北京城市实验室
Beijing City Lab

ID of the slides

49

Slides of BCL

www.beijingcitylab.com

How to cite

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

北京城市实验室2015年会 “新数据环境下的城市规划与研究”

2015年6月6日 @ 北京交通大学建筑与艺术学院

城市光谱：用数据理解空间

Urban Spectrum: Understanding Space by Data

李栋 博士

中国城市规划设计研究院

- 万物互联
  - Internet of WWW → Internet of PEOPLE → Internet of EVERYTHING
- 数据汹涌
  - 每18个月新增数据量是有史以来全部累积数据量的总和
  - 到2020年全球每年产生的数据量将达 40ZB（GB → TB → PB → EB → ZB → …）
- 城市研究如何“进化”？
  - 如何进入IT (Information Technology)时代？如何进入DT (Data Technology)时代？

Gantz J and Reinsel D (2009). As the Economy Contracts, the Digital Universe Expands. <http://www.emc.com/collateral/leadership/digital-universe/2009DU_final.pdf>

- 众包数据城市分析框架的四个方向

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), pp. 49-57.

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

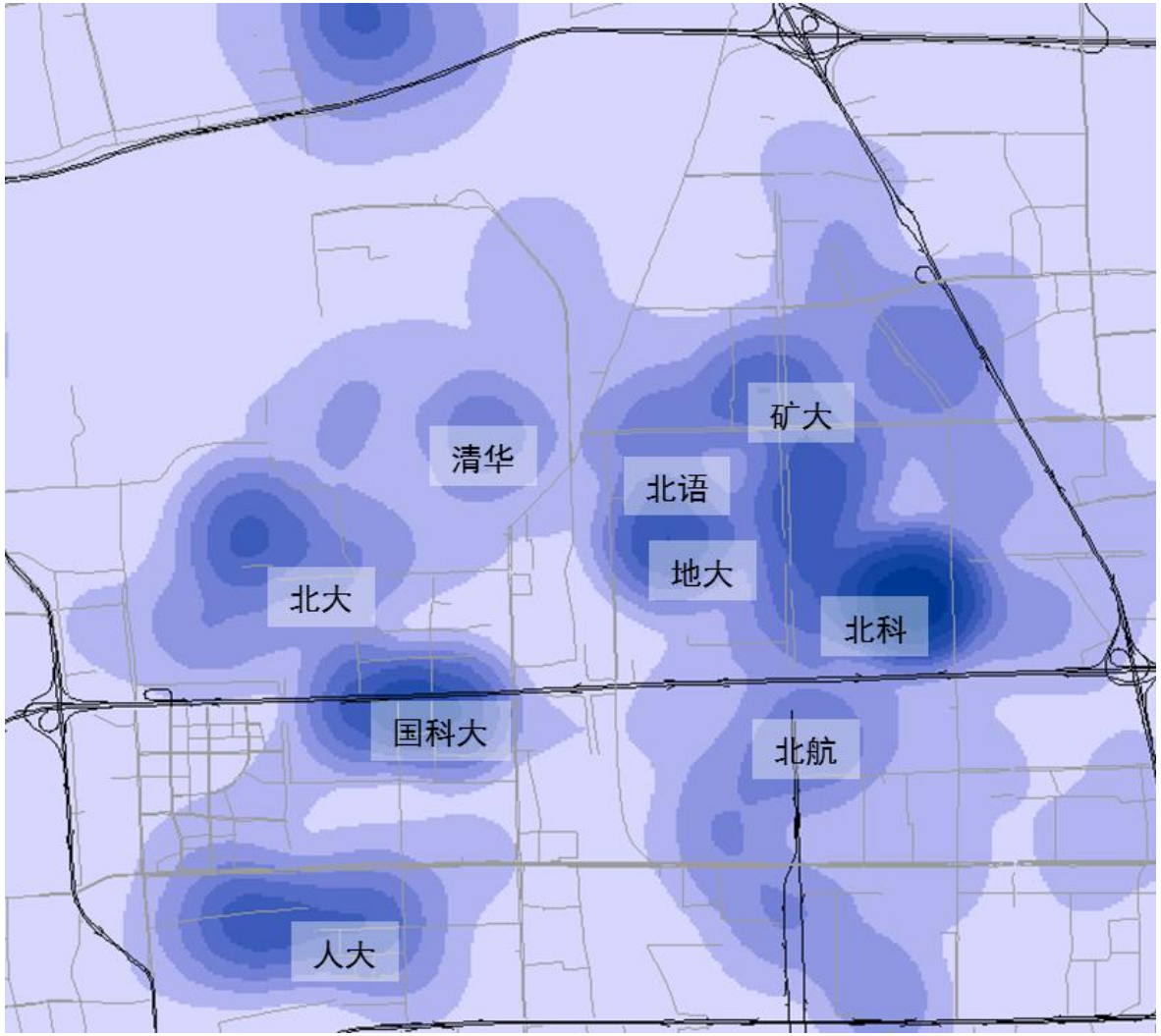
“今天的空气非常不错”

“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”

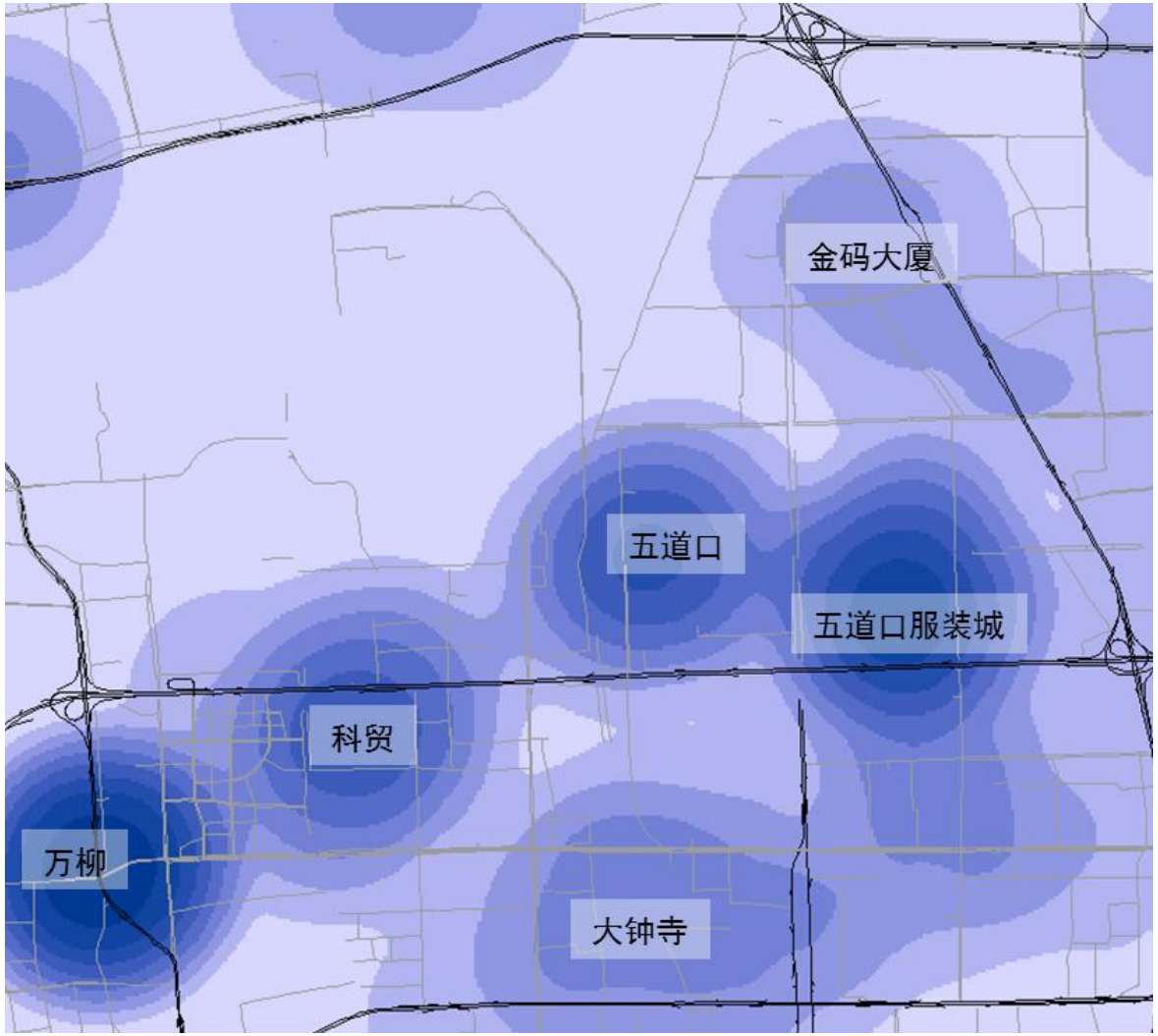
“中国家庭收入差距明显，收入最多的20%的家庭和收入最少的20%的家庭相差19倍左右，流动家庭和留守家庭已经成为家庭的常规模式”

