

北京市统计局  
国家统计局北京调查总队

# 基于移动通信大数据的 人口动态监测



2016年9月

# 主要内容

- 一、利用大数据开展人口动态监测的背景
- 二、移动数据的相关定义
- 三、利用移动数据的监测原理
- 四、大数据测算结果的主要应用
- 五、问题与思考

# 一、利用大数据开展人口动态监测的背景

## 1. 北京人口的特点：

1.流动性大，暂住管理、入户调查等传统方式很难跟上人口变化;

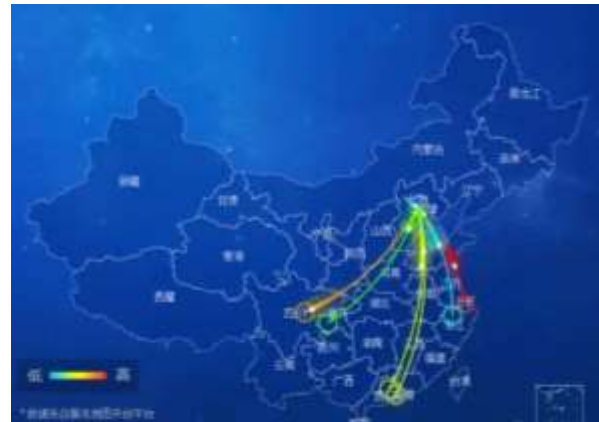
2.行为多变，各种节假日、来京探亲及商务活动等影响因素众多;

3.潮汐效应，工作和居住地集中，上下班路程远影响分区域统计;

4.磁吸效应，工作机会多，拆迁及疏导是否缓解人口压力难评估;

## 2.城市的发展对人口统计提出了新的课题

北京特殊的地位和人文、经济环境，京津冀一体化以及高铁、航空等交通工具的飞速发展，导致目前传统的人口统计方法面临巨大的挑战。



**如何更加精细的洞察人口微观变化？如何为调控人口提供辅助决策？当前的形势发展为人口统计提出了新的课题。**

## 二、移动数据的相关定义

### 1. “信令数据”的基本概念

信令（ Signaling ）是指通信系统中的控制指令，它可以指导终端、交换系统及传输系统协同运行，在指定的终端之间建立临时的通信信道，并维护网络本身正常运行。信令系统是通信网的重要组成部分，是通信网的神经系统。

## 2. “信令数据” 包含的内容

2/3/4G 网络的手机信令数据主要包含以下内容：

- **加密后的用户标识**：可理解为加密脱敏后手机用户的ID，且全网唯一；
- **时间戳**：记录了信令事件的发生时间，可表征用户的时间信息；
- **小区编号**：信令事件发生时所占用的基站小区，可表征用户的位置信息；
- **信令事件类型**：记录了用户手机业务类型，如挂机、发短信、接短信、主叫、被叫、正常位置更新、上网流程等。

### 3. 对各类用户的界定

- **在网用户**：当月有信令产生（捕捉到至少一条位置信息）。
- **有效用户**：一天内在北京停留10小时以上。
- **稳定用户**：一个月内有15天以上，每天在北京停留10小时以上。
- **白天用户**：每天7点至19点间，有效用户在某行政区内停留时间最长，则计入该区的白天用户。
- **夜晚用户**：每天21点至次日7点间，有效用户在某行政区内停留时间最长，则计入该区的夜晚用户。

### 三、利用移动数据的监测原理

利用移动公司覆盖全市16个区的4万余个基站，每隔一定时间，所捕获到的手机信号，对持机用户在我市及各行政区的停留时间、位置、停留时长等进行监测，再按照不同的定义标准，计算相应的手机用户数。



