



北京城市实验室  
Beijing City Lab

ID of the slides

16



## Slides of BCL

[www.beijingscitylab.com](http://www.beijingscitylab.com)

### How to cite

Author(s), Year, Title, Slides at Beijing City Lab, <http://www.beijingscitylab.com>

E.g. Long Y, 2014, Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest, Slides at Beijing City Lab, <http://www.beijingscitylab.com>



城市模型学术报告会暨 2014北京城市实验室年会  
Urban Modeling Workshop & 2014 BCL Conference



# 志愿者地理大数据分析与可视化

## 中国案例研究

---

王江浩

([wangjh@reis.ac.cn](mailto:wangjh@reis.ac.cn))

中国科学院地理科学与资源研究所  
资源与环境信息系统国家重点实验室

2014-06-11 @ 清华大学建筑学院

# 主要内容

---

- 1** 志愿者地理信息(VGI)
- 2** VGI数据挖掘技术
- 3** 中国案例研究
- 4** 未来展望

# 1 志愿者地理信息(VGI)



# 志愿者地理信息 (VGI)

Citizens as sensors: the world of volunteered geography.

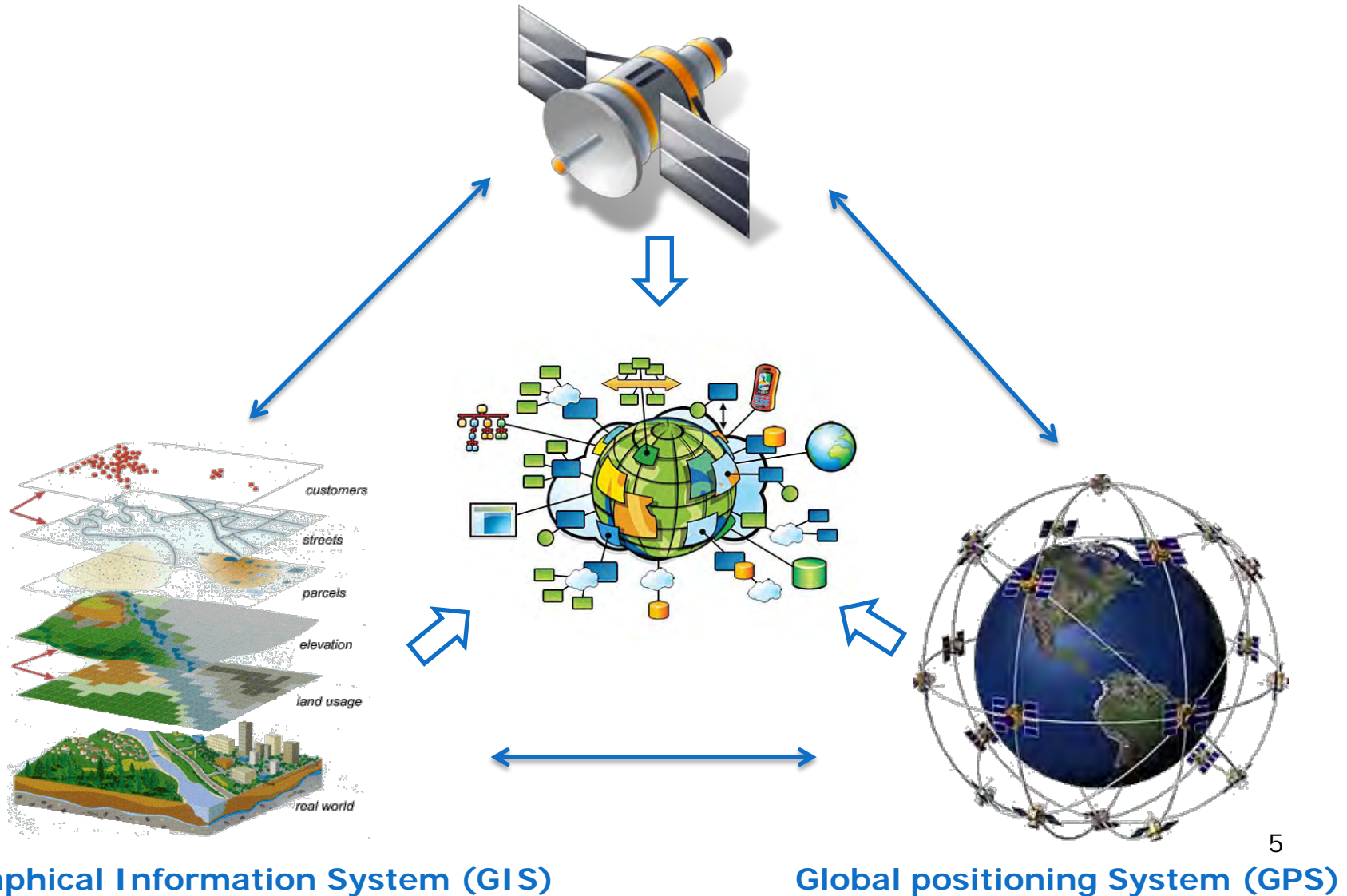
—— Michael F. Goodchild , 2007

- 随着计算机技术，GPS，移动终端技术发展，在互联网Web2.0驱动下，传统地理信息由单向方式逐渐向**交互双向协作**方向发展。
- **维基百科**中对VGI的定义是指用户通过在线协作的方式，以普通手持GPS终端、开放获取的高分辨率遥感影像，以及个人空间认知的地理知识为基础参考，创建、编辑、管理、维护的地理信息。



# 传统GIS数据获取方式

Remote Sensing (RS)



# Web2.0下地理信息获取

## □ Web2.0特点

- 多人参与
- 人是灵魂
- 信息可读可写
- 可定制、自由度大



## □ VGI地理信息来源

- 移动定位 LBS
- IP地址识别
- 用户生成内容(UGC)



