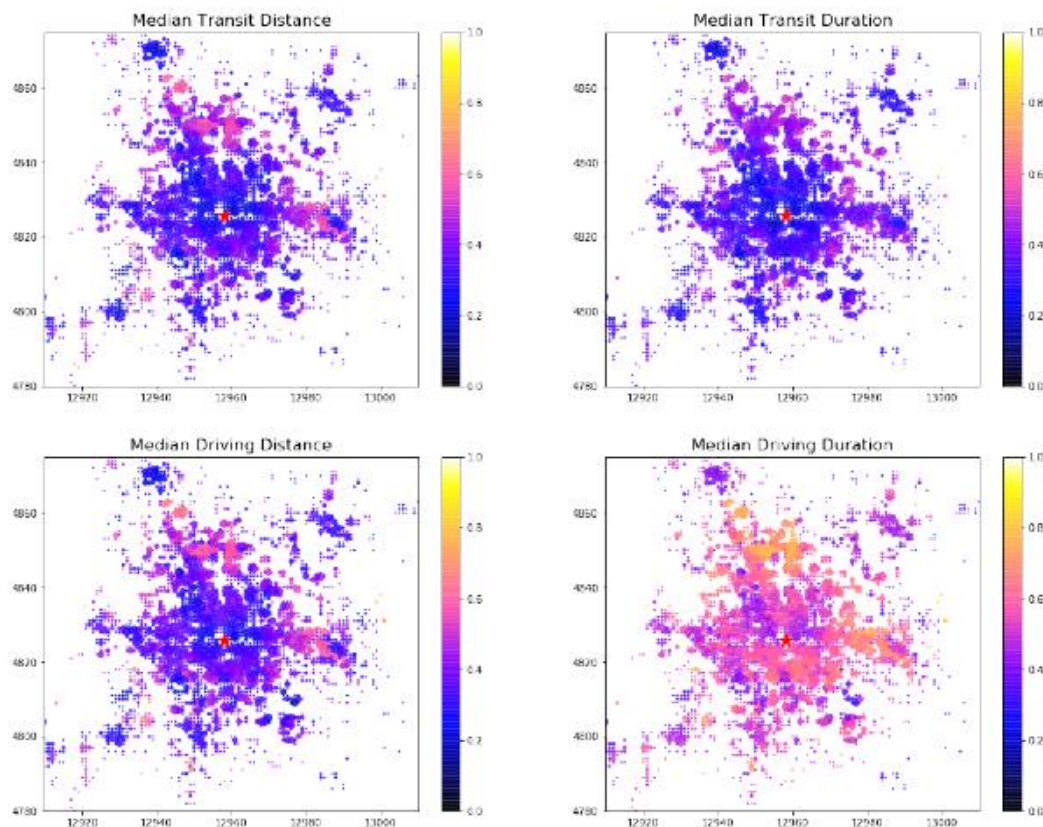
目录

CONTENTS

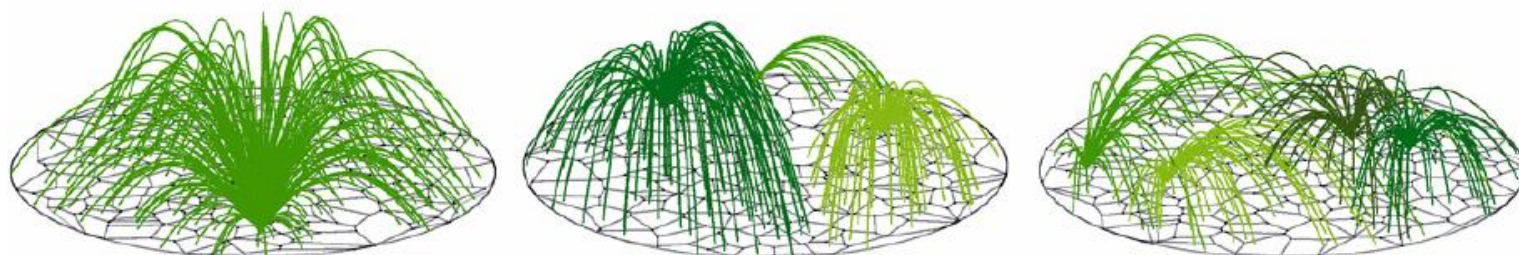
- 1/ 大数据与社会感知
- 2/ 社会感知方法
- 3/ 社会感知应用
- 4/ 总结



