城乡规划系列课程之大数据在规划设计中应用

2016年8月

城市众包数据分析框架研究

李栋 博士

北京清华同衡规划设计研究院 技术创新中心

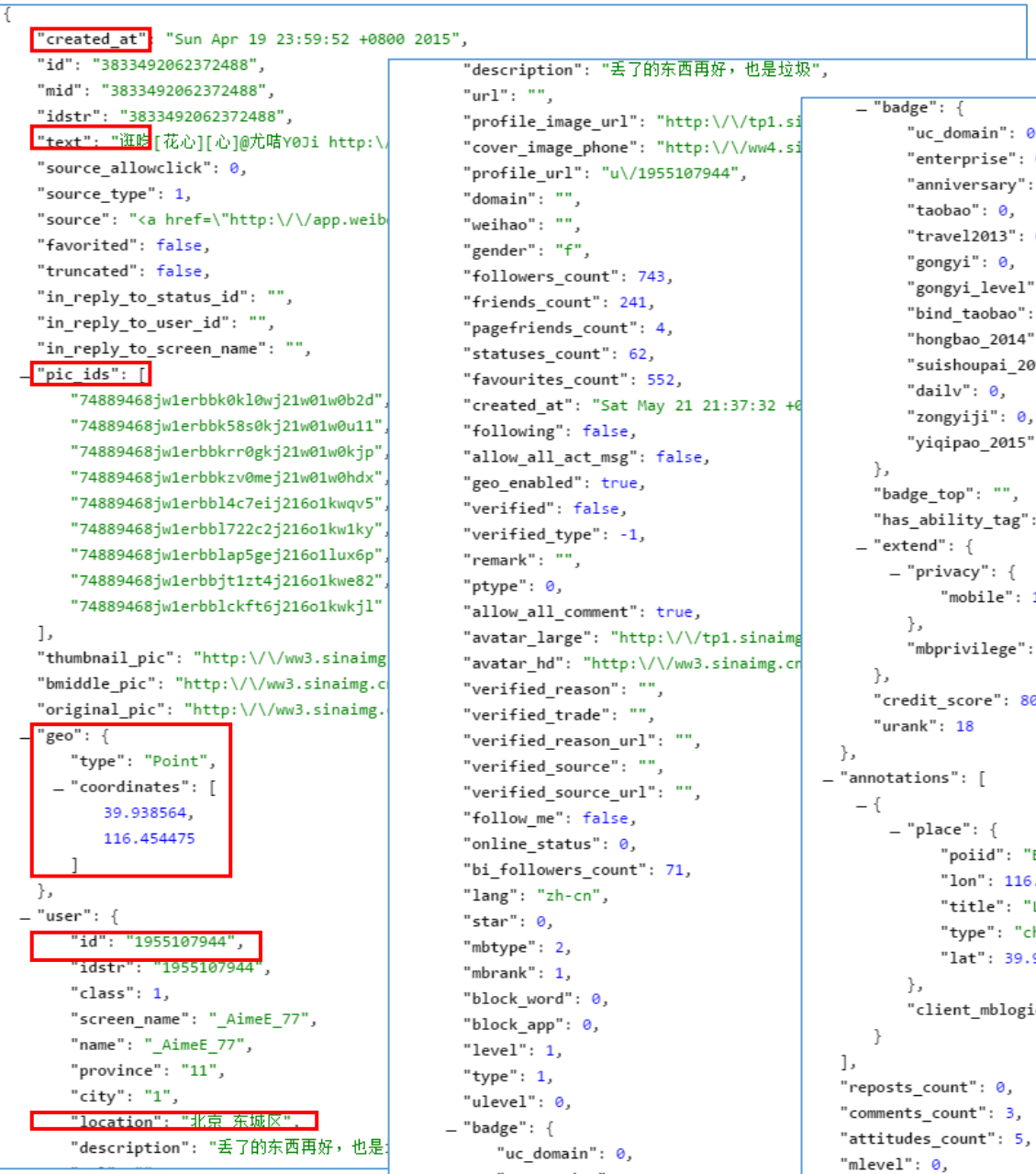
内容

- 起因
- 分析框架主要内容
- 两个示例
- 一点思考

起因

- 基于可获得的数据开展分析
  - 字段的属性、内容

数据属性和内容：以位置微博为例



起因

- 基于可获得的数据开展分析
  - 手机信令、浮动车轨迹

COLUMN	TYPE	DESCRIPTION
MSID	VARCHAR2（100）	移动用户唯一识别号
TIMESTAMP	NUMBER（17）	信令的时间戳，单位：毫秒（ms）
LAC	NUMBER（5）	位置区编号
CELLID	NUMBER（6）	LA内基站小区编号
EVENTID	NUMBER	信令事件类型

字段名	字段含义
序号	记录的编号，对每天所保存的记录进行排序编号，为数字型字符
调度中心ID	浮动车所属调度中心的编号，使用不超过4个ASCII字符，可变长
车辆ID	sim卡号形式的11位数字型字符
时间标签	使用GMT时间格式共14个数字字符，格式为（YYYYMMDDHHMMSS），包含年月日时分秒。
WGS84经度	最长为11个数值型字符，精度为（xxx.xxxxxx），变长，单位度。
WGS84纬度	最长为10个数值型字符，精度为（xx.xxxxxx），变长，单位度。
速度	单位为公里/小时，整数，为1-3个数值型字符，精确到个位。
方向	单位度，整数，为1-3位数值型字符，0~360角度，以正北为0度，顺时针旋转。
状态	浮动车的运营状态，为1位数字，数值型字符，包括5个状态，分别是： 0：空载；1：满载；2：驻车；3：停运；4：其他；
事件	浮动车的运营事件，为1数字，数值型字符，包括5个事件，分别是： 0：客人下车；1：客人上车；2：锁车门；3：开锁车门；4：其他；
高度	为1-3位数字，数值型字符

分析框架主要内容

- 城市众包数据分析框架，四个方向

1. 分布格局
2. 移动轨迹
3. 语义认知
4. 社会关系

A Crowd-Sourced Data Based Analytical Framework for Urban Planning

Li Dong, Long Ying

Abstract Aimed at the challenges faced by the current urban development and urban planning, along with the research opportunities brought by "big data," this paper proposes an analytical framework based on crowd-sourced data for urban planning by reviewing related literature and practice. The framework is mainly oriented towards three major requirements of analysis in urban planning: the physical spaces, the user communities, and the social relationships. This analytical framework can be regarded as a preliminary attempt for future data-intensive applications in urban planning and assessment.

Keywords location-based services (LBS); crowd-sourced; natural language processing; quantitative urban study

1. Introduction

1.1 Challenges for urban development and urban planning

After 30 years of rapid urban development, China currently has an urbanization rate of more than 50%. Many negative impacts of urbanization, i.e., so-called "urban diseases," are emerging, including traffic congestion, excessive population concentration, heavy consumption of resources, environmental pollution, poor safety and disaster prevention, and so on. Urban diseases are testing the capabilities of urban management and sustainability, with a tendency of spreading out from mega cities to less developed small and medium-sized cities. Since the *1996 Istanbul Declaration on Human Settlements* proposed a statement of "making human settlements safer, healthier, and more livable, equitable, sustainable and productive," the gap between the goal and the reality has become even larger, and the living quality of urban residents has faced serious challenges.

Meanwhile, as a leading player, urban planning itself also faces many threats. Since the conditions and status of cities have changed significantly and are becoming extremely complex, the effectiveness of traditional tools for urban planning such as technical standards and analytical methods has been declining. Planners' abilities of analyzing, diagnosing, and assessing the status of urban development are doubted, let alone their abilities to guide future development, to implement proactive solutions to issues in real world, or to enhance the feasibilities of targets. One of the reasons is the insufficiency of traditional data and analysis. Planners have to extract information based on a relatively small amount of data and try to seek an overall conclusion. In other words, the mode of analysis requires an effective transition from fragmented, low-frequency statistics to a complex overview picture of a city, as well as a shift from a rough aggregation of figures to a finer profiling of individuals (Zhang, 2014).

1.2 The "big data" wave

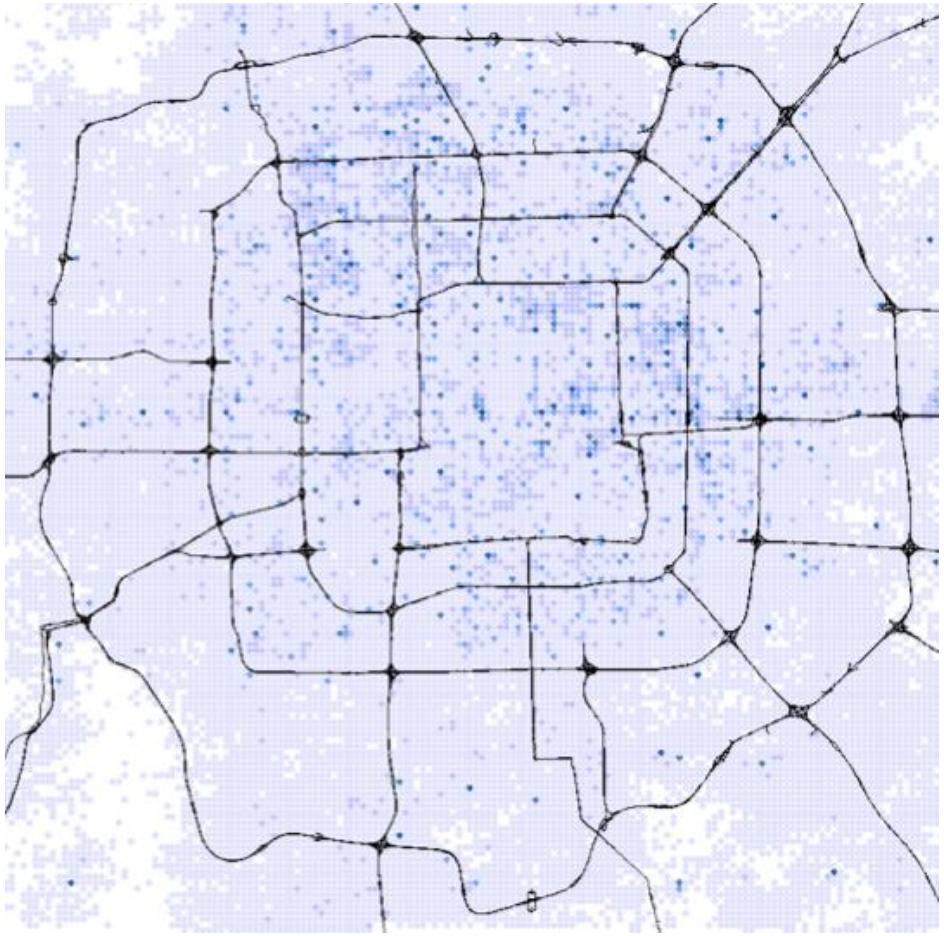
With the booming of information and communications technology (ICT), incredible amounts of data are produced and available in our cities and on our planet, via various chips, sensor networks, positioning systems, mobile communications, and high-performance computing and storing technologies. Urban daily life, such as transportation and recreation, has also been impacted by the evolving ICT. For example, Baidu.com processes 6 billion search requests every day; over 500 million people talk via WeChat app and compose over 100 billion relationships online; the bus-pass card in Beijing is used up to 20 million times per day, and so on. Human and various types of operation sensors will produce more and more data. According to the estimates from the white paper of the International Data Corporation (IDC) (Gantz and Reinsel, 2009), for every 18 months the volume of new data is equal to the sum of all data in the past, while the total amount of data generated each year will reach 40 ZB by 2020.

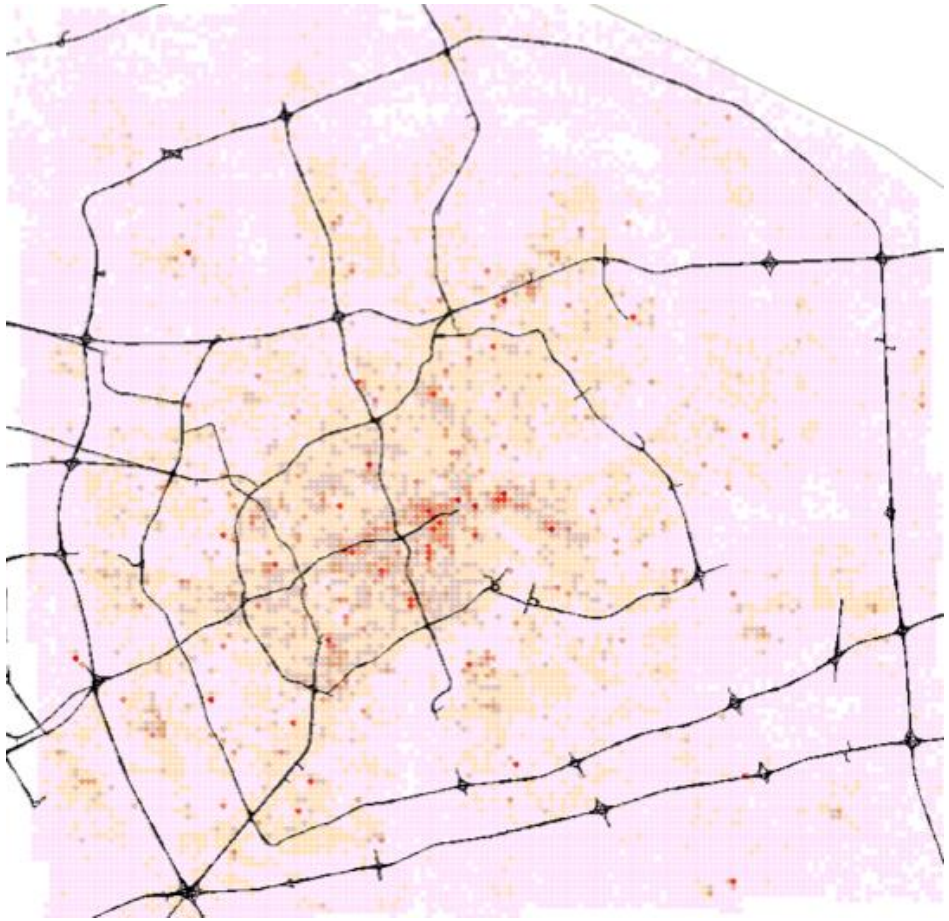
To respond to the "big data" wave and further reveal its impacts on cities and human society, the academic community has carried out a considerable amount of research works, represented by two special issues published respectively by *Nature* (2008) and *Science* (2011). Data deposits will be increasing gradually from the level of GB (GigaByte) to PB (PetaByte) and EB (ExaByte), while effective meta-analyses on these data with complex

Dong LI, Ying LONG. A Crowd-Sourcing Data Based Analysis Framework for Urban Planning[J]. China City Planning Review. 2015, 24 (1),