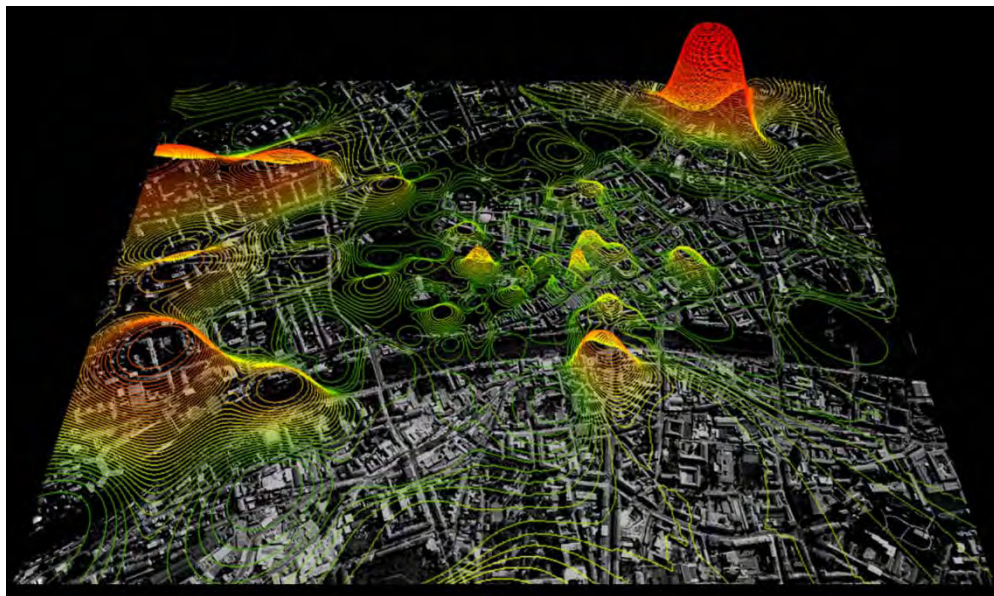
# VGI的应用进展

## □ GIS产业界



## □ 行业应用

- 新型的重要空间(大)数据来源
- VGI与LBS，社交网络
- 时空间行为
- 公共环境监测
- 城市规划，管理
- 智慧城市
- .....





# 2 VGI数据挖掘技术



# VGI信息获取源

## □ 社交网络



## □ 专业网站



## □ 其他来源



# VGI 分析流程与工具

## □ 数据获取

- 爬虫系统



## □ 空间数据库



## □ 时空数据分析与挖掘



## □ 数据可视化



CartoCSS

# 3 中国案例研究

- ① 全国（重点）城市小区单元房价时空信息系统
- ② 全国公共交通信息采集与可视化
- ③ 百度LBS人口迁徙研究
- ④ 全国感兴趣点（POI）数据采集与挖掘
- ⑤ 全国微博签到信息分析与挖掘

# 小区单元房价时空信息系统

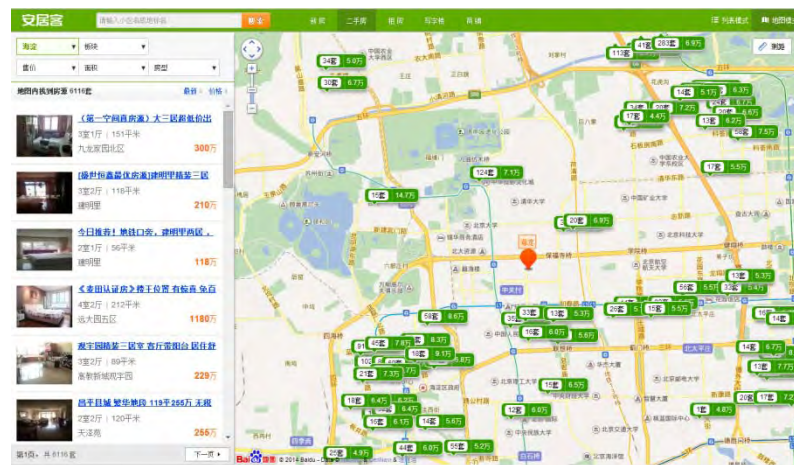
## □ 研究目的

- 收集全国城市小区尺度房价信息及其关联因子
- 分析城市内部房价空间格局
- 时空GWR模型案例
- .....

## □ 研究方法

- 抓取UGC网站提供的信息
- 构建小区房价时空信息系统

## □ 数据来源



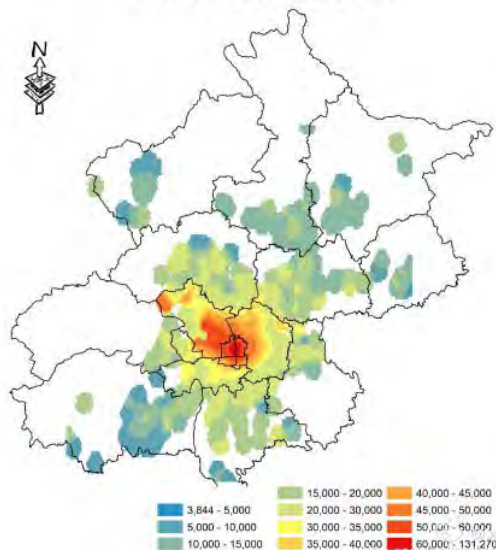
# 小区单元房价时空信息系统

## 房价信息

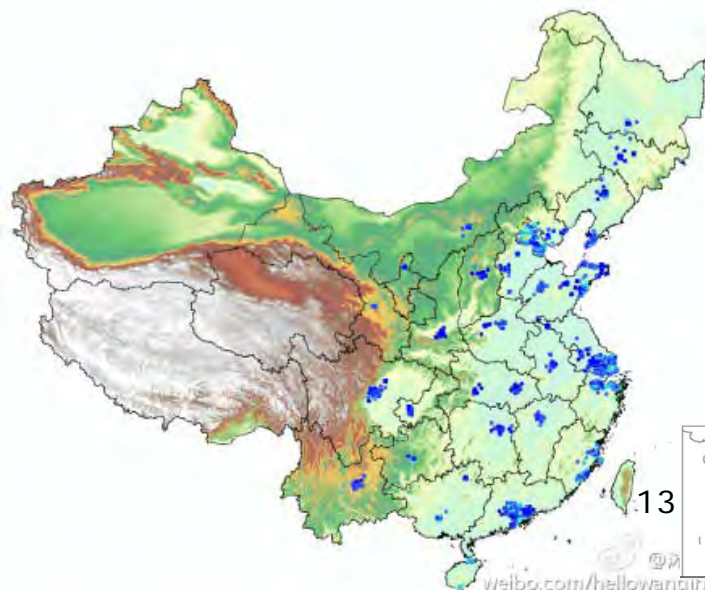
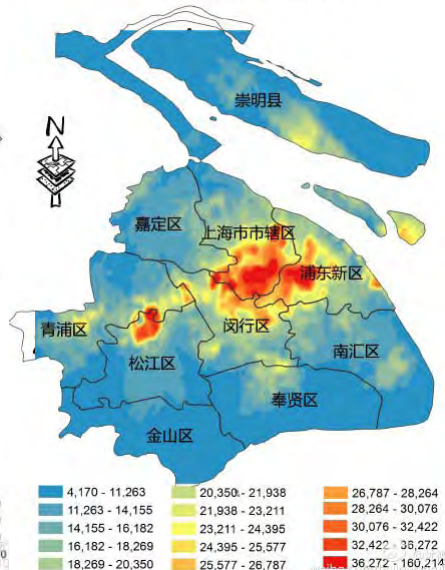
- <http://jianghao.github.io/HousePrice/>
- 全国超过10万的小区房价信息、物业信息（包括容积率，绿化率，停车位等等）
- 部分城市包括房价时间序列
- BCL上共享北京房价数据