人口分布场

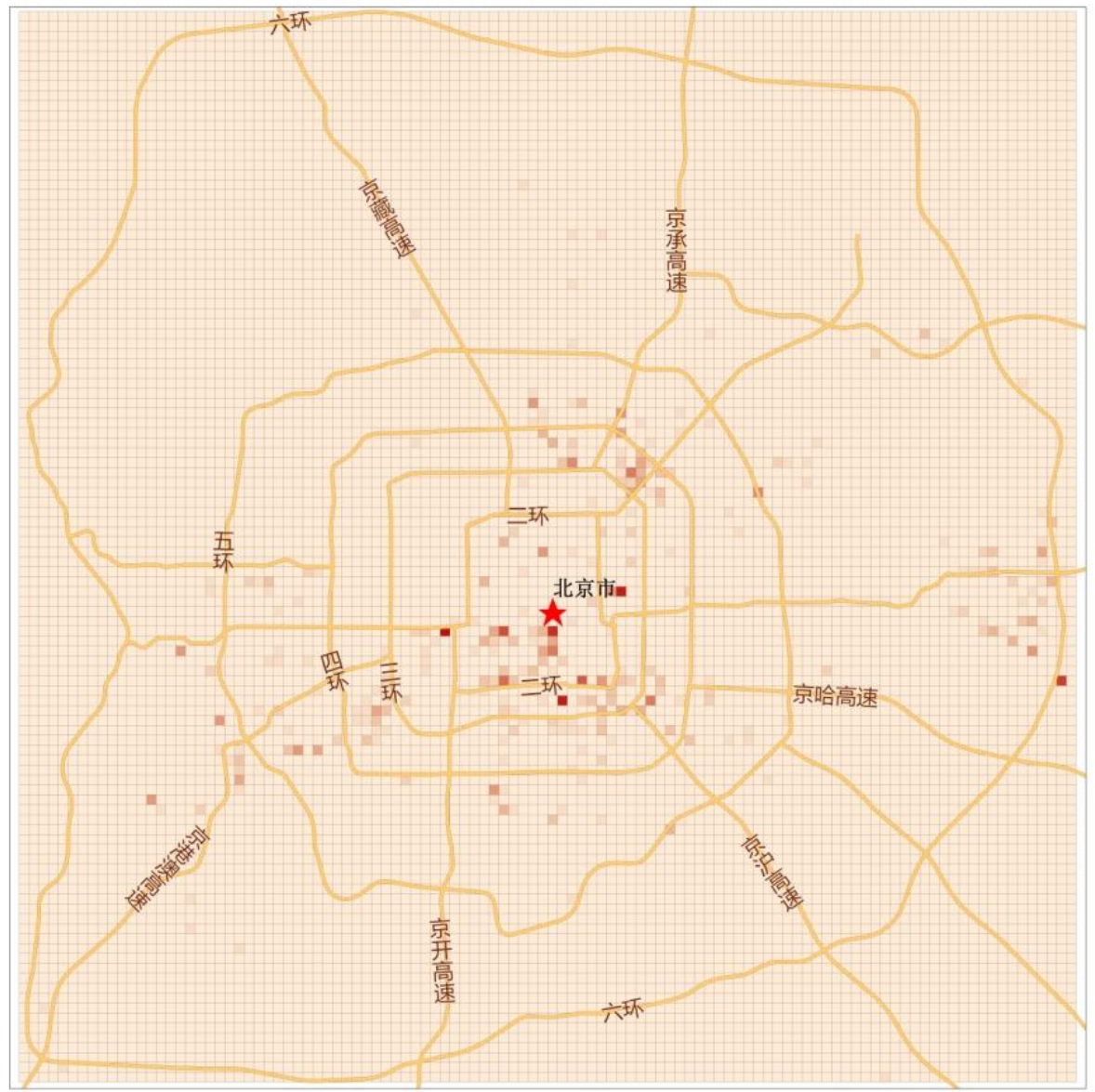


城市的通勤流场

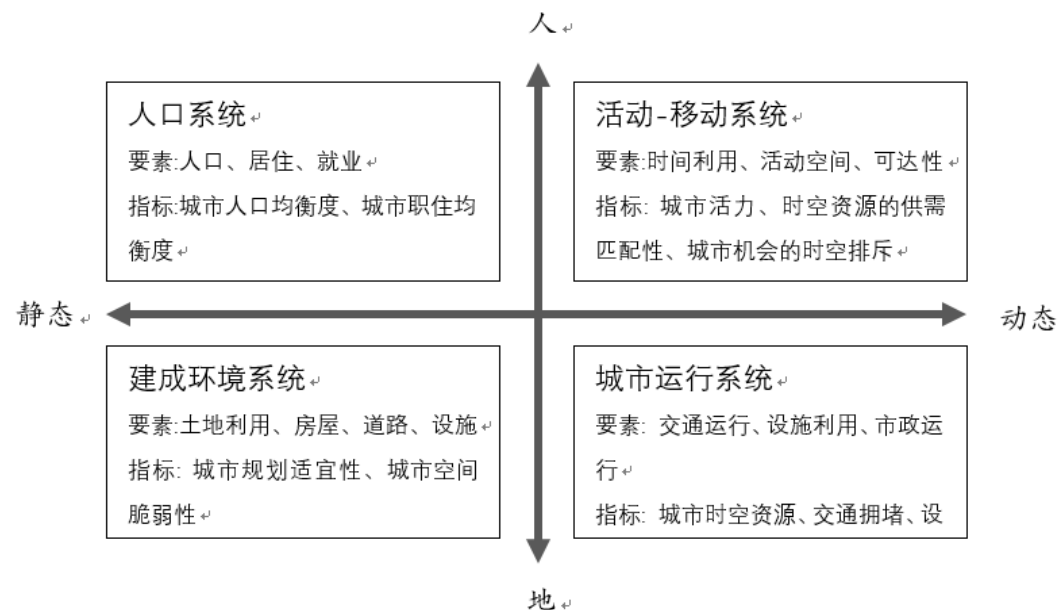




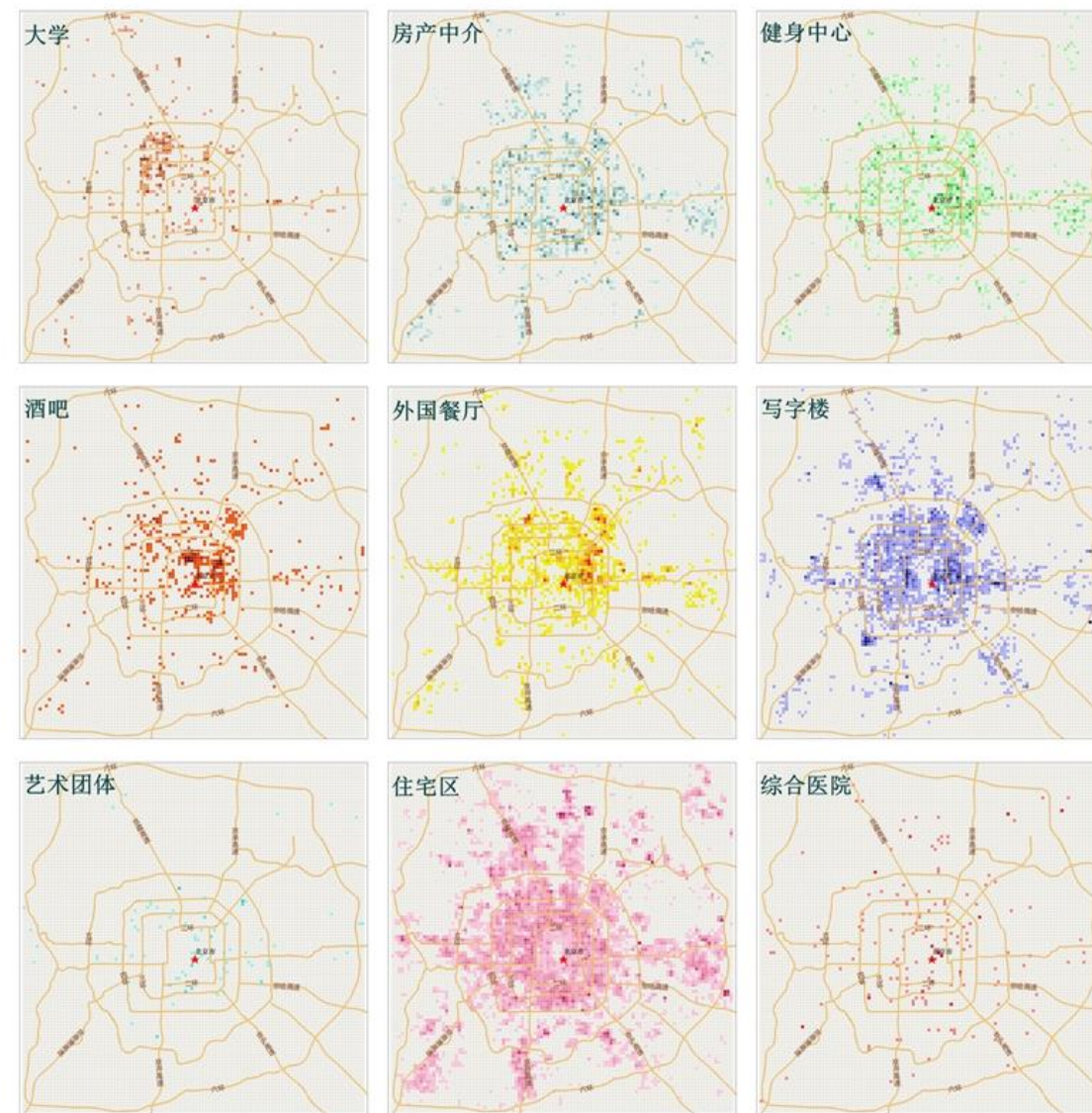
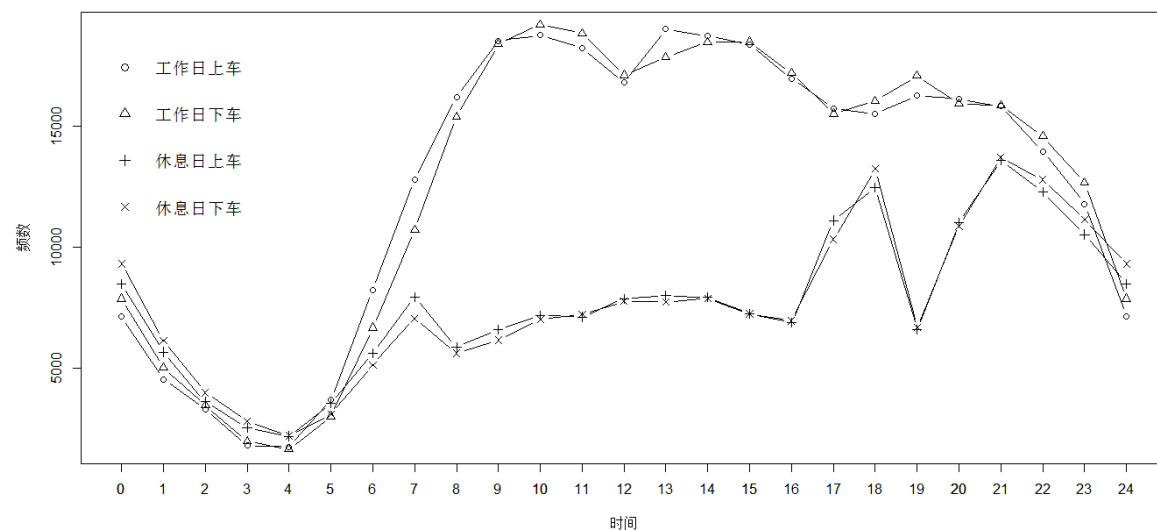
案件类型	部门名称	执法日期	经度	纬度
无照经营	西城区城管执法局	2016-01-26	116.370	39.892
非法营运	丰台区城管执法局	2016-01-06	116.393	39.861
违法广告	石景山区城管执法局	2016-01-24	116.232	39.915
擅自摆摊设点	东城区城管执法局	2016-01-25	116.419	39.875
擅自设置地锁	丰台区城管执法局	2016-01-14	116.368	39.871
未按规定清运垃圾渣土	朝阳区城管执法局	2016-01-18	116.568	39.916