同一个中关村、不一样的生活方式

“**大学**”名词n：分布热图

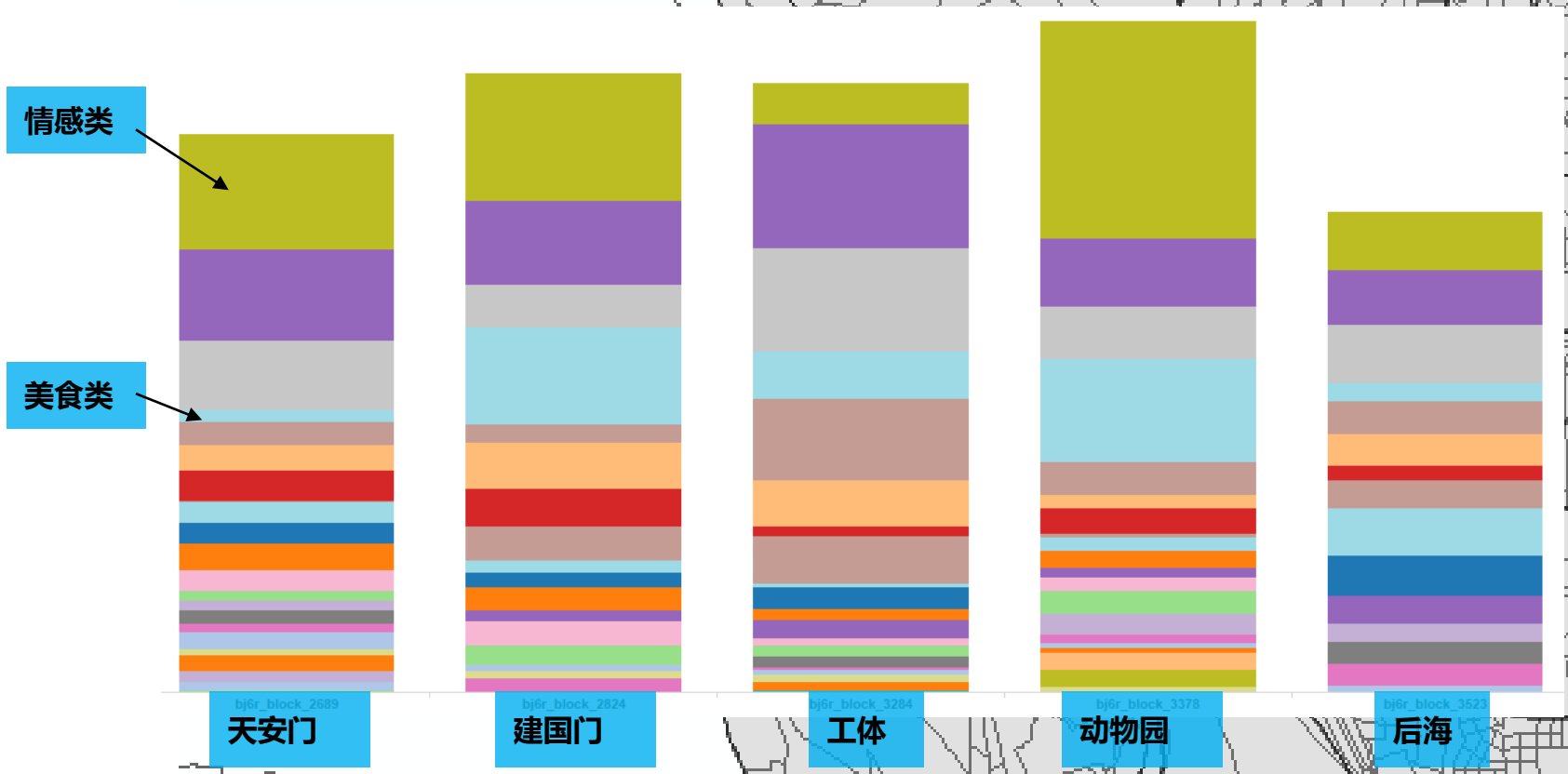


“**购物**”动词v：分布热图



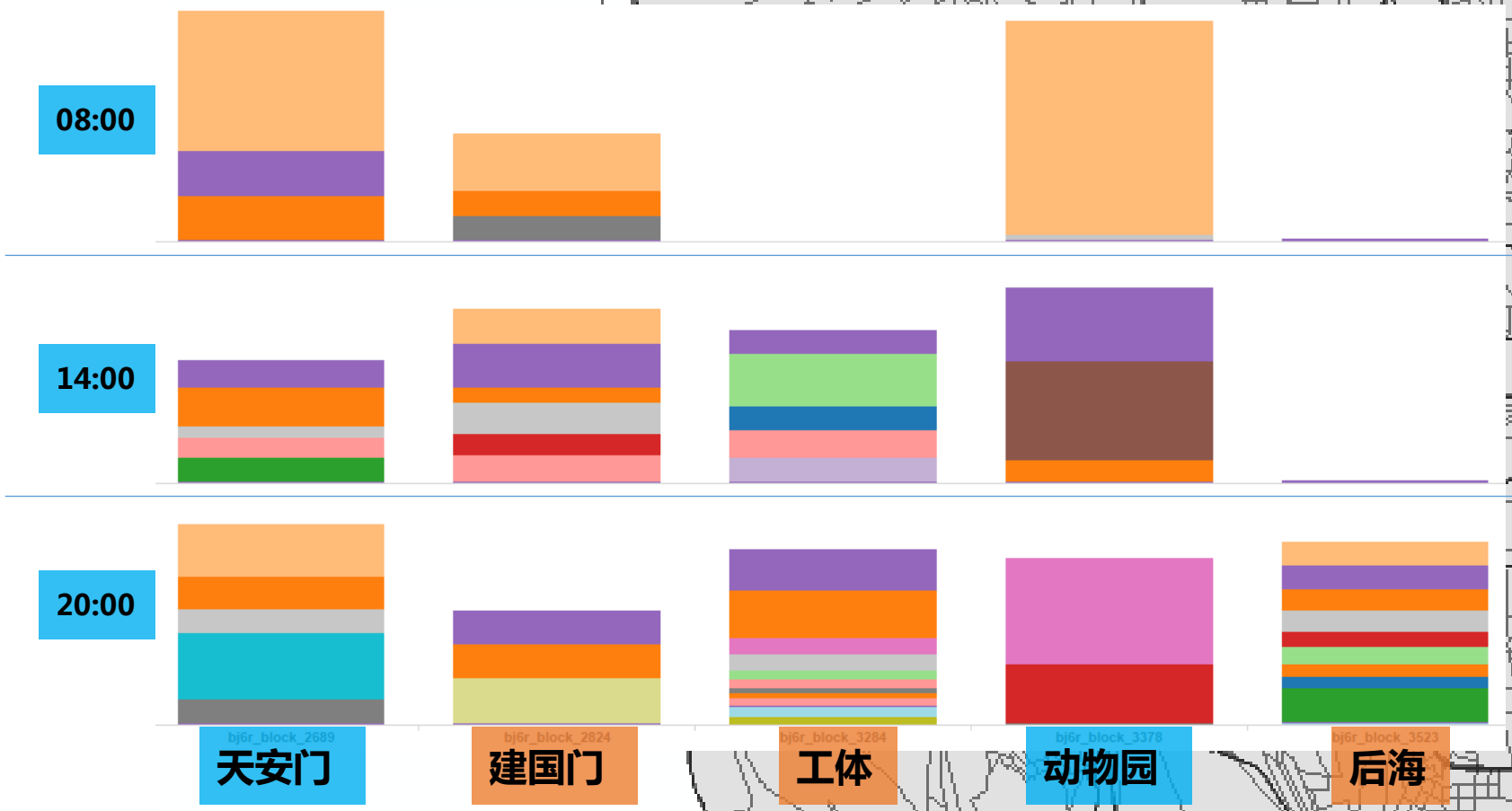
Data: Weibo

不同地块语义主题的差异



Data: Weibo, OSM

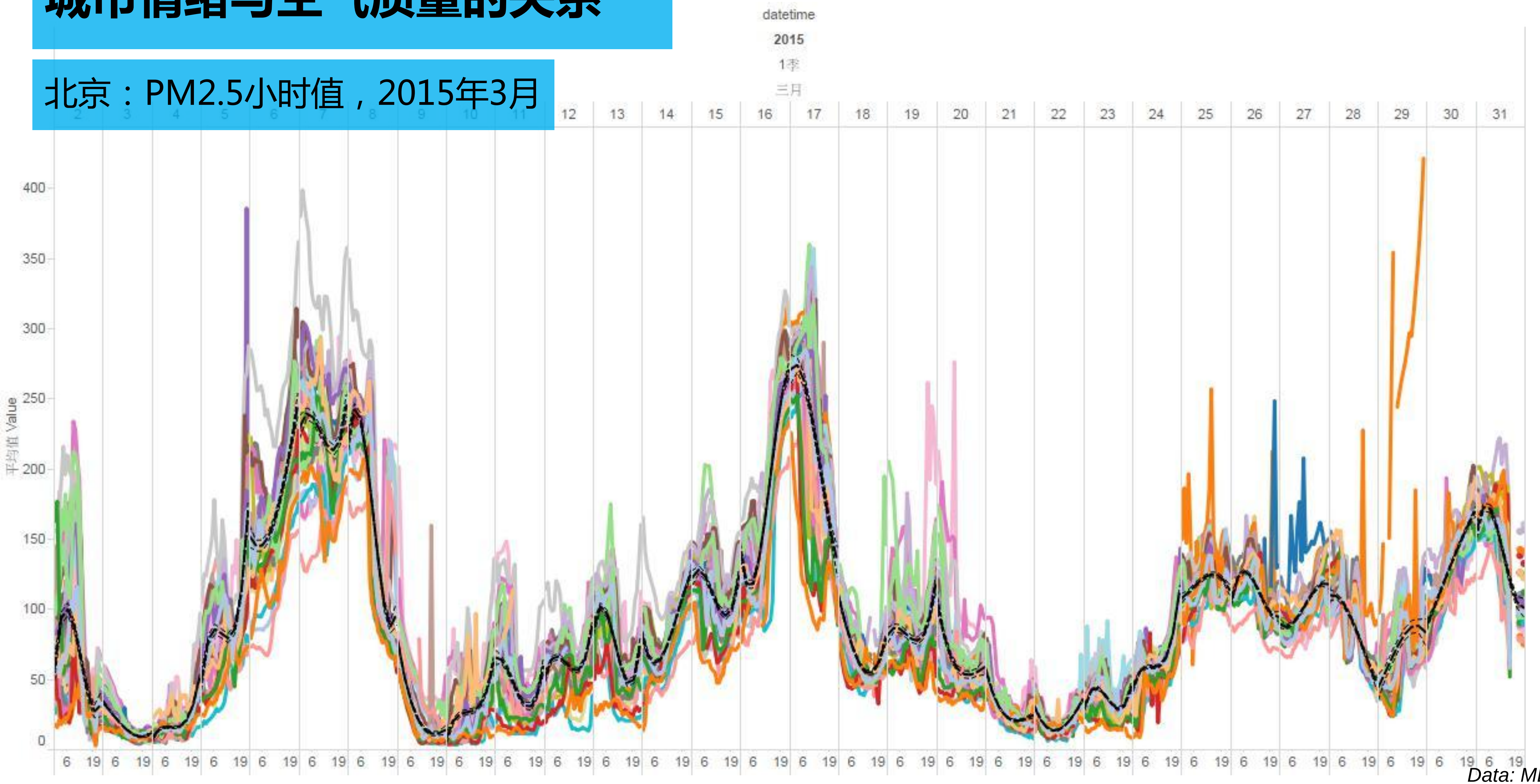
不同地块语义主题的差异：不同时段



Data: Weibo, OSM

城市情绪与空气质量的关系

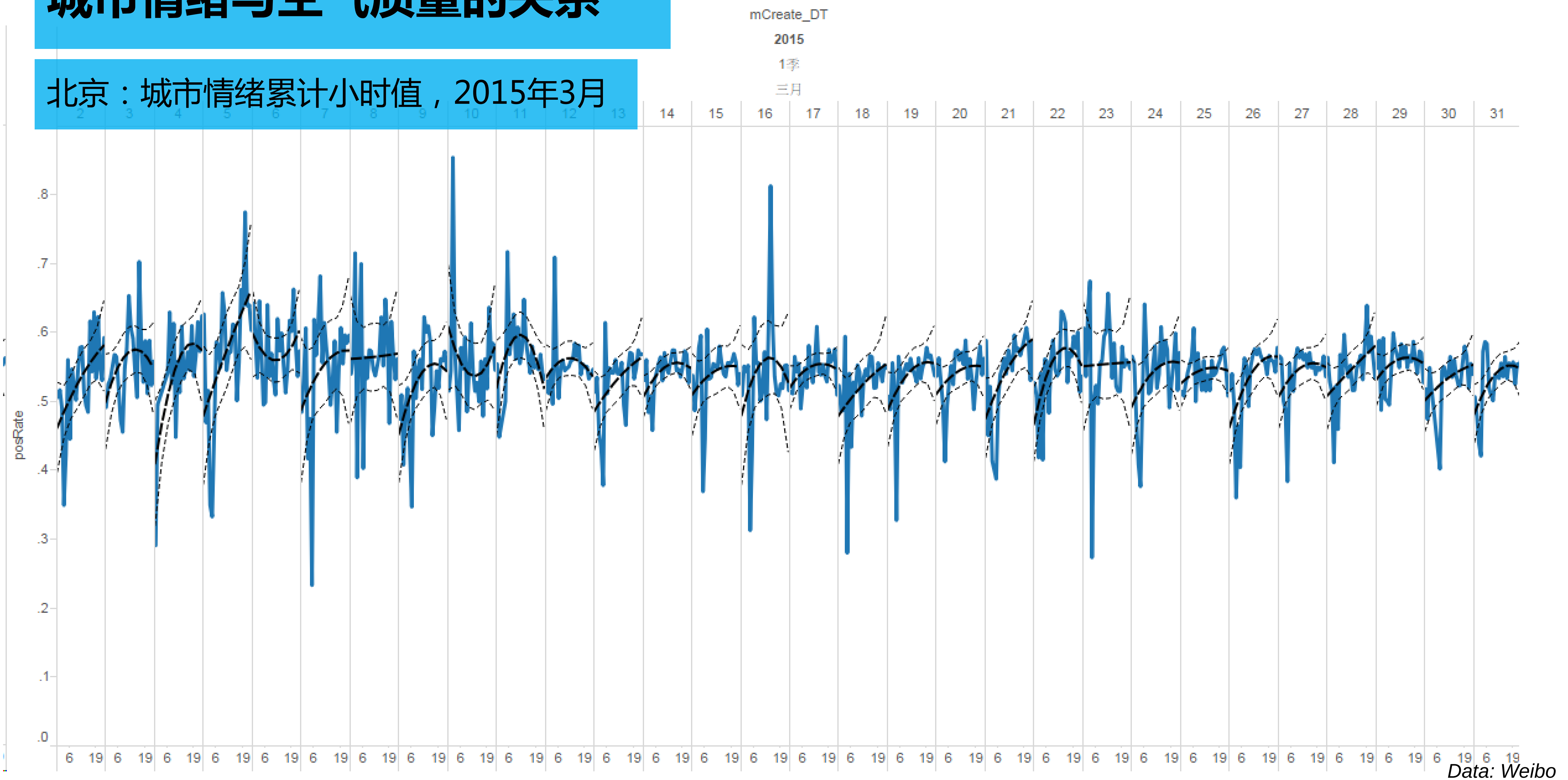
北京：PM2.5小时值，2015年3月



Data: MEP

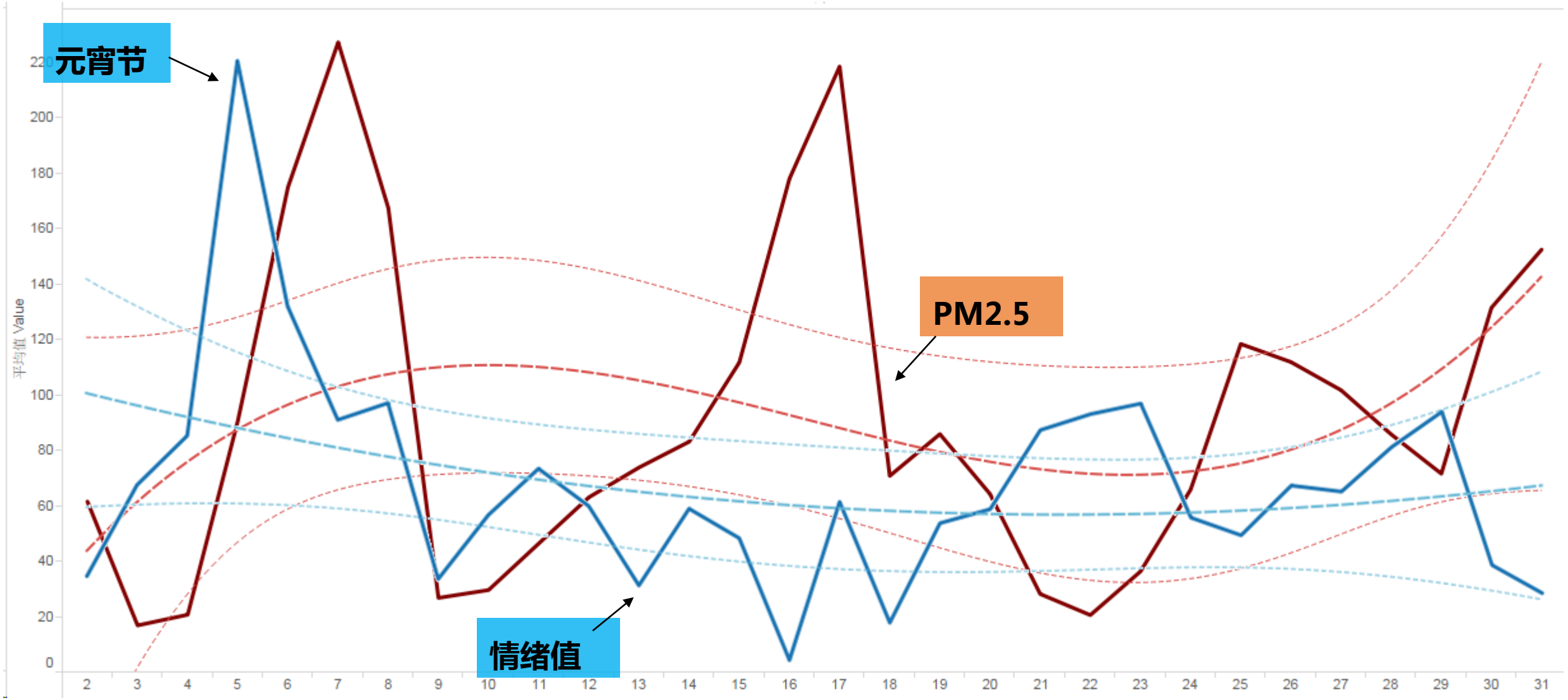
城市情绪与空气质量的关系

北京：城市情绪累计小时值，2015年3月



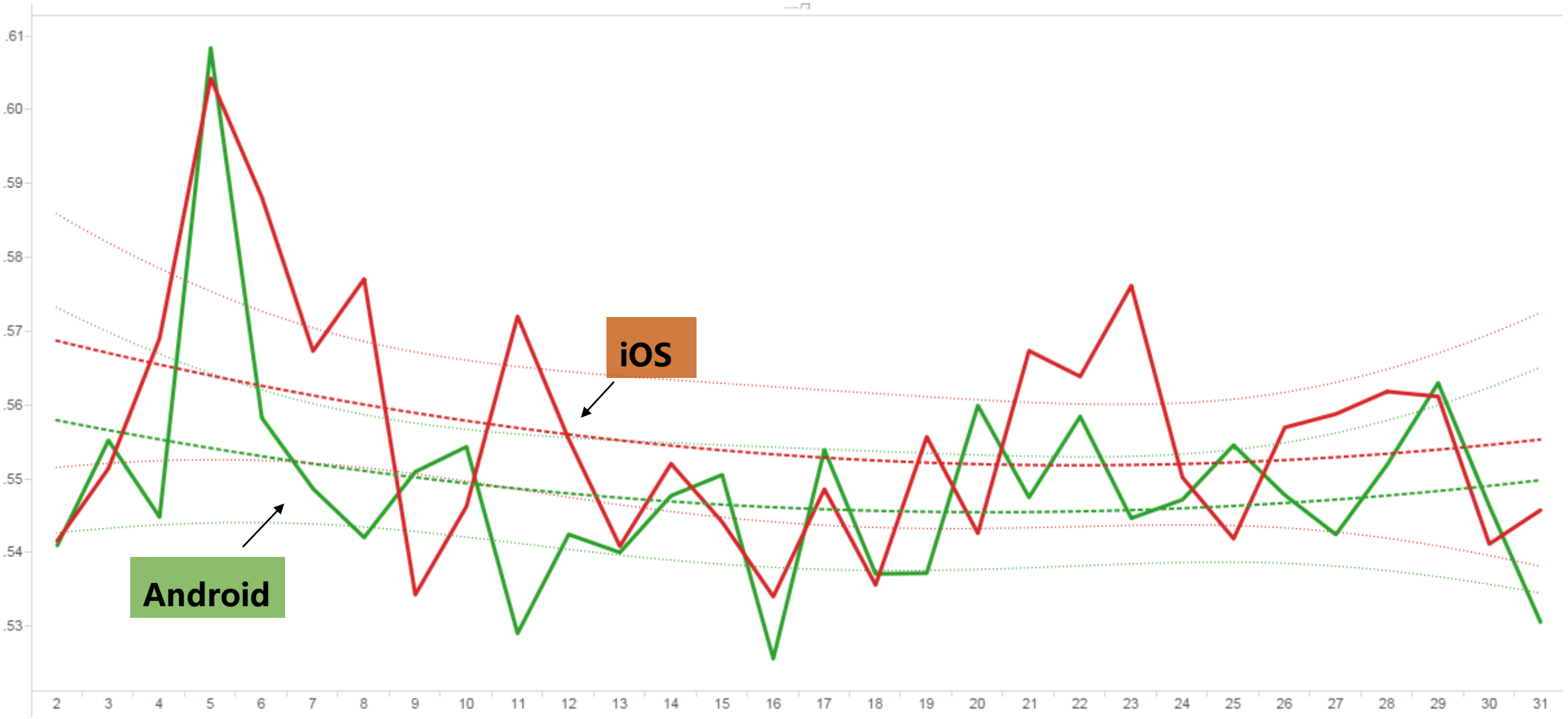
Data: Weibo

城市情绪与空气质量的关系



Data: MEP, Weibo

城市情绪与手机品牌的关系？



Data: Weibo

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

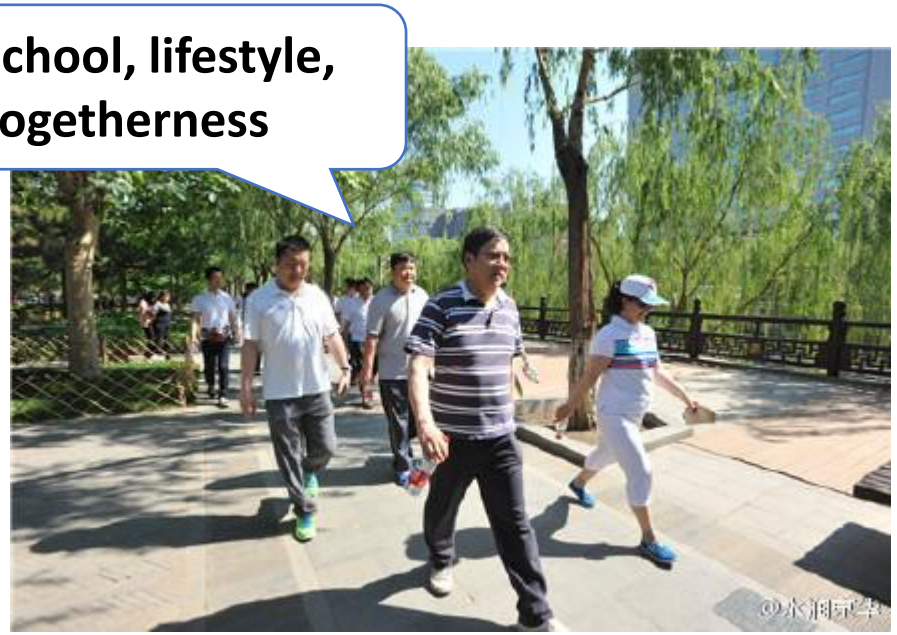
阅读，并理解：图像


```
"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