2013年10月北京房价地图



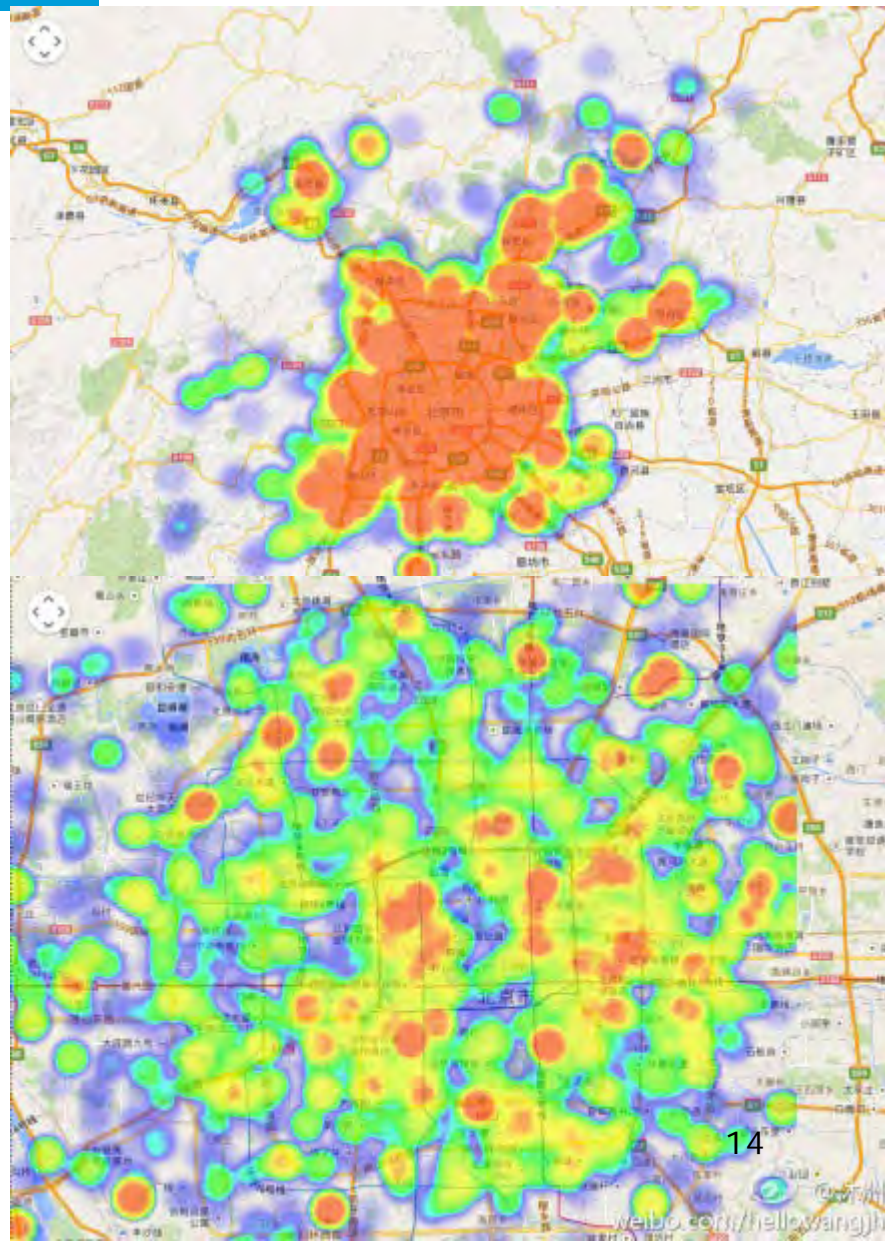
上海2013年8月房价地图



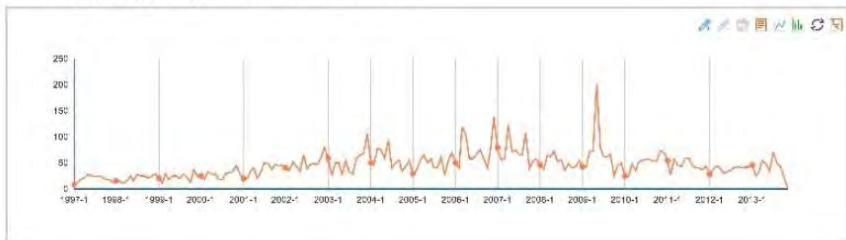
# 规划许可用地热点分析

- BCL获取并空间化北京规划委员会公布的1997 - 2013年 9061个规划许可用地，并进行地理编码
- 多尺度热点分析。间接反映近年来北京城市建设的热点和方向