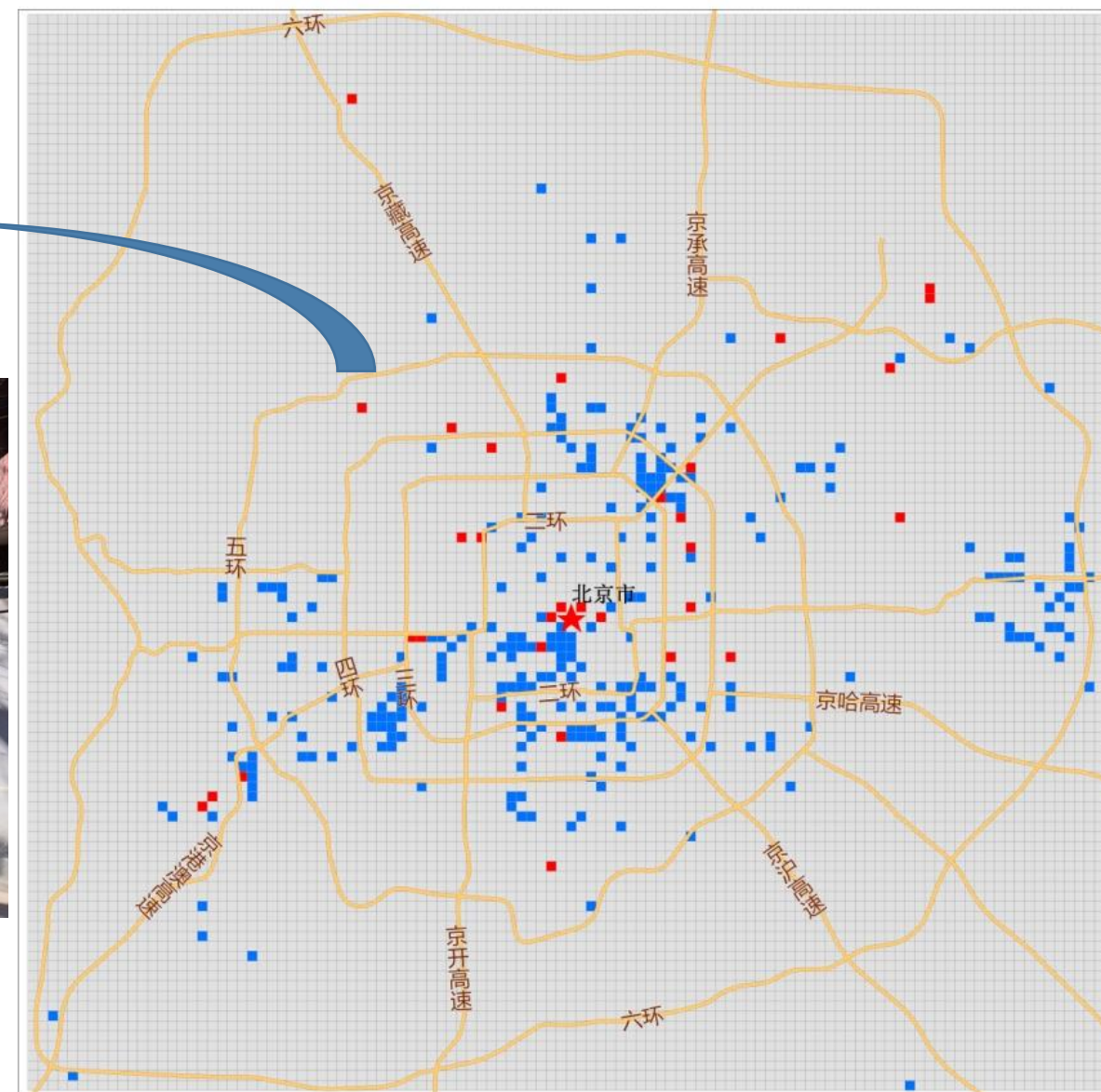
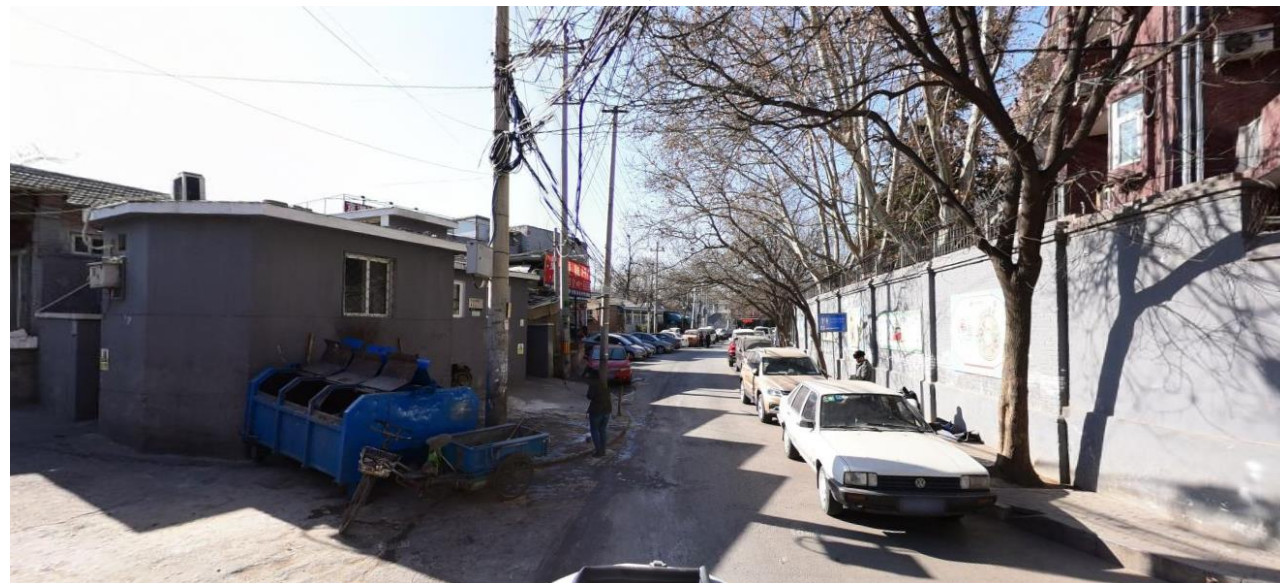