分析框架主要内容

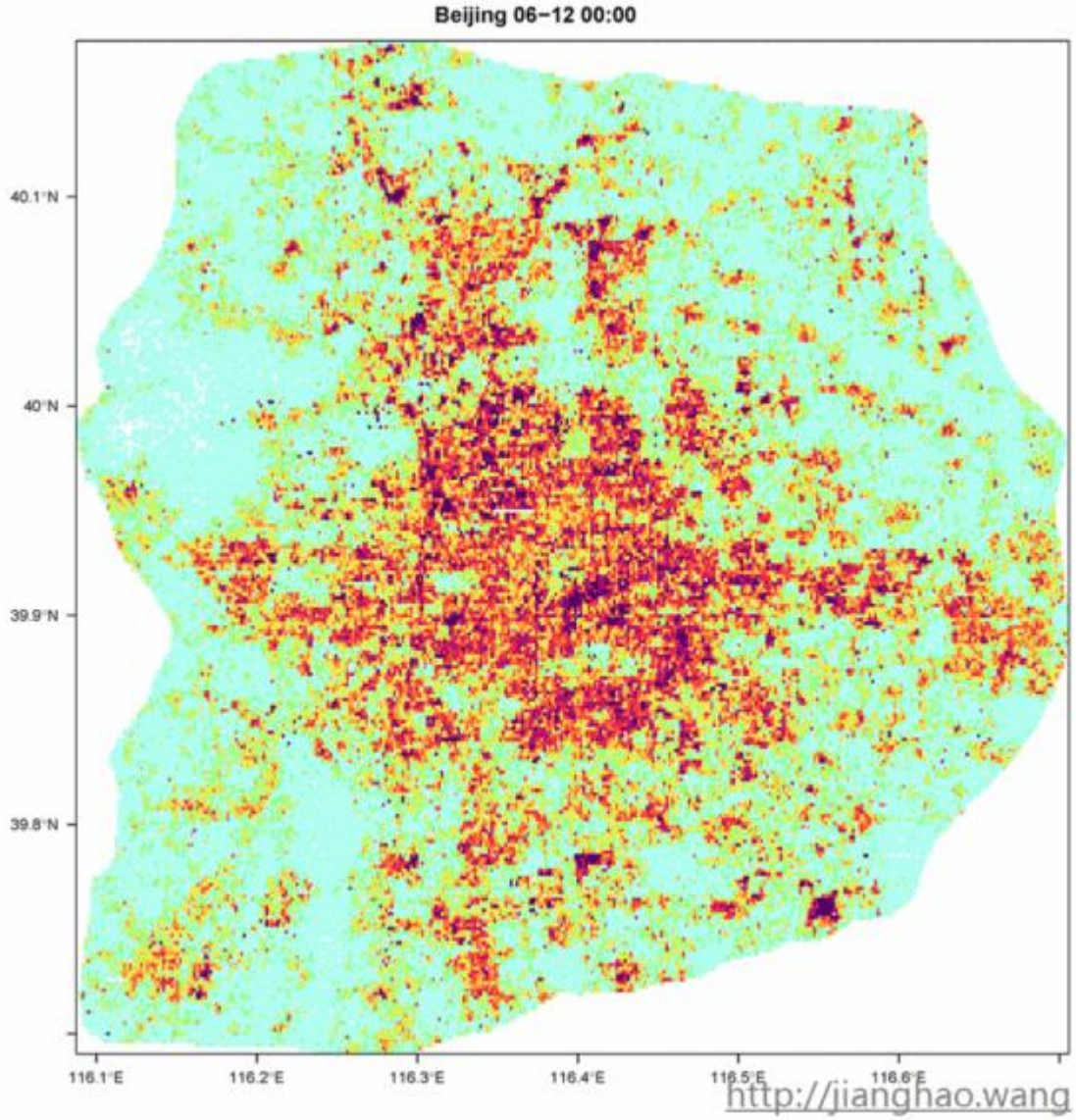
1. 分布格局





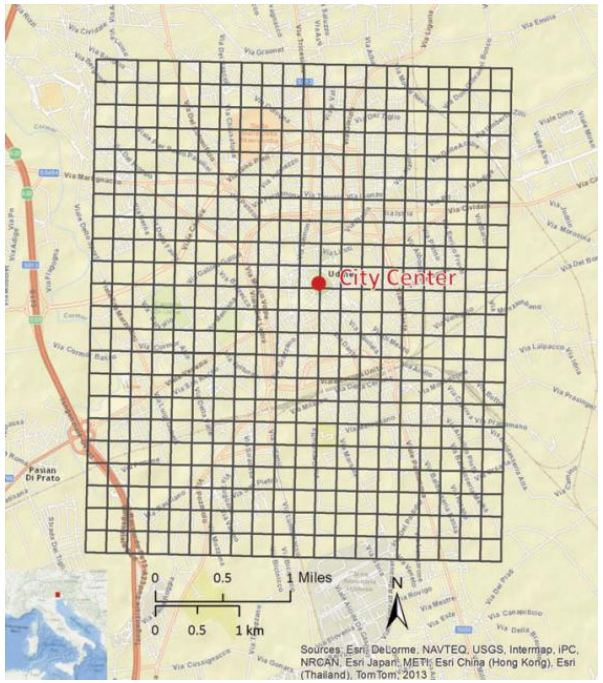
分析框架主要内容

1. 分布格局



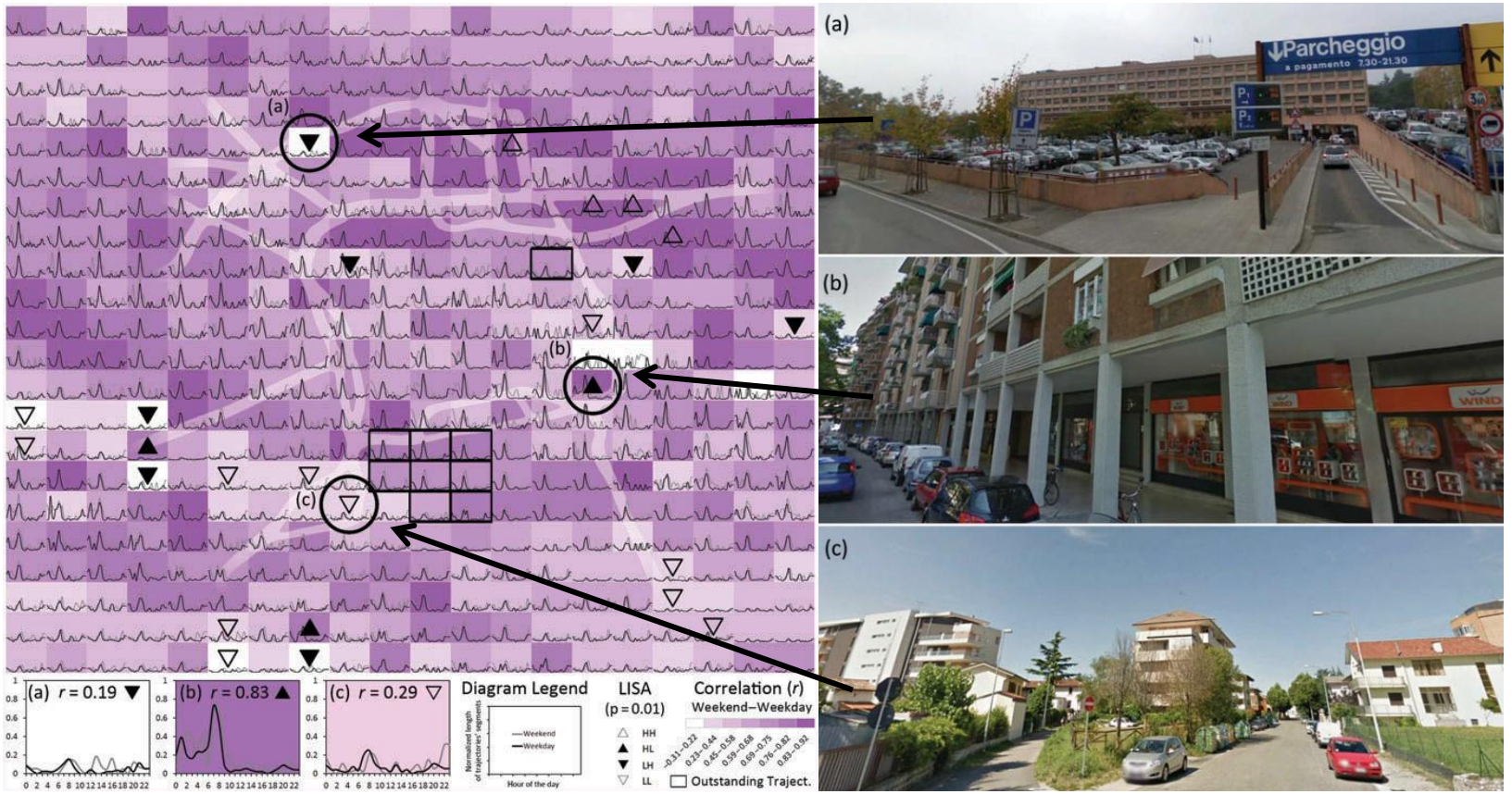
分析框架主要内容

时间格局：分时段对比



意大利乌迪内Udine

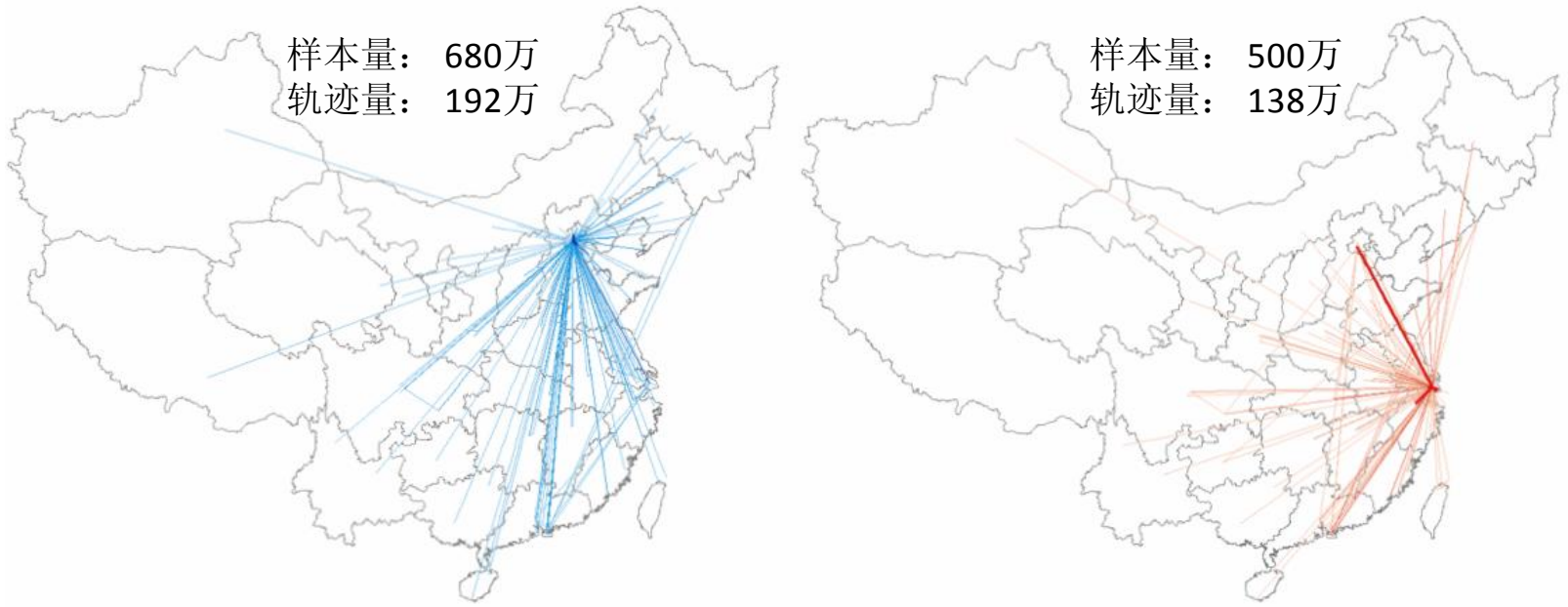
平日-周末：时序相关性+空间聚集性

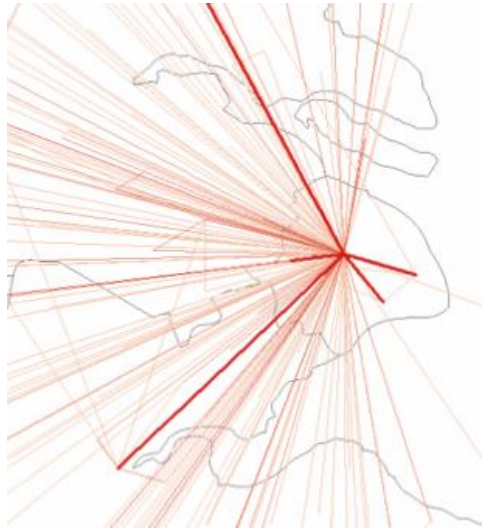


Guenther Sagla*, Eric Delmelleb & Elizabeth Delmelleb, *Mapping collective human activity in an urban environment based on mobile phone data,* CaGIS

分析框架主要内容

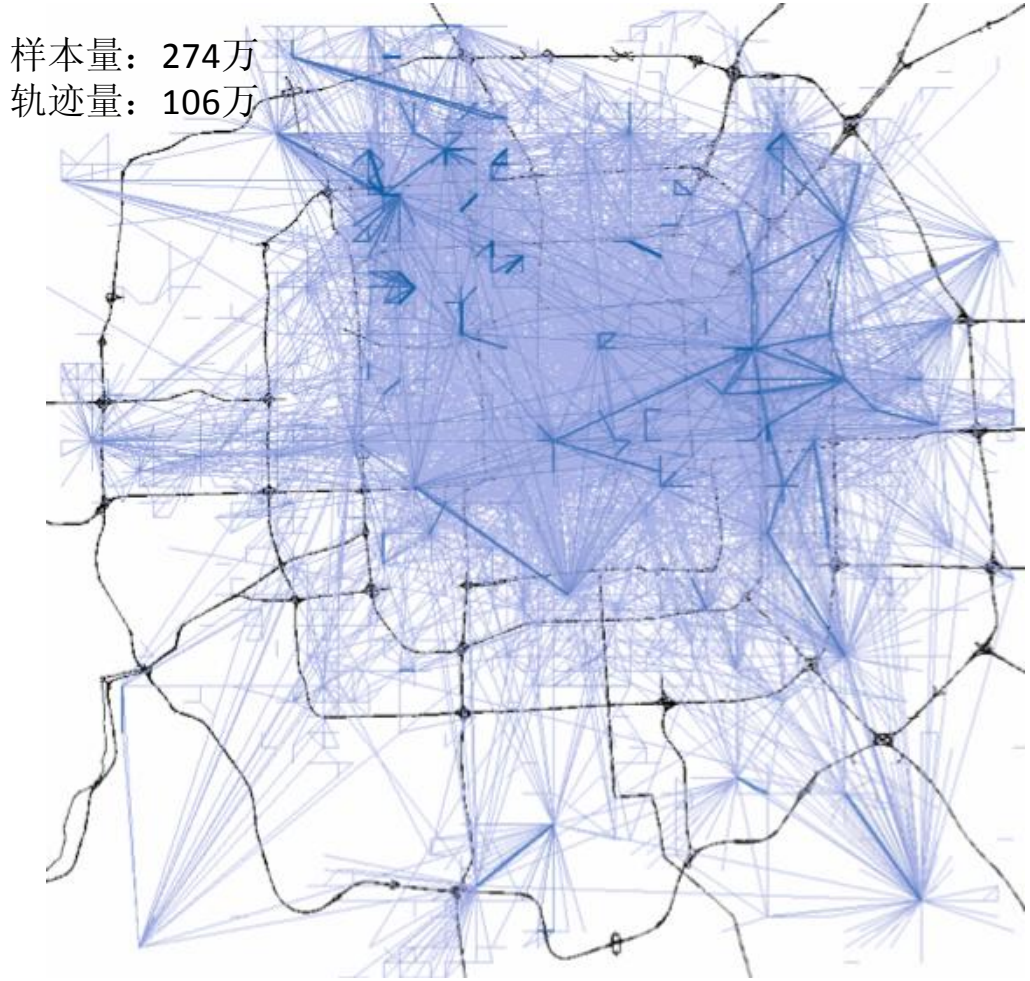
- 北京、上海区域联系格局的对比
  - **统计范围：全国**
  - **对比尺度：区县**

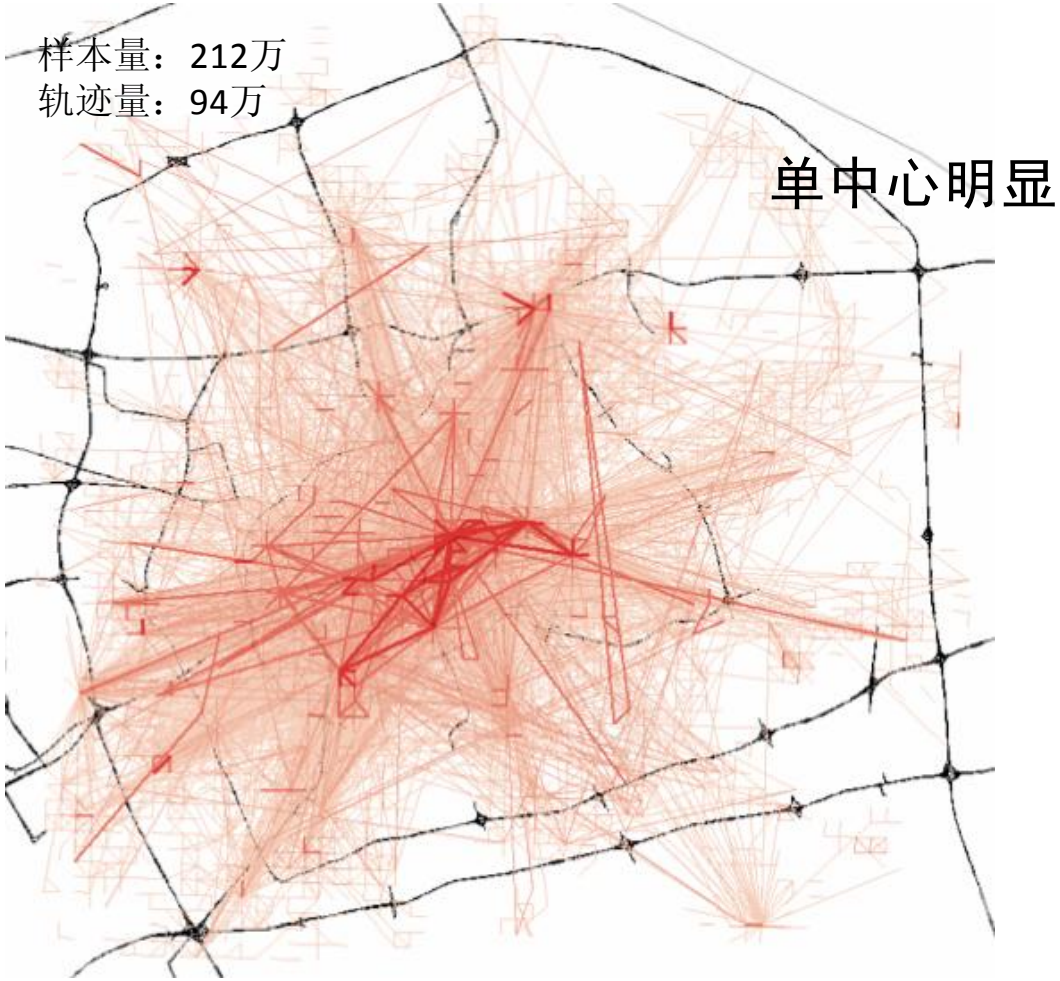




分析框架主要内容

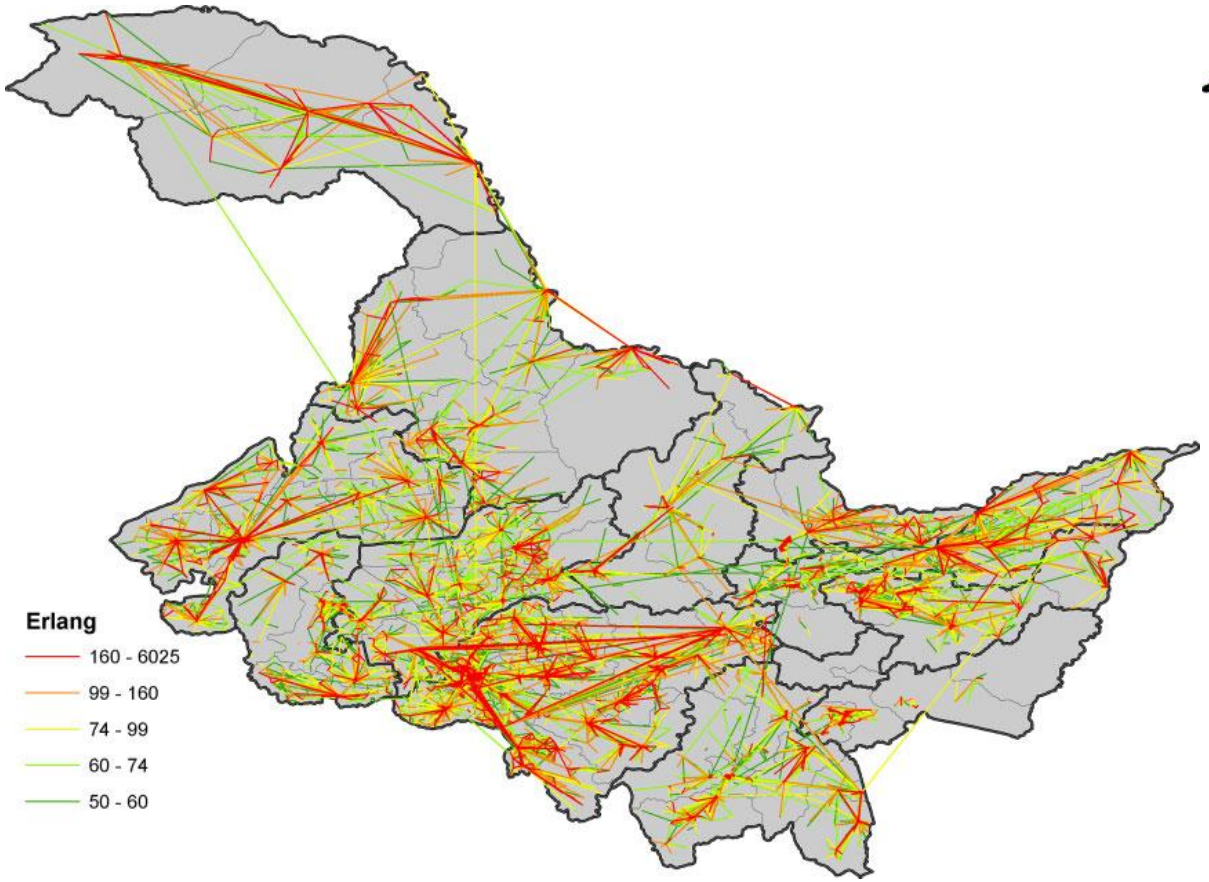
- 北京、上海城区联系格局的对比，**对比尺度： 500m均匀网格**

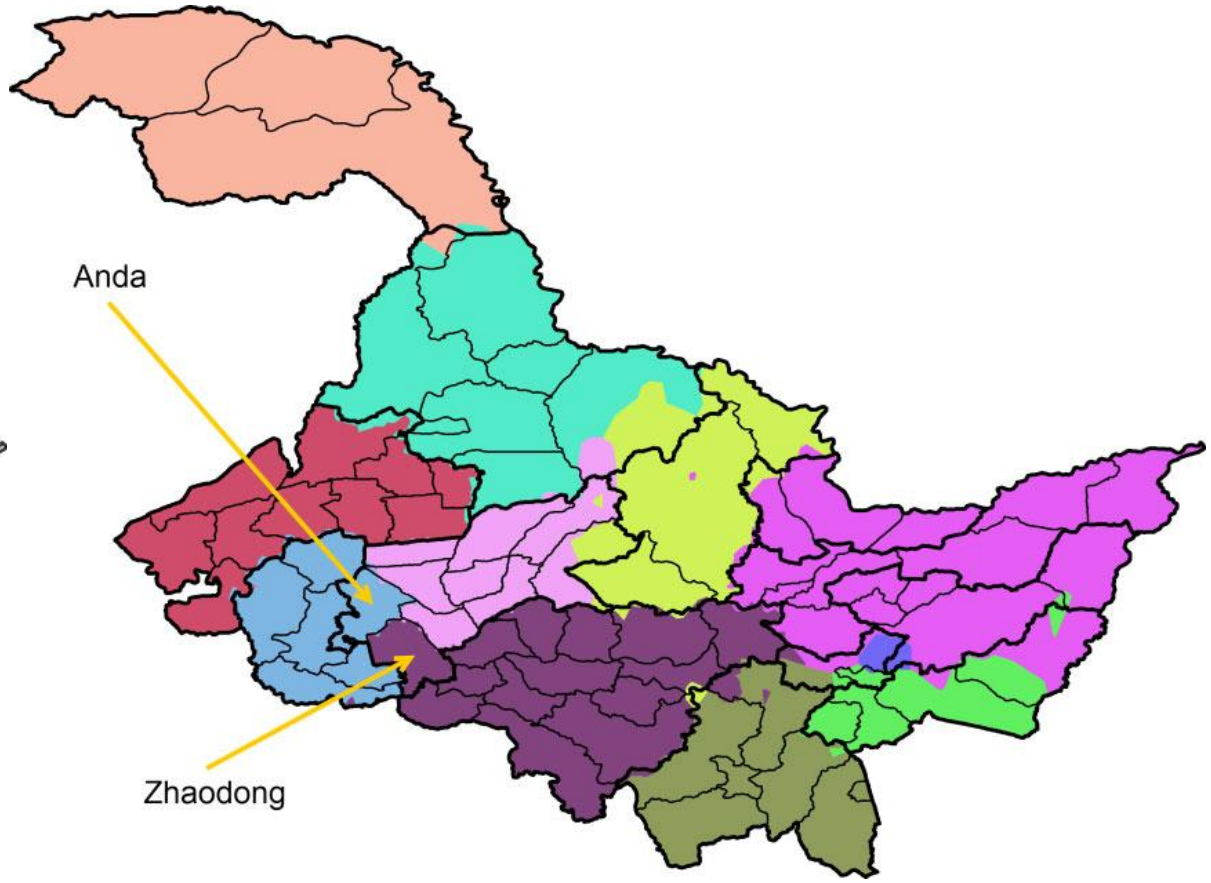




分析框架主要内容

- 社会关系





Guanghua Chi, Jean-Claude Thill, Daoqin Tong, Li Shi, Yu Liu (2014). Uncovering regional characteristics from mobile phone data: a network science approach. Papers in