# 1.手机信号监测——每日有效停留人口

由于手机使用有关机、停电等行为，故对手机信号日监测时长做适当调整，累计停留时间10小时以上，即为该区域的当日有效停留人口。

我们分别对6、8、10、12、15小时进行了测算，10小时后基本趋于稳定。

## 2.手机信号监测——月度驻留人口

与年度常住人口相对应，时间取半，定义一个月内，在一定区域有效停留15天以上的手机用户为手机监测月度驻留人口（实际监测中，我们对各停留天数均有统计）。

# 移动手机监测有效停留手机用户数分布情况(某月)

单位：部

驻留天数	手机用户数	累计	驻留天数	手机用户数	累计
31	6576506	6576506			
30	1553237	8129743	15	432497	17045520
29	1157635	9287378	14	433629	17479149
28	934732	10222110	13	440038	17919187
27	794246	11016356	12	453769	18372956
26	704052	11720408	11	473848	18846804
25	623480	12343888	10	449358	19296162
24	566486	12910374	9	597865	19894027
23	502951	13413325	8	520809	20414836
22	539949	13953274	7	615762	21030598
21	452732	14406006	6	724379	21754977
20	450928	14856934	5	908527	22663504
19	445204	15302138	4	1224014	23887518
18	443978	15746116	3	1649941	25537459
17	435300	16181416	2	2472485	28009944
16	431607	16613023	1	3601935	31611879

### 3. 关于“常住人口”的定义

“常住人口”是指实际经常居住在某地区半年以上（含半年）的人口。主要包括：除离开本地半年以上（不包括在国外工作或学习的人）常住本地的户籍人口；户口在外地但在本地居住半年以上者；或离开户口登记地半年以上而调查时在本地居住的人口。

# 基于移动通信大数据的人口研究

以移动通信数据为基础，结合传统统计学的入户调查、数据插补分析等方法，通过全新的人口数据挖掘、行为和交通通勤轨迹链分析，为北京市人口进行全面的时间、空间、行为三维“画像”。



