

# 大数据与城市空间研究

同济大学建筑与城市规划学院  
大数据与城市空间分析实验室

王德

2016年3月

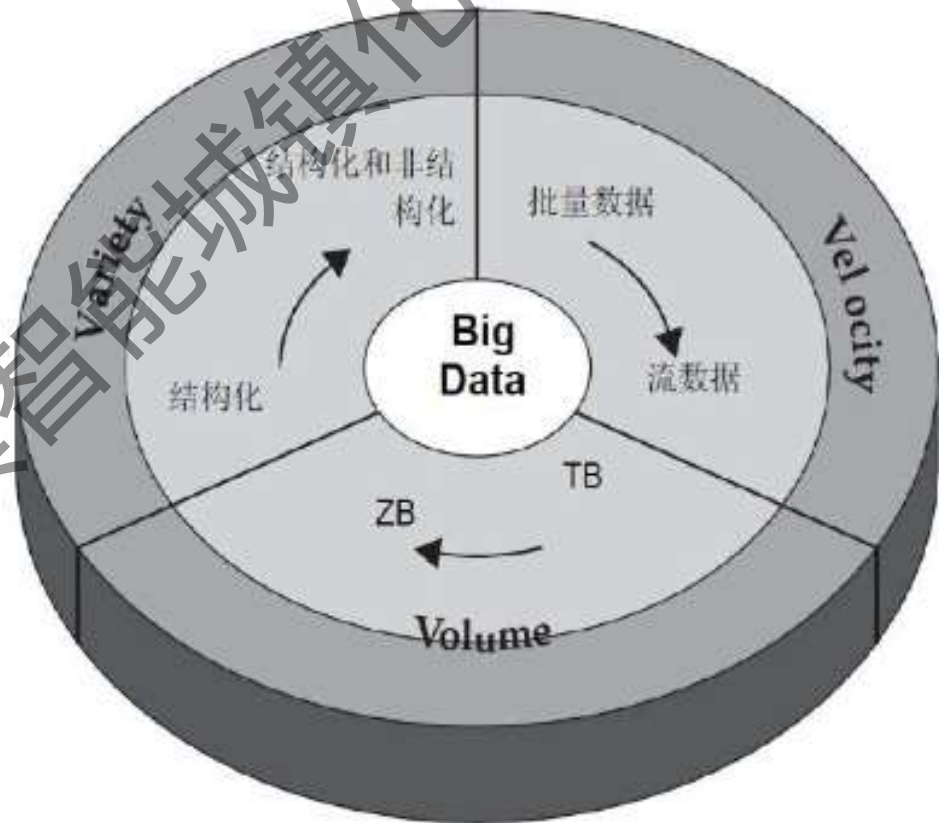
# 大数据

- 百度百科、维基百科：
- 所涉及的资料量规模巨大到无法通过目前主流软件工具，在合理时间内达到获取、管理、处理、并整理成为帮助企业经营决策更积极目的的资讯。
- 大数据是痕迹数据汇集的并行化、在线化、生活化、社会化（邱泽奇）
- 大数据为叫**BIG DATA**，不叫**large**，这不仅仅是强调数据量本身大，是思维方式的改变。

## 大数据的特征

IBM用3V特征来定义大数据：数量、种类和速度

在此基础上，IDC又增加了巨大的数据价值（value）（4V）



年份	数字信息比例		存储容量	备注
	数据信息比例	非数据化数值信息比例		
1960				出现“信息时代”、“数字村镇”的概念
2000		75%	300亿字节	报纸、胶片、黑胶唱片、盒式磁带等信息占3/4
2007	93%			只有7%的数据存储在报纸、数据、图片等媒介
2013		98%	1.2泽字节	

中文简称	英文简称	进率 (Byte=1)
比特	b	0.125
字节	B	1
千字节	KB	$2^{10}$
兆	MB	$2^{20}$
吉	GB	$2^{30}$
太	TB	$2^{40}$
拍	PB	$2^{50}$
艾	EB	$2^{60}$
泽	ZB	$2^{70}$

大数据——数据量增长

大数据——量变导致质变

大数据——海量数据计算预测事情发生的可能性

# 大数据时代的到来

大数据为居民生活带来了便捷，理念在智慧！

通过智能手机、社交网络、社区监控视频等设备对居民日常行为或活动数据进行收集、处理、分析及开发，能够深入了解居民关注热点，发现亟需解决的城市问题，智慧引导和服务居民出行、购物、娱乐、工作等，“智慧”理念已经成为大数据时代新的生活关键词。