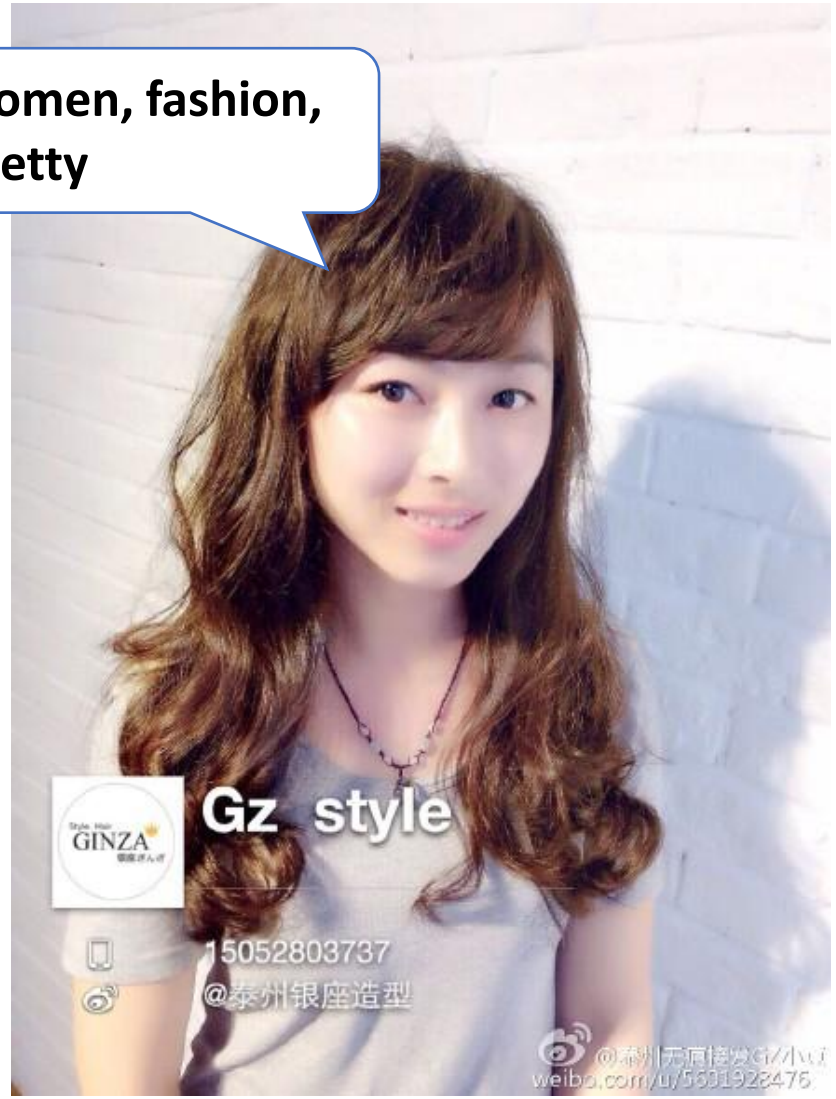
阅读，并理解：图像




```
"classes": [                    "probabilities": [
    • "school",                     • 0.9909061193466187,
    • "lifestyle",                  • 0.9887098670005798,
    • "togetherness",               • 0.9805930852890015,
    • "people",                     • 0.9792243242263794,
    • "education",                  • 0.9758434295654297,
    • "motion",                     • 0.9687234163284302,
    • "friendship",                 • 0.9673053026199341,
    • "daytime",                    • 0.9599980115890503,
    • "fun",                        • 0.9592617750167847,
    • "female",                     • 0.9485963582992554,