**完整**

手机用户只要一产生通信行为，就会在网络上留下“印记”，从而完整识别出北京境内任何时间、任何地点的常驻及流动人口。

**精准**

基于通信基站位置信息，可精准识别出手机用户的工作、居住地，分区县、分街区流动线路，确定拆迁、热点区域的人口流向。

## 四、大数据测算结果的主要应用

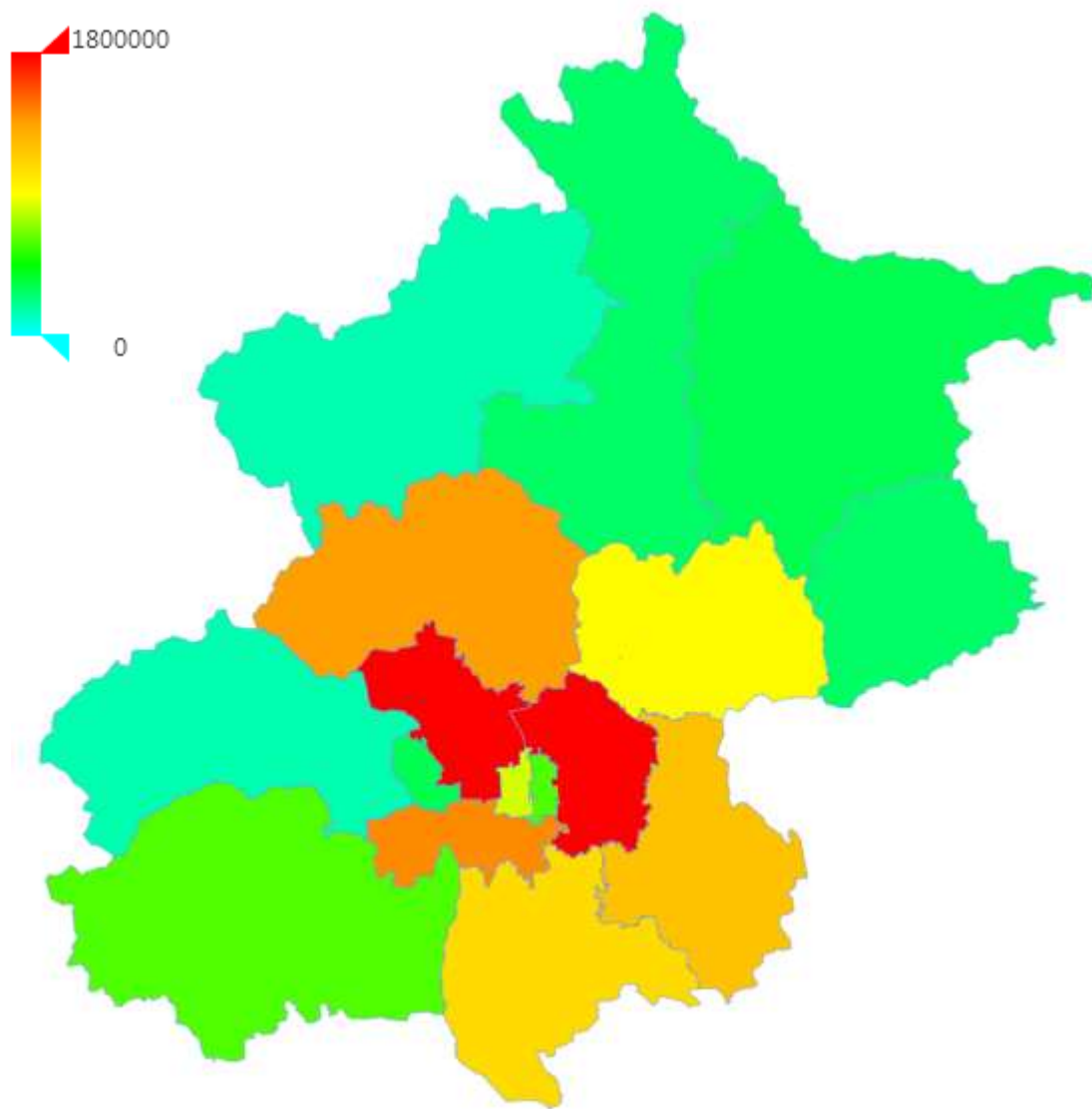


# 1.用于常住人口总量评估

## 评估前提假设

将每个月的移动手机稳定用户数量，视同为相应行政区划内的驻留人口，根据全市及各区驻留人口增减变化趋势，推算人口统计常住人口变动情况。