Figure 2.12 Journey On website screen shot

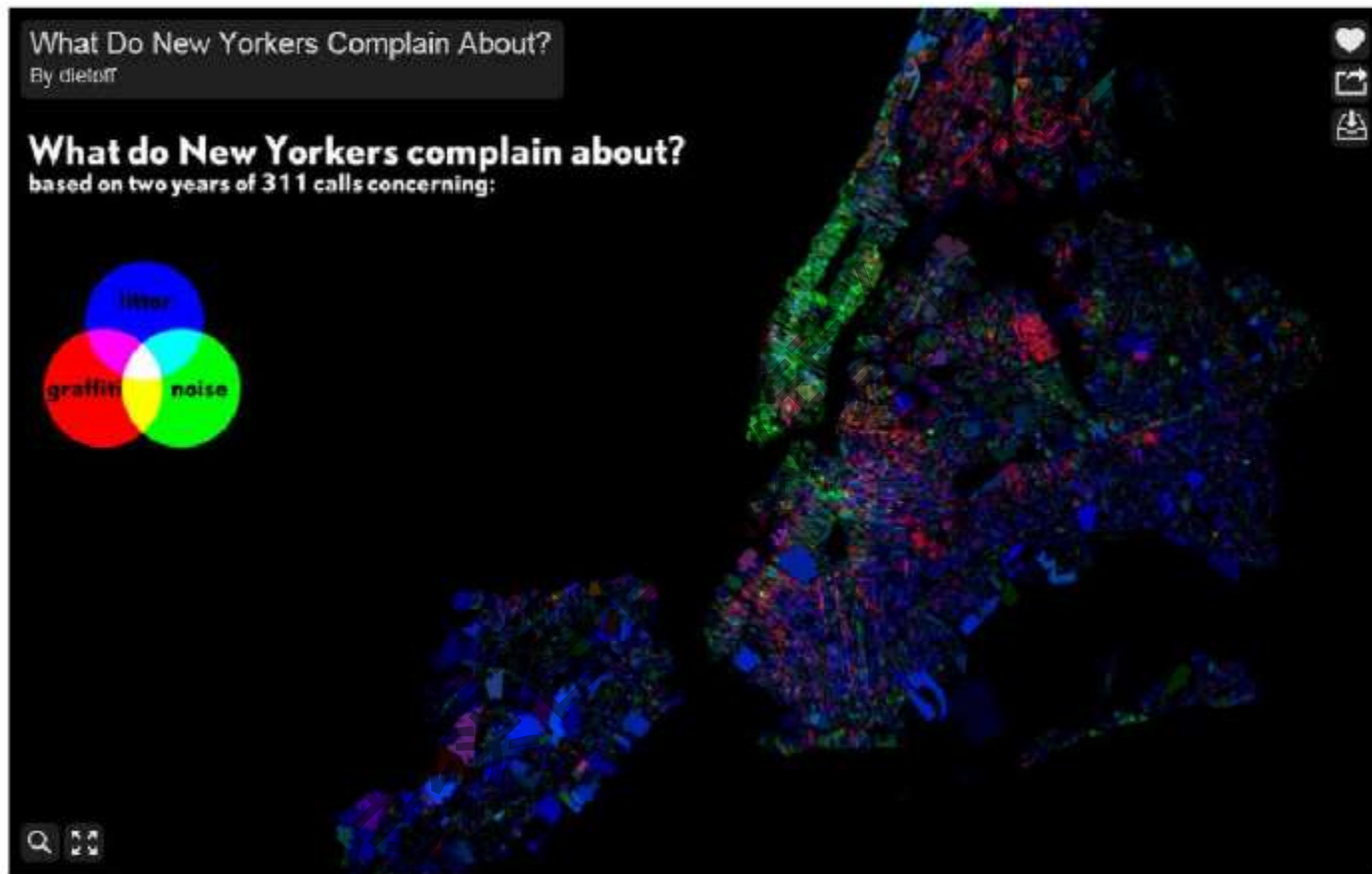


## 城市中的移动网络 - 日内瓦



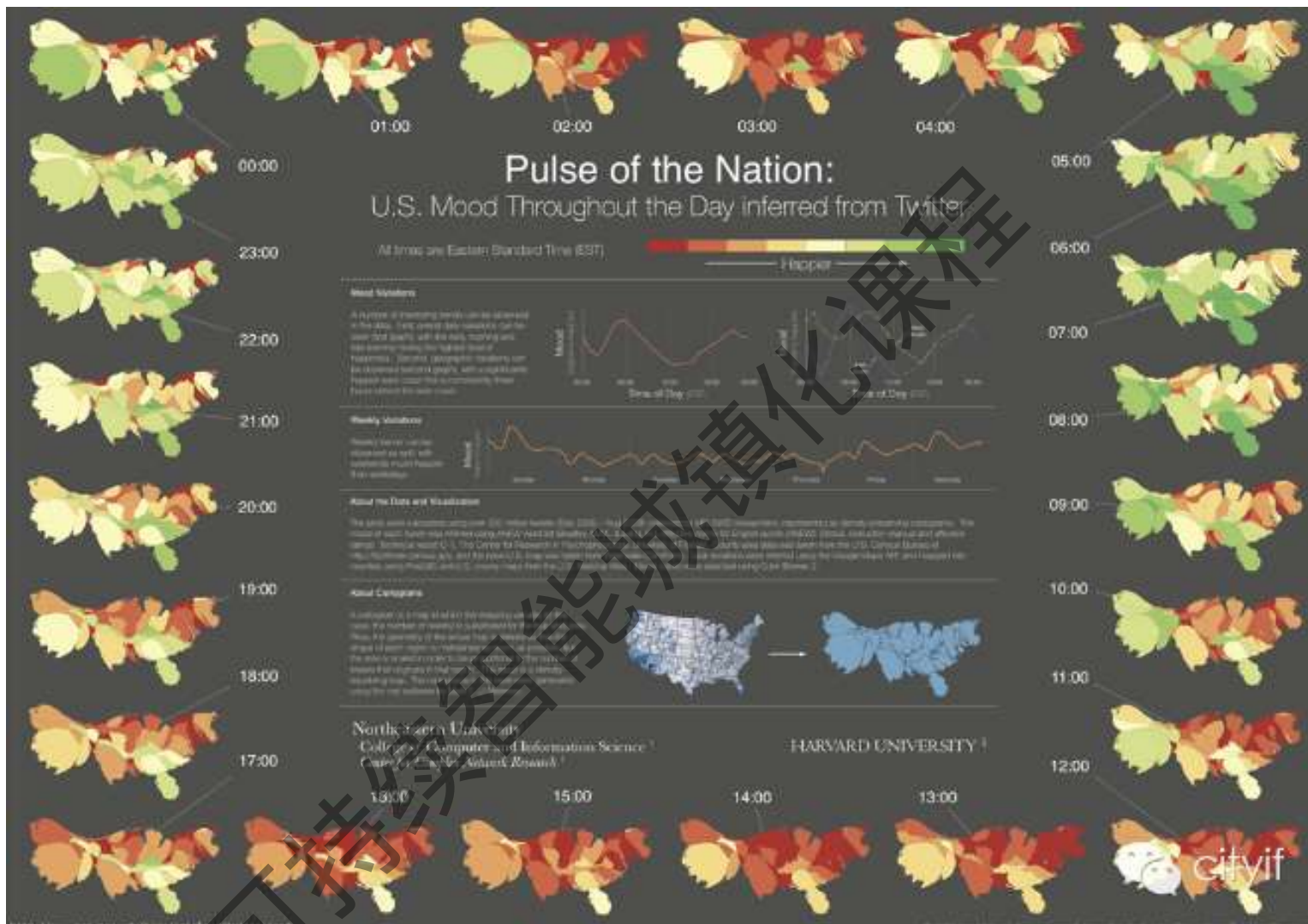
瑞士日内瓦与Lift conference合作，将Swisscom提供的一周内手机用户在通话中移动产生的地理数据进行视觉化，为我们描述了一座活动的城市。主要包括日/夜实时流量、各城区间的流量、各时段的通话流量、周流量、周末出城/入城流量等主题信息图。





纽约客抱怨地图：蓝色—抱怨乱扔垃圾；绿色—抱怨噪音；红色—抱怨街头涂鸦。曼哈顿抱怨噪音最多；布朗克思抱怨街头涂鸦最多；斯塔腾岛抱怨乱扔垃圾最多。（网站数据，特别是社交网站数据）

城市的心情  
2015-04-22  
放小浪、  
储小芽  
儿



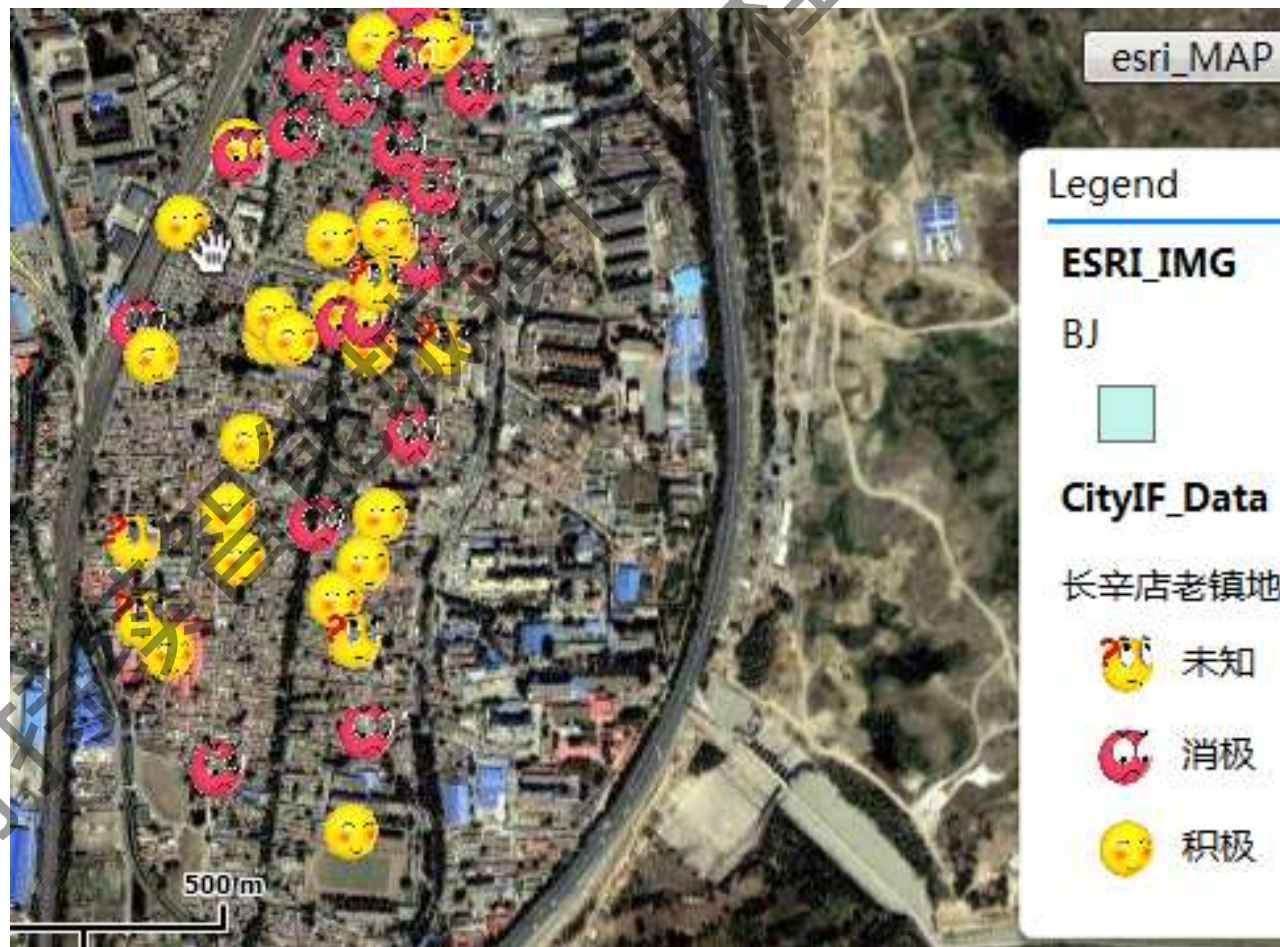
- 早在2010年，就有来自哈佛大学（Harvard University）和东北大学（Northeastern University）的研究人员对将近3亿条 tweets（推特博文）进行了关键字分析，制作出了下面这张反映美国人心情起伏波动的“晴雨表”：