    • "child",                      • 0.9463801383972168,
    • "recreation",                 • 0.9411329627037048,
    • "happiness",                  • 0.9400149583816528,
    • "road",                       • 0.9381043314933777,
    • "city",                       • 0.9364811182022095,
    • "north america",              • 0.9325109720230103,
    • "leisure",                    • 0.9296536445617676,
    • "family",                     • 0.9287199378013611,
    • "enjoyment",                  • 0.923075258731842,
    • "politics"                    • 0.9210243225097656
],                                  ]
```

Data: Weibo

阅读，并理解：图像



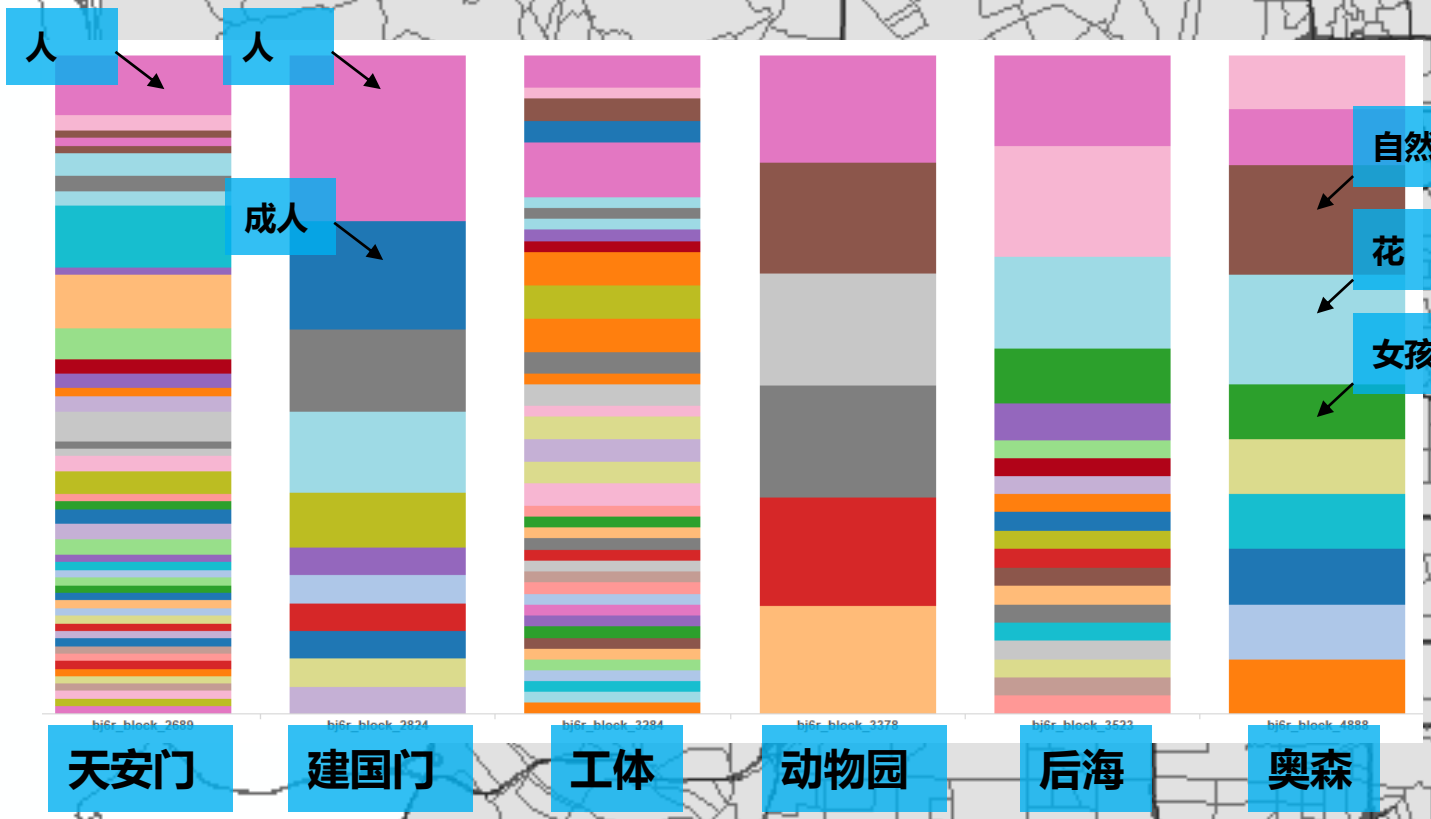