分析框架主要内容：语义认知

阅读，并理解：词汇

“今天去中关村修电脑，完事附近逛逛音像店，想找找有没有健哥的专辑。虽然没找到专辑，但是在新中关购物中心旁的步行街却看到了健哥的手印，而且是一低头就是健哥。哈哈”

@{"lng":116.3162,"lat":39.9978}

@1426379400 (2015-03-15 08:30:00)

名词

动词

Data: Weibo

阅读，并理解：话题

“北京从6.1日起电动车不限行，以前还担心路上更堵。早上听新闻，说电动车的充电桩不多，影响买车。从这看，政府还是想缓解交通拥堵的，电动车你可以买，但是能上路的不多，口碑有了，销售收入有了，对交通影响还不大，政府这作法一举多得啊”

话题1 未分类： 20.0%

话题2 广告： 33.4%

话题3 汽车： 46.6%

@{"lng":116.4904,"lat":39.8869}

@1427976300 (2015-04-02 20:05:00)

Data: Weibo

阅读，并理解：情绪

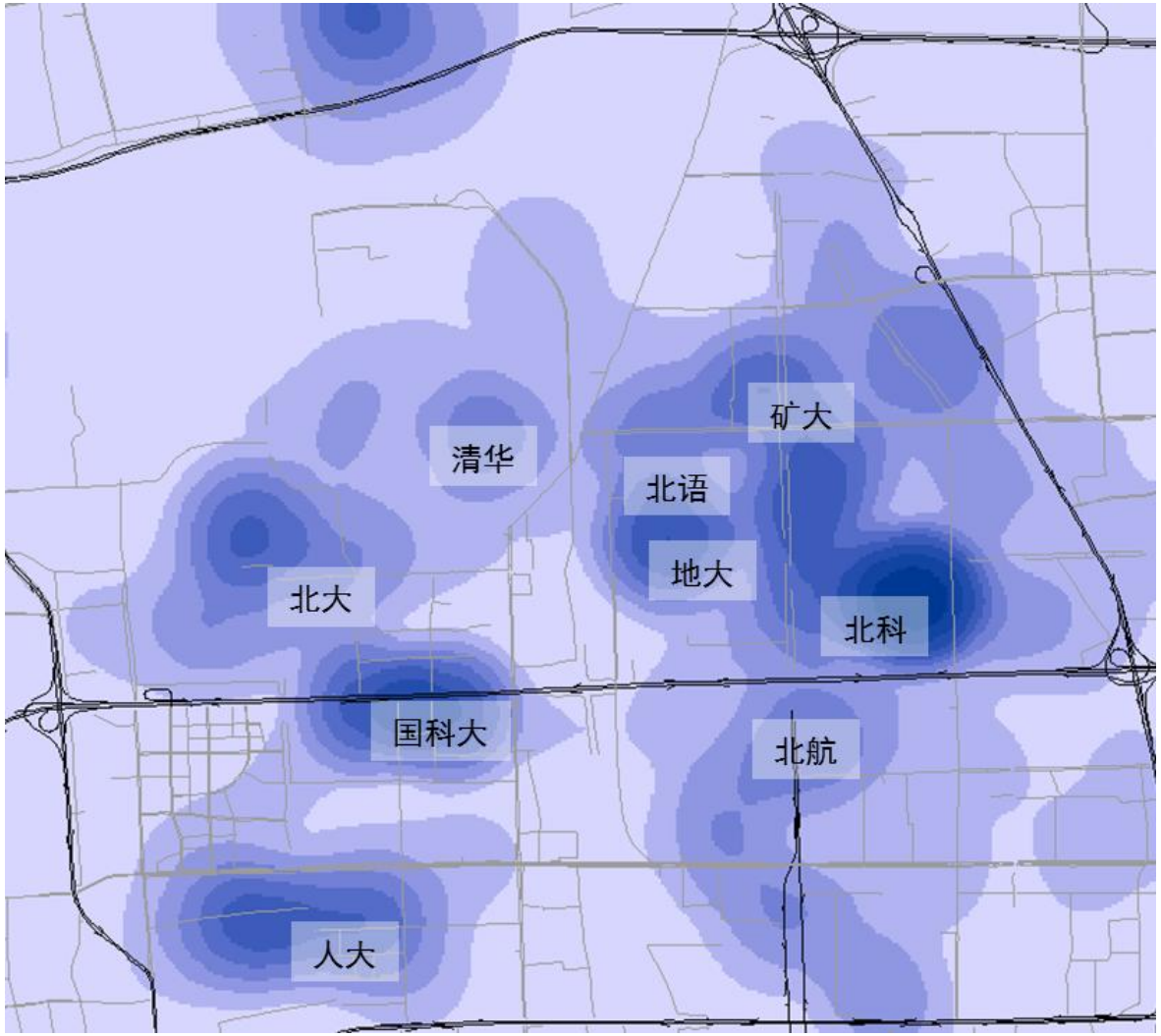
“今天的空气非常不错”

“当习近平夫妇来到工业园时，卢卡申科亲自来到下车处迎接。习近平夫妇同卢卡申科共同参观工业园沙盘，认真听取园区负责人介绍，询问两国企业合作和园区企业生产等情况”

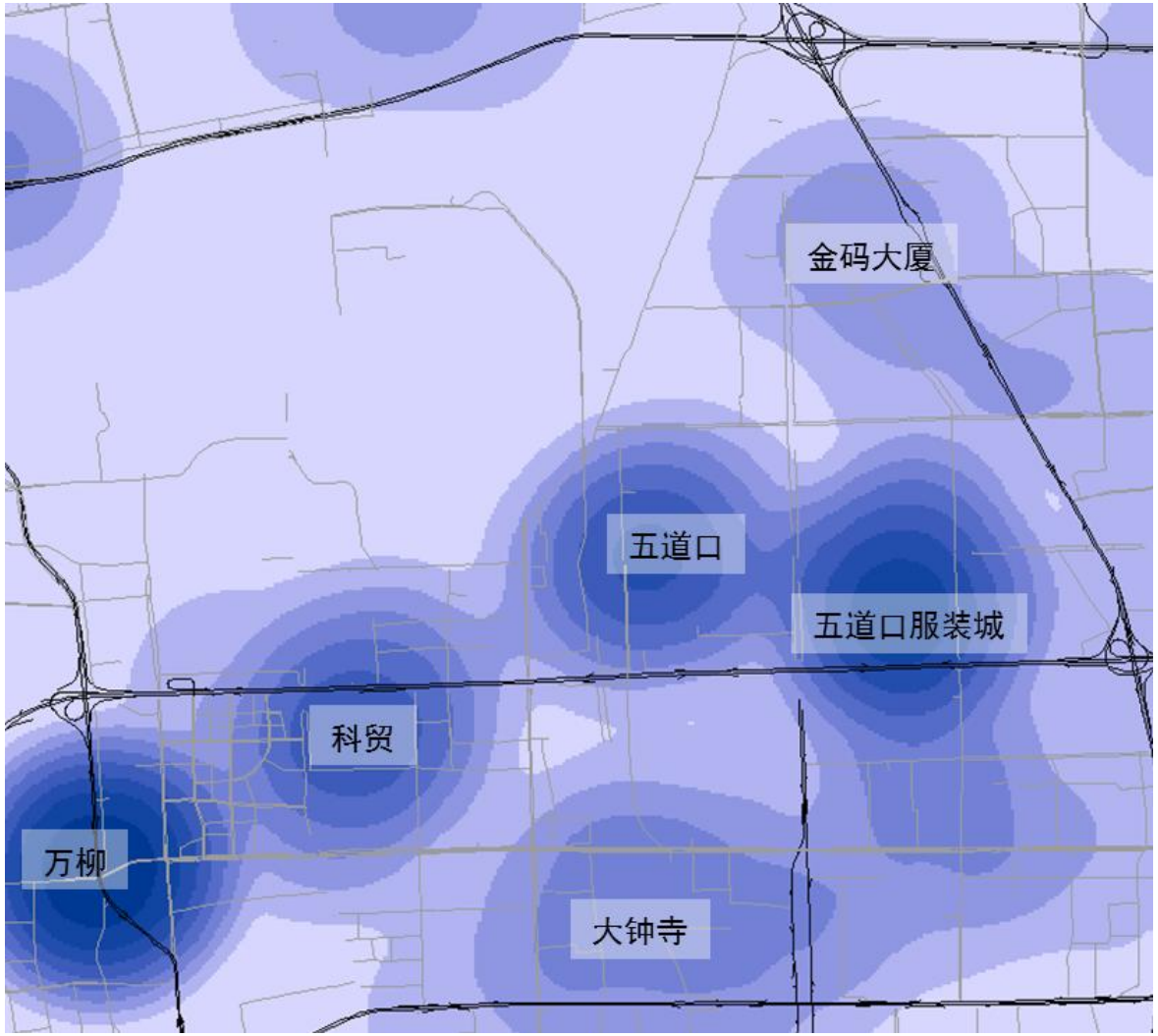
“中国家庭收入差距明显，收入最多的20%的家庭和收入最少的20%的家庭相差19倍左右，流动家庭和留守家庭已经成为家庭的常规模式”

同一个中关村、不一样的生活方式

“**大学**”名词n：分布热图

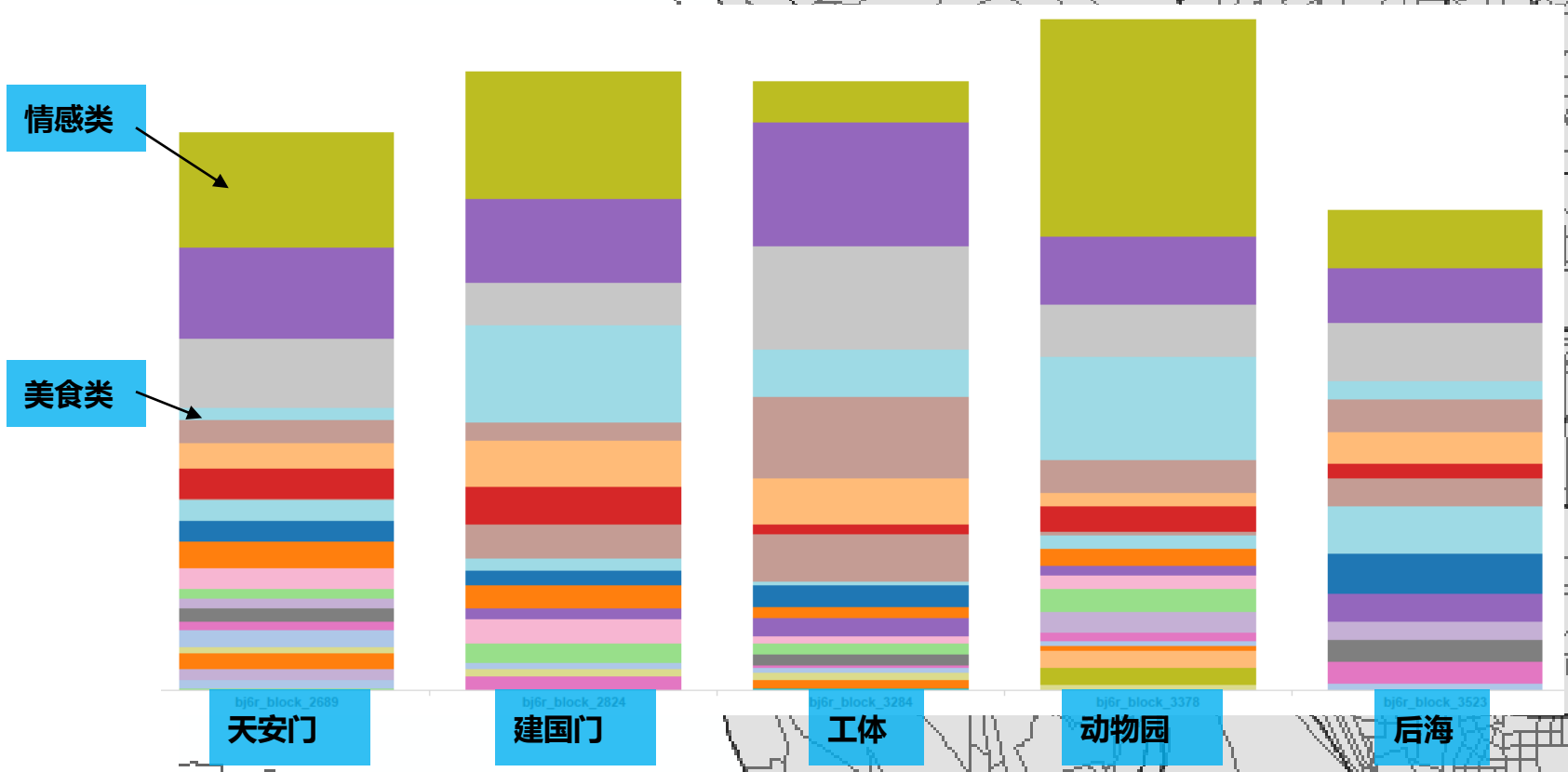


“**购物**”动词v：分布热图



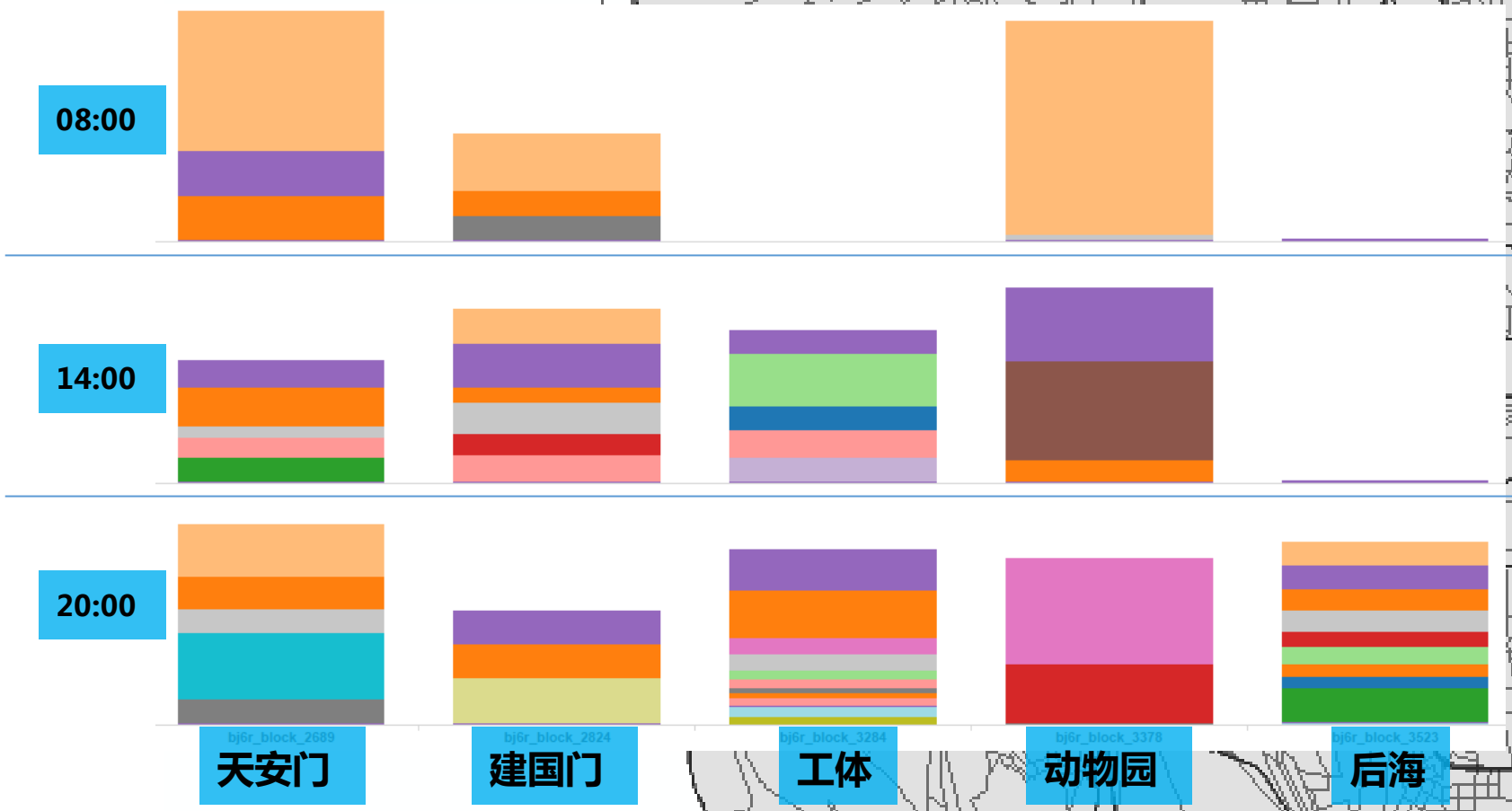
Data: Weibo

不同地块语义主题的差异



Data: Weibo, OSM

不同地块语义主题的差异：不同时段



Data: Weibo, OSM

阅读，并理解：图像


```
"classes": [
    • "nobody",
    • "road",
    • "landscape",
    • "travel",
    • "vehicle",
    • "city",
    • "water",
    • "river",
    • "transportation",
    • "architecture",
    • "tree",
    • "outdoors",
    • "middle east",
    • "highway",
    • "indochina",
    • "energy",
    • "north america",
    • "industry",
    • "traffic",
    • "bridge"
],
```

```
"probabilities": [
    • 0.9956164360046387,
    • 0.9825328588485718,
    • 0.979576826095581,
    • 0.9788038730621338,
    • 0.9557791948318481,
    • 0.9533190131187439,
    • 0.9472818374633789,
    • 0.9342042207717896,
    • 0.9326856136322021,
    • 0.9322887659072876,
    • 0.9165146350860596,
    • 0.8839698433876038,
    • 0.8822923898696899,
    • 0.8820817470550537,
    • 0.8766086101531982,
    • 0.8655034303665161,
    • 0.8607180118560791,
    • 0.843806803226471,
    • 0.8406920433044434,
    • 0.8345965147018433
]
```