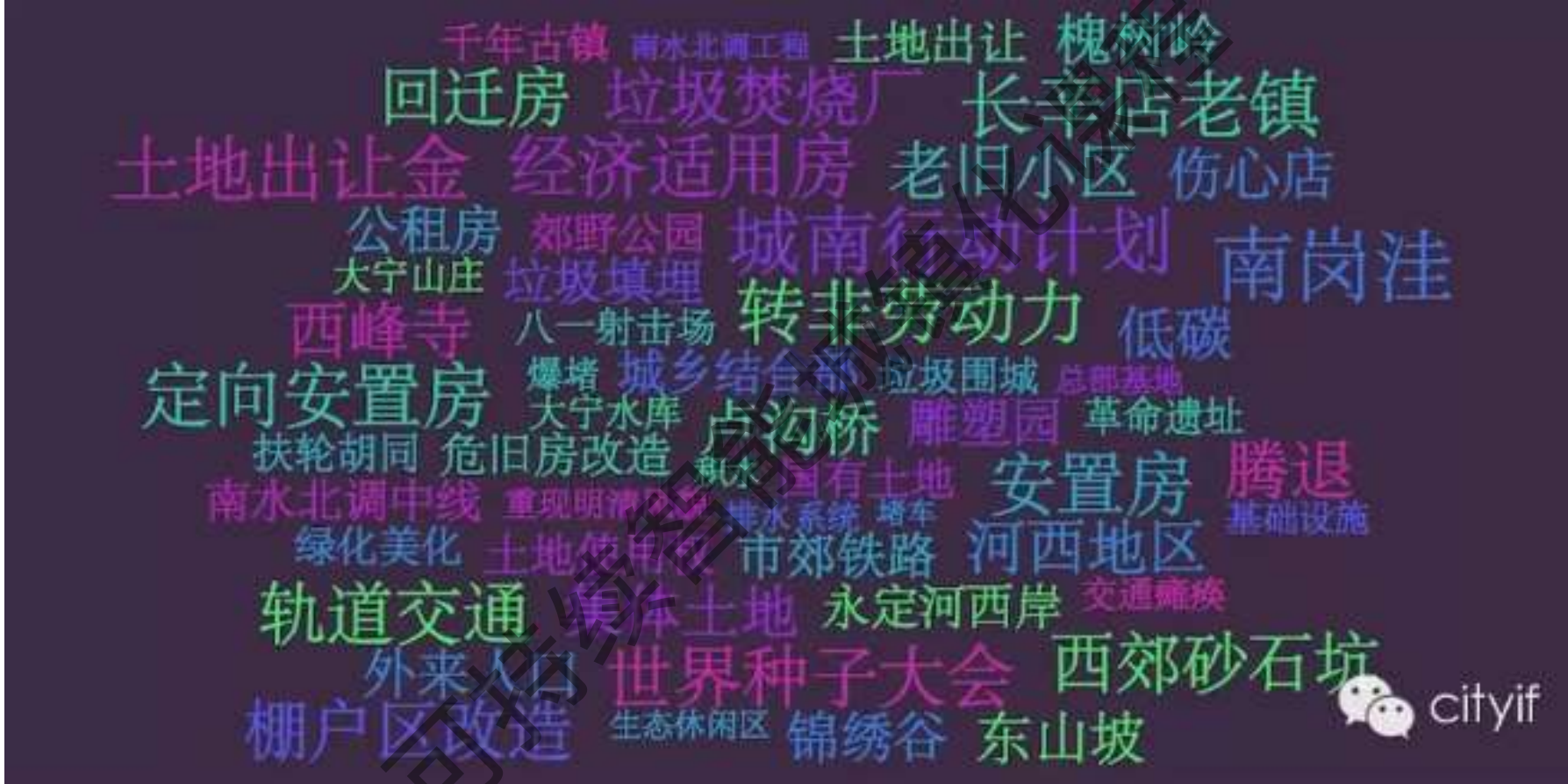




# 心情地图简介

- 以涉及长辛店地区的网络和新闻评论为对象，通过分析公众对地点的情绪，复参
- 及社交评论内容，通过情感倾向分析，获取对于该地点的情绪，复参
- 长辛店地区网络和新闻评论内容，通过情感倾向分析，获取对于该地点的情绪，复参
- 店网评论内容，通过情感倾向分析，获取对于该地点的情绪，复参





千年古镇 南水北调工程 土地出让 槐树岭

回迁房 垃圾焚烧厂 长辛店老镇

土地出让金 经济适用房 老旧小区 伤心店

公租房 郊野公园 城南行动计划 南岗洼

大宁山庄 垃圾填埋 转非劳动力 低碳

西峰寺 八一射击场 爆堵 城乡结合部 垃圾围城 总部基地

定向安置房 大宁水库 卢沟桥 雕塑园 革命遗址

扶轮胡同 危旧房改造 南水北调中线 重现明清 排水系统 堵车 安置房 腾退 基础设施

绿化美化 土地使用 市郊铁路 河西地区

轨道交通 集体土地 永定河西岸 交通瘫痪

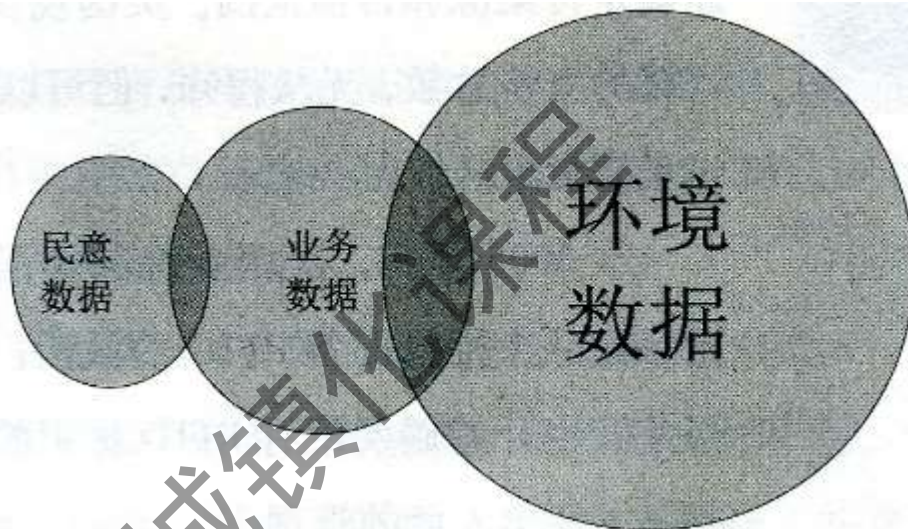
外来人口 世界种子大会 西郊砂石坑

棚户区改造 生态休闲区 锦绣谷 东山坡





- 民意数据
- 业务数据
- 环境数据



— 传感器—RFID，环境数据收集

联邦政府三种数据源以及收集方式的对比

数据种类	收集对象	收集方法	里程碑
业务数据	下级部门和各类社会组织	以基层上报、被动接收为主	1973 年卫生部引进最小数据集
民意数据	单个公民或组织	投入人力财力、主动收集	1940 罗斯福总统引进民意调查
环境数据	自然环境、动植物、物体	以传感器自动采集为主	1962 年的海浪监测计划