```
"classes": [
    • "women",
    • "fashion",
    • "pretty",
    • "cute",
    • "glamour",
    • "sexy",
    • "eyes",
    • "children",
    • "teens",
    • "elegant",
    • "looking",
    • "makeup",
    • "young",
    • "long",
    • "sensuality",
    • "casual",
    • "one",
    • "precious",
    • "femininity",
    • "skin"
],
```

```
"probabilities": [
    • 0.9893248677253723,
    • 0.9867547750473022,
    • 0.985087513923645,
    • 0.9745194315910339,
    • 0.9724042415618896,
    • 0.9660674929618835,
    • 0.965543270111084,
    • 0.9647226333618164,
    • 0.963720440864563,
    • 0.9463629722595215,
    • 0.9438561201095581,
    • 0.9430670738220215,
    • 0.9316335916519165,
    • 0.9229427576065063,
    • 0.9212859869003296,
    • 0.9196590185165405,
    • 0.9190796613693237,
    • 0.9188071489334106,
    • 0.918481707572937,
    • 0.9108995199203491
]
```

Data: Weibo

不同地块图像主题的差异

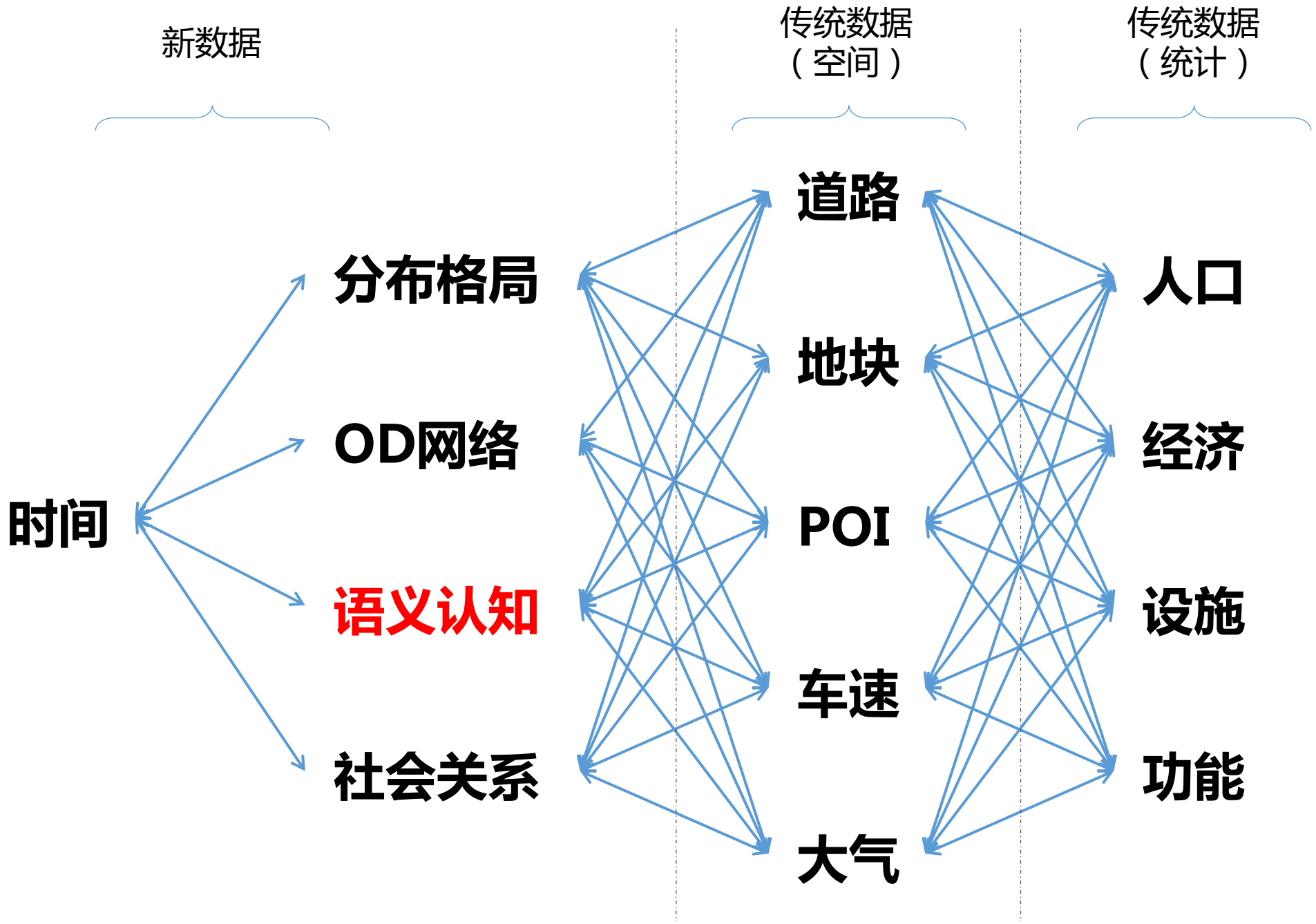
北京：2天大约2.8万张照片

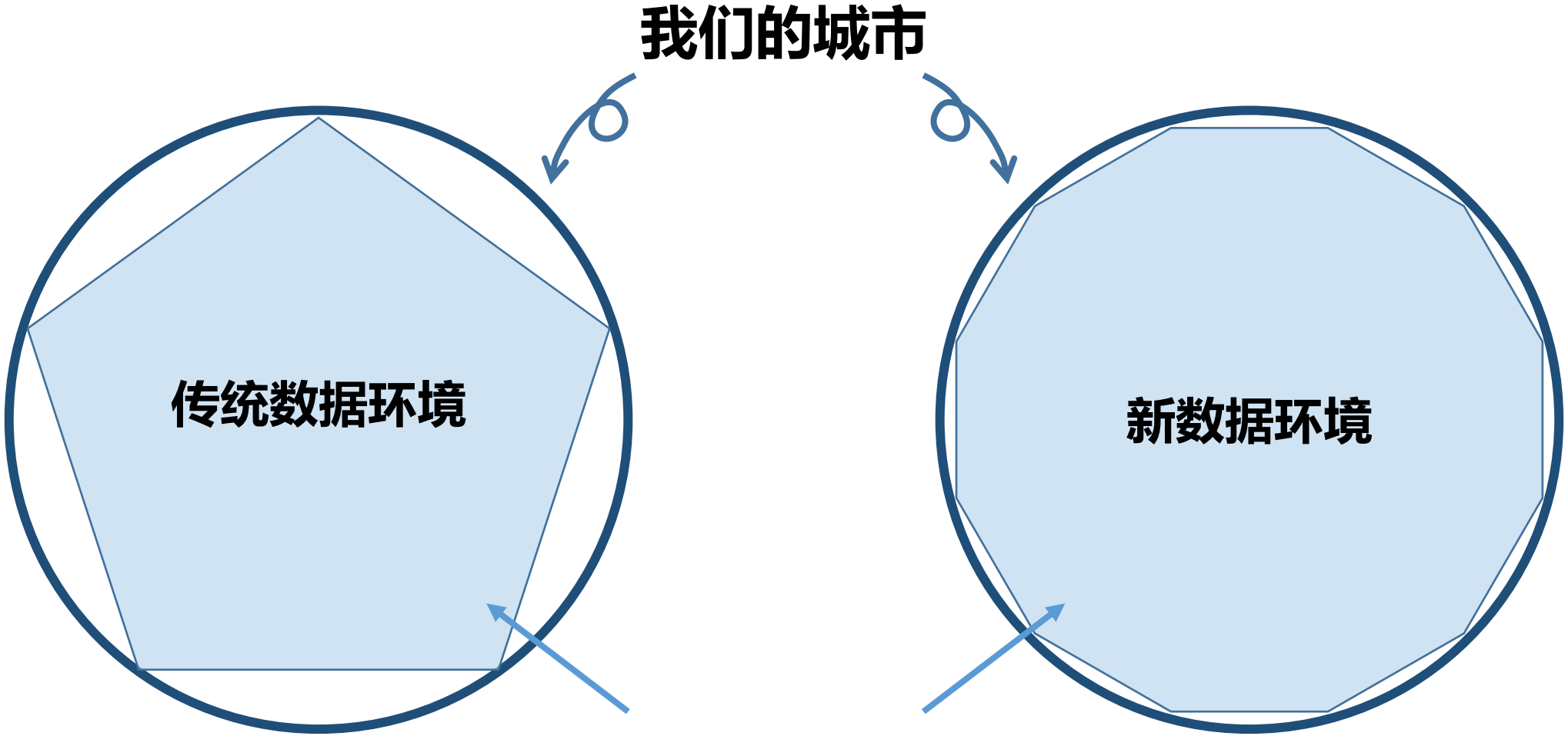


Data: Weibo

- 众包数据城市分析框架的四个方向

1. 分布格局
2. 移动轨迹
3. 语义认知：文本、图像
4. 社会关系


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
社会关系
道路
地块
POI
车速
大气
人口
经济
设施
功能


我们的城市
传统数据环境
新数据环境
理解的手段：研究、
认识和分析