Data: Weibo

阅读，并理解：图像

Car, transportation, travel


```
"classes": [
    • "car",
    • "transportation",
    • "travel",
    • "road",
    • "traffic",
    • "vehicle",
    • "water",
    • "automobile",
    • "tourism",
    • "speed",
    • "street",
    • "horizontal",
    • "flood",
    • "city",
    • "taxi",
    • "tree",
    • "auto",
    • "drive",
    • "nature",
    • "asphalt"
],
```

```
"probabilities": [
    • 0.9982941746711731,
    • 0.9931734204292297,
    • 0.9930871725082397,
    • 0.9909658432006836,
    • 0.9903783798217773,
    • 0.9878414869308472,
    • 0.9720317125320435,
    • 0.9533627033233643,
    • 0.9490107297897339,
    • 0.9484788179397583,
    • 0.9468980431556702,
    • 0.9452767372131348,
    • 0.939334511756897,
    • 0.9327230453491211,
    • 0.9229570627212524,
    • 0.9218120574951172,
    • 0.9155992269515991,
    • 0.9019625186920166,
    • 0.8980552554130554,
    • 0.8920489549636841
]
```

Data: Weibo

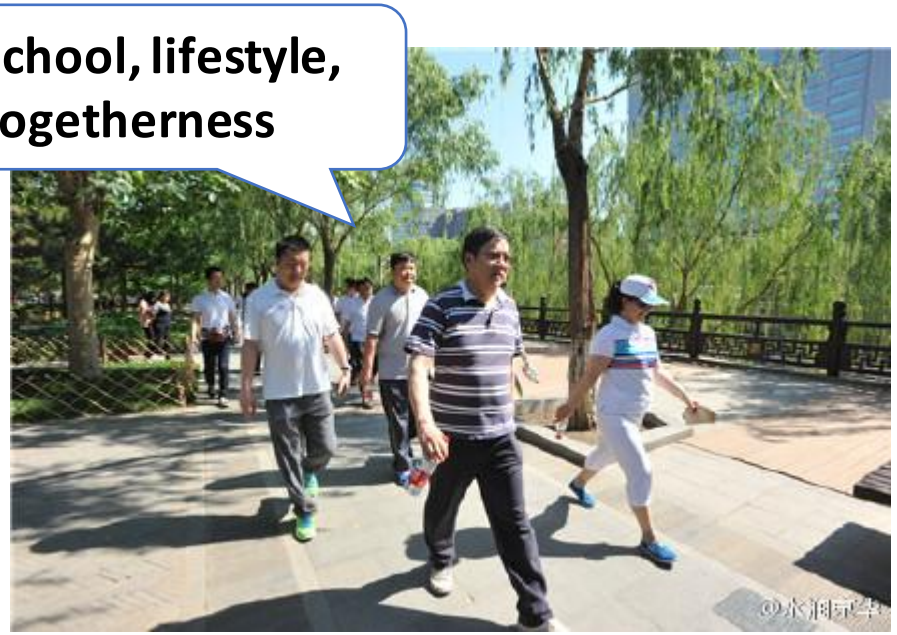
阅读，并理解：图像


```
"classes": [
    • "dog",
    • "pet",
    • "puppy",
    • "cute",
    • "outdoors",
    • "nature",
    • "summer",
    • "grass",
    • "sitting",
    • "small",
    • "park",
    • "yard",
    • "canine",
    • "friendly",
    • "mammal",
    • "funny",
    • "portrait",
    • "fur",
    • "animal",
    • "nobody"
],
```

```
"probabilities": [
    • 0.9944829940795898,
    • 0.9793223738670349,
    • 0.9763623476028442,
    • 0.9629213809967041,
    • 0.957022488117218,
    • 0.9402827620506287,
    • 0.9338744878768921,
    • 0.9336146712303162,
    • 0.9298191070556641,
    • 0.9178627729415894,
    • 0.9112381339073181,
    • 0.8879156708717346,
    • 0.8793382048606873,
    • 0.8657771944999695,
    • 0.8612759113311768,
    • 0.8590993881225586,
    • 0.8539232015609741,
    • 0.8477699756622314,
    • 0.8475475311279297,
    • 0.8225152492523193
]
```

Data: Weibo

阅读，并理解：图像

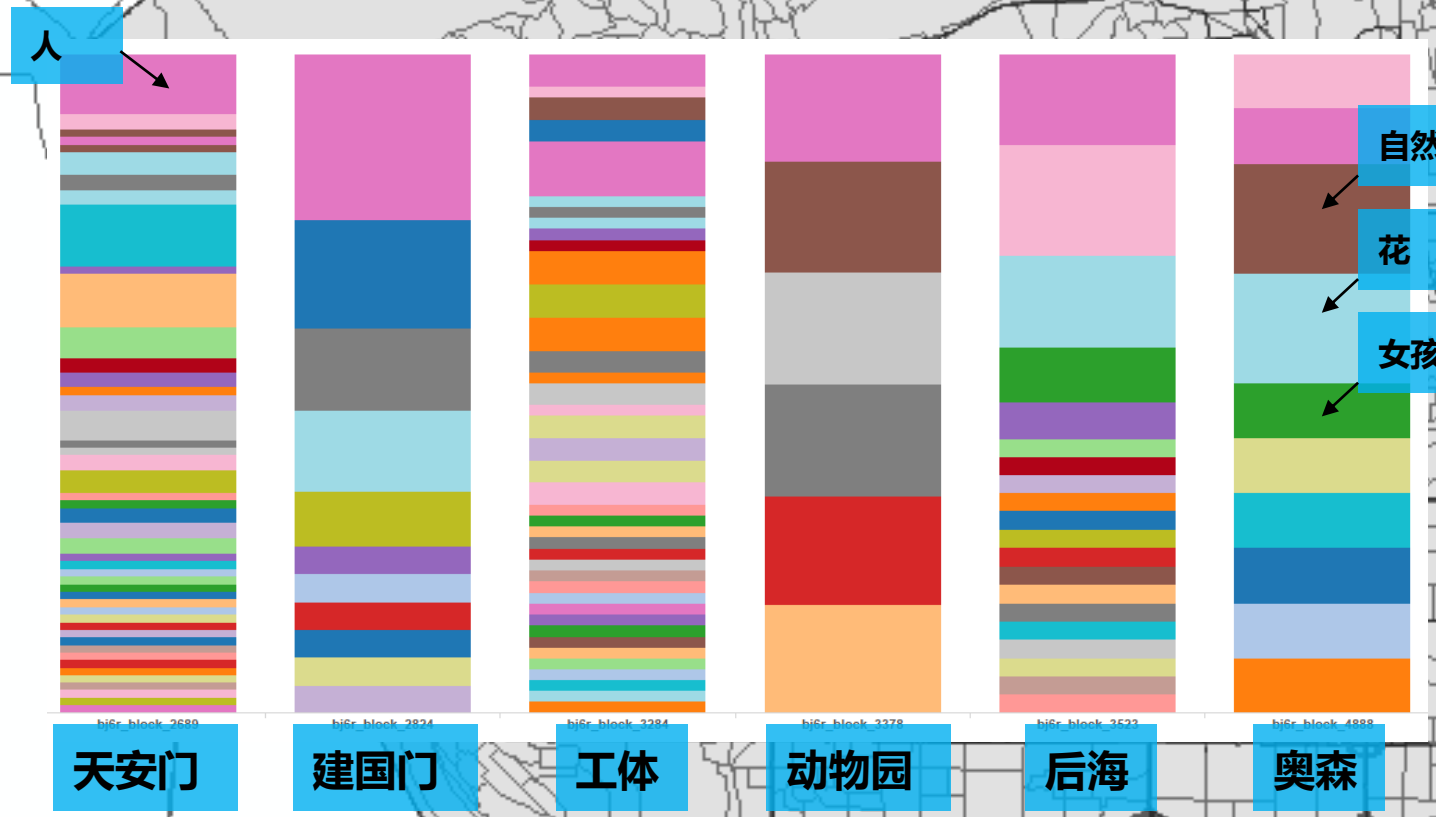



```
"classes": [
    • "school",
    • "lifestyle",
    • "togetherness",
    • "people",
    • "education",
    • "motion",
    • "friendship",
    • "daytime",
    • "fun",
    • "female",
    • "child",
    • "recreation",
    • "happiness",
    • "road",
    • "city",
    • "north america",
    • "leisure",
    • "family",
    • "enjoyment",
    • "politics"
],
```

```
"probabilities": [
    • 0.9909061193466187,
    • 0.9887098670005798,
    • 0.9805930852890015,
    • 0.9792243242263794,
    • 0.9758434295654297,
    • 0.9687234163284302,
    • 0.9673053026199341,
    • 0.9599980115890503,
    • 0.9592617750167847,
    • 0.9485963582992554,
    • 0.9463801383972168,
    • 0.9411329627037048,
    • 0.9400149583816528,
    • 0.9381043314933777,
    • 0.9364811182022095,
    • 0.9325109720230103,
    • 0.9296536445617676,
    • 0.9287199378013611,
    • 0.923075258731842,
    • 0.9210243225097656
]
```