## 维度一：数据样本比例

# 数据源分类

### 小样本数据vs. 大样本数据：

- 小样本数据：数据取样比例低，间接反映整体数据，e. g. 抽样问卷
- 大样本数据：数据取样比例高，较直观反映整体数据，e. g. 在线地图

### 大样本数据的理想情景——全样本数据

- 全样本数据：直观研究完整数据，e. g. 手机信令数据





## 维度二：数据更新频率

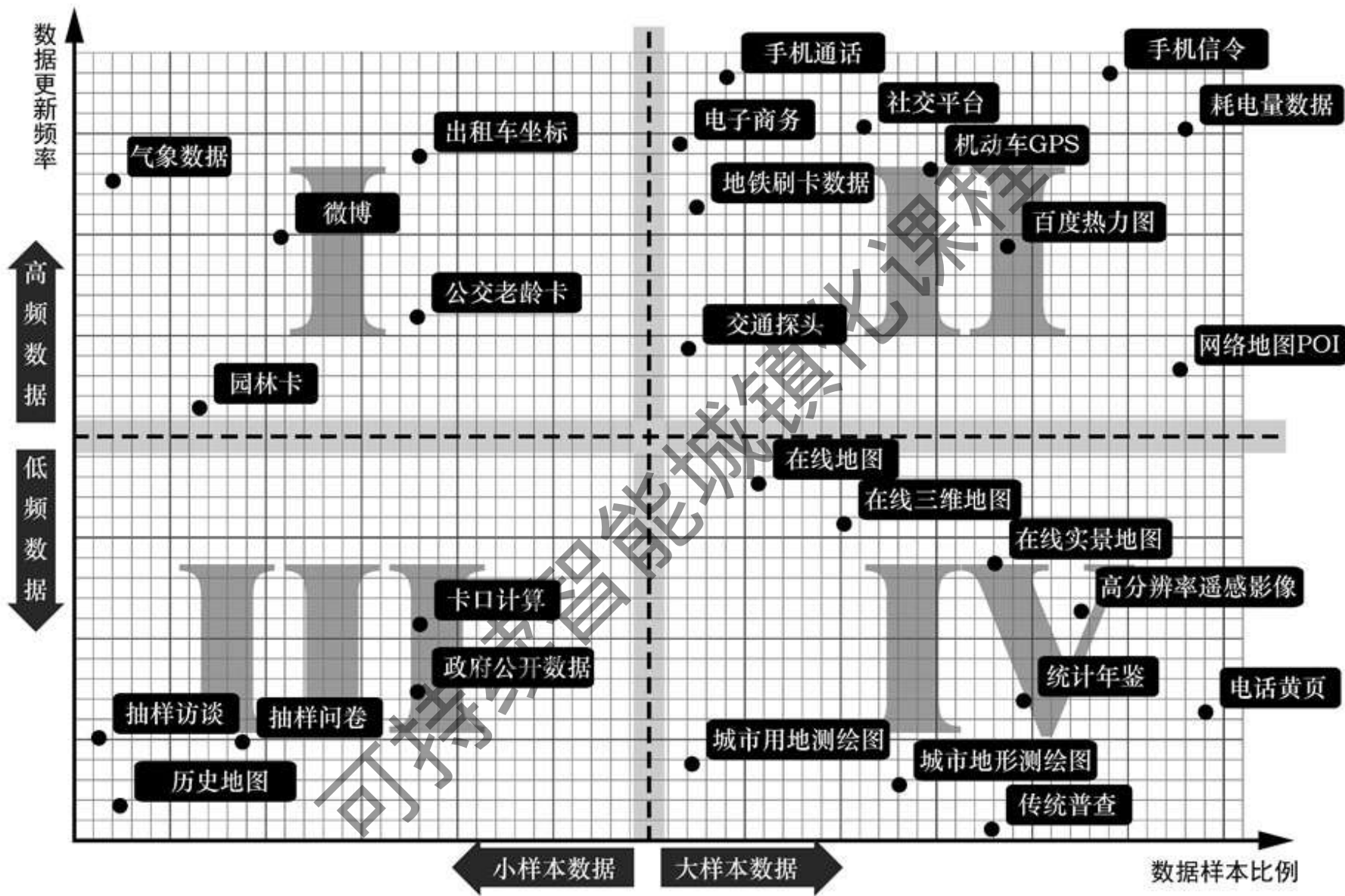
### 低频数据vs. 高频数据：

- 低频数据：数据更新速度慢，更新周期：年，e. g. 统计年鉴
- 高频数据：数据更新速度快，更新周期：分钟或秒，e. g. 社交网站

### 高频数据的理想情景——瞬频数据

瞬频数据：实时反映数据变化，e. g. 电子商务（淘宝，铁路购票系统）





# 数据是什么？数据分析是什么？

- 现代社会中，使用手机、点击网页、切换电视频道、开车穿过自动收费站、用借记卡或信用卡购物.....我们**每时每刻都制造数据**
- 雅虎、Google、360。。。。正在捕获它们，不会意识到一群新兴的数学精英正千方百计以惊人的准确度剖析它们。这群如狼似虎的数据恐怖分子可以洞悉我们的每个举动，预测我们的行动计划，他们神不知鬼不觉地将我们买了什么、对什么感兴趣、与谁坠入爱河尽收眼底。





图2 武汉某时段的实时交通

资料来源：百度地图 <http://map.baidu.com/>



图5 2014年春节期间人口迁徙图

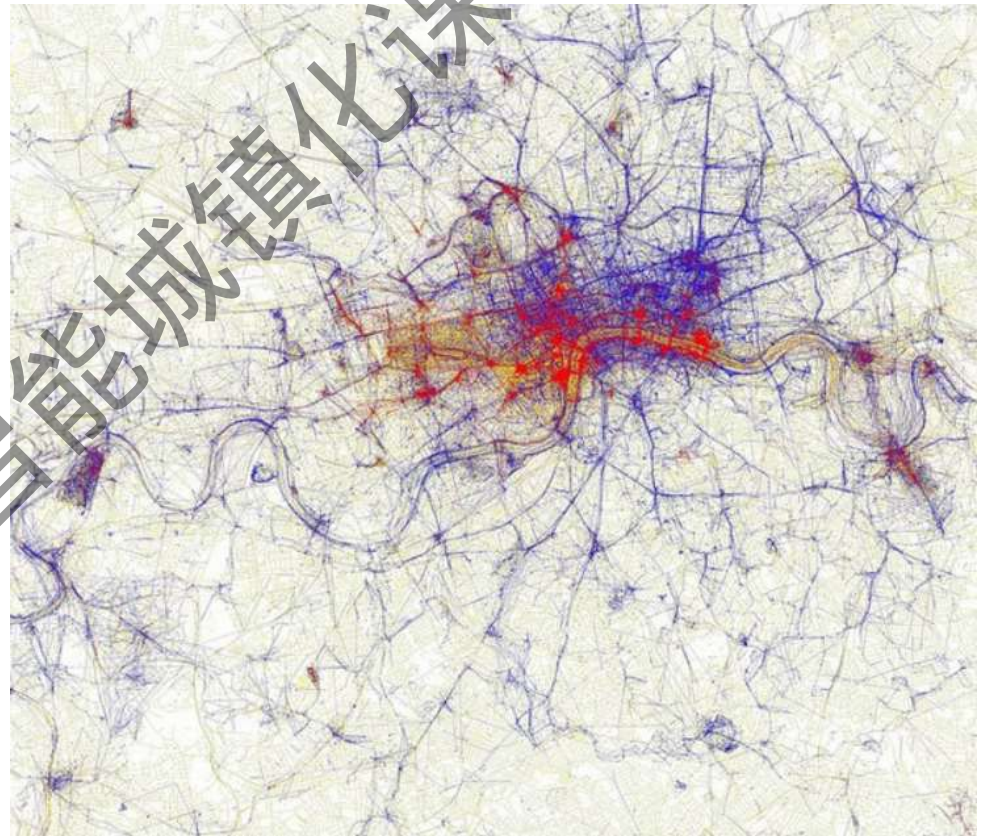
资料来源：百度迁徙 <http://qianxi.baidu.com/>



百度大数据实践

# “map geek”（地图极客）的Eric Fischer用城市数据制作的地图

- 我所做的一切，只是试图将我们透过大数据看到的真实世界朦胧化，然后放大某些特质，进而去理解城市的结构
- Local and Tourists（当地人与游客），当地人与游客的地理标记照片具象出伦敦、纽约、旧金山、巴黎、台北等124个城市
- 这是一个简单的概念，但的确揭示出人们探索一个城市时止步的边缘





- See something or say something
- 将人们上传到Flickr的城市照片与发表在tweet上的城市推文做个比较。从Fischer的地图可以发现，城市被拍与被推特的量不完全相仿。



San Francisco. "See something or say something". Eric Fischer

微信号: dashuju36

# Strava都市项目

- Strava都市项目是一个数据服务提供商，主要为那些骑行和步行用户提供“路面真实状况”的信息。每周全球范围内数以百万计的基于GPS轨迹的出行活动上传到Strava。在高密度的都市区域内，这类活动接近一半是通勤出行。这些活动创造了数十亿的数据点，通过聚合分析、数据挖掘尝试帮助城市的管理者更好的理解现实世界中自行车和行人的路径选择。