城管执法事件空间分析(2)



上下车点均值



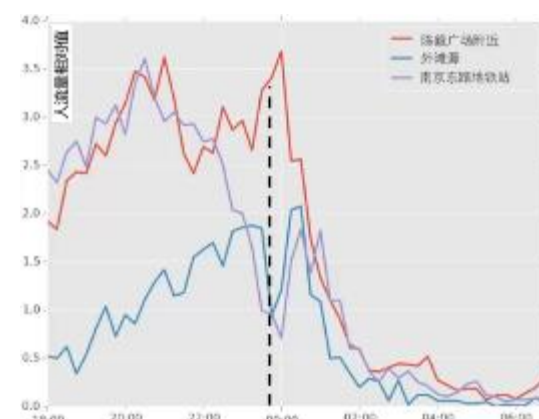
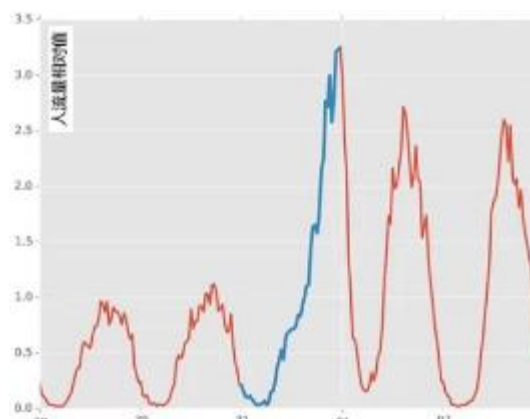




- 2014. 12 • 31上海外滩踩踏事件
 - 23时35分许，跨年夜活动
 - 外滩陈毅广场进入和退出的人流对冲
 - 发生踩踏事件
- 大数据分析（百度研究院大数据实验室）
 - 当时的人流量大到什么程度？事发当时是否是当晚人流量最大的时候？
 - 当时人流的对冲到底是什么样的程度？
 - 群体聚集是突发情况，可以预警吗？