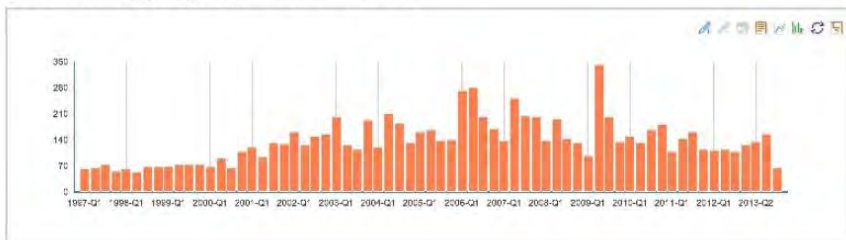
<http://jianghao.github.io/bjPublic/>



Monthly aggregated time series



Quarter aggregated time series



# 全国公共交通信息采集与可视化

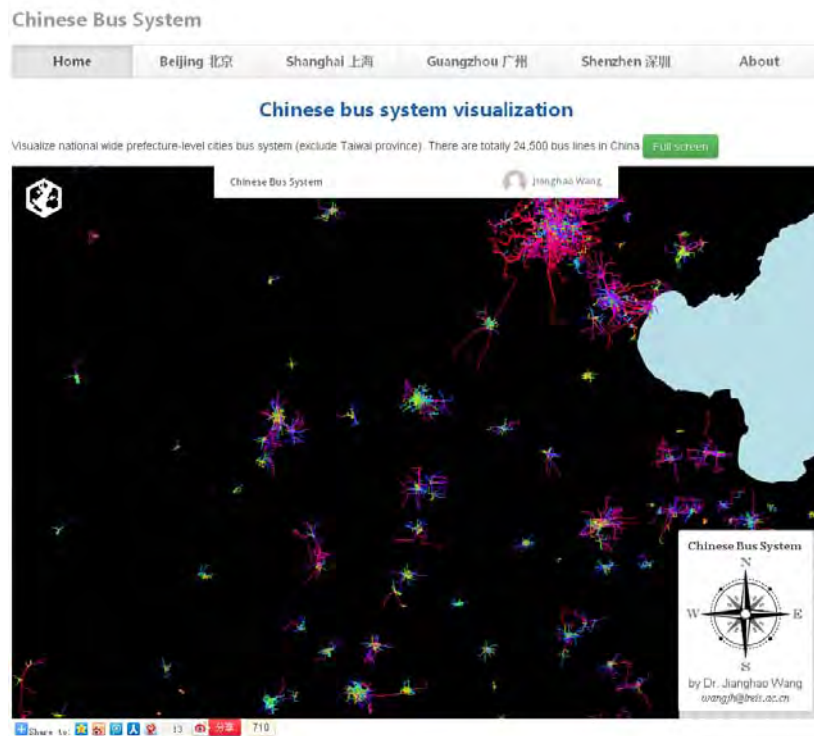
## □ 研究目的

- 采集全国尺度数据
- 城市规划中基础设施评价
- 通勤信息以及智能卡信息匹配
- 可达性分析
- 生活质量评价重要指标

## □ 研究方法

- 采用爬虫系统采集公交线路和站点信息
- 建立公共交通空间数据库
- 大数据可视化研究

## □ 数据来源



<http://jianghao.github.io/bus/>

338个地级以上城市  
24,500 条公交线路  
600,000个公交站点



# 全国公共交通信息可视化

北京