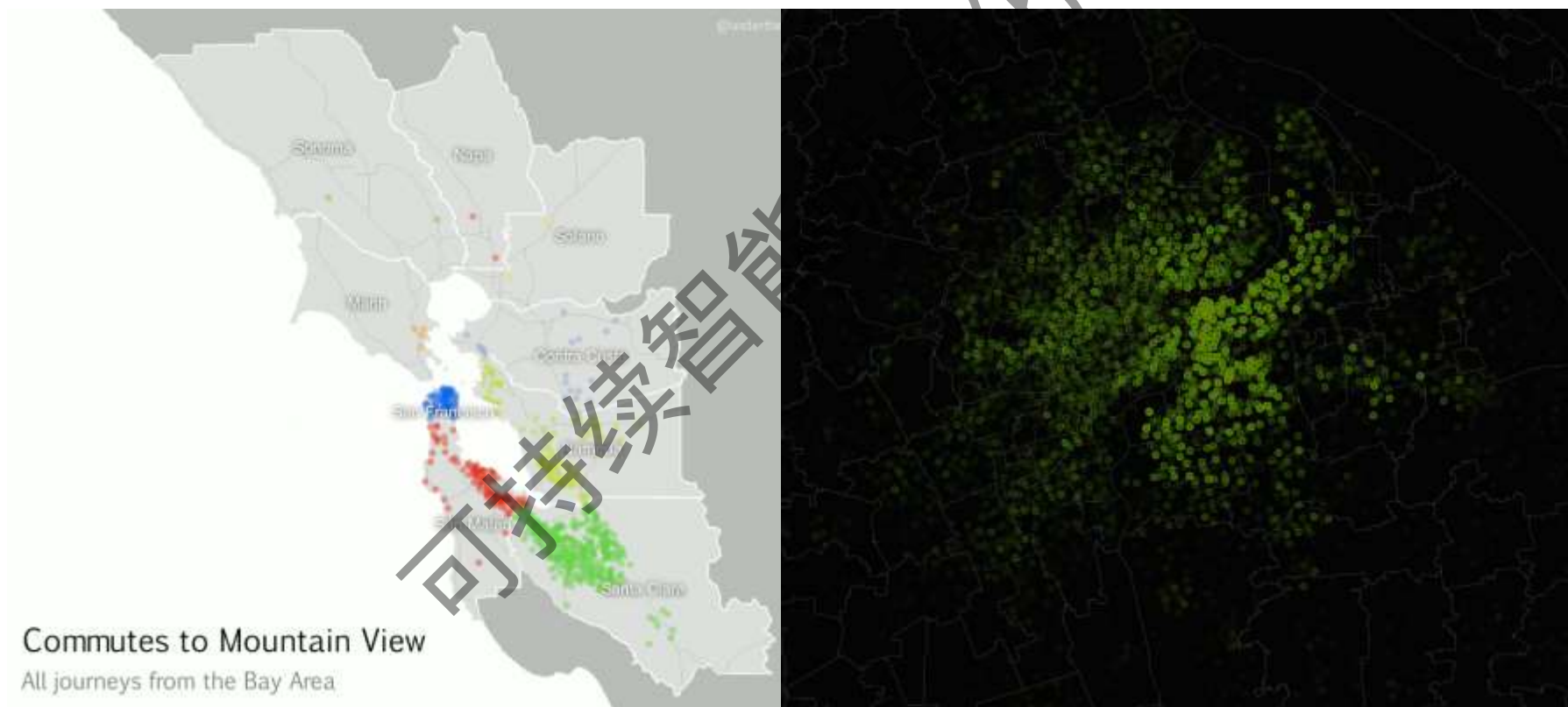
# Living Singapore



Imagine a *living* city  
where you know in real time  
all that is happening around you...



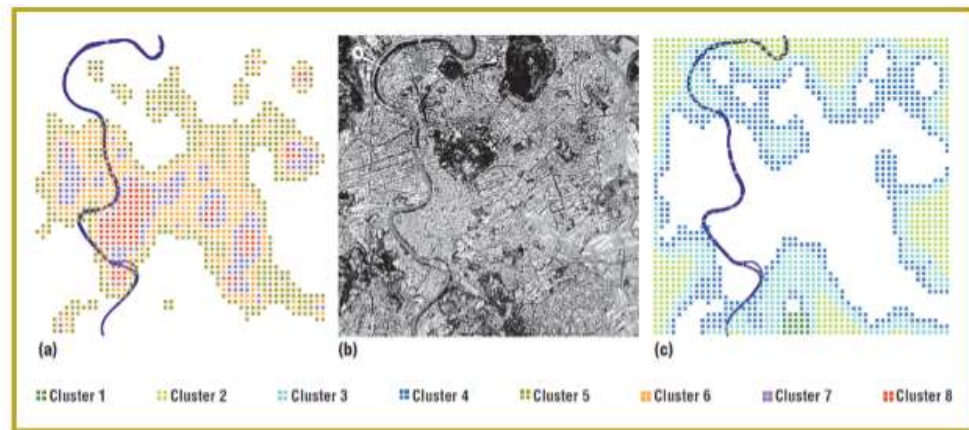
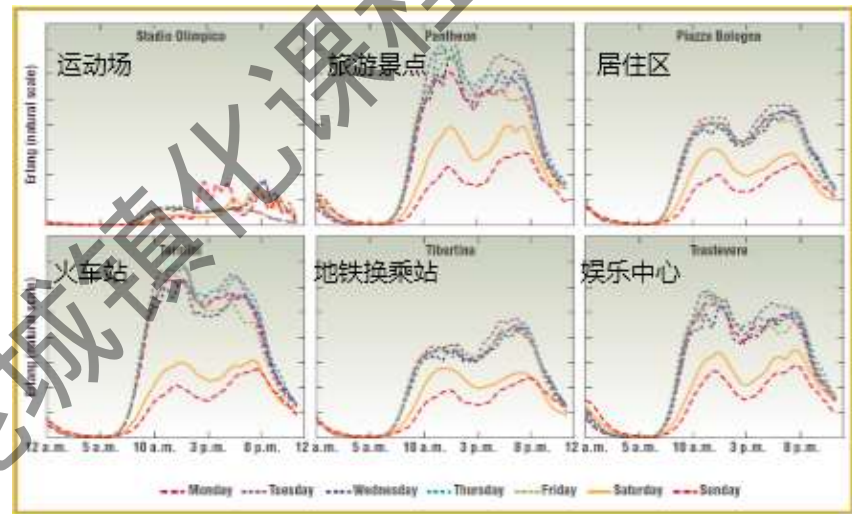
# 通勤动态可视化





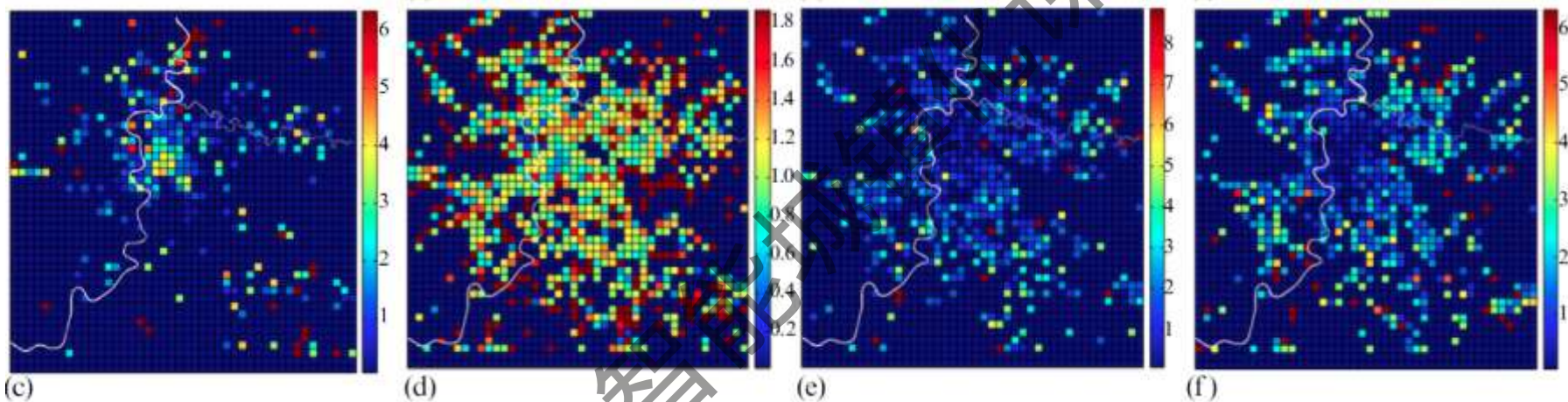
# 研究者的关注

- 1、我们是否可以找到重复出现的人类行为模式？
- 2、这些模式受特定事件的影响如何？
- 3、我们是否可以区分特定的空间区域，发现活动模式在空间上的特征？
- 4、我们是否可以比较多种模式在世界各主要城市间的差异？