# 1.用于常住人口总量评估



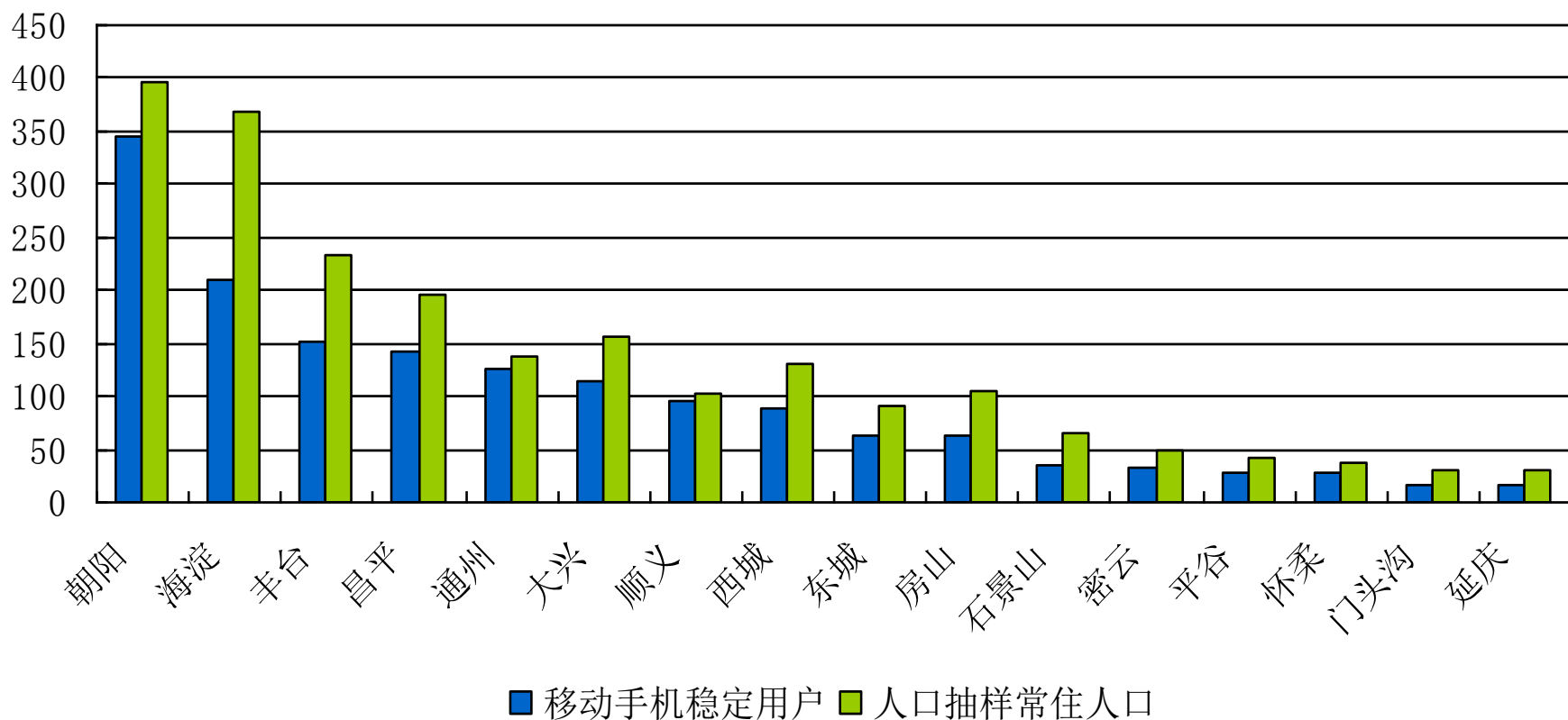
行政区	移动手机 稳定用户	比重 (%)
全市	15592206	100
朝阳	3450698	22.1
海淀	2105396	13.5
丰台	1517495	9.7
昌平	1433054	9.2
通州	1265677	8.1
大兴	1134192	7.3
顺义	958026	6.1
西城	888417	5.7
东城	639142	4.1
房山	634349	4.1
石景山	346331	2.2
密云	332846	2.1
平谷	281747	1.8
怀柔	279777	1.8
门头沟	164339	1.1
延庆	160720	1.0