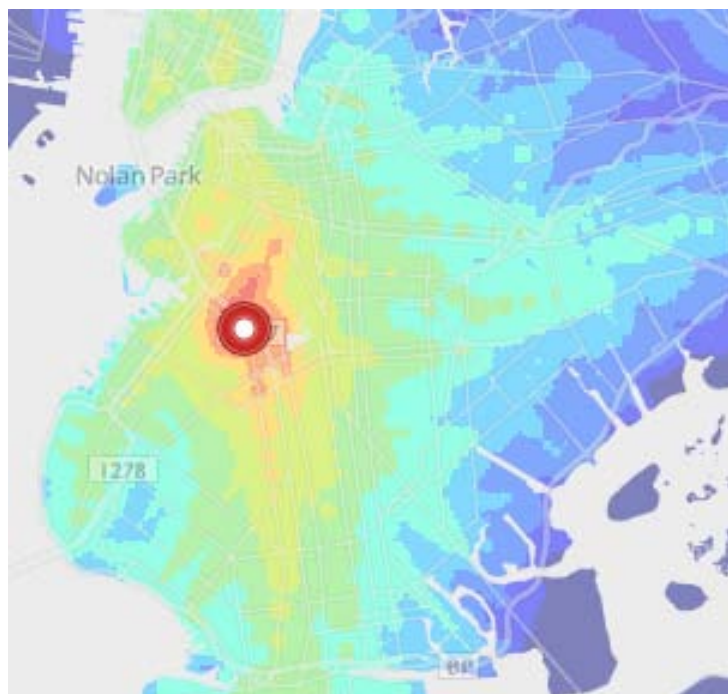


上海



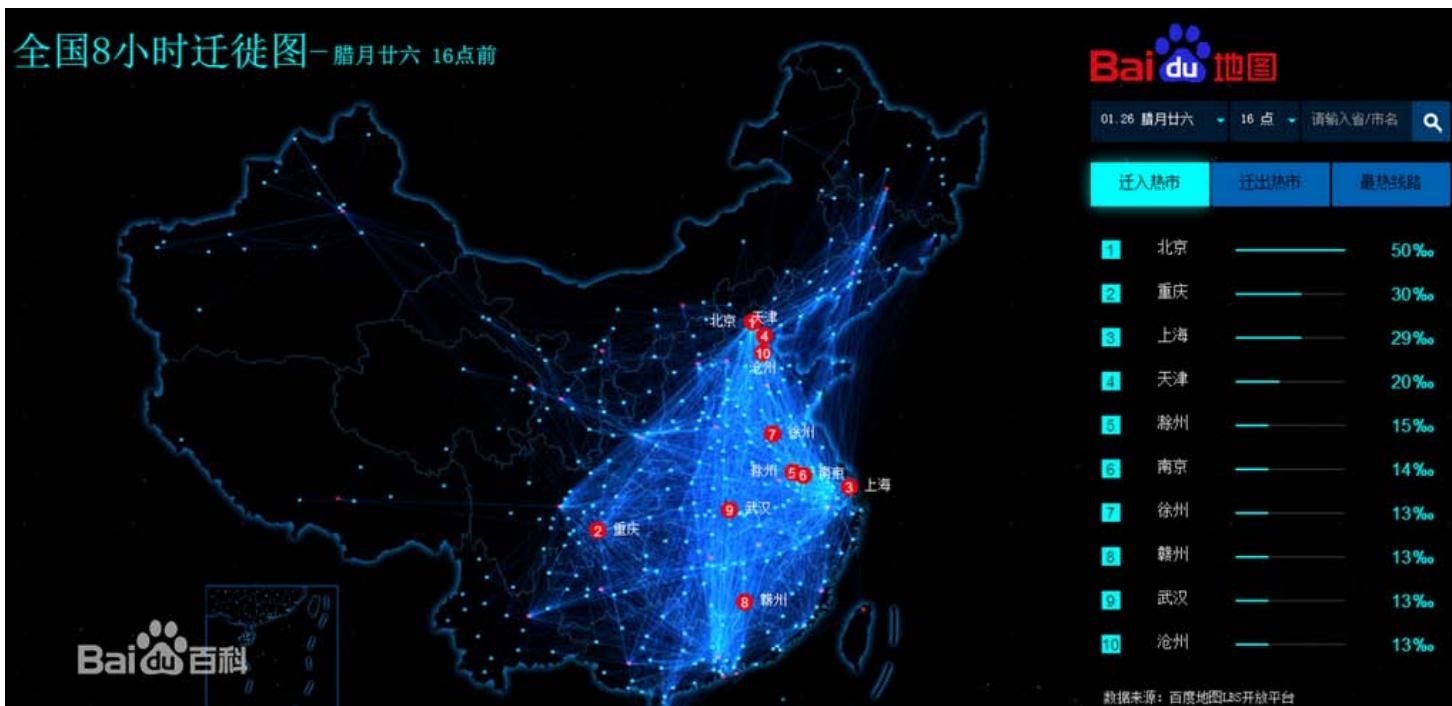
# 基于Web 地图的可达性分析

- Web地图包括最完整的路网信息（等级，速度，拥堵状况等）
- Web地图包括公共交通信息（公交，地铁）
- 贴近现实的出行时间计算

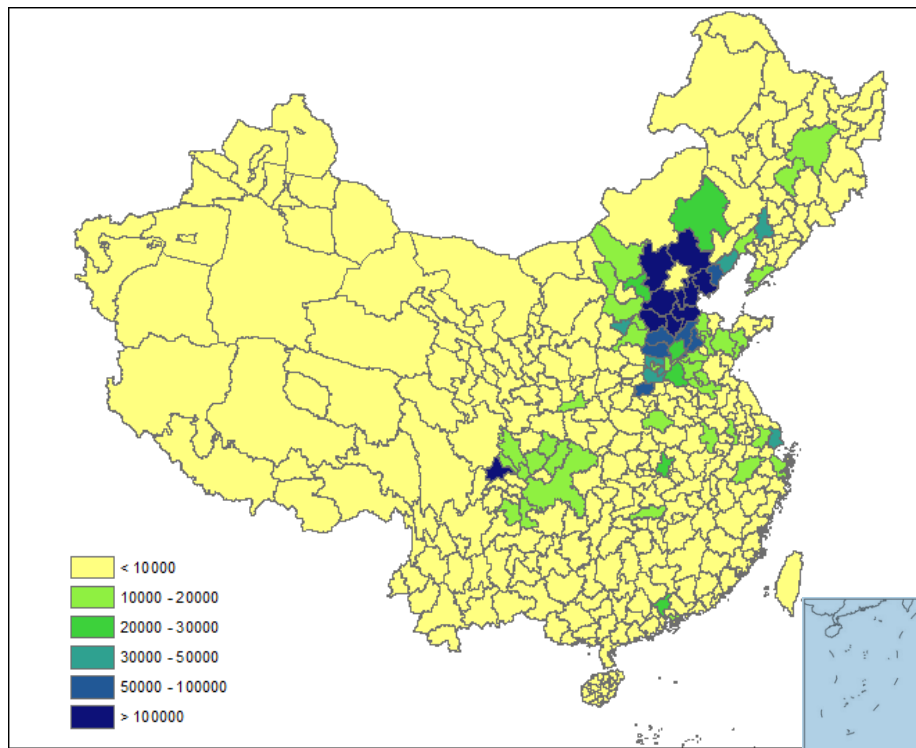


# 百度LBS人口迁徙研究

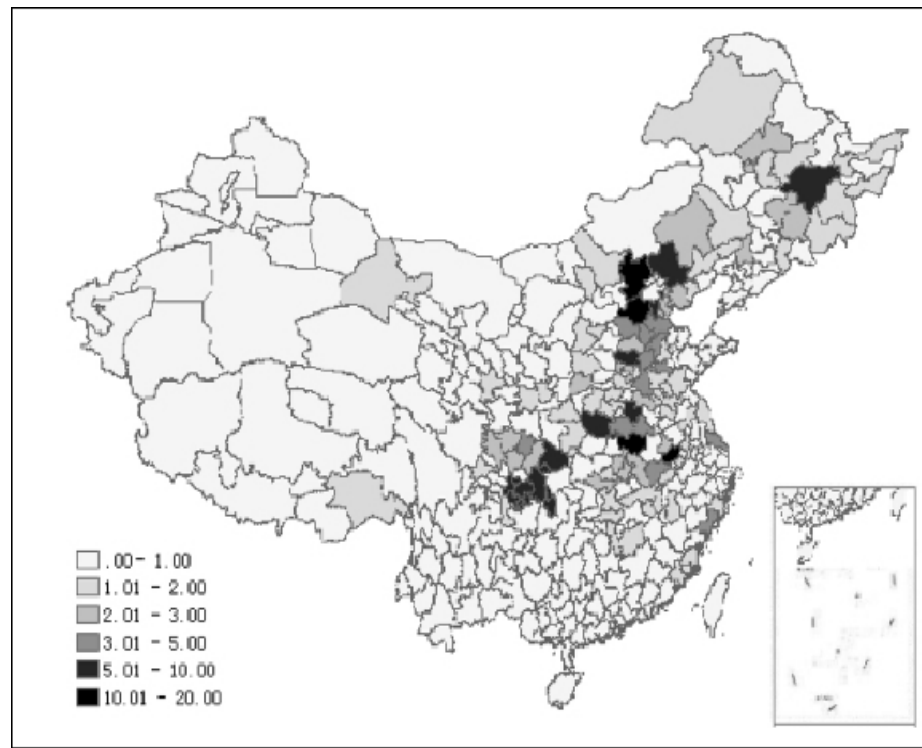
- 百度在春运期间，对其拥有的LBS大数据进行计算分析。实现了全程、动态、即时、直观地展现中国春节前后人口大迁徙的轨迹与特征。
- 我们在百度迁徙网站运行期间，获取春运期间，每小时人口流动统计数据，研究中国地级市尺度人口流动格局和影响因素。



# 基于百度迁徙的北京市流动人口分析



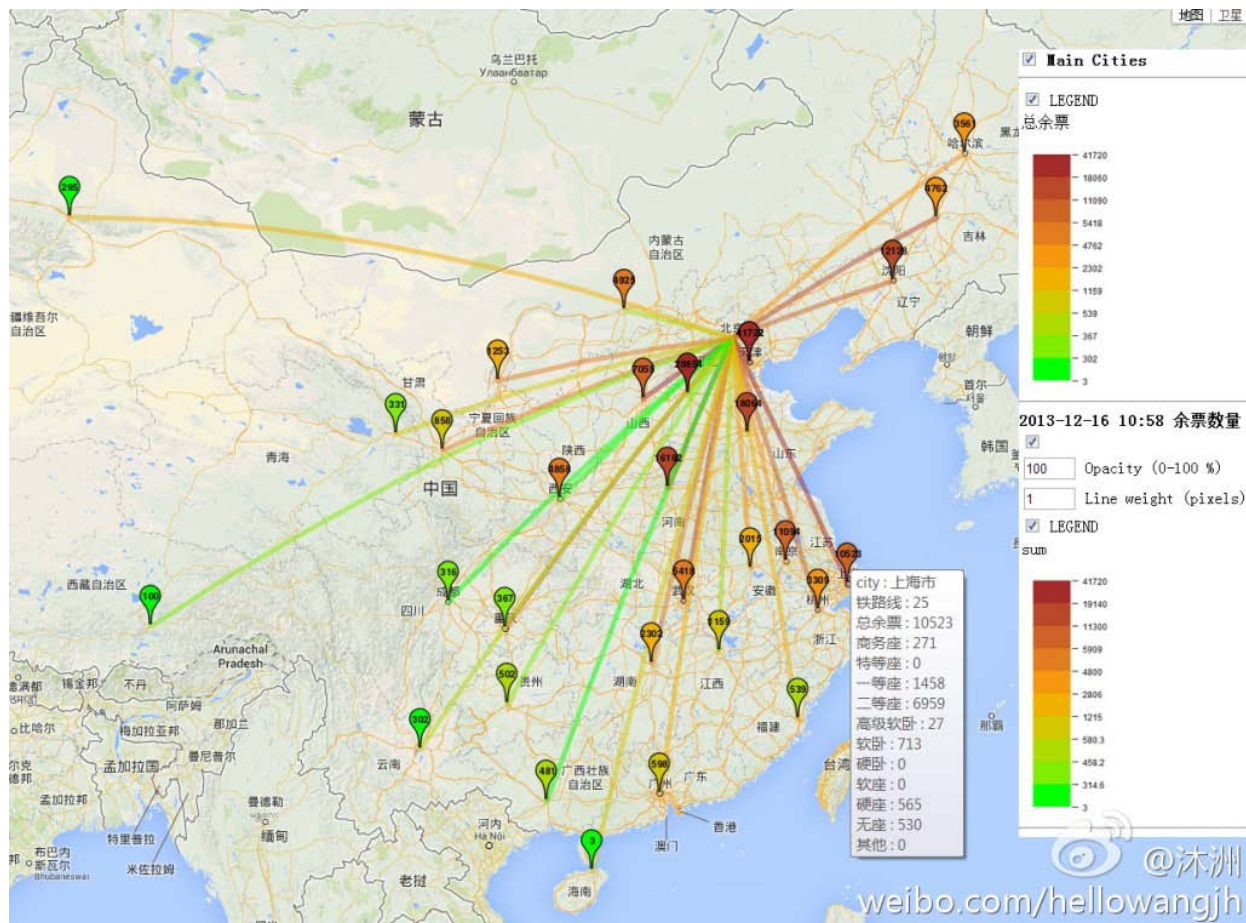
2014年1月16日-除夕15天内“百度迁徙”