城市光谱：空间边界有限、内部变化无穷

- ***Spectrum*** is a condition that is not limited to a specific set of values but can vary infinitely within a continuum.

一个案例/故事

http://en.wikipedia.org/wiki/Spectrum

向左走,向右走 (2003)

城市里的人为什么不是完全理性的人？

——忍受污染、房价、拥堵……

——寻求职业、发展、交友等机会

如何度量潜在的社会交往？

——以及与空间/场所的关系？

向左走,向右走 (2003)

基于时空约束的共现搜索

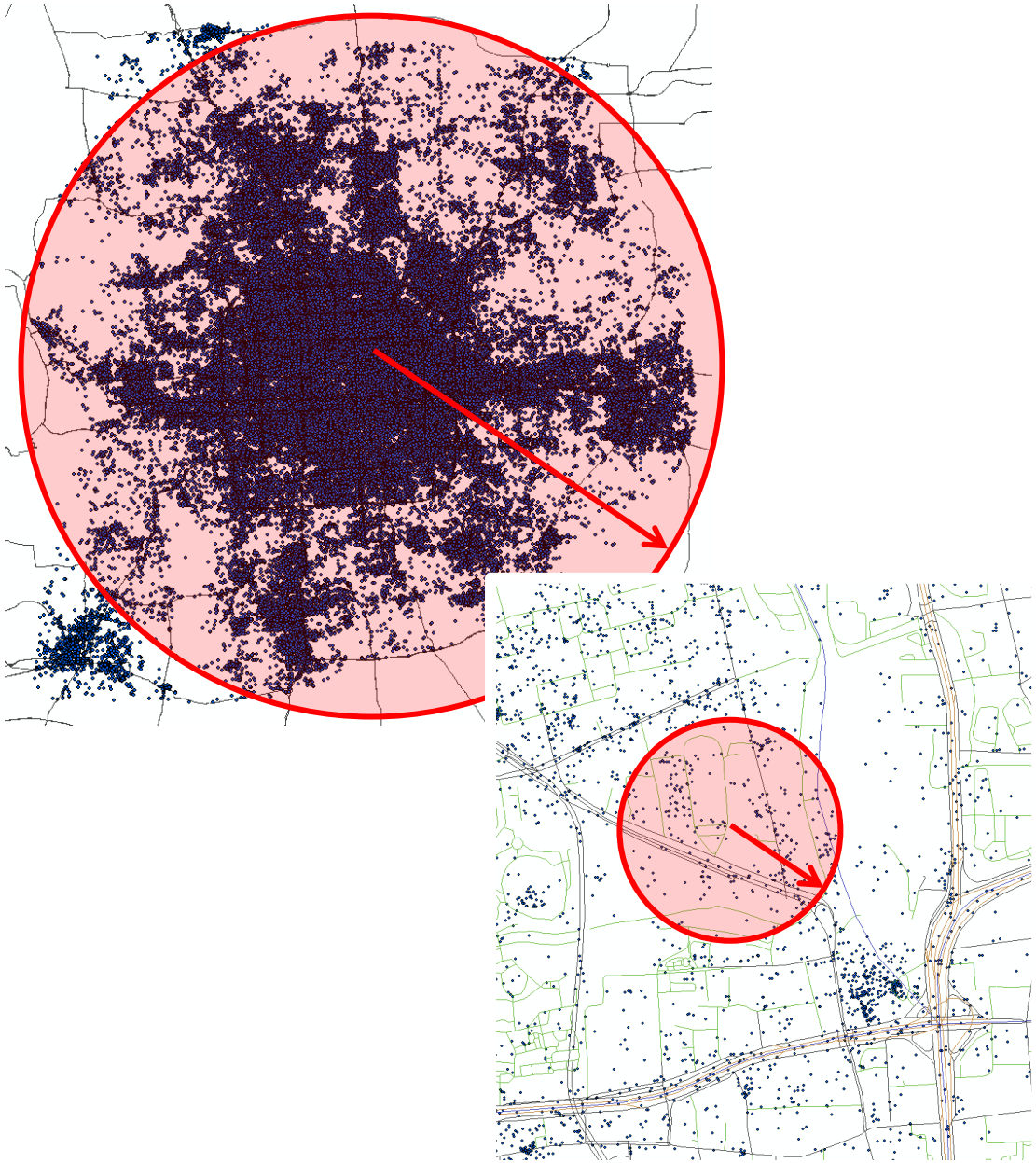
全量计算无法实现：

e.g. 500000 * 500000 = 2500亿

$$PAIR_{ij}$$

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

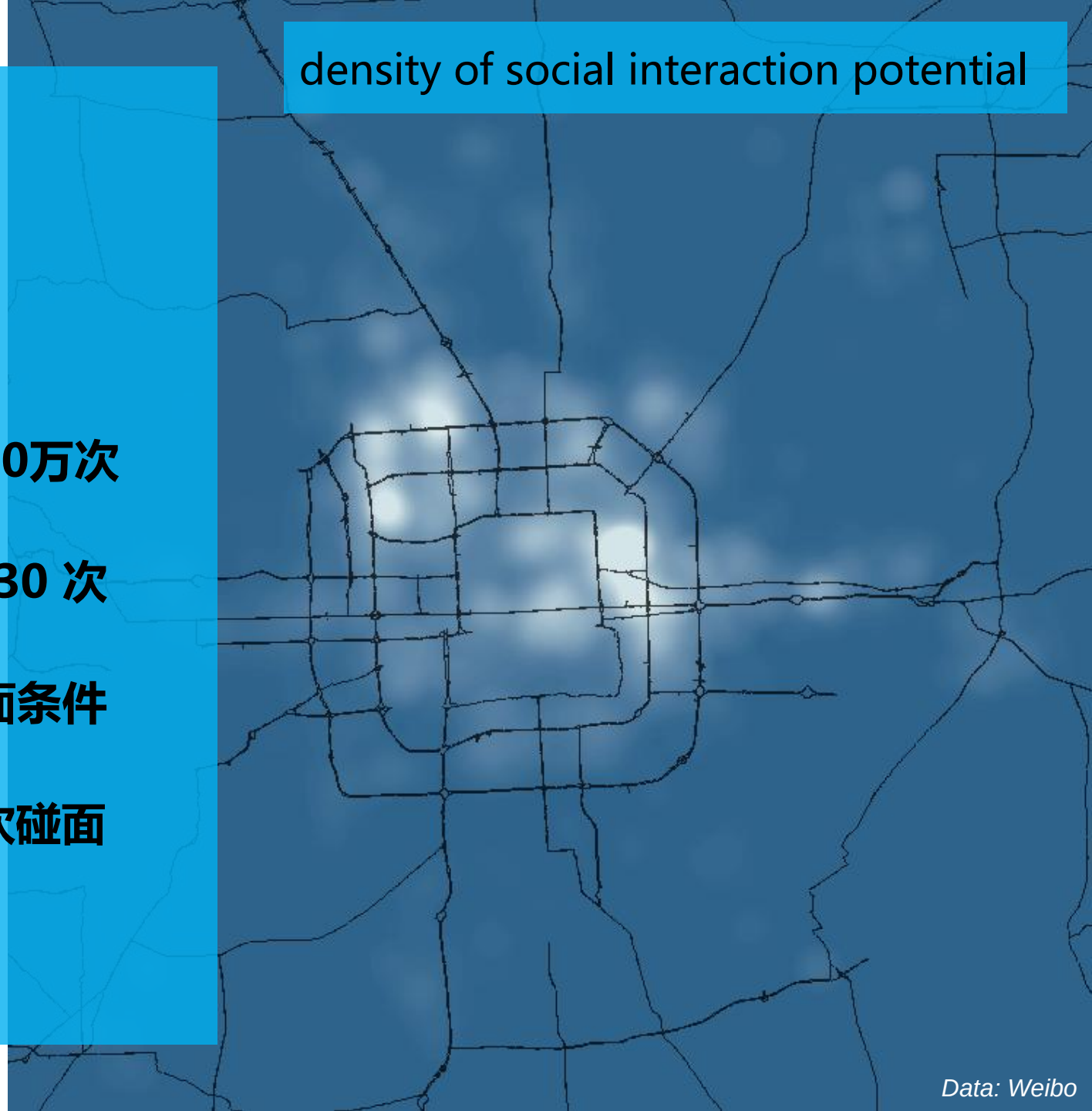
$$T_{diff} = |T_i - T_j| < 24\ h$$

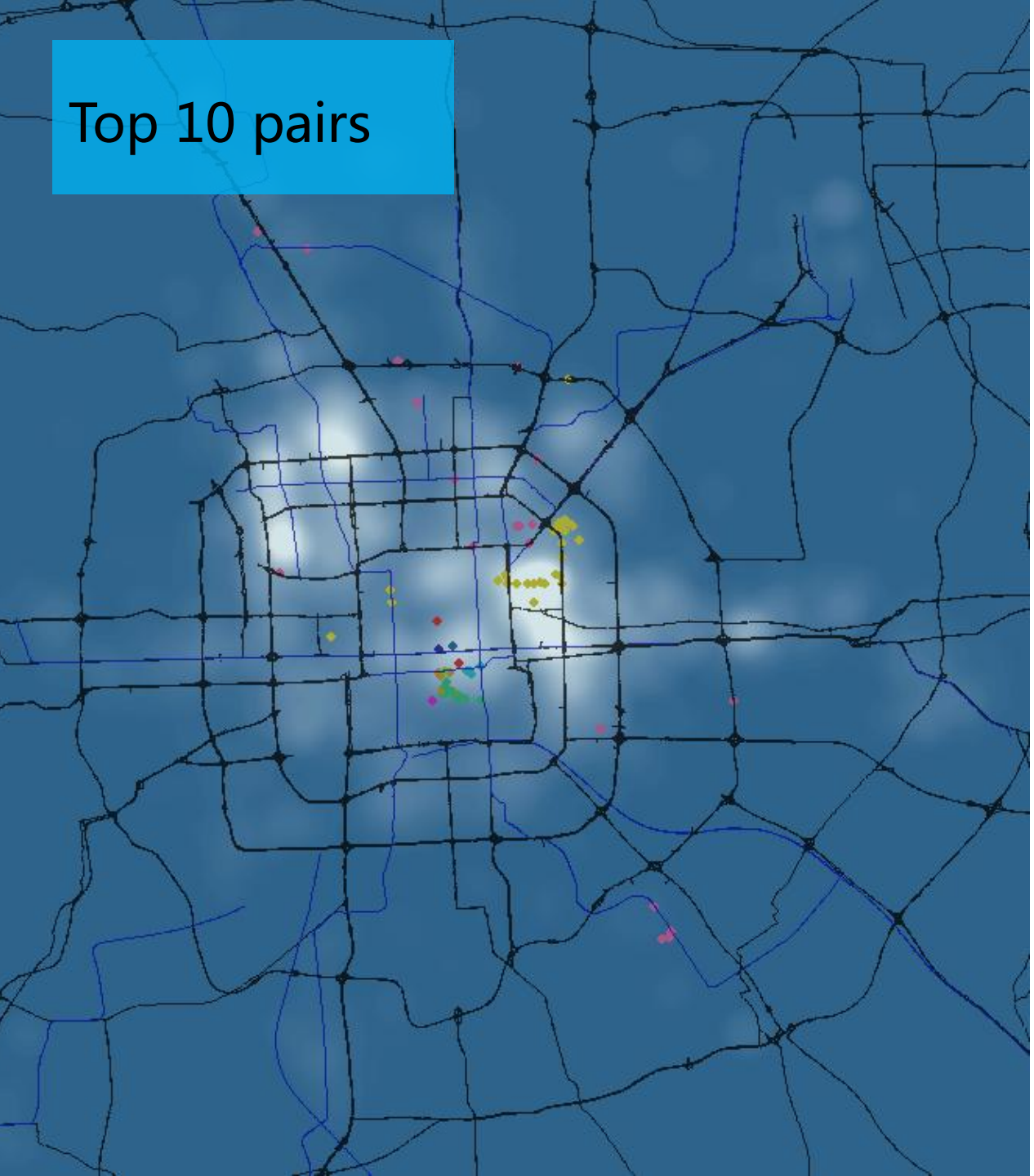


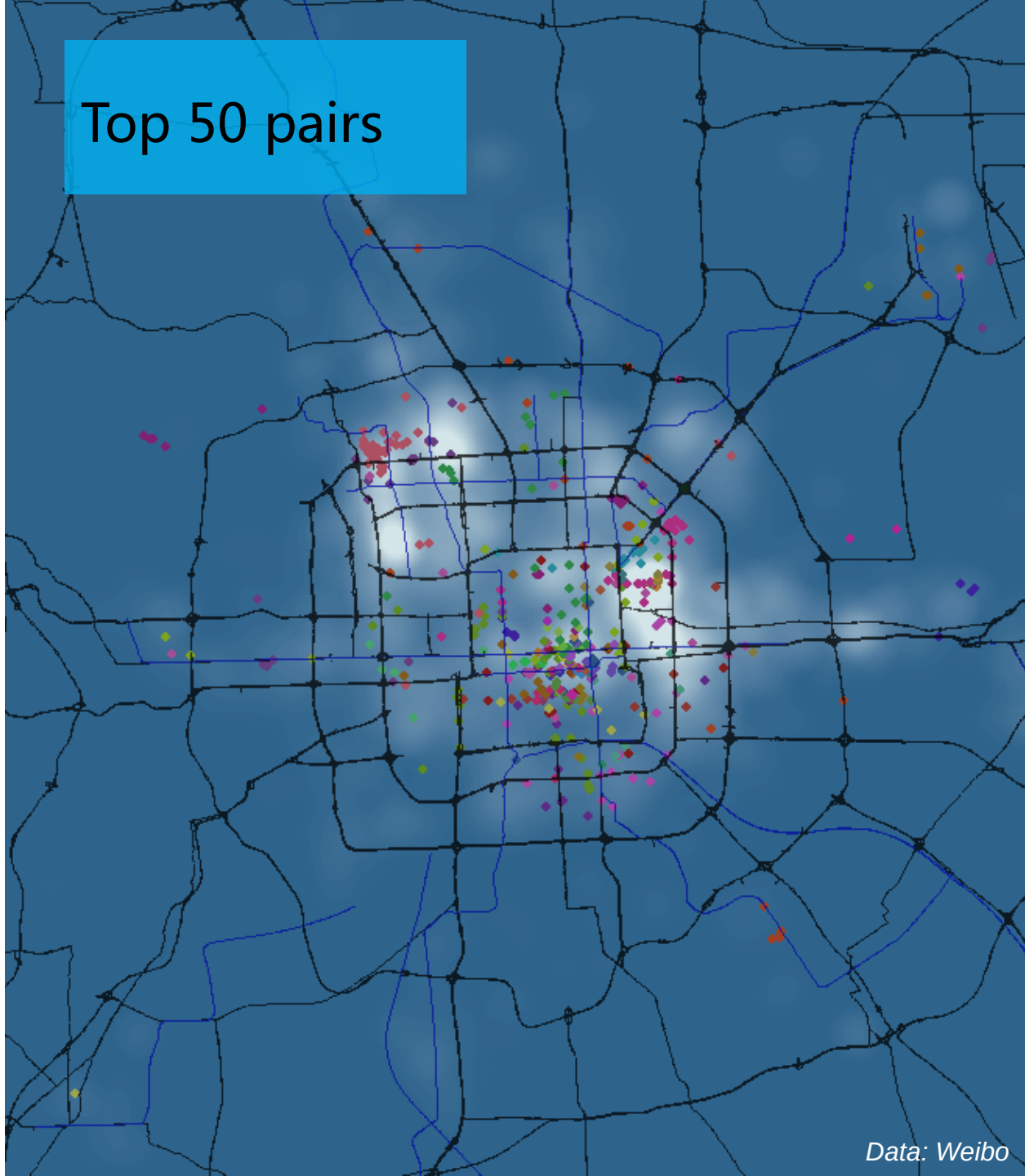