# 1.用于常住人口总量评估

图1：各区移动手机稳定用户数与人口抽样调查常住人口数对比

单位：万人

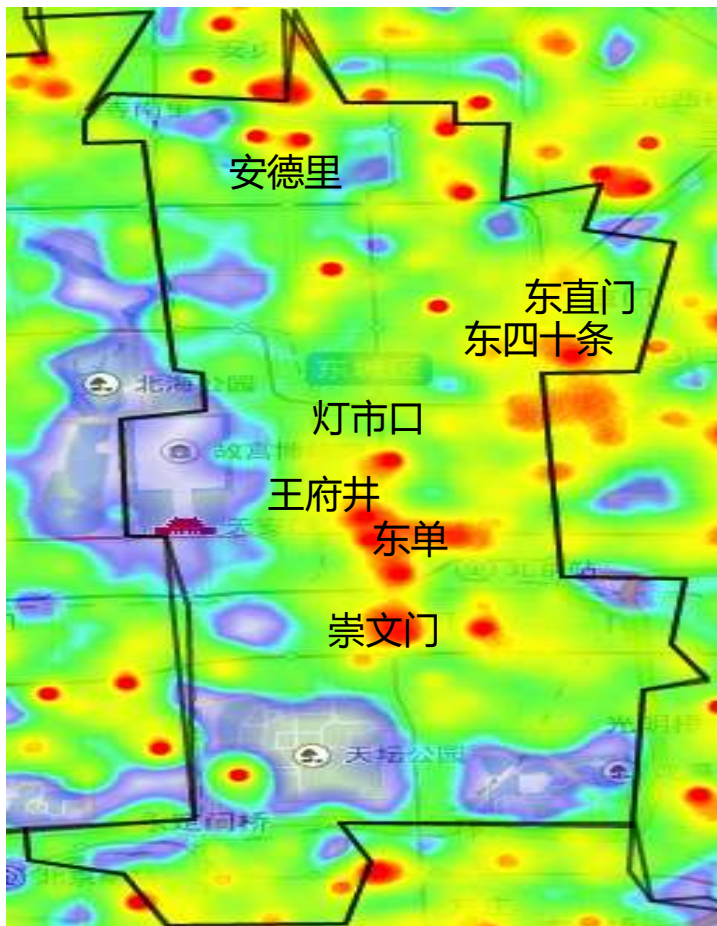


移动手机稳定用户数量与各区常住人口数不尽相同，但各区比重排名大体一致。