2005 年全国 1%人口抽样调查

# 铁路余票信息查询与可视化

春运期间，实时获取并可视化铁路12306 网站中北京到省会城市间每日余票信息



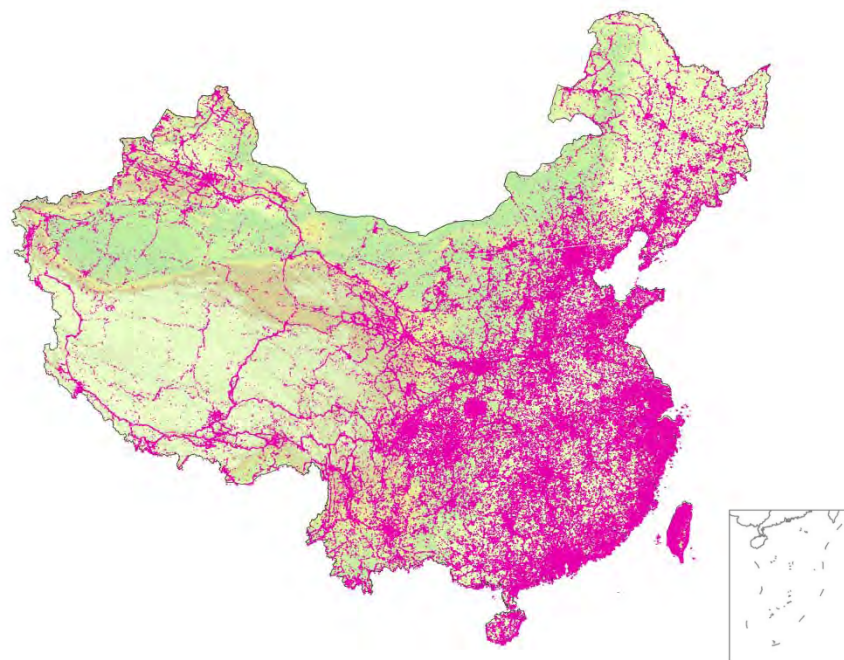
<http://jianghao.github.io/train/test/2013-12-16>