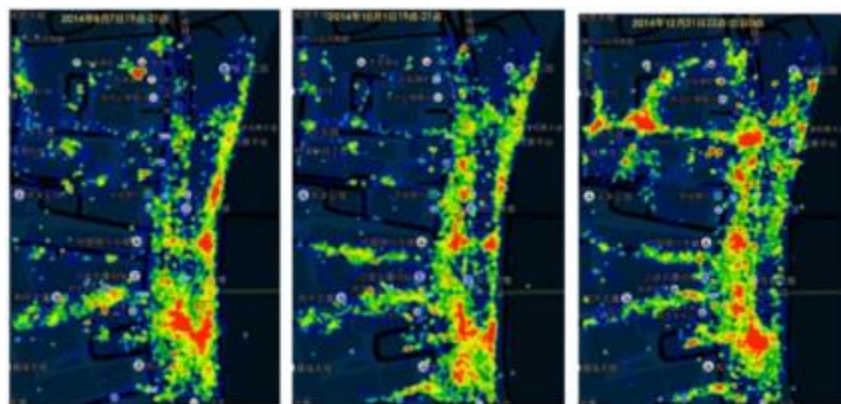
人群热力图



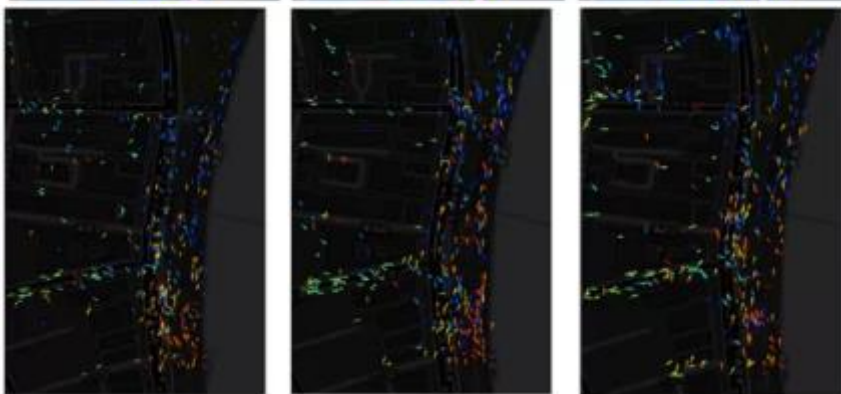
人群流量趋势图



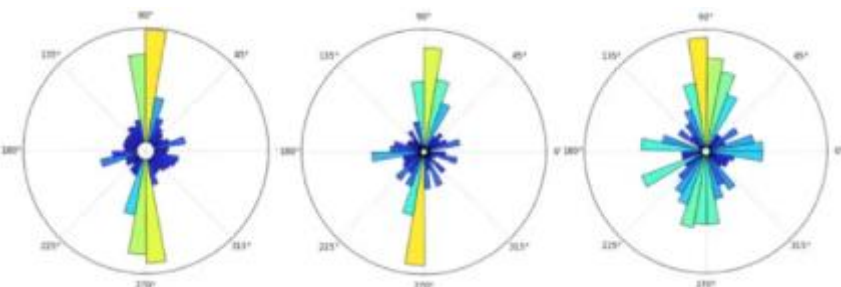
人群分布热力图



人群流动方向图



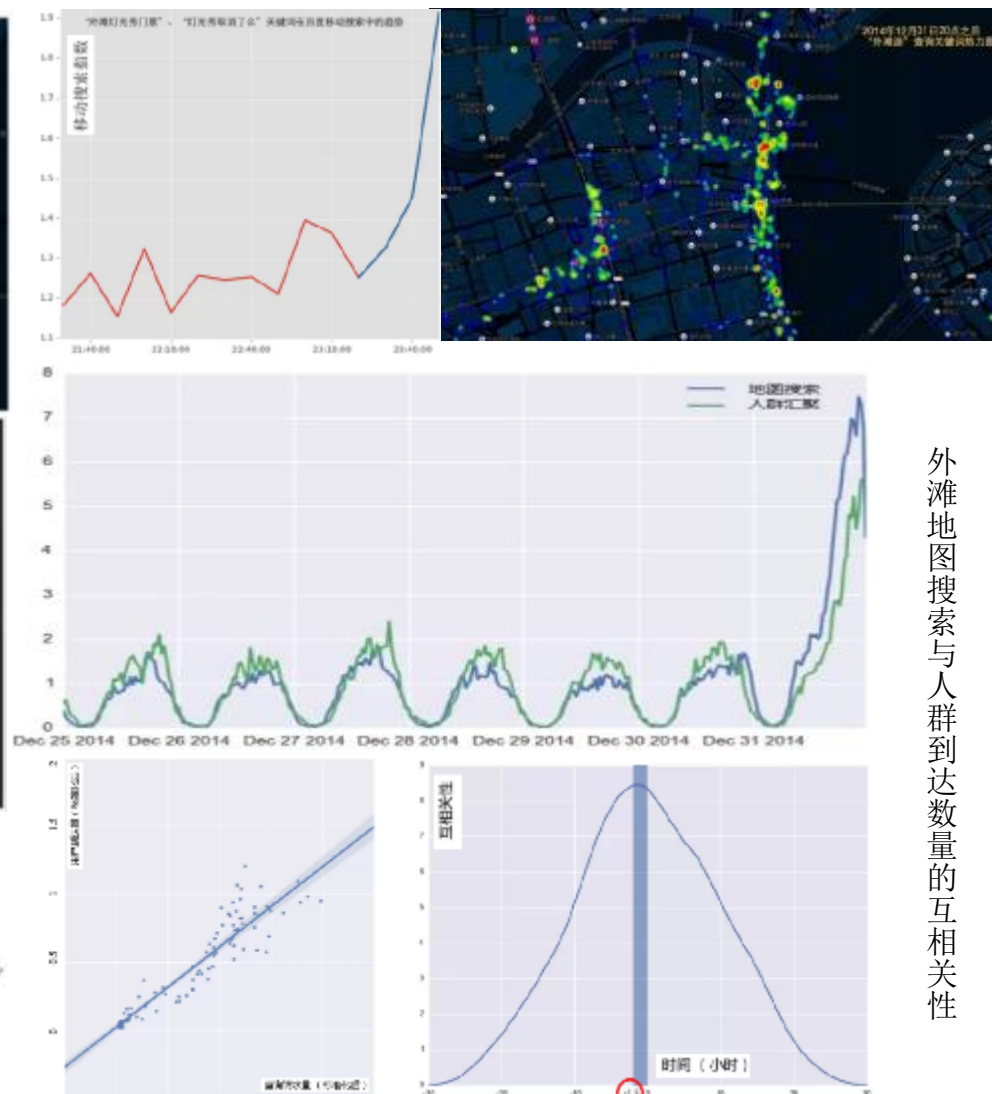
人群流动方向分布



(1) 中秋前夜

(2) 国庆当晚

(3) 跨年当晚



外滩地图搜索与人群到达数量的互相关性

互相关性曲线在-1.5小时的时候达到峰值：根据地图上相关地点搜索的请求量，至少可能提前几十分钟预测出人流峰值的到来



2015年11月6日

参 政 借 息

科技前沿 · 7 ·

手机定位数据揭秘中国“鬼城”