Data: Weibo

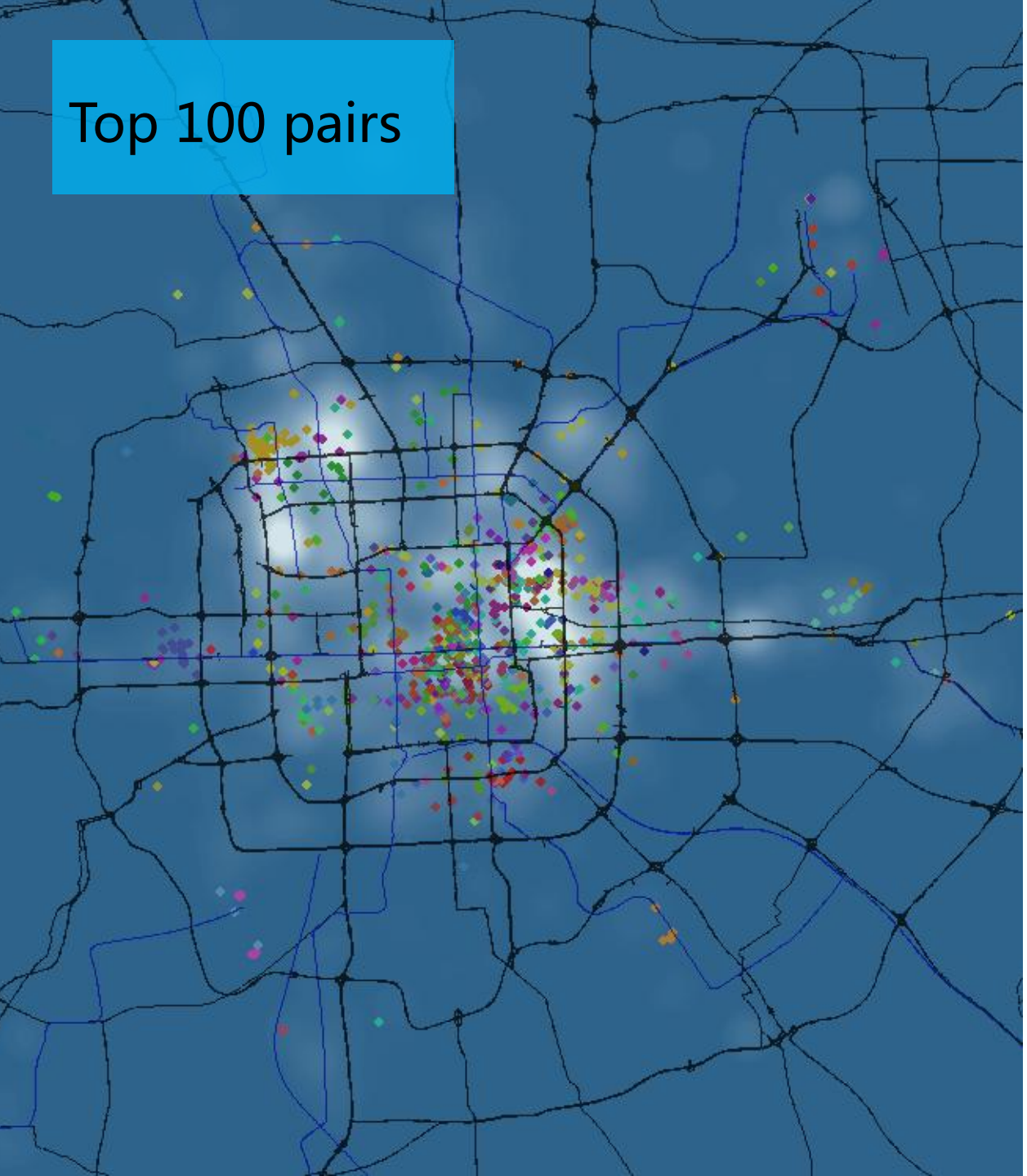
两个月里、六环内：

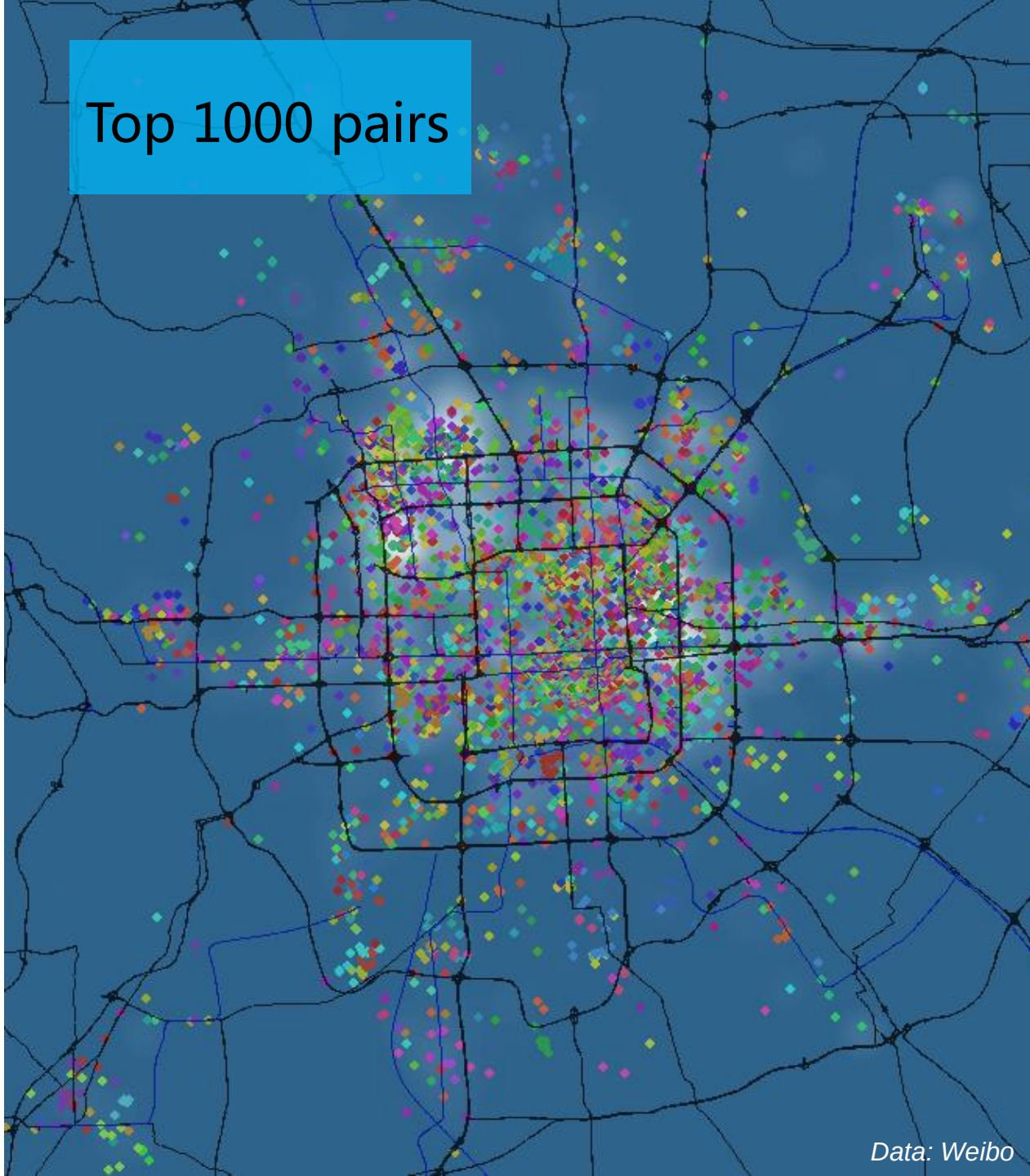
- 潜在碰面 14,004,320 组次
- 潜在碰面 18,137,910 人次，每天约30万次
- 最有“缘分”的两人“擦肩而过” 2330 次
- 最多时一个人身边有 6376 人满足碰面条件
- 最多的一个block里发生了 201108 次碰面
- ……

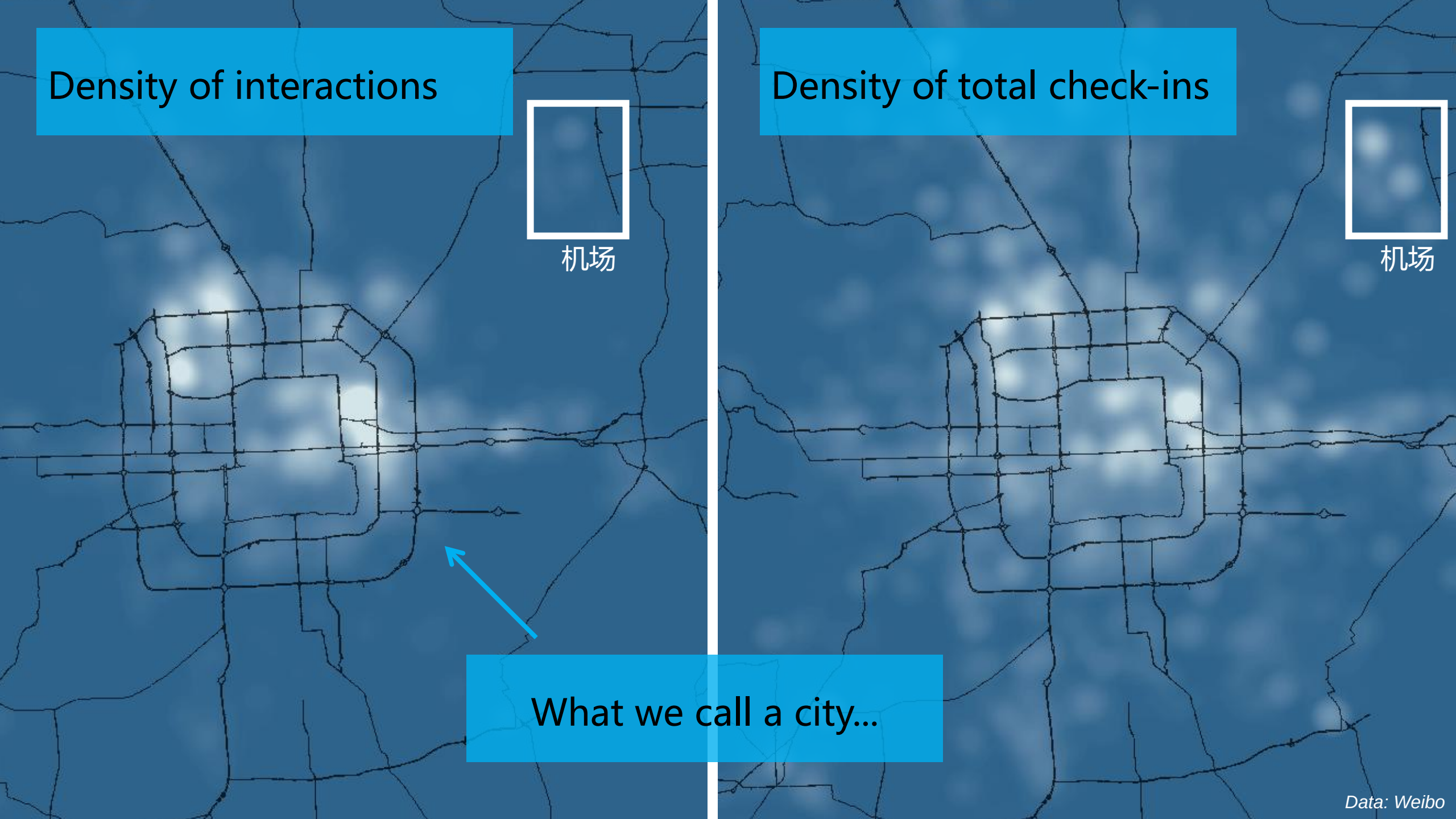



Top 10 pairs


Top 50 pairs
Data: Weibo


Top 100 pairs


Top 1000 pairs
Data: Weibo


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.

跨过基本技术门槛后，数据挖掘需要回归城市研究的本质：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，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

http://weibo.com/uplanning