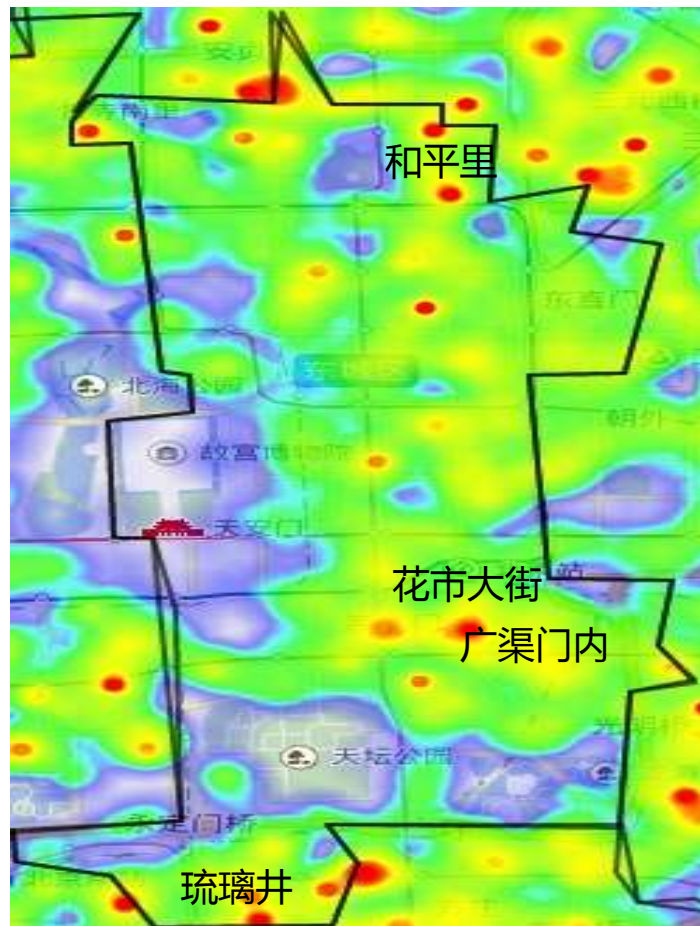
## 2.用于白天夜晚人群分布

### 东城区

白天



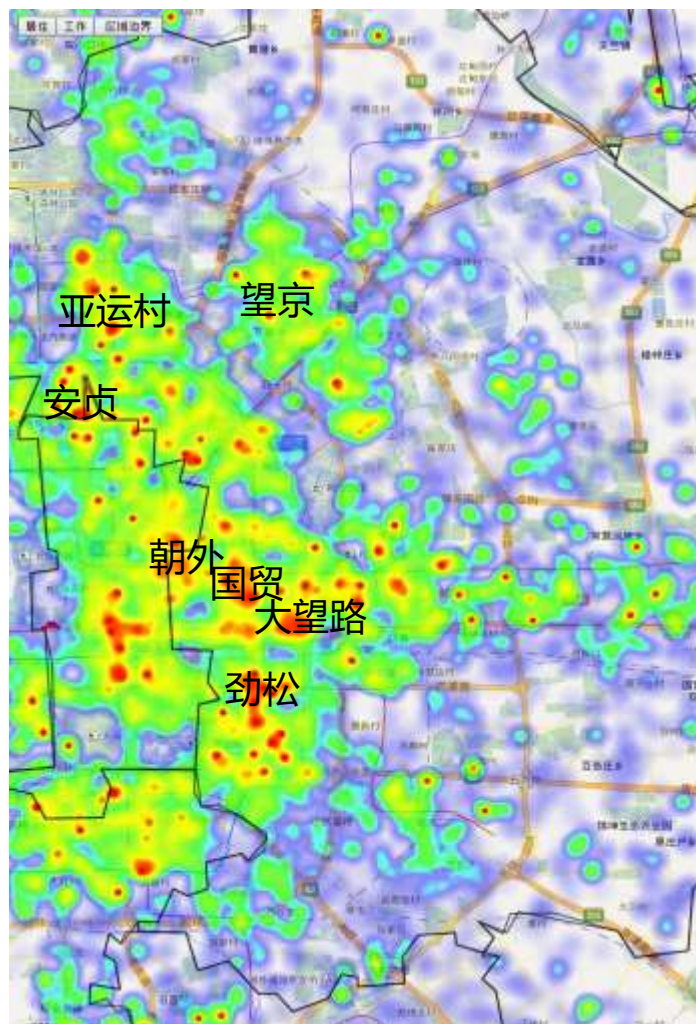
晚上