# 城市内部空间结构

**研究结果：**与business location data ( Pagine Gialle数据 ) 对比。商业类型划分四类：**住宿、餐饮、日用品零售和食品零售**



**住宿：**相对集中分布，集聚在游客出现出现频繁的地方；在外围交通设施（机场和轨道点）集聚；

**餐饮：**非常分散（意大利人喜爱美食的性格）

**日用零售和食品零售：**在市中心相对较少，普遍外围分布，指向居住区

# 大数据的应用

运用信息和通信技术手段感测、分析、整合城市运行核心系统的各项**关键信息**

从而对包括民生、环保、公共安全、城市服务、工商业活动在内的各种需求做出**智能响应**

**为人类创造更美好的城市生活**

# 城市空间信息推送



# 外滩踩踏事故分析

百度大数据实验室报告

可持续智能城市化课程



# 外滩踩踏事故简介

2014年12月31日晚，上海外滩陈毅广场发生拥挤踩踏事故，造成**36人**死亡，**49人**受伤的惨剧



死亡：**36人**

男性**10人**

女性**26人**

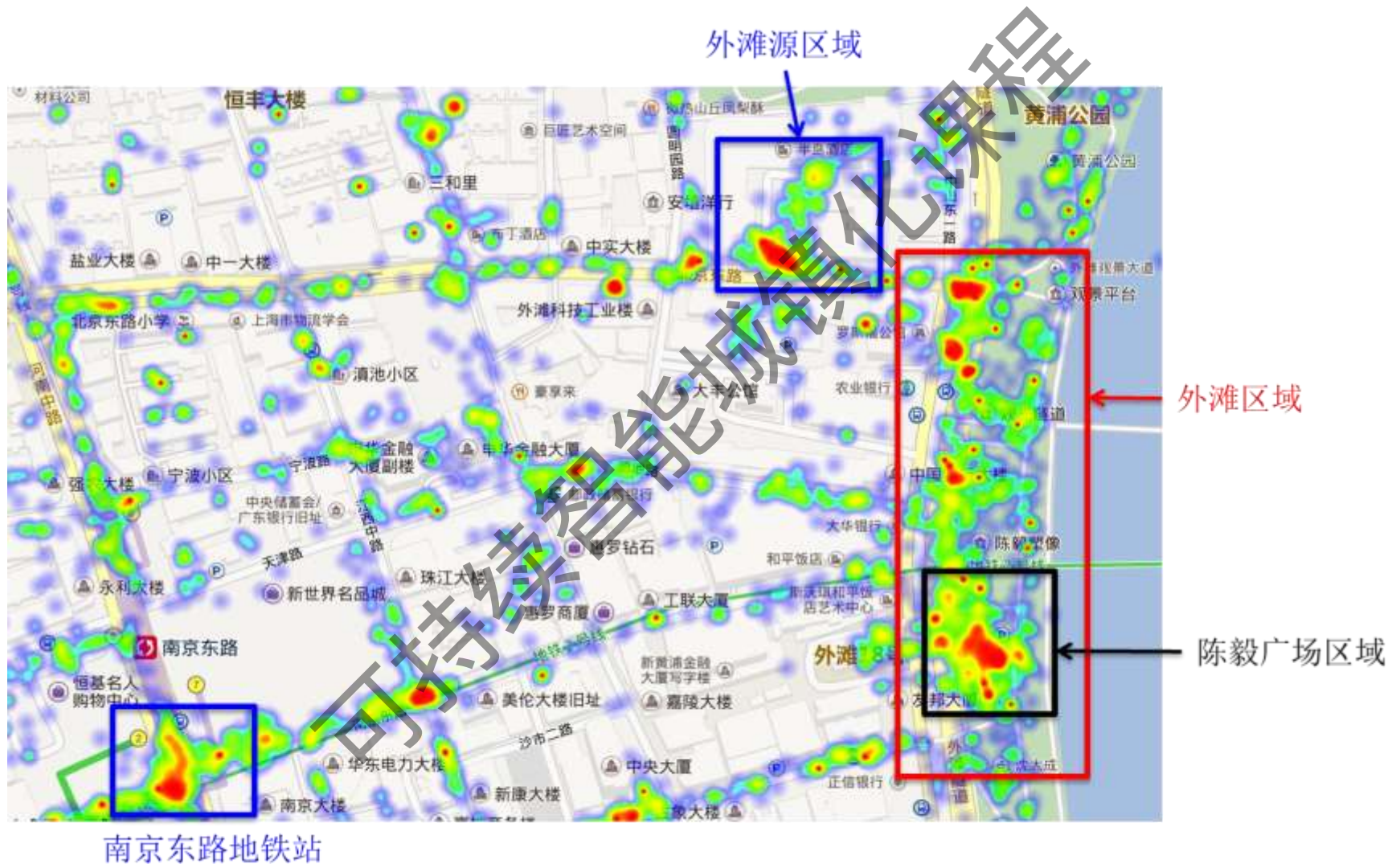
受伤：**49人**

重伤**13人**

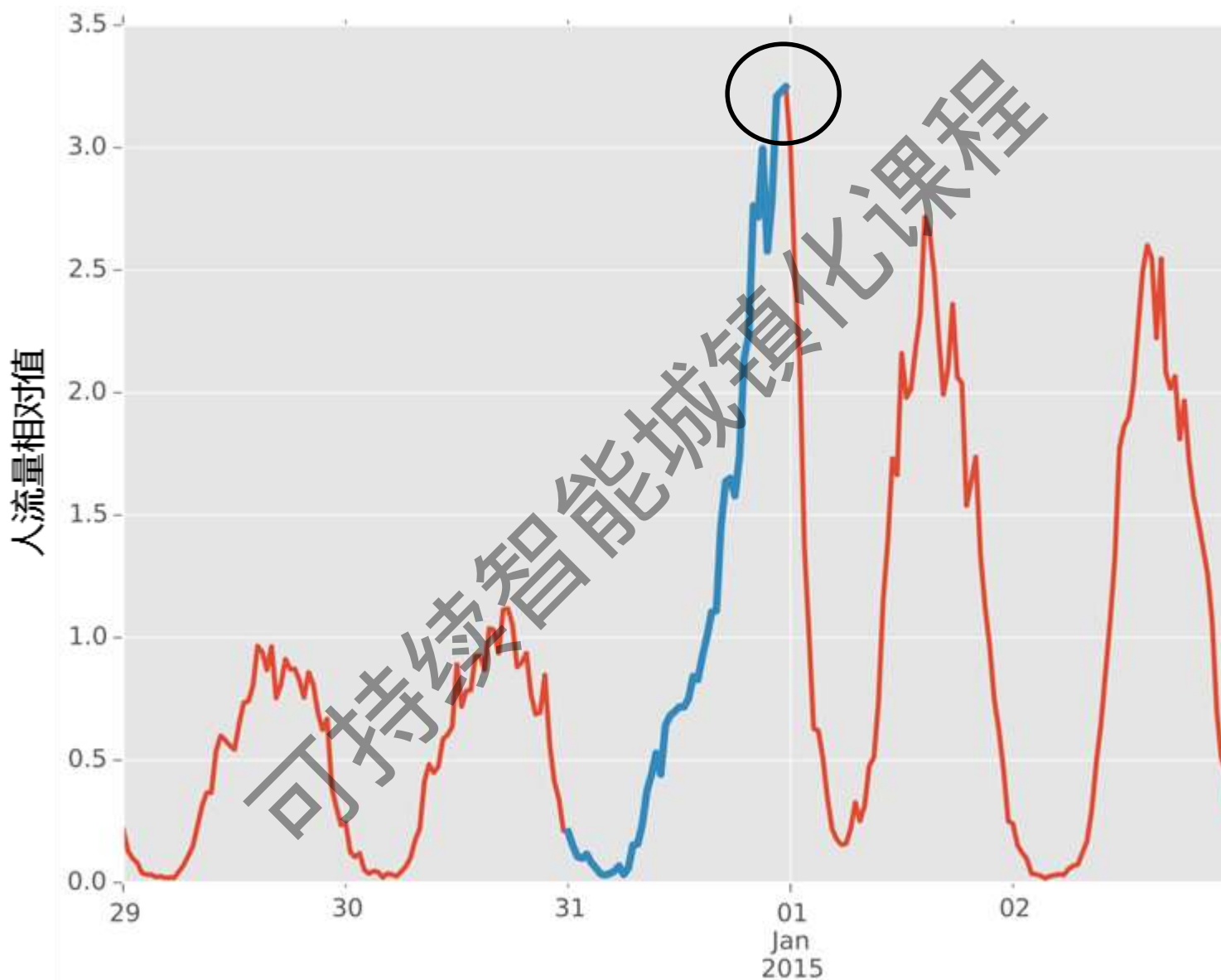
轻伤**36人**

最大**36岁**，最小**16岁**；平均年龄仅**22岁**

# 2014.12.31 事发时外滩区域人群热力图

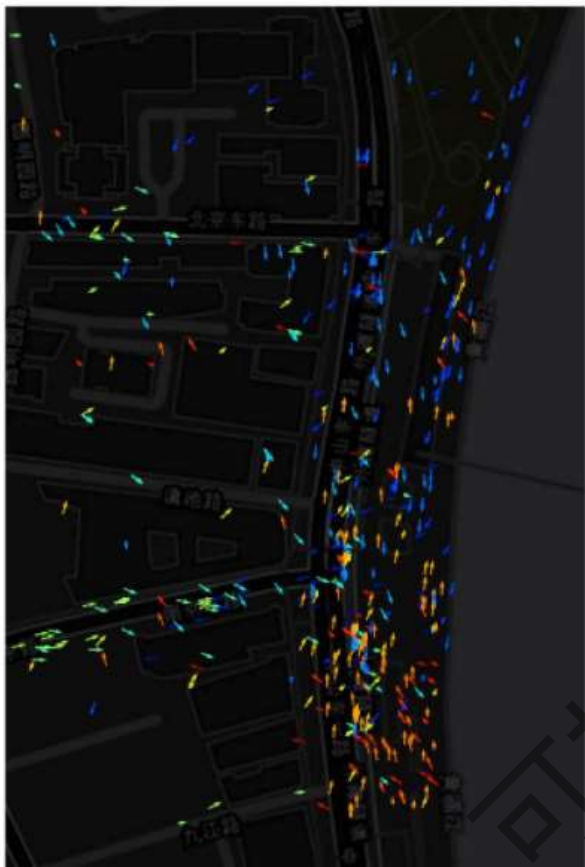


# 2014.12.29-2015.1.2外滩区域人流量趋势

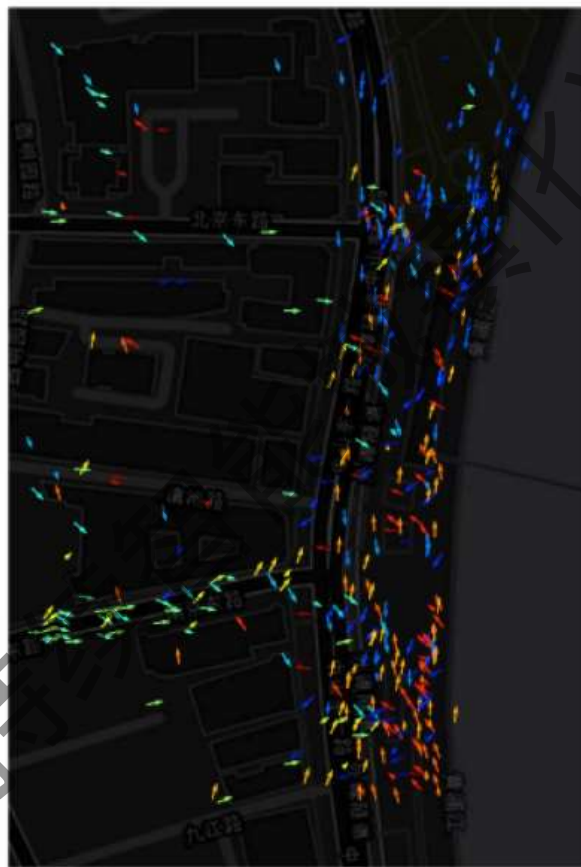




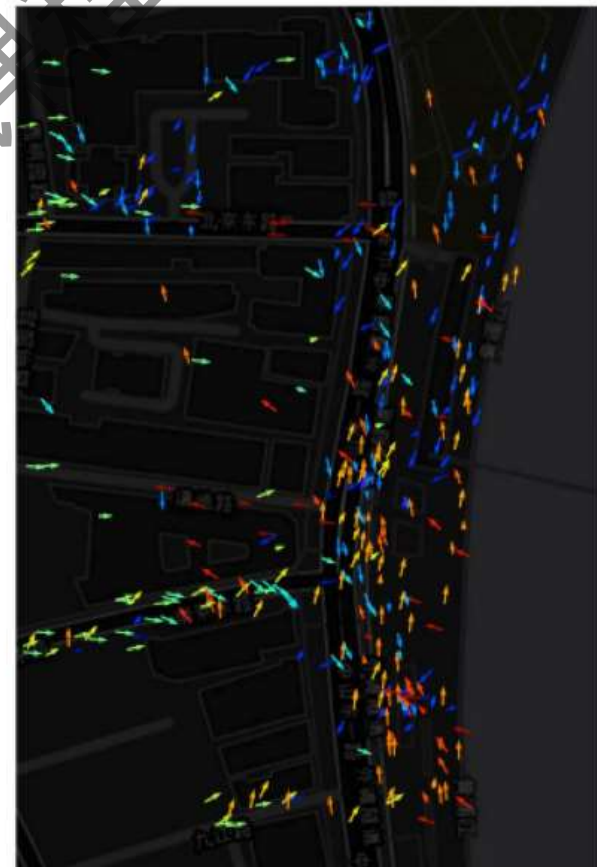
# 外滩和外滩源区域人群流动方向示意图 (部分采样)