Data: Weibo

不同地块图像主题的差异



Data: Weibo

示例1：潜在社会交往

城市里的人为什么不是完全理性的人？

如何度量潜在的社会交往？

——以及和空间/场所的关系？

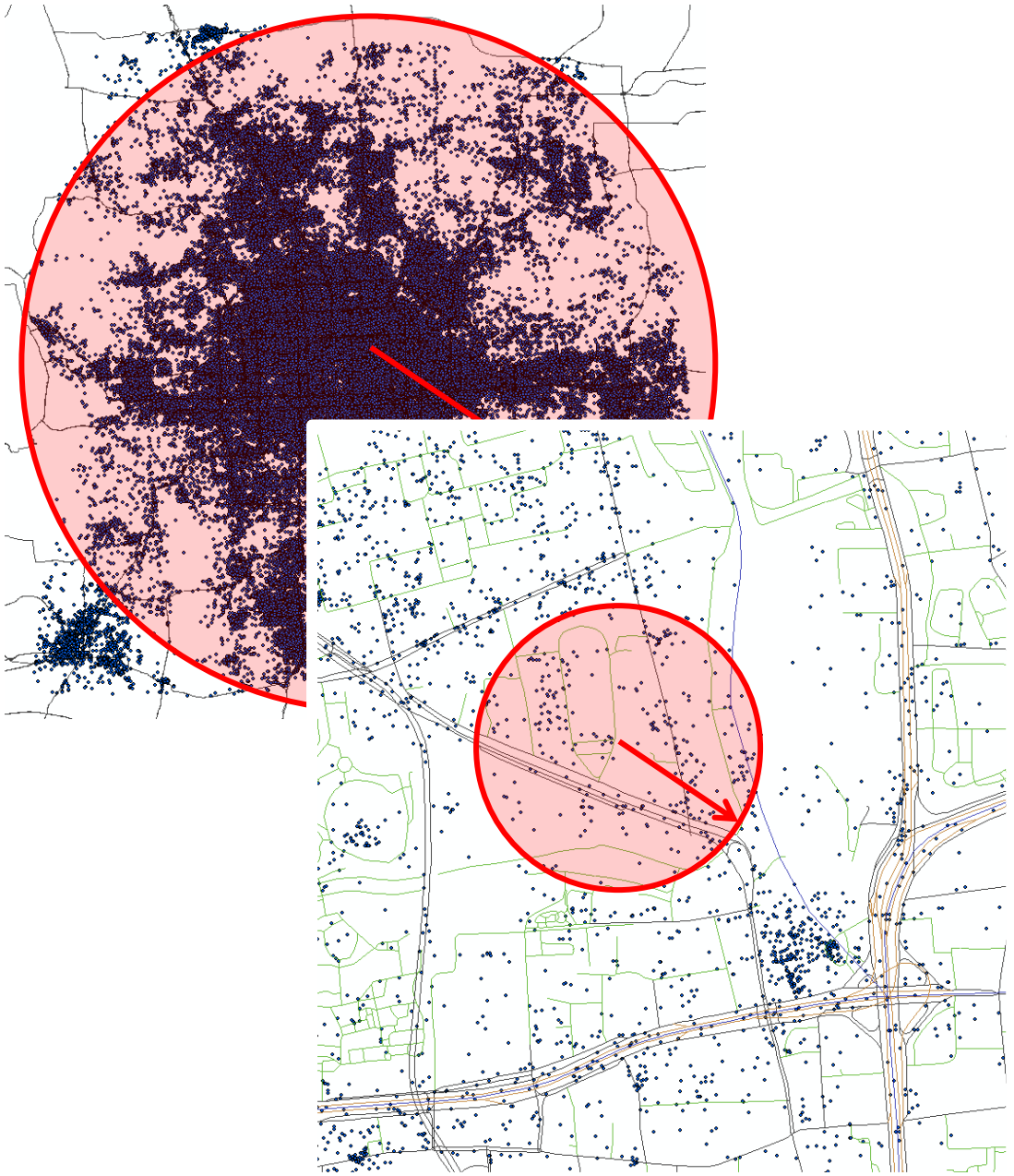
基于时空约束的共现搜索

全量：500000 * 500000 = 2500亿

$$PAIR_{ij}$$

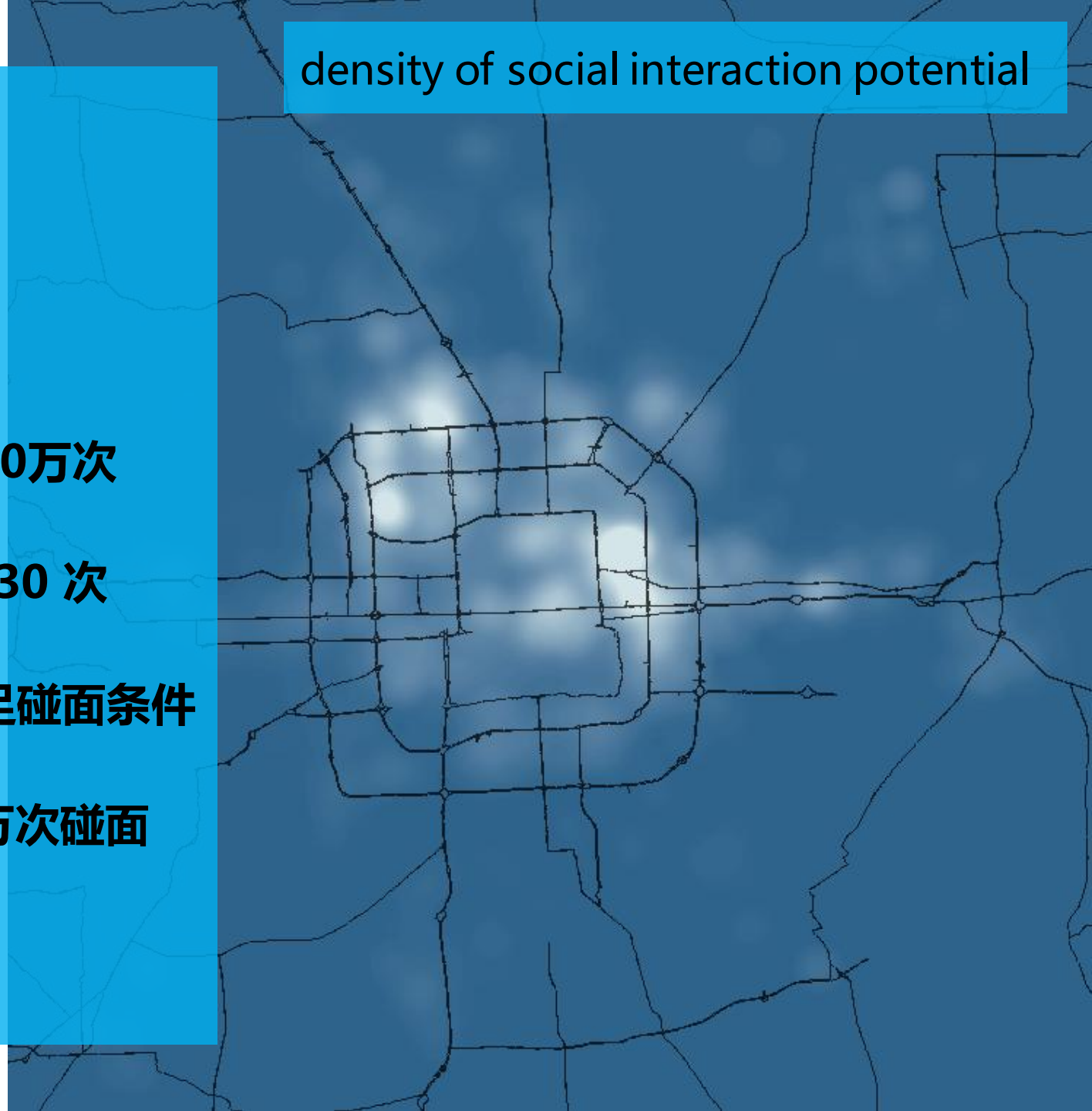
$$where\ dist_{ij} < 2\ km$$

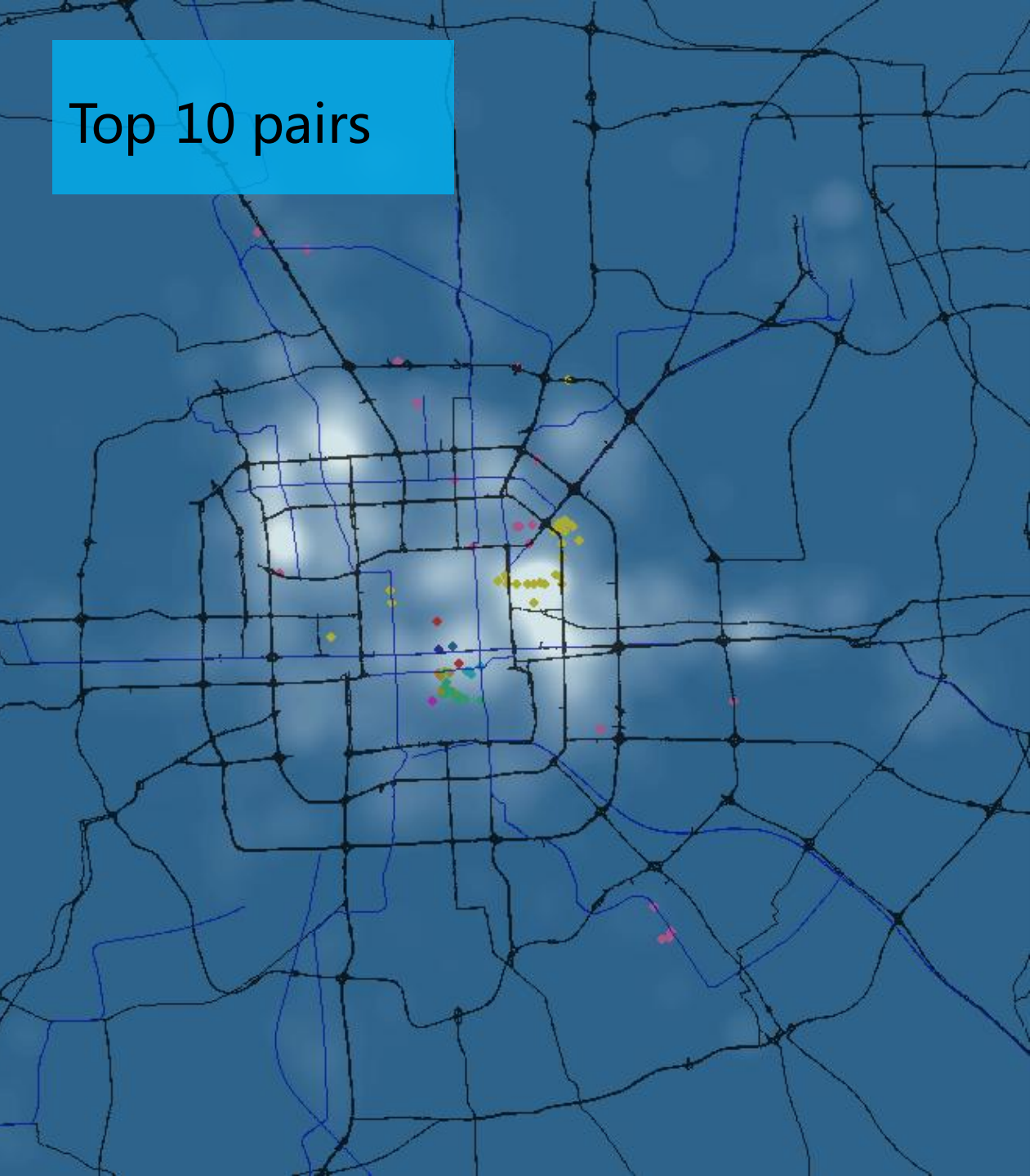
$$T_{diff} = |T_i - T_j| < 24\ h$$

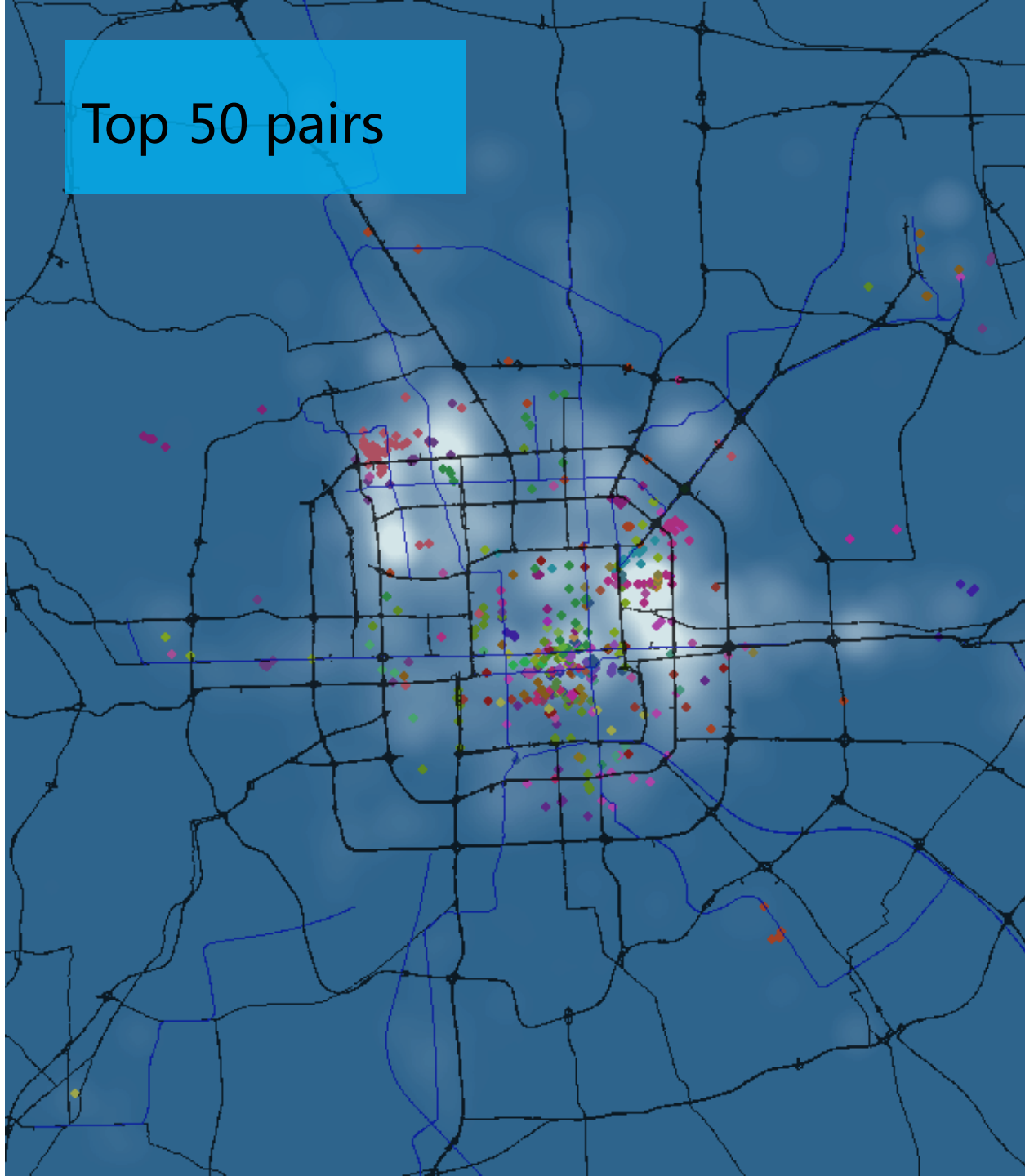


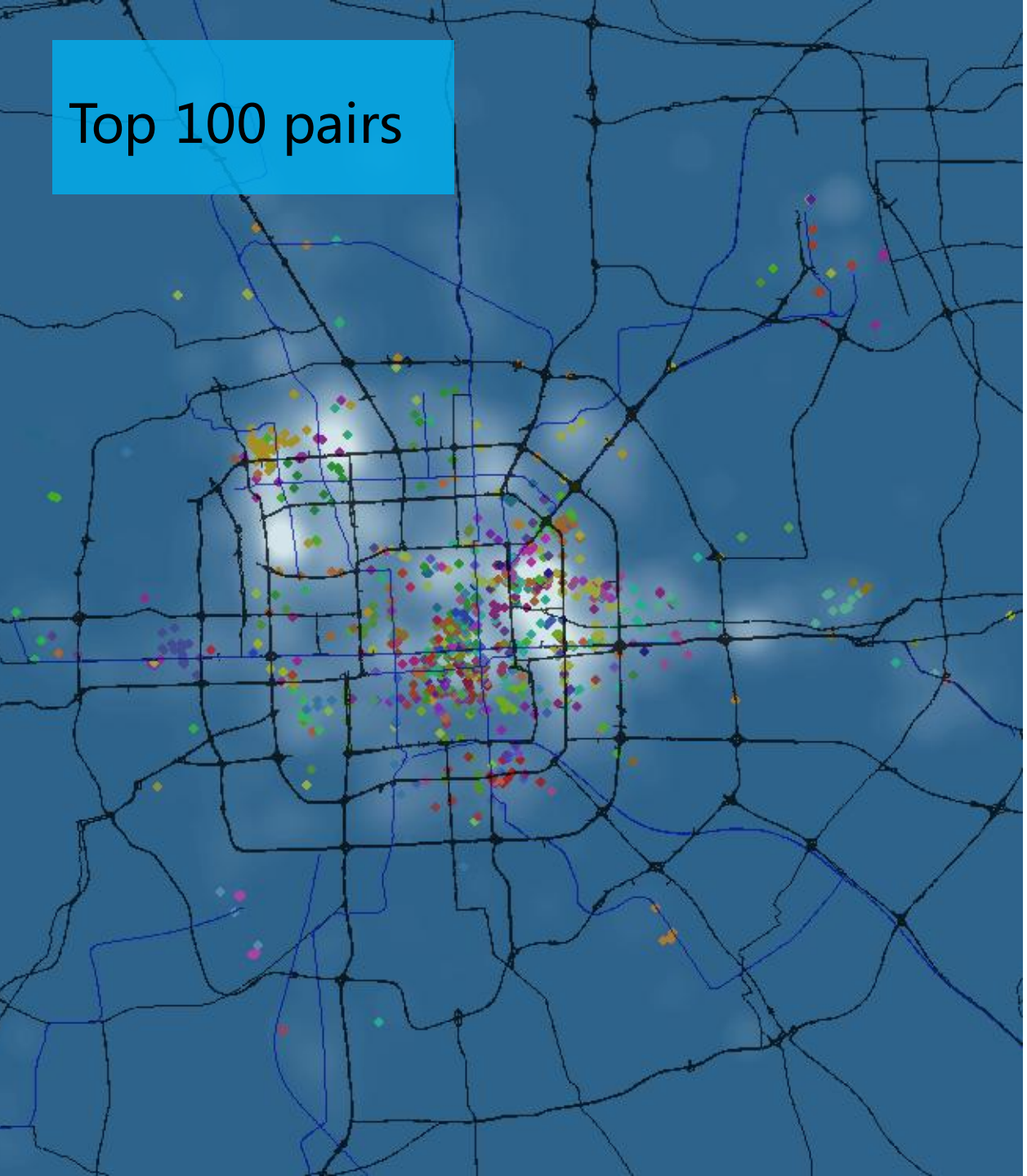
两个月内、六环内：

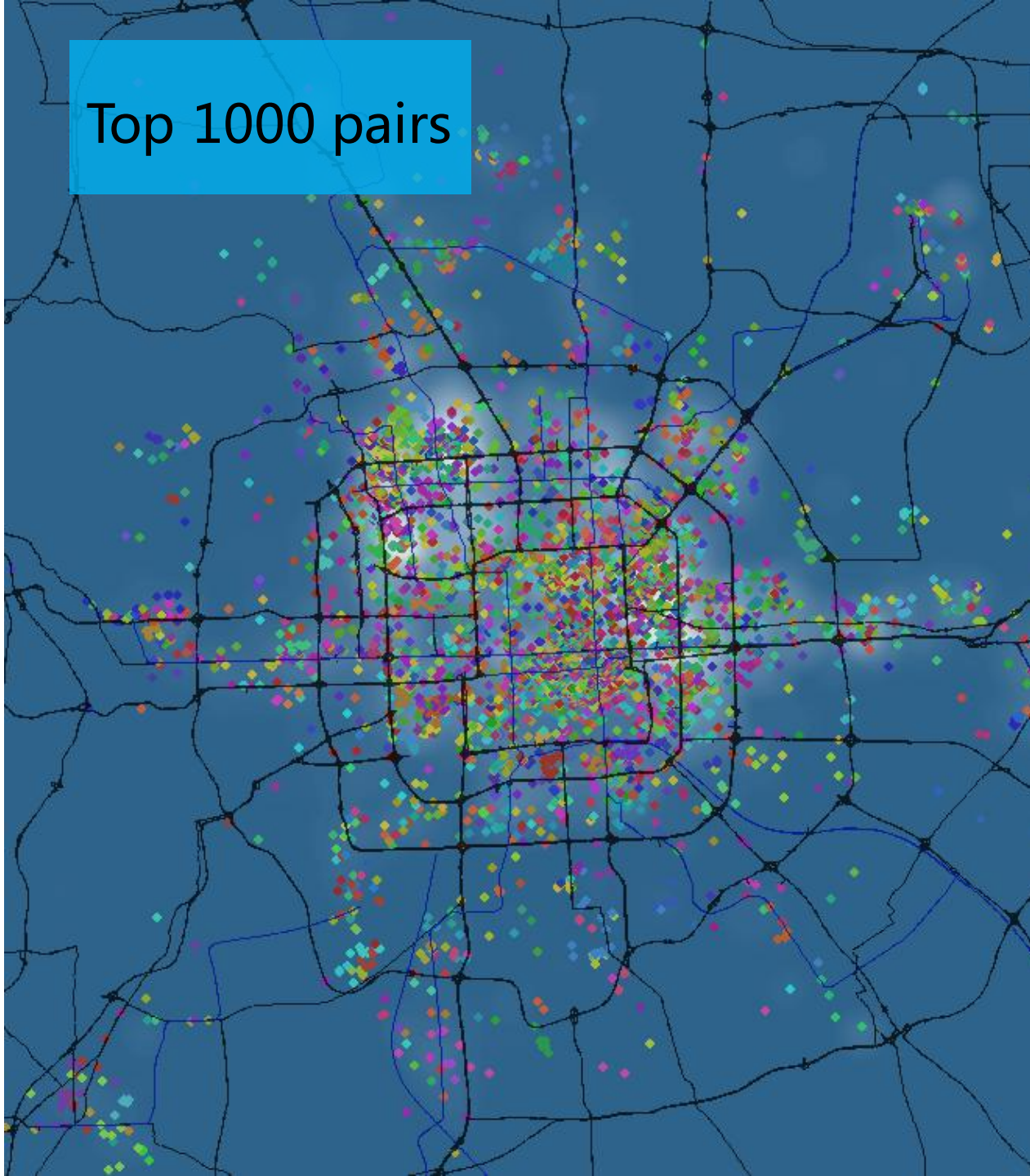
- 潜在碰面 14,004,320 组次
- 潜在碰面 18,137,910 人次，每天约30万次
- 最有“缘分”的两人“擦肩而过” 2330 次
- 最多的时候一个人身边有 6376 人满足碰面条件
- 最多的一个block里发生了 201108 万次碰面
- ……

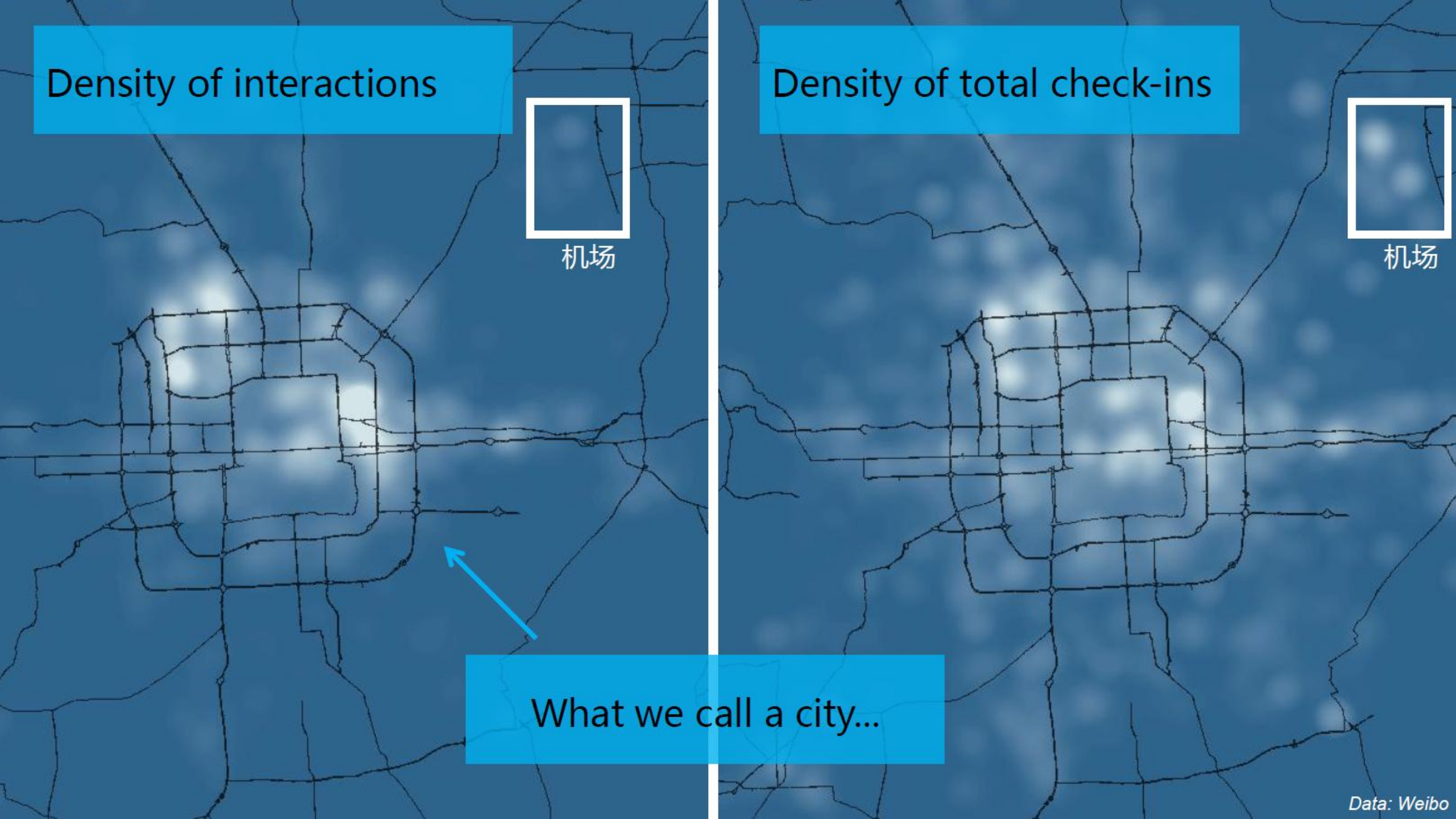



Top 10 pairs


Top 50 pairs


Top 100 pairs


Top 1000 pairs


Density of interactions
机场
Density of total check-ins
机场
What we call a city...
Data: Weibo

Social + Interaction + Potential

- Lijun Sun: **Understanding metropolitan patterns of daily encounters**, PNAS 13774–13779.
- Markus Schlapfer, Luis M. A. Bettencourt, Sebastian Grauwin *et al.*: **The scaling of human interactions with city size**. arXiv:1210.5215v3
- T. Neutens *et al.*: **Spatial variation in the potential for social interaction**: A case study in Flanders (Belgium). Computers, Environment and Urban Systems 41 (2013) 318–331.
- S. Farber, X. Li: **Urban sprawl and social interaction potential**. Journal of Transport Geography 31 (2013) 267–277.
- J.K. Brueckner, A.G. Largey: **Social interaction and urban sprawl**. Journal of Urban Economics 64 (2008) 18–34.

Social + Interaction + Potential

- Lijun Sun: **Understanding metropolitan patterns of daily encounters**, PNAS 13774–13779.