# 全国感兴趣点(POI)分析与可视化

- 目的：收集全国所有POIs
- 数据来源：谷歌地图、新浪微博、点评网



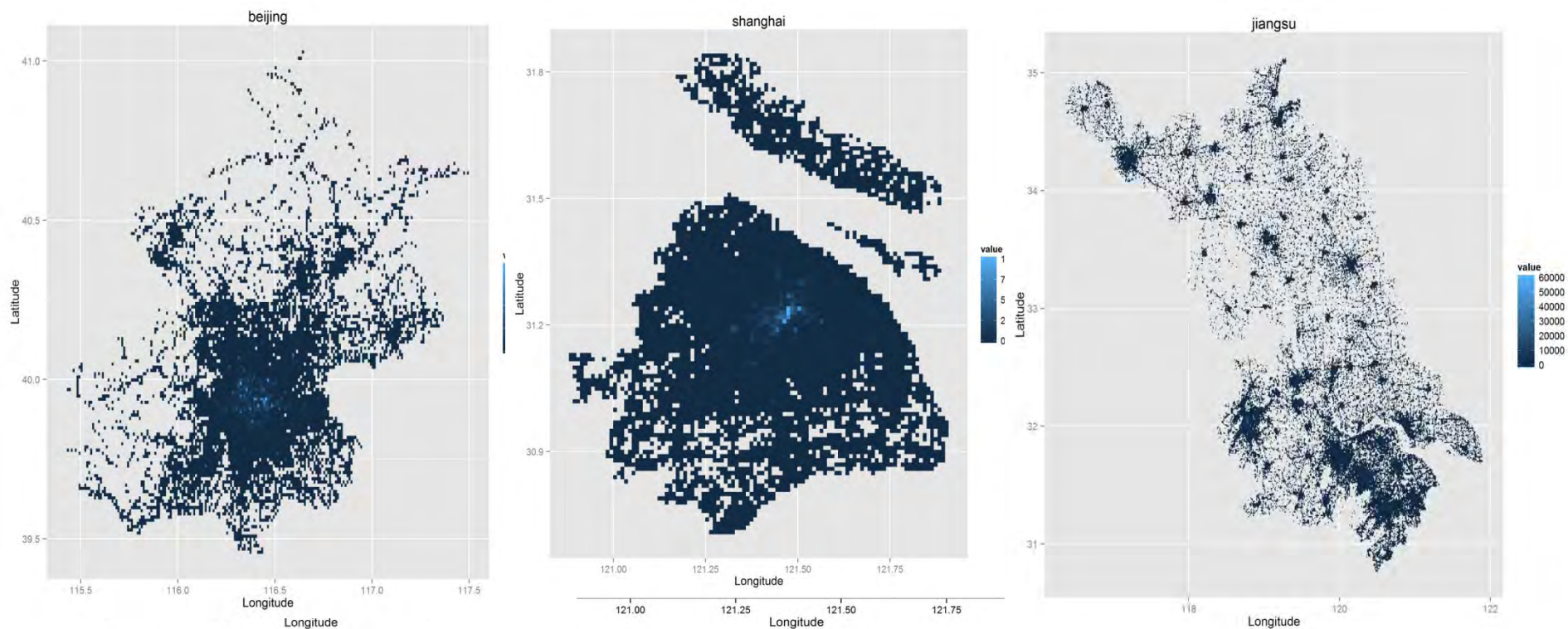
全国医院分布热力图



Google map geotagged photos

# 新浪微博POIs

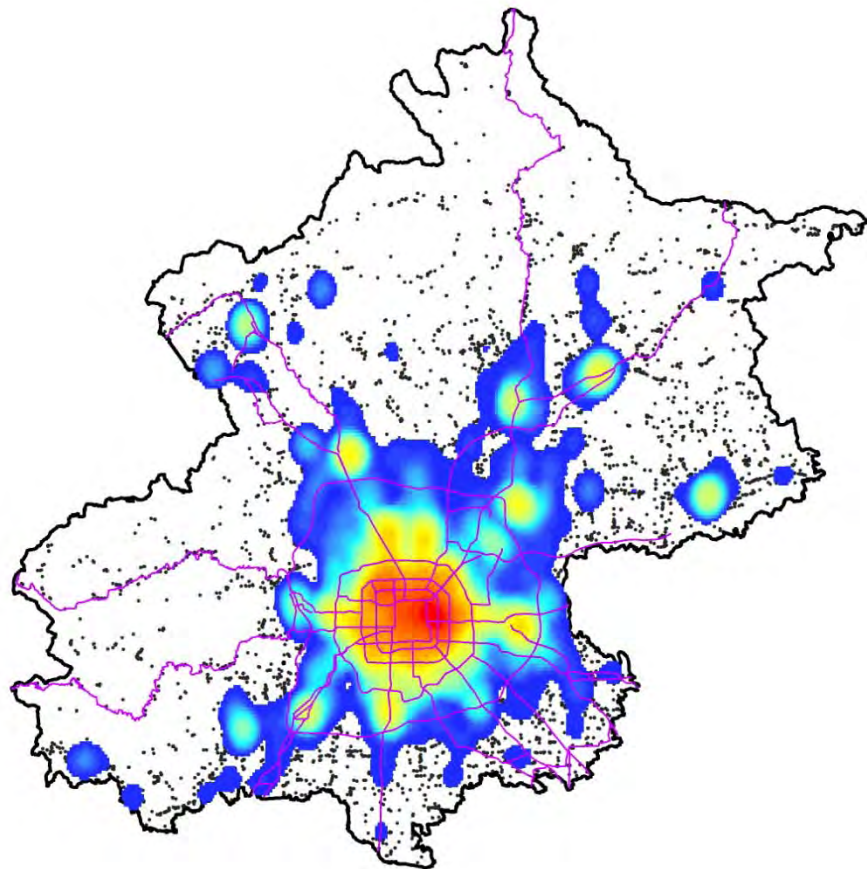
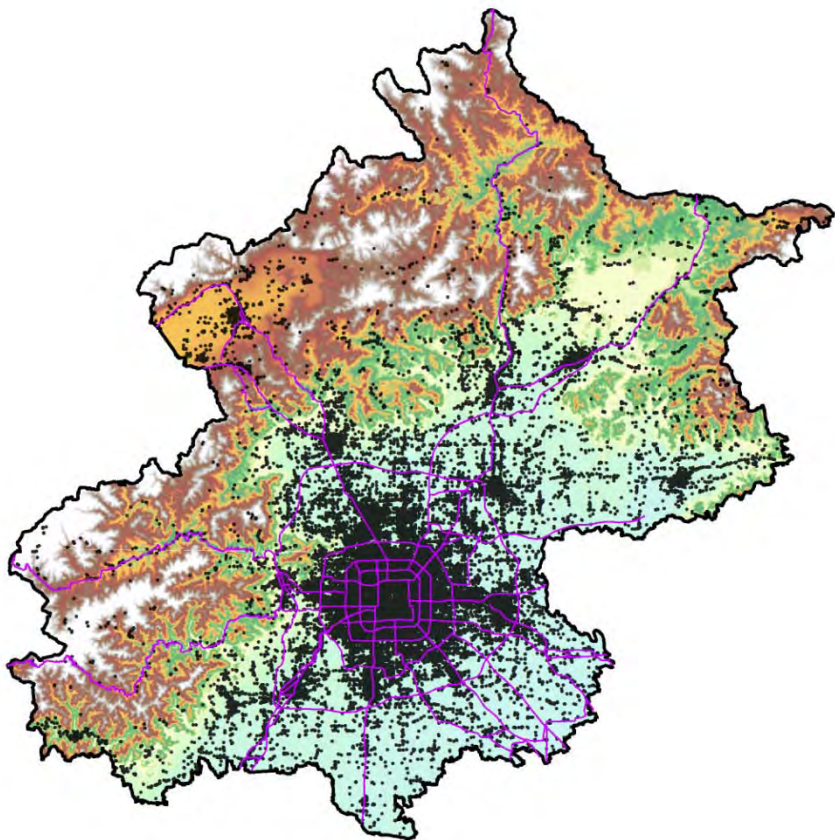
- 总量：约800万个兴趣点
- 优点：可统计POIs上check-in的数量和签到人数，照片数量等人类活动因子。



单位格网微博签到数量

# 点评网POIs

- 优点：UGC，1200万POIs，100多属性列(包括更多信息)