【美国《华盛顿邮报》网站11月4日报道】题：智能手机如何揭开中国最大的一个谜

关于中国“鬼城”的大多数研究都带有道听途说的色彩。就连最严谨的研究方法也有许多值得改进的地方，比如投资调研公司将刚入职的员工派到偏远地区，让他们去数晚上有多少楼房会亮起灯。

如今，中国最大的科技企业之一百度公司采取了一种新的研究方法，让人们首次对中国的“鬼城”有了系统性认识。

百度大数据实验室和北京大学的科研人员利用由手机和全球定位系统接收装置收集的定位数据，在历时6个月的时间里，实时并在全国范围内追踪人们在有可能是“鬼城”的地方进出的情况。

为了确定哪些地方有可能是“鬼城”，研究人员创造了一种计算方法，能够确定人口相对稀少的区域。他们定义的“鬼城”是，每平方公里的人口少于5000人的城市区域。这样的人口密度差不多是中国住房和城乡建设部所推荐的人口密度的一半。

研究人员追踪了50个存在大范围空置住宅区的城市。（为了不影响房地产价格，研究人员没有披露所有50个城市的名称。）不过，这些城市大多是中国东部的中型城市，空置住宅区主要位于城市的外围地带或新开发的区域。

研究人员遇到了一个问题：或许这些城市有的并非“鬼城”，而是旅游城市。就像全球其他地方一样，旅游目的地城市在一年当中的某些时候基本上处于空置状态，但在其他一些时候又会变得满满的。

研究报告用两个城市的情况对比了这种差别：位于中国东部沿海的山东乳山和位于内蒙古自治区鄂尔多斯市郊的康巴什。为了排除旅游目的地城市，研究人员追踪了一些大型空置住宅区在不同时间段的人口密度变化。研究人员说，康巴什的人口在工作日增多，周末和假期则减少，表明人们是在这里工

作。

科研人员说，这样的活动模式表明康巴什符合居住者稀少的“鬼城”特点，而乳山则更应被称为旅游城市。他们说，在这50个城市中，最终看起来像是旅游目的地的有26个，像“鬼城”的有24个。

参与研究工作的吴海山说，这些数据有助于地方政府未来在规划方面作出更明智的决策，同时了解应该改善哪些地区的基础设施，从而有助于让“鬼城”变成能够发挥城市功能的区域。吴海山以郑州新区为例，之前这里曾被认为是“鬼城”，但现在越来越多的人正在被吸引到这个地区。

这些数据在供中国城市规划者借鉴的同时，也为有眼光的房地产投资者提供了参考。据科研人员说，要成为正常运作的社区，城市需要起来提供就业、工业、娱乐、医疗和学校。单靠一片无人居住的房屋，远不能称之为城市。

宇宙微波背景现神秘亮点 科学家称或存在平行宇宙

【合众社美国加利福尼亚州帕萨迪纳11月4日电】美国加州理工学院宇宙哲学学家加拉姆·查理认为，他或许发现了平行宇宙的存在证据。

查理在美国《天体物理学杂志》季刊上发表

论认为，持续膨胀的宇宙产生一团团膨胀更为迅速的能量，它们形成自己的小宇宙。在有些科学家看来，宇宙膨胀概念需要多重宇宙貌似合理。

麻省理工学院研究员艾伦·古思说：“我要

也许只是一个不断膨胀的超大区域的一部分。”