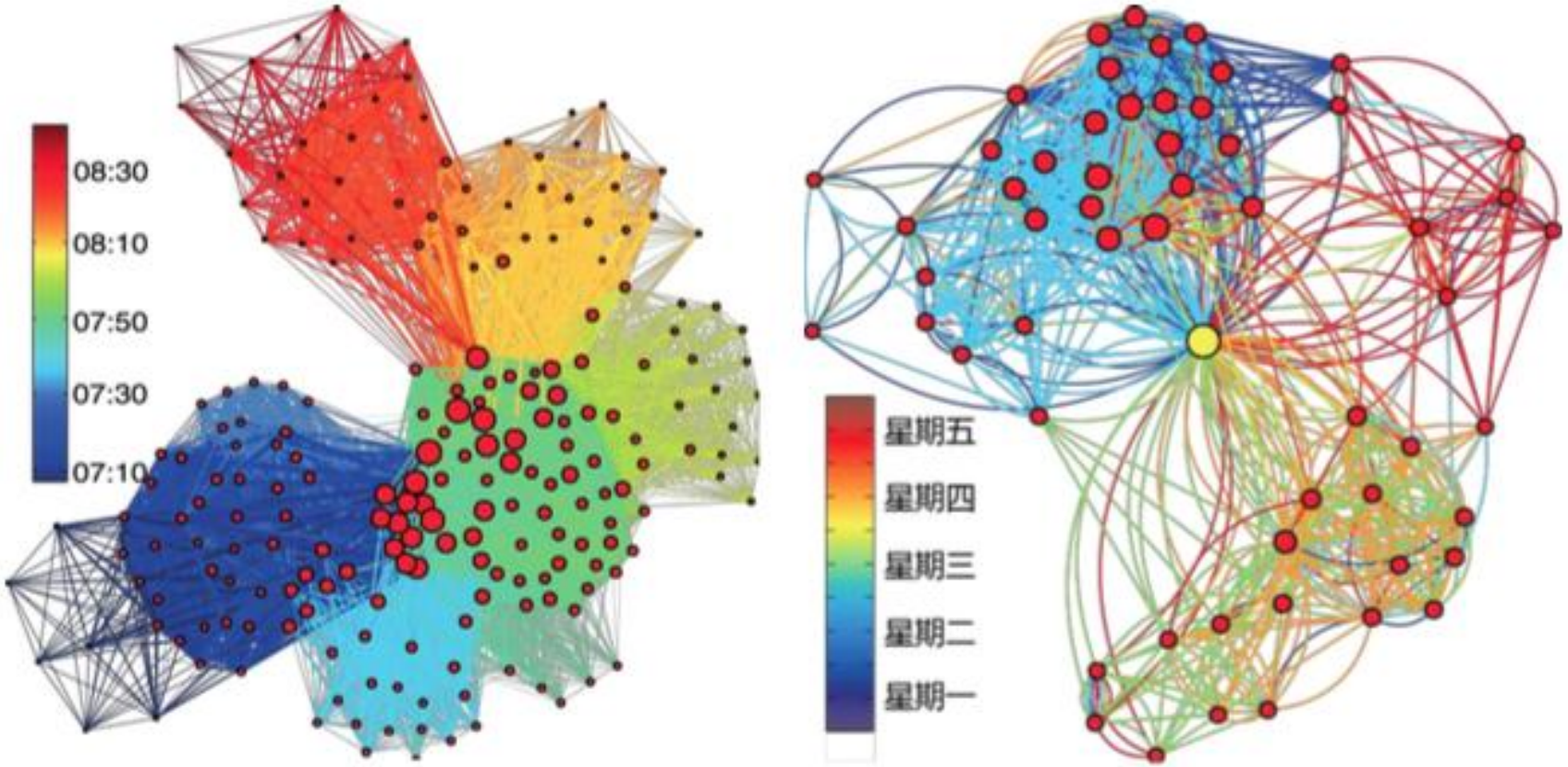
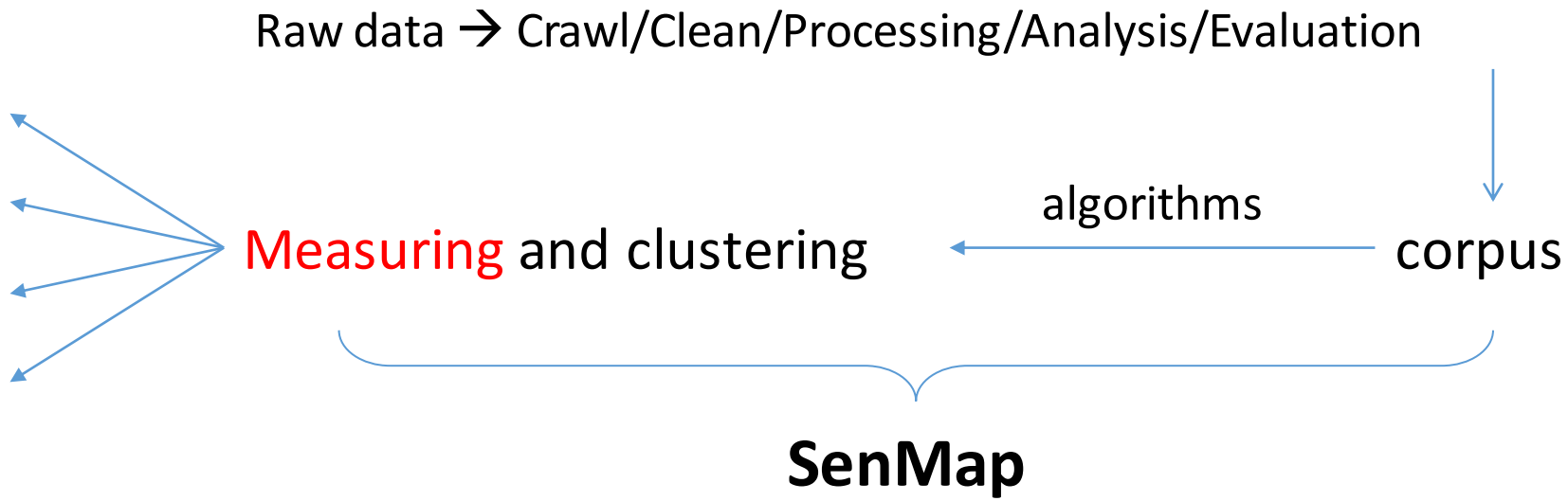


图 4：在一辆公交车上的乘客相遇网络（a）及个体一周内"熟悉的陌生人"网络（b），来源：Sun 等(2013)

示例2：城市情绪地图

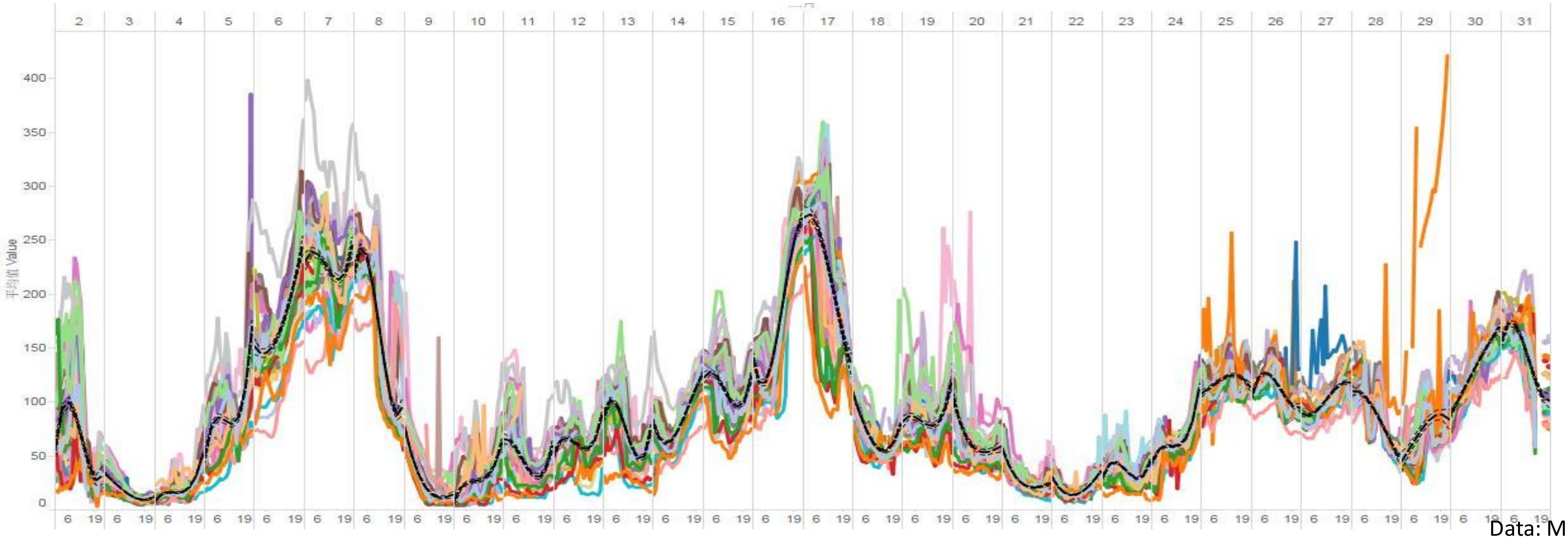
- Formation and contents of sentiments
  - Attitudes
  - Believes
  - Judgments
  - Emotions