(1) 中秋前夜

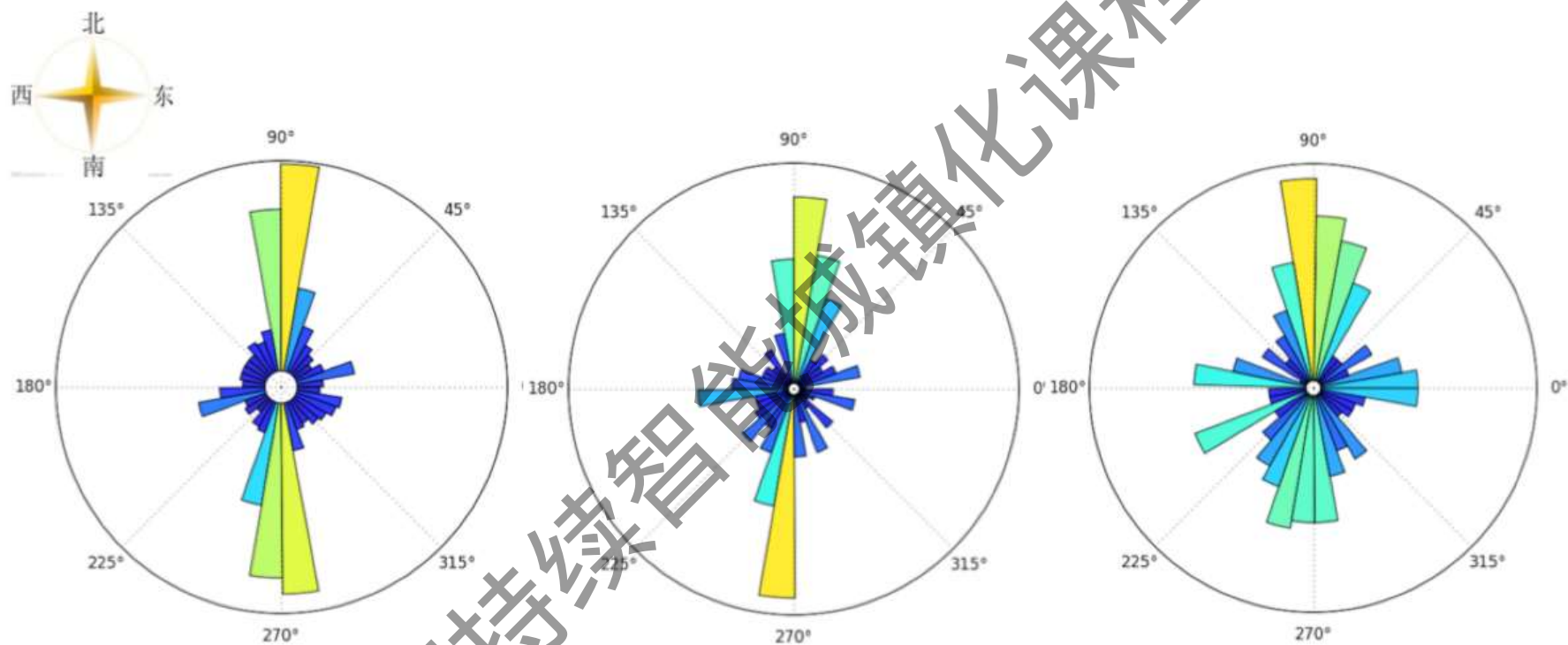


(2) 国庆当晚



(3) 跨年当晚

# 外滩区域人群流动方向分布图



(1) 中秋前夜

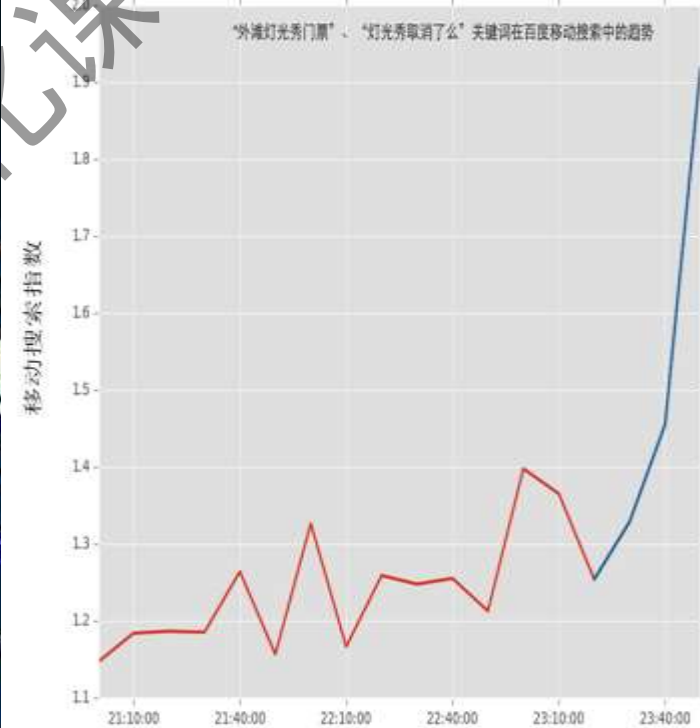
(2) 国庆当晚

(3) 跨年当晚



# 相关搜索行为分析

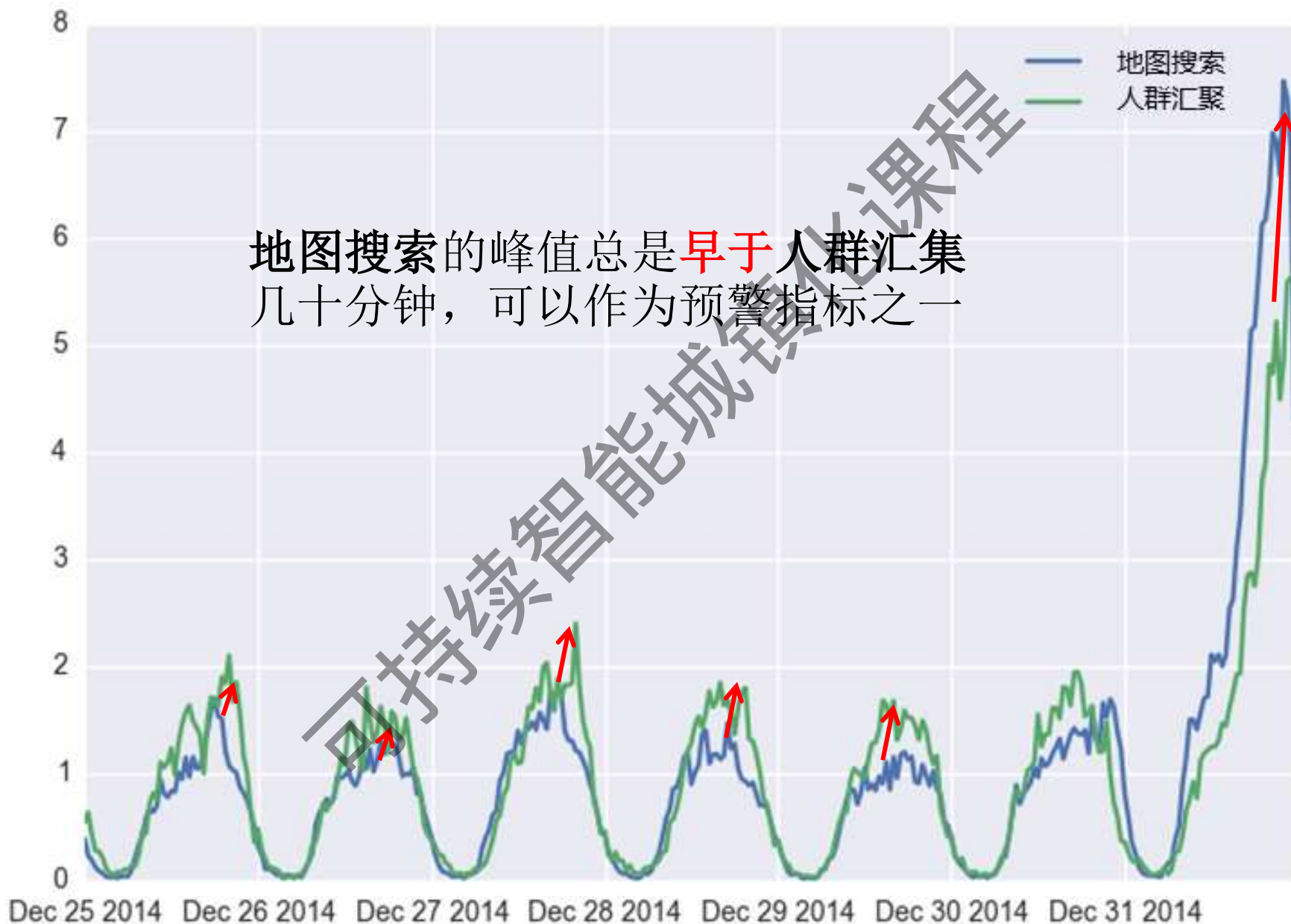
课程



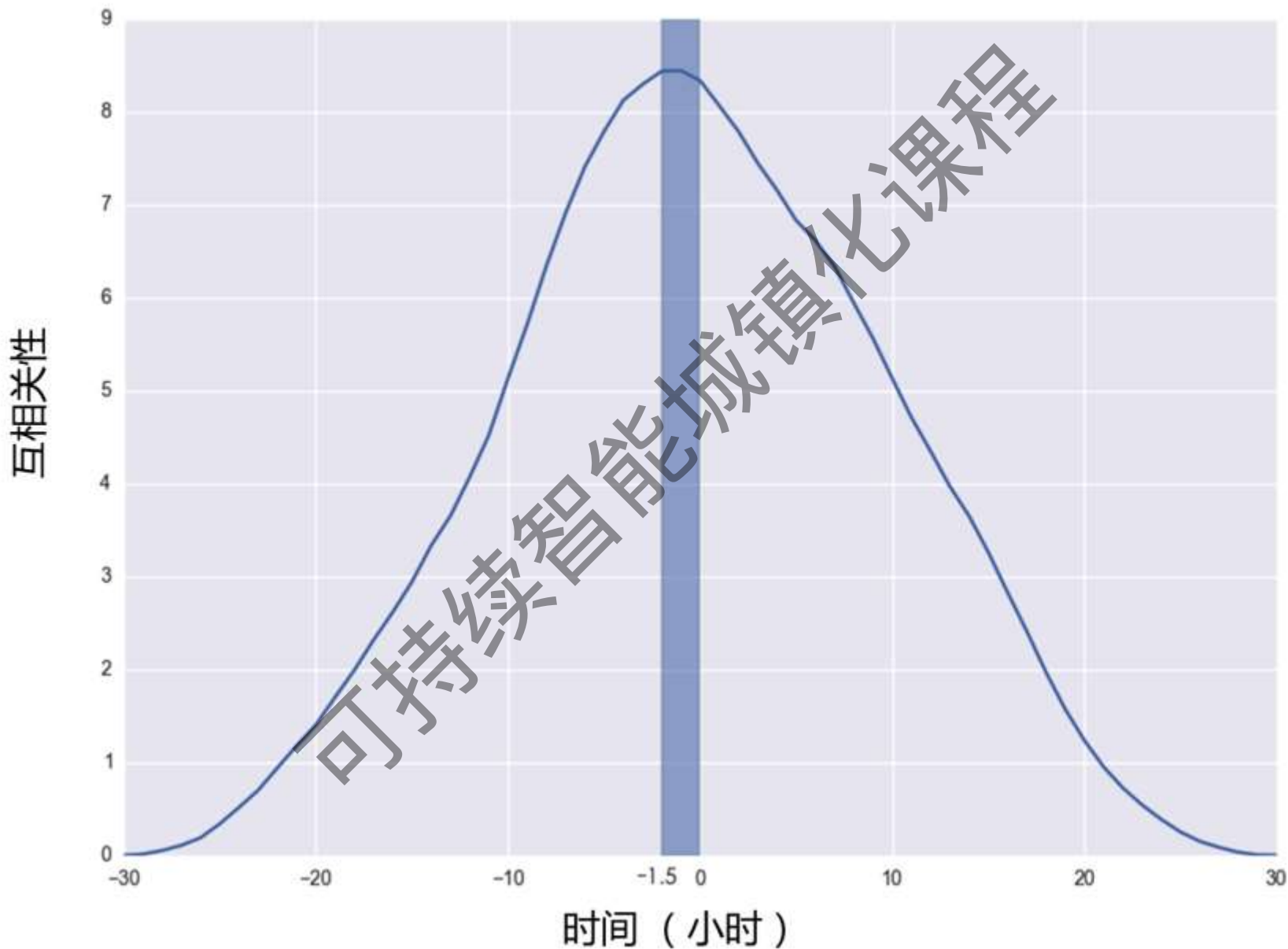
“外滩源”为目标的地图搜索发起点热力图

搜索“灯光秀取消了么”  
和“灯光秀门票”的关键词  
指数

# 外滩地图搜索与人群汇聚情况趋势图



# 外滩地图搜索与人群数量的互相关性曲线



- 谢谢聆听

[dewang@tongji.edu.cn](mailto:dewang@tongji.edu.cn)

王德 Wang De

同济大学建筑与城市规划学院 CAUP, Tongji University