多重宇宙理



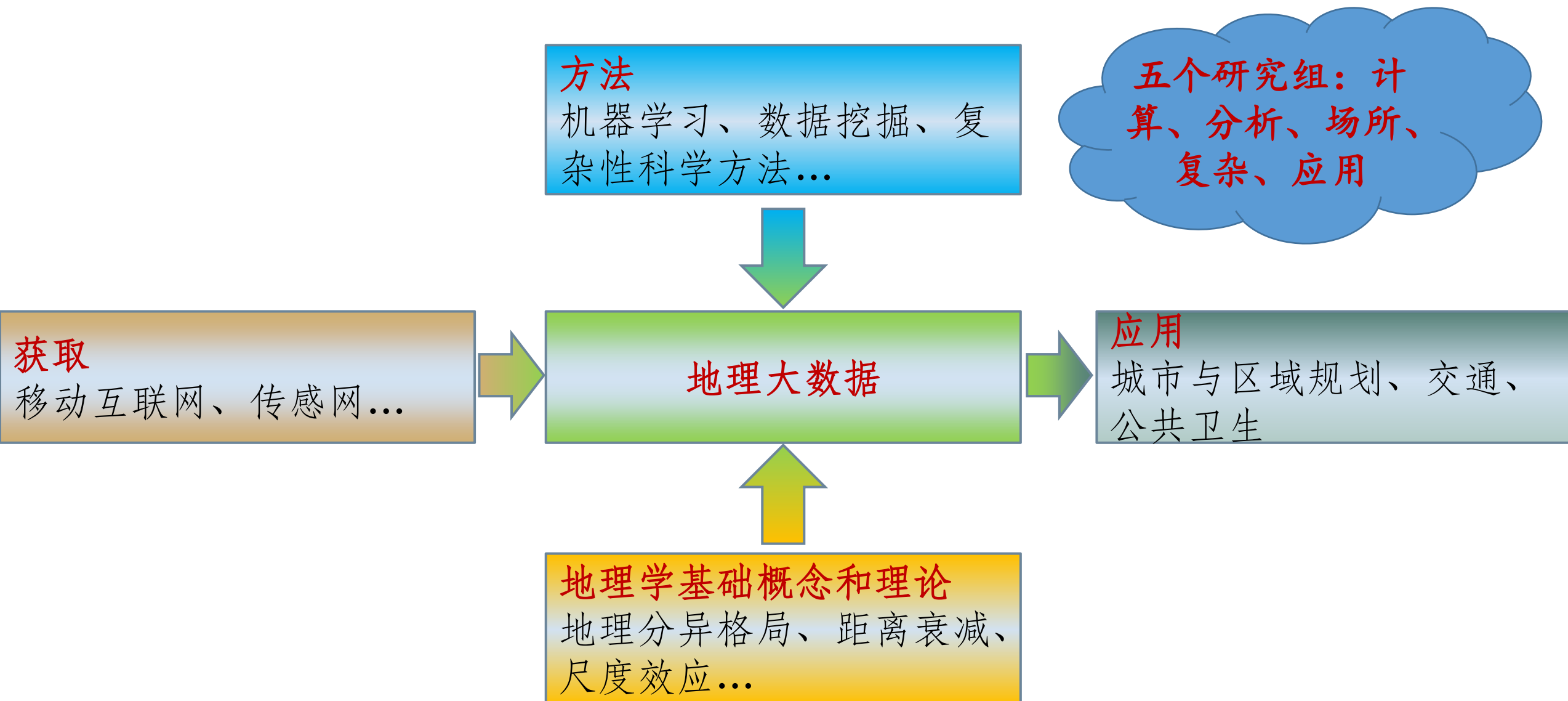
目录

CONTENTS

- 1/ 大数据与社会感知
- 2/ 社会感知方法
- 3/ 社会感知应用
- 4/ 总结



- 社会感知：基于大数据研究人的时空间行为模式，进而揭示其背后的社会经济特征，有助于我们从一个全新角度审视地理学的基本问题
- 社会感知：由“人”及“地”，场所→交互→过程
- GIS的使命：需要与应用领域研究人员紧密结合，提供更多的时空分析方法和更强的计算能力





- 地学计算（针对地理大数据）
 - 实时流计算
 - 并行计算
 - 可视化
- 地理分析（基于地理大数据）
 - 空间模式提取
 - 重访可变面积单元、距离衰减等地理基本规律
 - 时间序列分析
- 建模场所语义（从地理大数据）
 - 表达场所地理范围
 - 增强场所语义
- 模拟复杂性（在地理大数据中）
 - 人类移动模型和复杂网络
 - 标度率、迸发
 - 基于Agent的模型
- 应用研究（利用地理大数据）
 - 旅游、犯罪、公共卫生



Yu Liu

Peking University

在 urban.pku.edu.cn 的电子邮件经过验证 - 首页

[GIScience](#) [Social Sensing](#)

关注

创建我的个人资料

标题

引用次数

年份

[Geographic information system: Principles, methods and applications](#)

L Wu, Y Liu, J Zhang, XJ Ma, ZY Wei, Y Tian
Beijing: Science Press

1829 *

2001

[Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data](#)

Y Liu, Z Sui, C Kang, Y Gao
PLoS ONE 9 (1), e86026

250

2014

[Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai](#)

Y Liu, F Wang, Y Xiao, S Gao
Landscape and Urban Planning 106, 73-87

218

2012

[Social Sensing: A New Approach to Understanding Our Socioeconomic Environments](#)

Y Liu, X Liu, S Gao, L Gong, C Kang, Y Zhi, G Chi, L Shi
Annals of the Association of American Geographers 105 (3), 512-530

213

2015

[Understanding intra-urban trip patterns from taxi trajectory data](#)

Y Liu, C Kang, S Gao, Y Xiao, Y Tian
Journal of Geographical Systems 14 (4), 464-483

205

2012

[Correlating mobile phone usage and travel behavior-A case study of Harbin, China](#)

Y Yuan, M Raubal, Y Liu
Computers, Environment and Urban Systems 36 (2), 118-130

181

2012

[Intra-urban human mobility patterns: An urban morphology perspective](#)

C Kang, X Ma, D Tong, Y Liu

150

2012

引用次数

[查看全部](#)

总计

2014 年至今

引用

6033

3468

h 指数

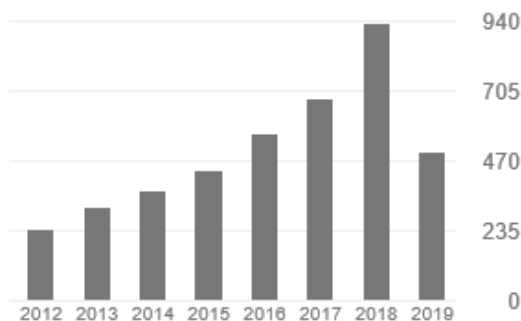
32

27

i10 指数

75

65



合著作者

[查看全部](#)



Chaogui Kang
Assistant Professor, School of R...



Song Gao
Assistant Professor of Geographi...



Yihong Yuan
Assistant Professor, Texas State...





多谢!
liuyu@urban.pku.edu.cn

未名·时空

Spatio-Temporal Social Sensing Lab
Peking University



感时空，知未名

关注我们: 未名时空
联系我们: geosoftpku@tom.com



版权声明:
原创文章版权归S³-Lab所有, 转载请注明出处。