示例2：城市情绪地图

- Demonstration: on the air quality

Concentration value of PM 2.5, Beijing, March 2015

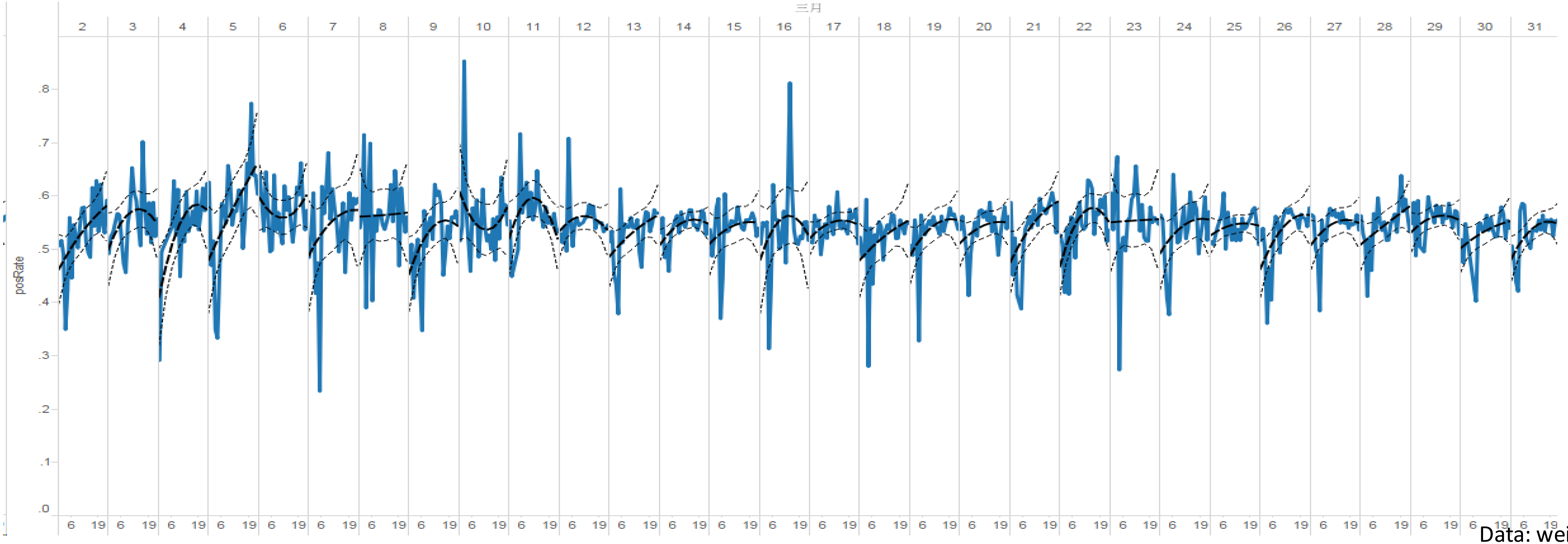


Data: MEP

示例2：城市情绪地图

- Demonstration: on the air quality

Aggregated value of positive emotions, Beijing, March 2015

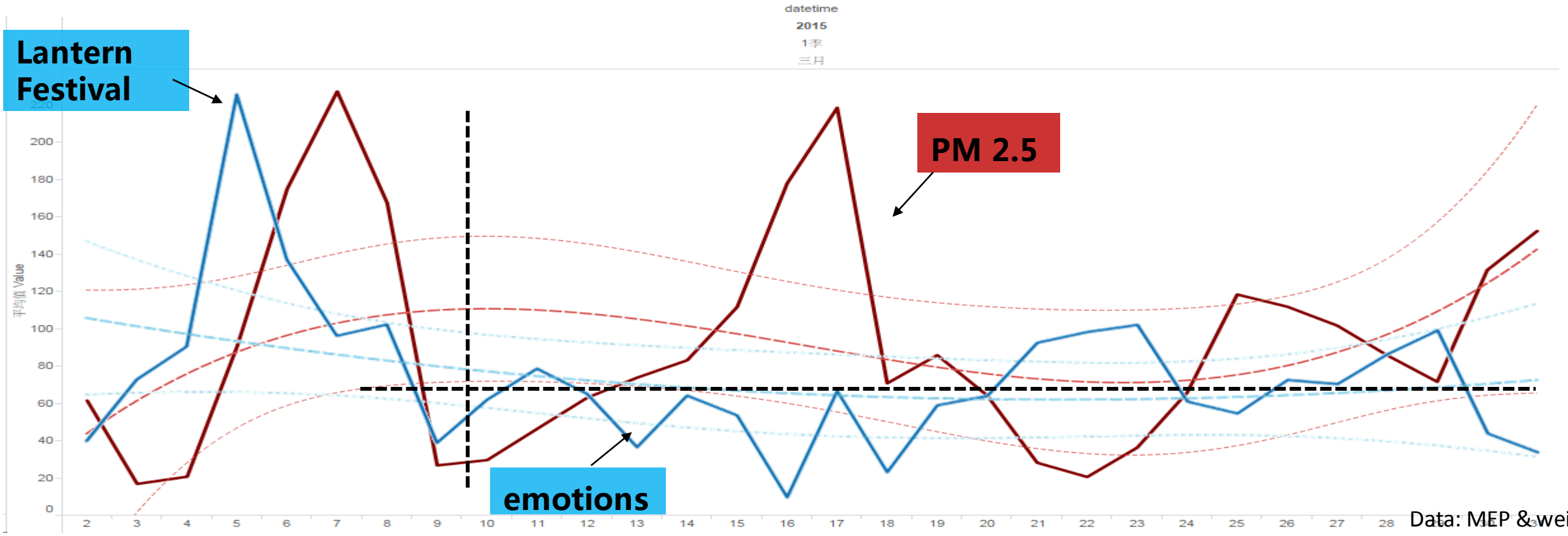


Data: weibo

示例2：城市情绪地图

- Demonstration: on the air quality

daily emotions VS. air quality

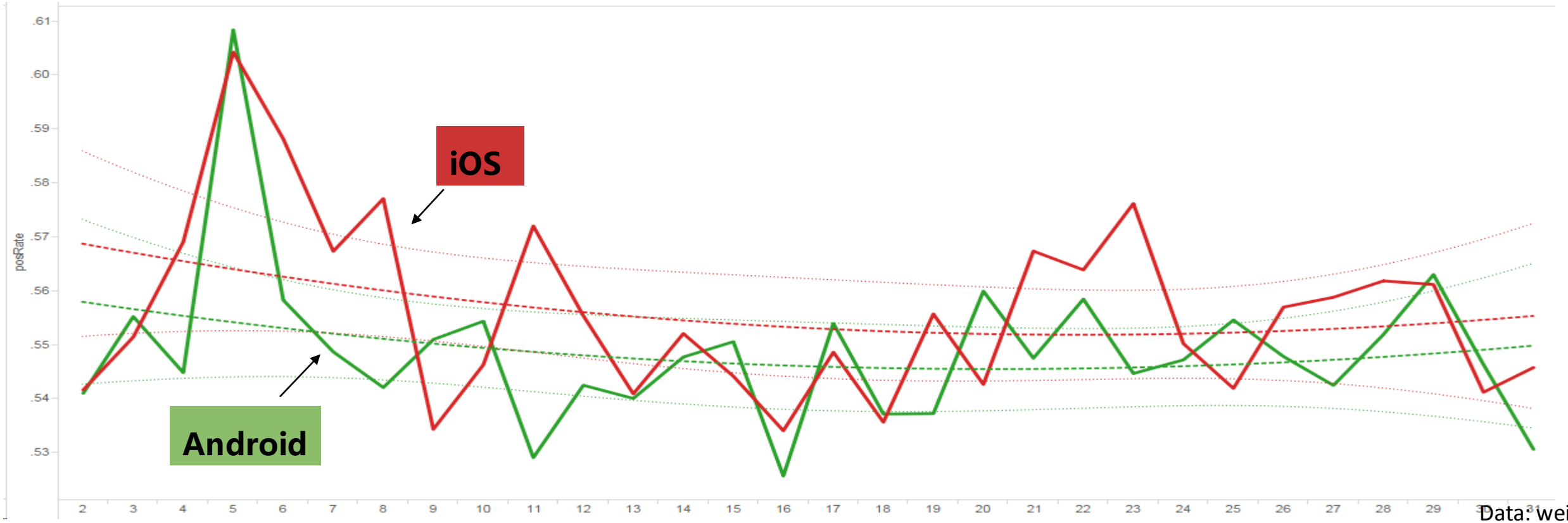


Data: MEP &weibo

示例2：城市情绪地图

- Demonstration: and on your cell phone brands

Aggregated value of positive emotions, by types of sources, Beijing, March 2015

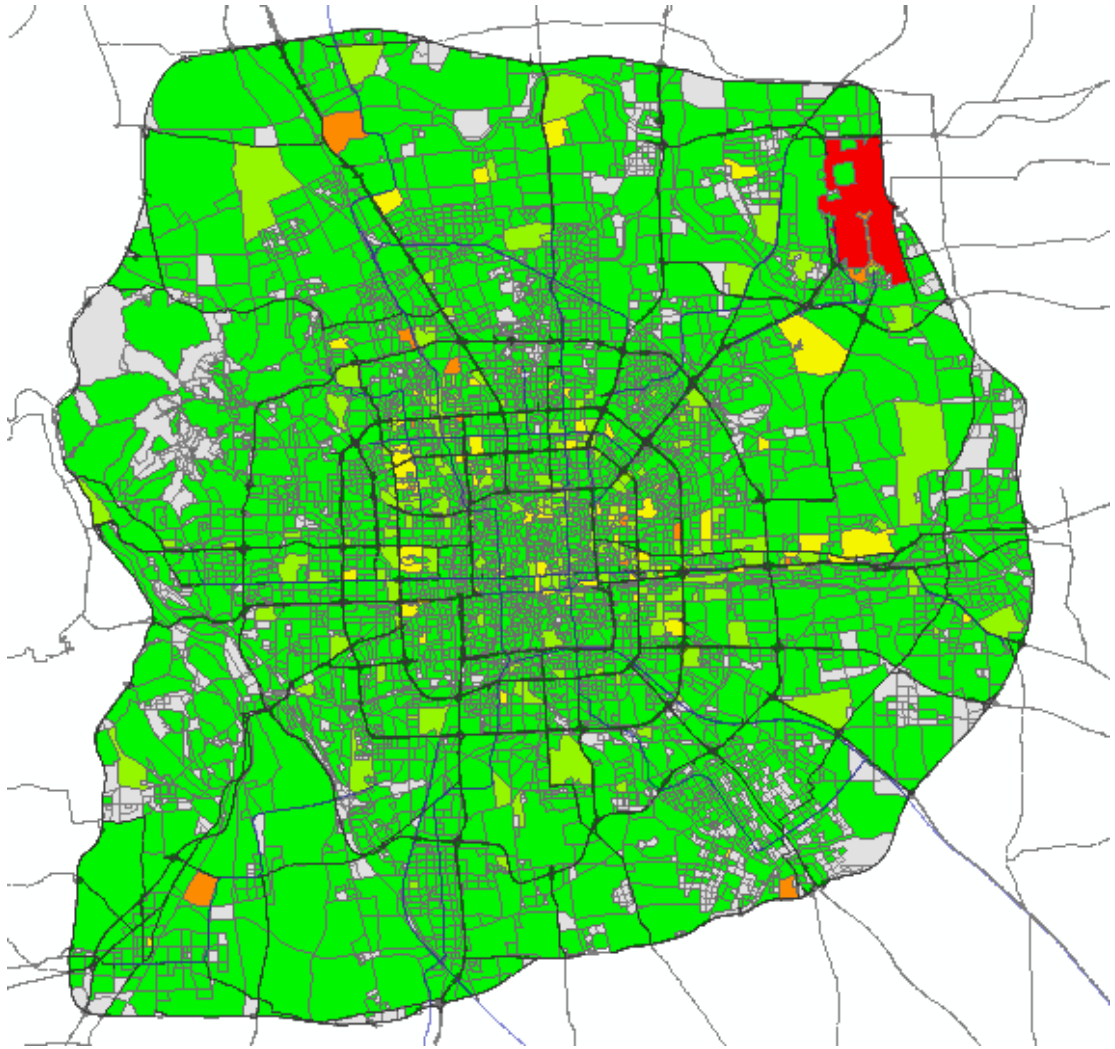


Data: weibo

示例2：城市情绪地图

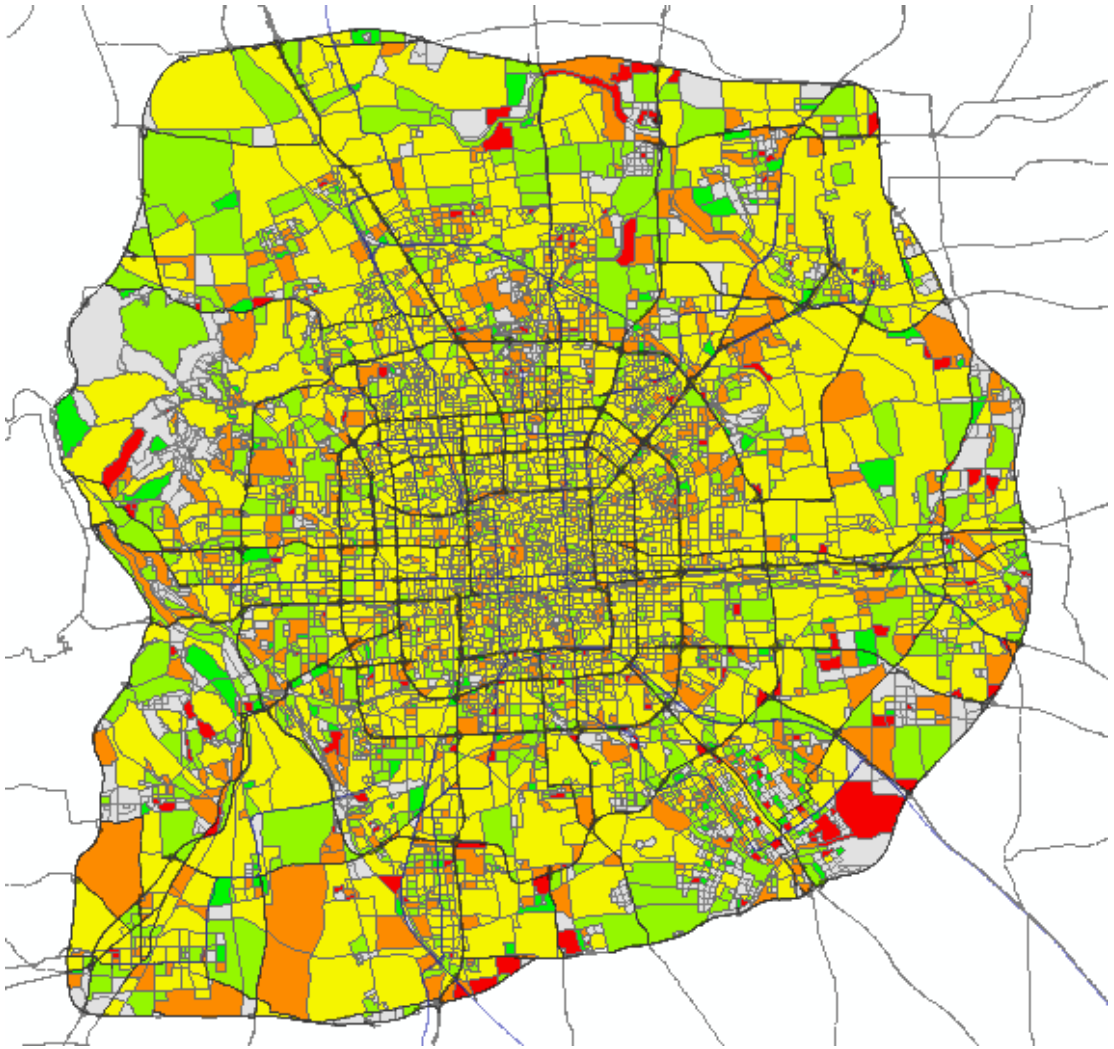
- Research questions
  - Is there a spatial pattern of emotions in a city?
  - Where are happy emotions located ?

Total volume of positive values, per block



Head/tail breaks

Rate of positive values, per block



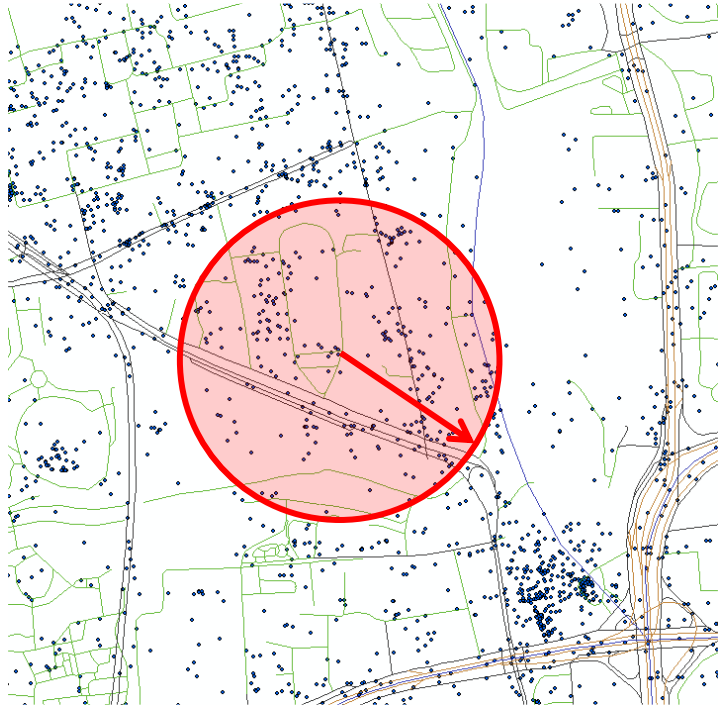
Natural breaks

Data: weibo

示例2：城市情绪地图

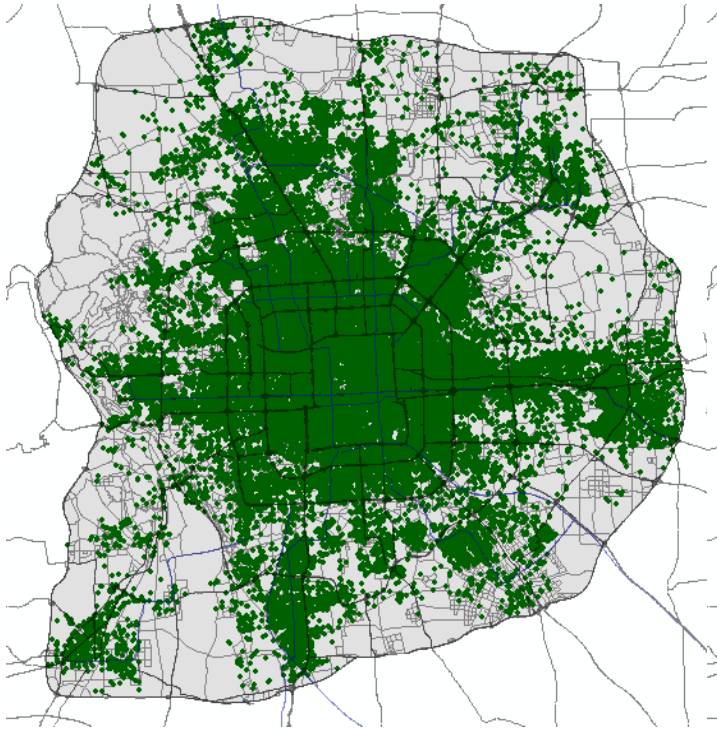
- Define “very”: how much happy you are
  - Higher than other sentiments located within a certain boundary of distance and time: 2km and 2h/8h/24h
- Happy Emotion Index (HEI)

$$\Delta EMO_i = EMO_i - \overline{EMO_j}, where\ i \neq j$$

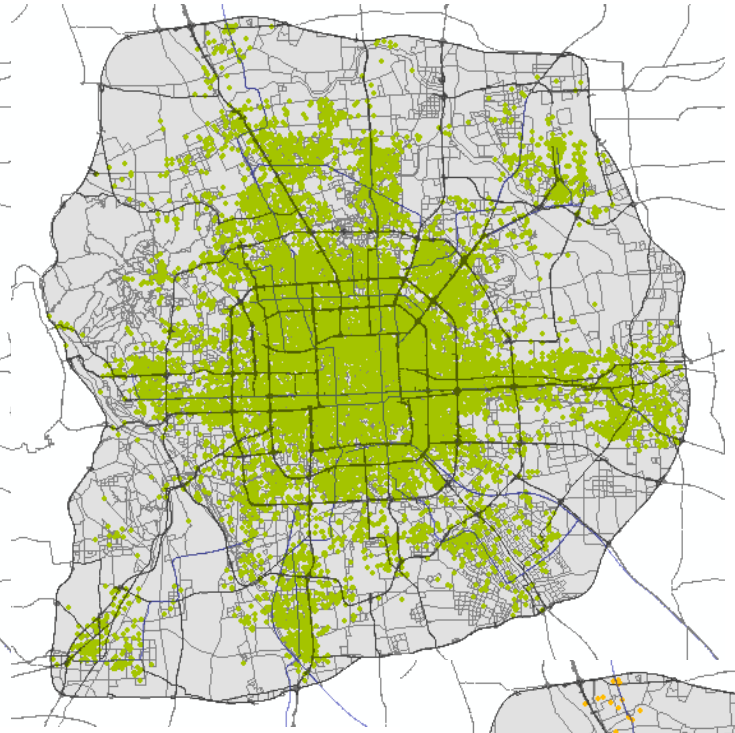


Level zero (usual): higher than the average， Level one: higher than the average plus one σ

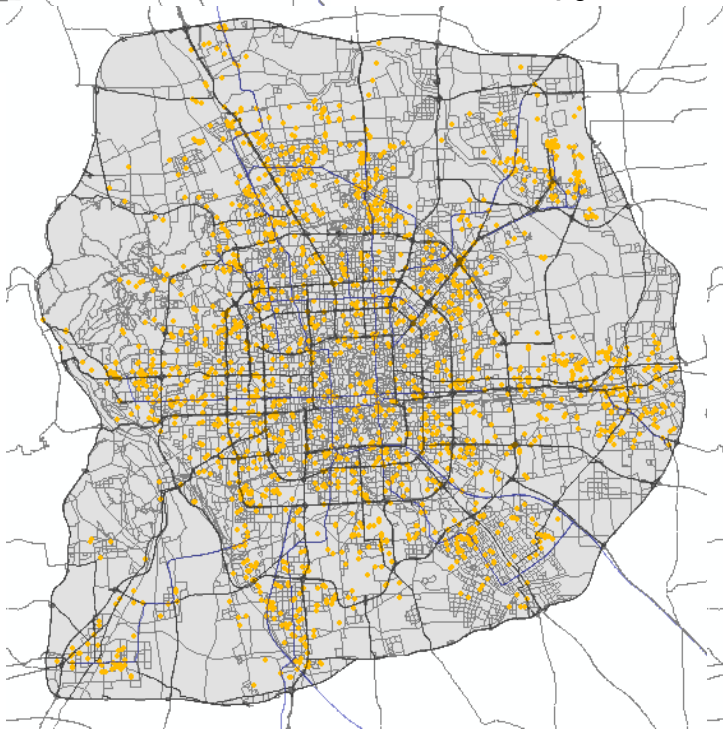
Level two: higher than the average plus two σ， Level three: higher than the average plus three σ