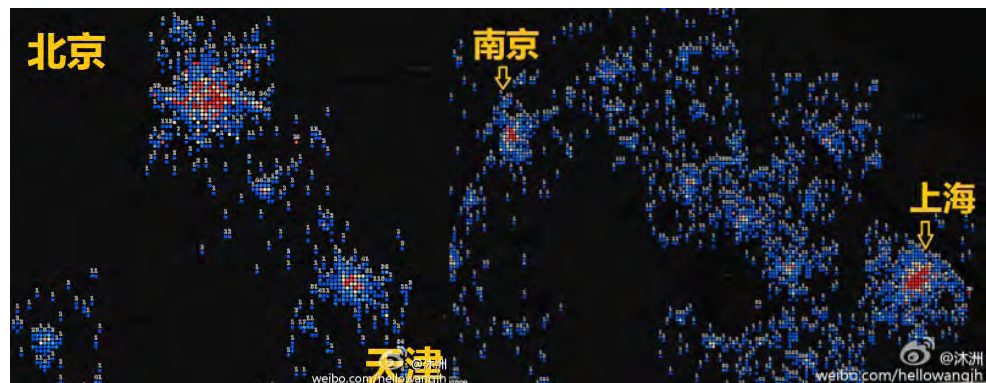
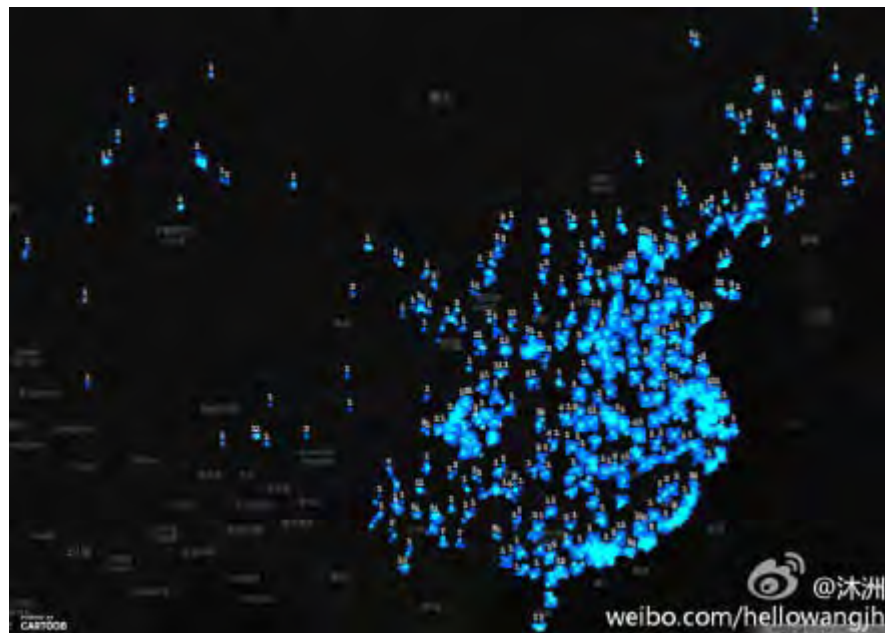
# 微博签到信息分析与挖掘

## □ 研究目的

- 社交网络空间大数据挖掘
- 人类活动空间分布模式
- 人的时空间行为研究

## □ 数据源

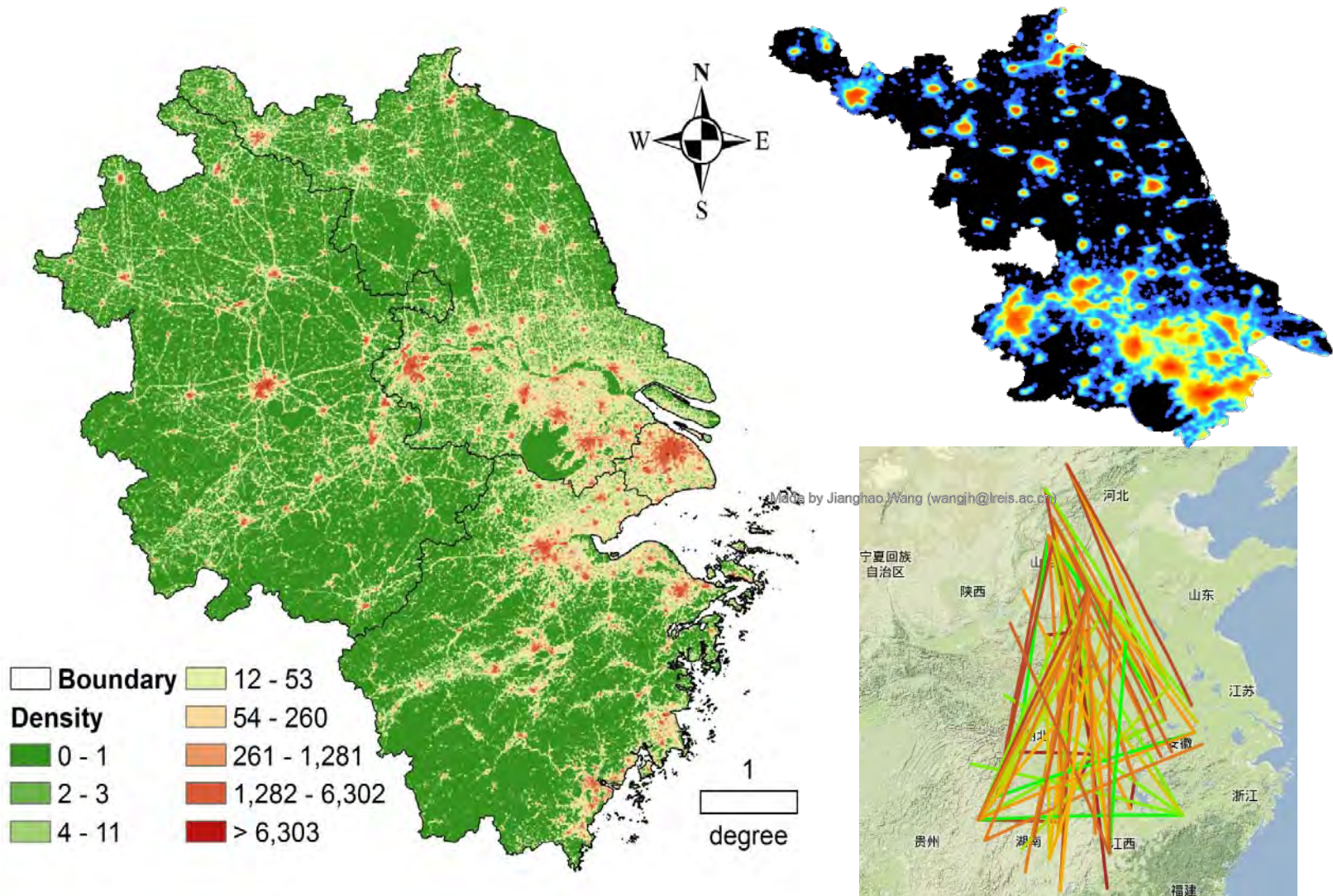
- 新浪微博checkin
- 2亿条签到记录
- 100多个属性描述人的属性和微博的信息



每小时微博签到数量空间分布

# 微博签到信息分析与挖掘

江苏省2010年DMSP/OLS灯光指数



长三角Weibo checkin密度图

中部城市城际联系与影响范围

# 微博签到信息分析与挖掘



Crandall et al. 2009. Mapping the world's photos.

# 总结与展望

---

- VGI和UGC将是Web2.0时代地理空间数据的重要来源
- 时空（大）数据分析与可视化是一把利器
- BCL将持续关注和分享
  
- 展望1：挖掘挖下去
- 展望2：研究走出去
- 展望3：行业用起来

# Thanks!

## Q & A

 沐洲 wangjh@lreis.ac.cn