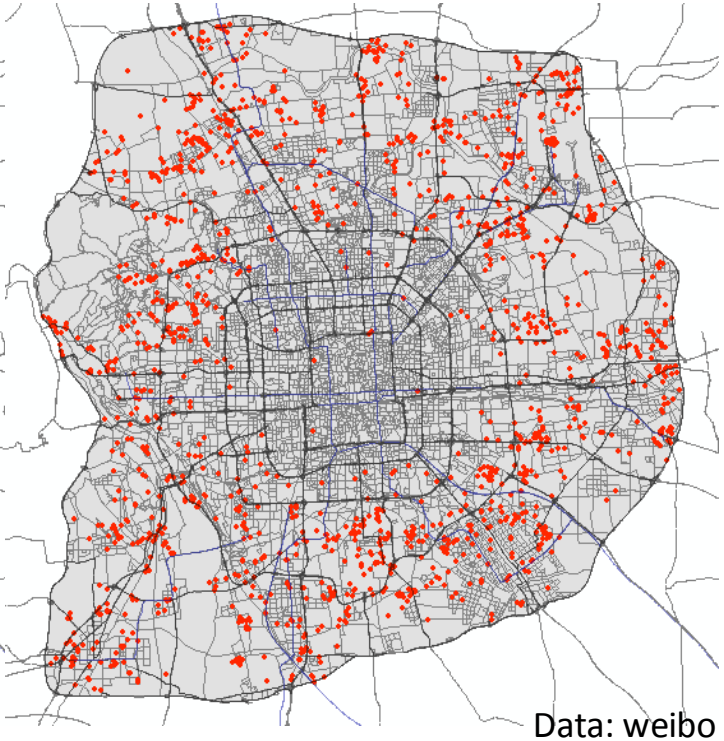
$HEI = 0$
75.2%



$HEI = 1$
21.5%

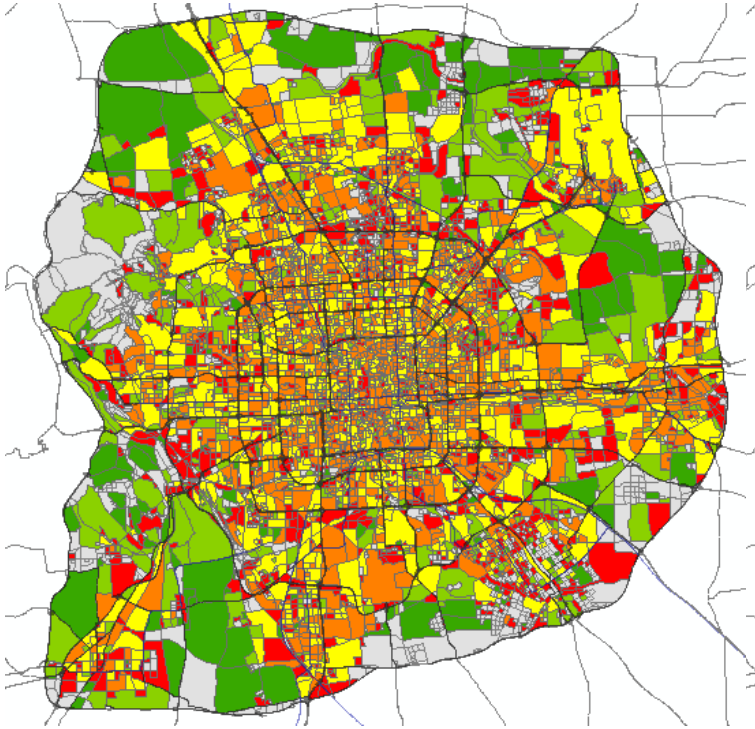


$HEI = 2$
1.4%

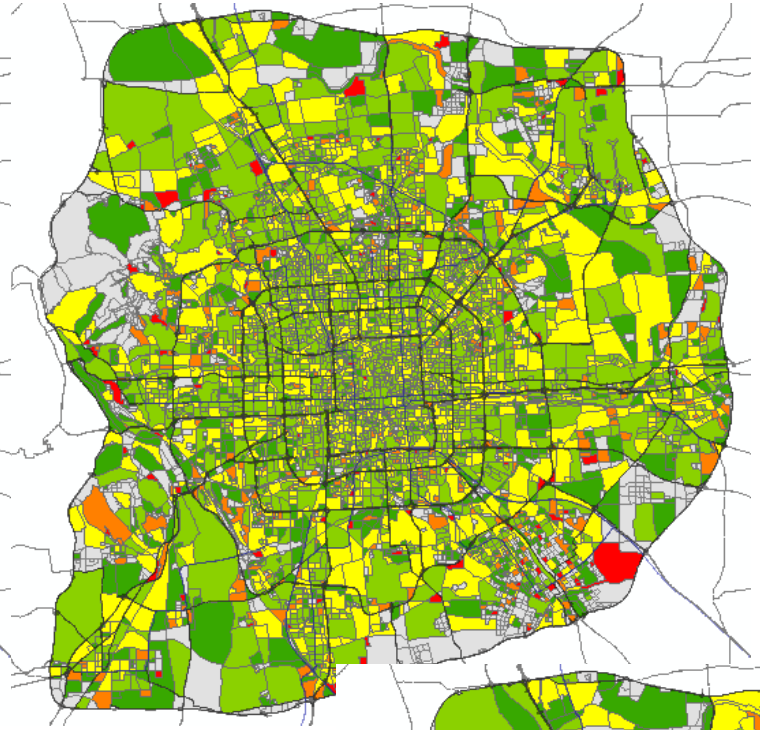


$HEI = 3$
1.8%

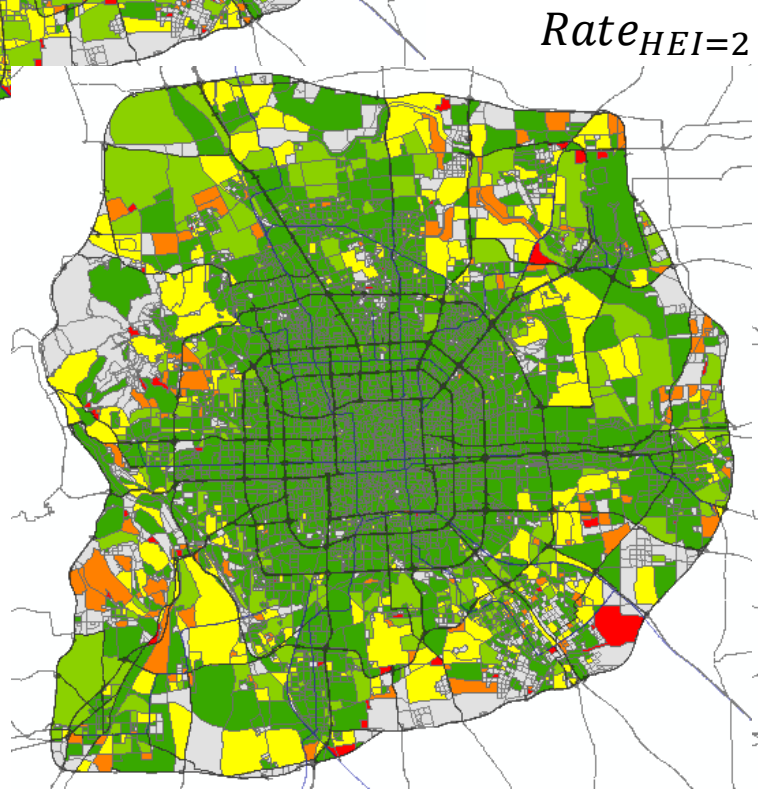
Data: weibo



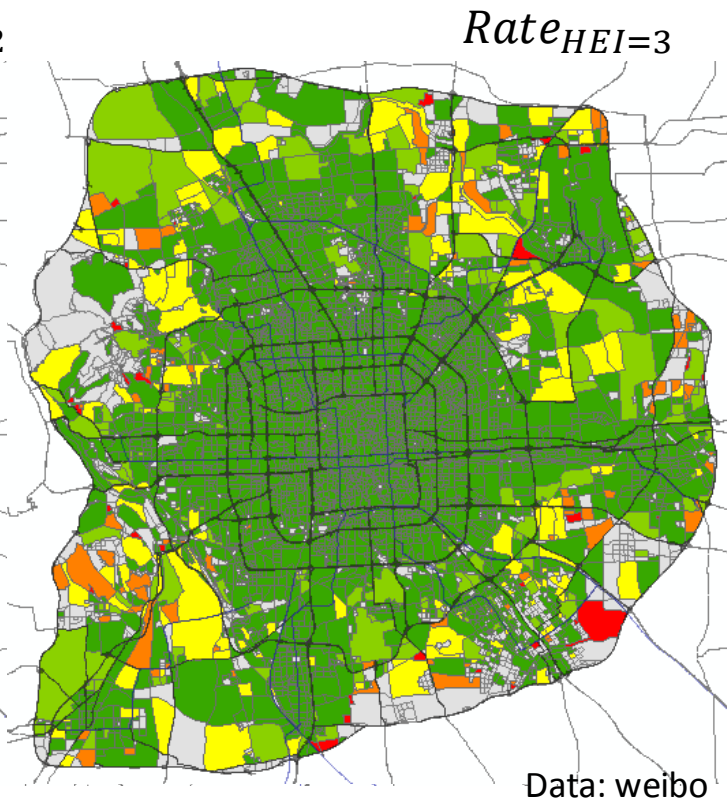
$Rate_{HEI=0}$



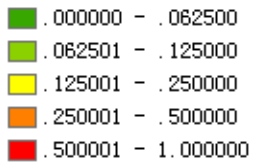
$Rate_{HEI=1}$



$Rate_{HEI=2}$



$Rate_{HEI=3}$



Data: weibo

示例2：城市情绪地图

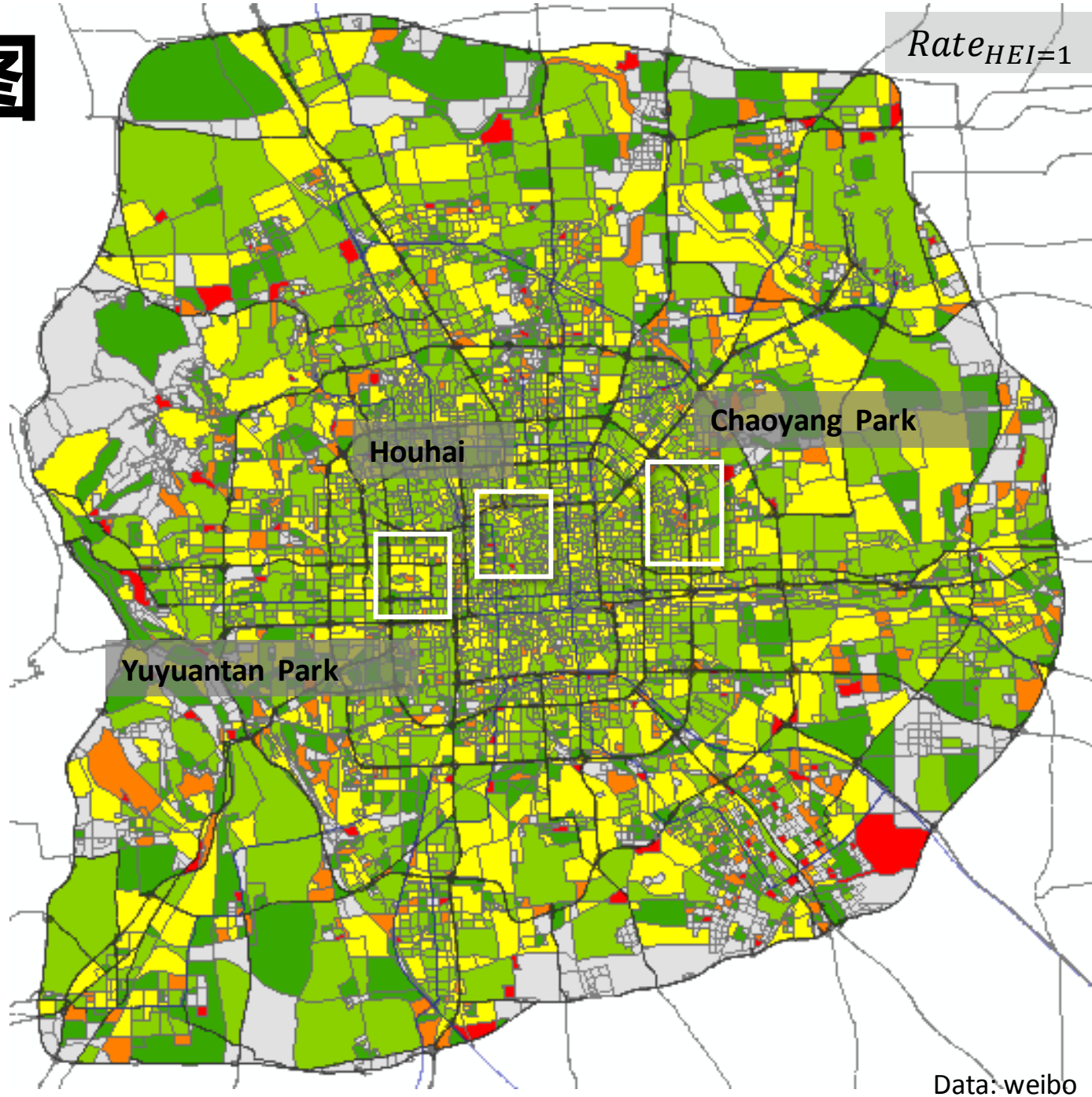
- Capability of distinguishment

$Block_x$

$Rate_{HEI} = {count_{HEI}}/{count_{all}}$

$HEI_i = EMO_i - \overline{EMO_j} - \sigma_{EMO_j} > 0$

$where\ i \neq j, dist_{ij} < 2km, \Delta T_{ij} < 24h$



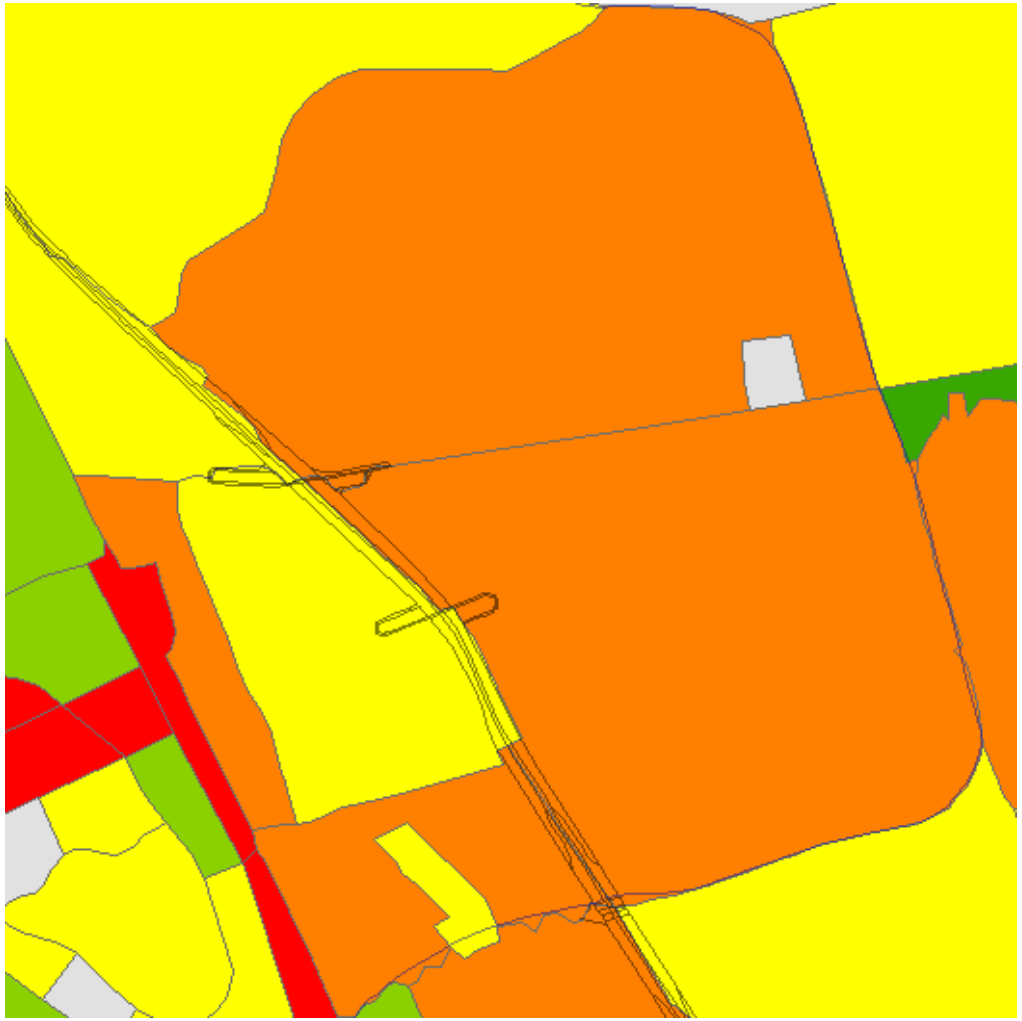
示例2：城市情绪地图

回龙观北四村

Data: Baidu BDL
ref: 刘瑜（北大）等

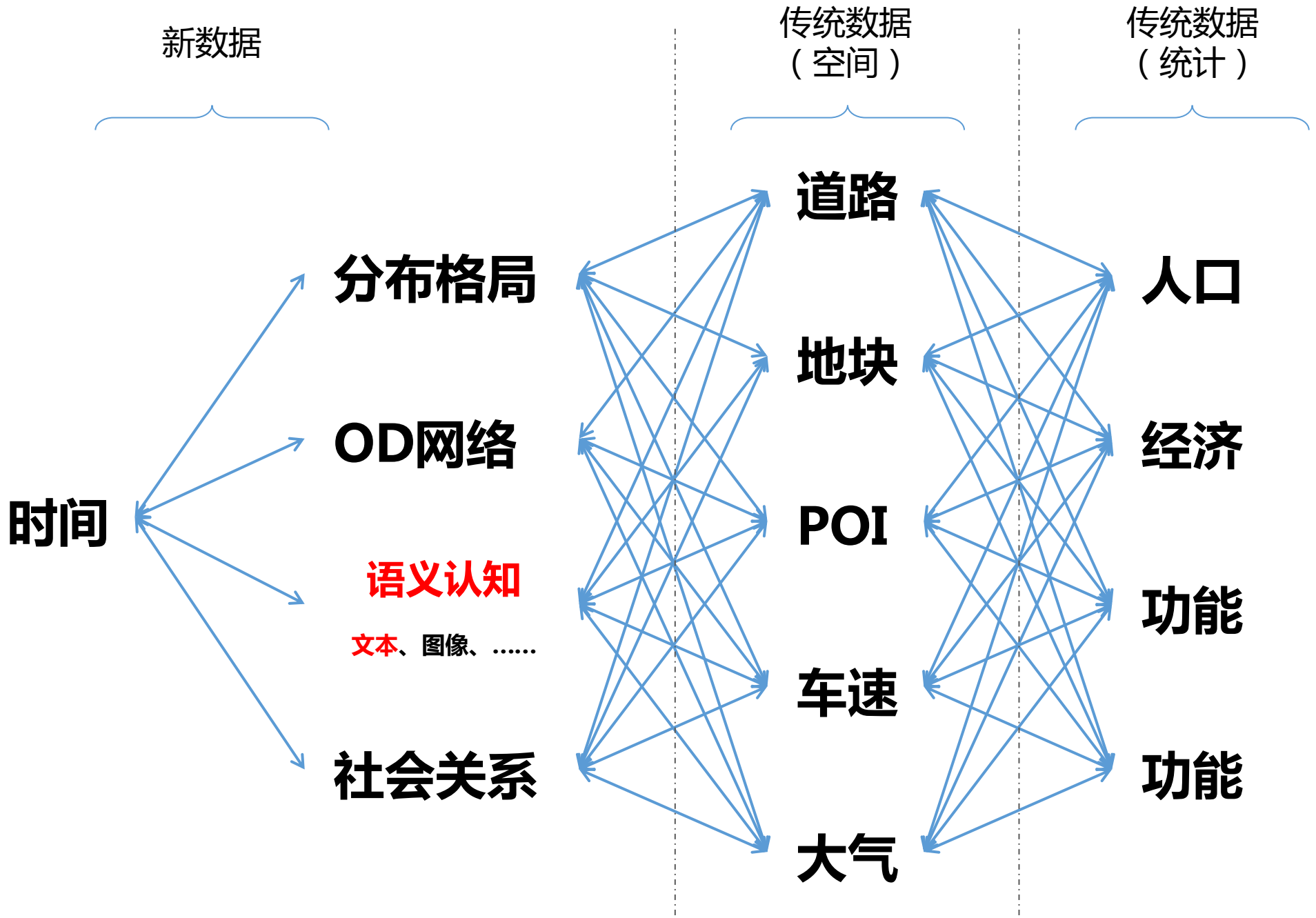
示例2：城市情绪地图

	蚁族	别墅
sum neg	54.61	62.59
sum pos	65.39	74.407
sum high	136	163
sum high3s	3	6
sum high2s	6	11
sum high1s	32	26
sum highAVG	95	120
Rate high3s	0.03	0.04
Rate high2s	0.05	0.08
Rate high1s	0.27	0.19
Rate highAVG	0.79	0.88
count	332	309
highRate	0.41	0.53

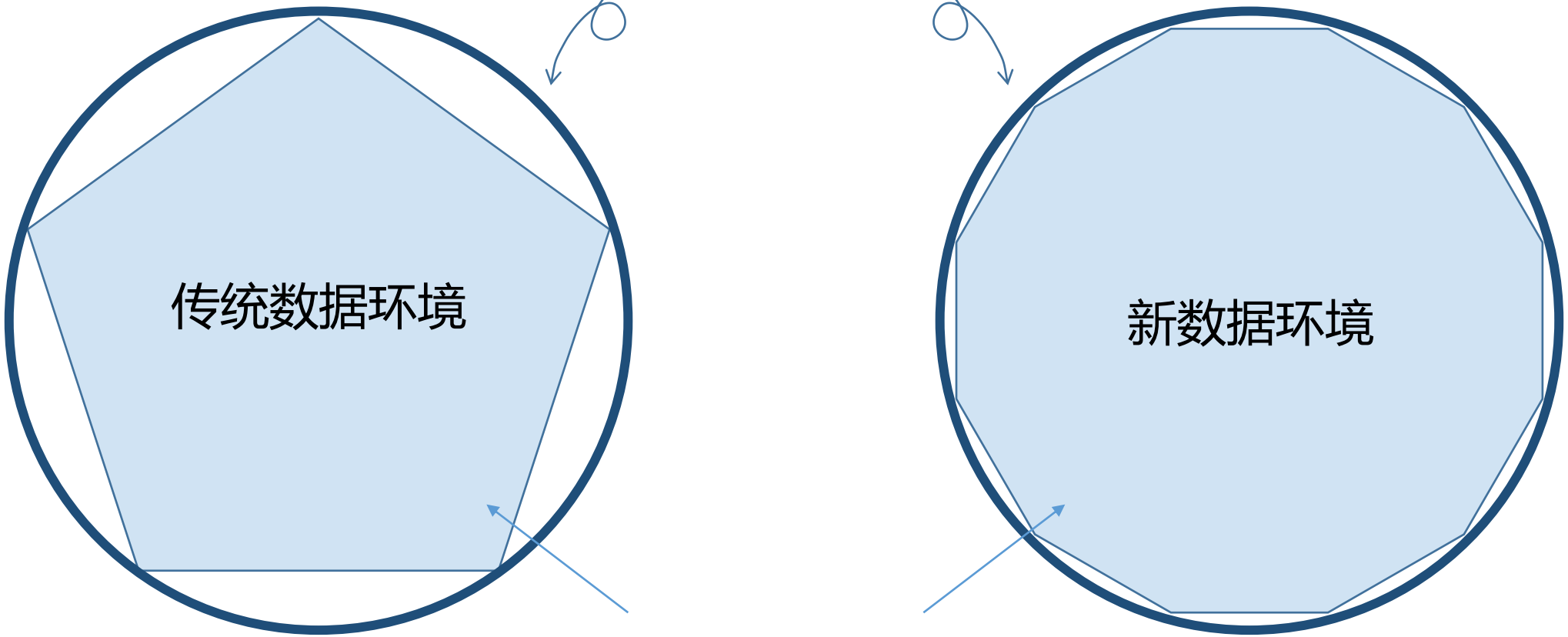


什么是“城市”

- 共同性：聚集
- 差异性：特征


新数据
传统数据
（空间）
传统数据
（统计）
时间
分布格局
OD网络
语义认知
文本、图像、……
社会关系
道路
地块
POI
车速
大气
人口
经济
功能
功能

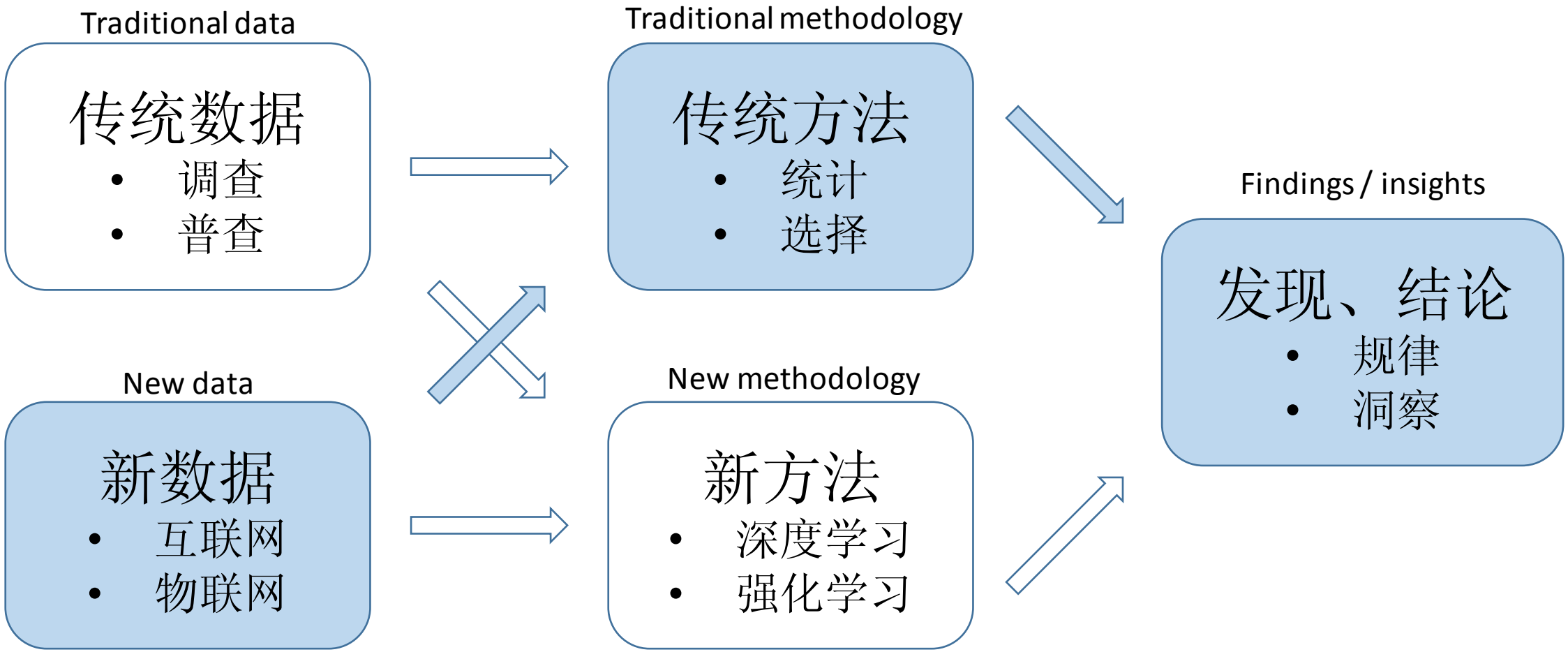
我们的城市



理解的手段：研究、

认识和分析

讨论


Traditional data
传统数据
• 调查
• 普查
Traditional methodology
传统方法
• 统计
• 选择
Findings / insights
发现、结论
• 规律
• 洞察
New data
新数据
• 互联网
• 物联网
New methodology
新方法
• 深度学习
• 强化学习

下一步的挑战

- 从表达、描述相关性到揭示因果和影响
- 对一个复杂系统如何开展整体性的研究和分析
  - 指标、评估的局限性
- 数据的挖掘侧重城市研究的本质：关注空间、关注人、关注人和人之间的联系。
- 大数据对城市整体性的描述不再重要，深入挖掘细分、特征和机理是下一步的方向。

感谢！欢迎任何意见和问题，请多提宝贵建议

规勒个划 @weibo