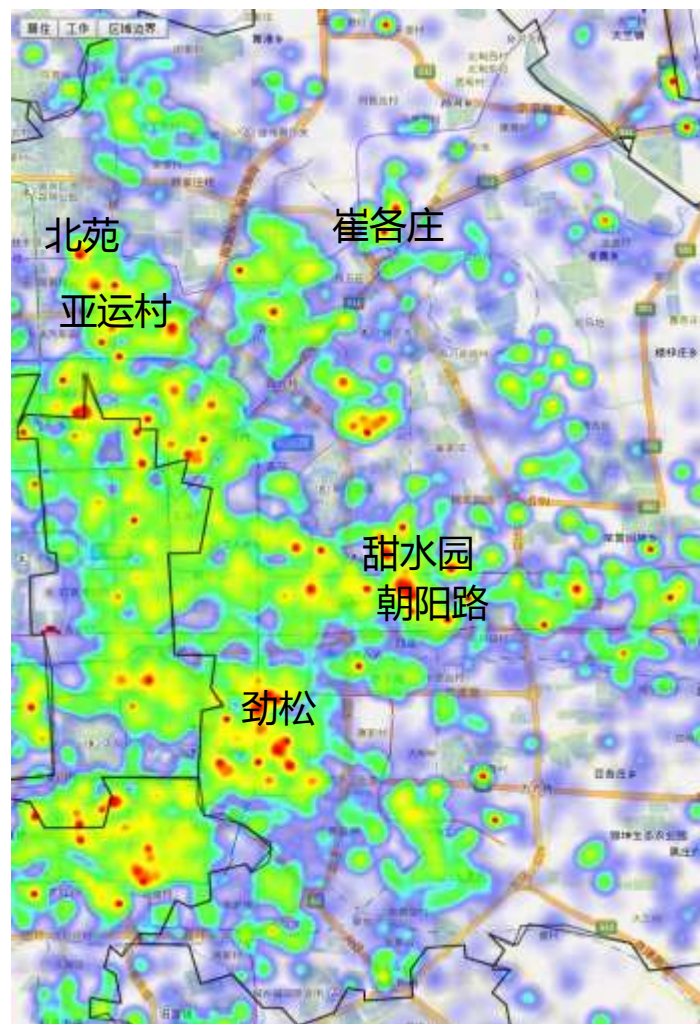
## 2.用于白天夜晚人群分布

### 朝阳区

白天



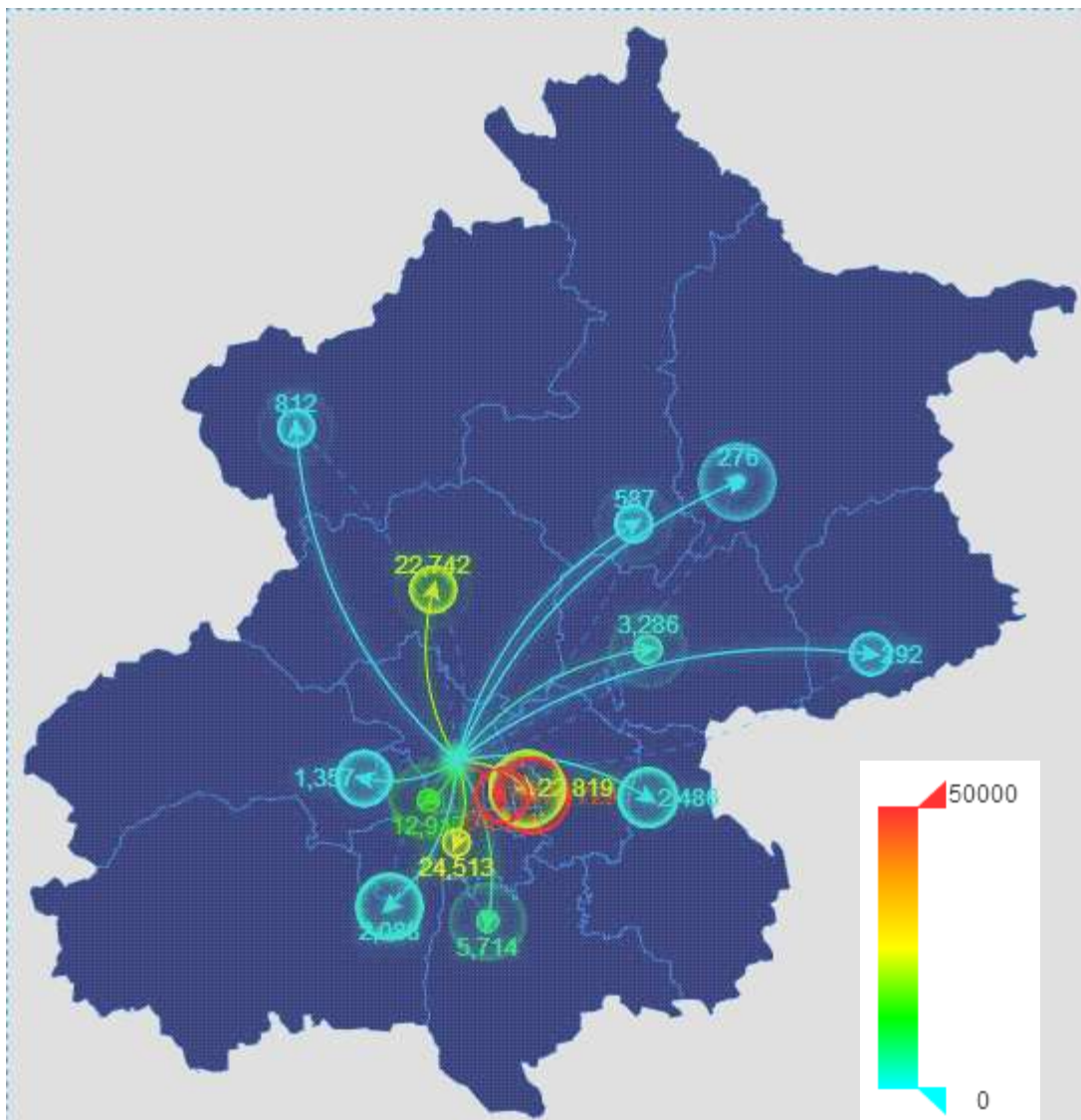
晚上



### 3.用于职住分离情况（海淀区）

海淀区夜晚用户白天通勤目的地

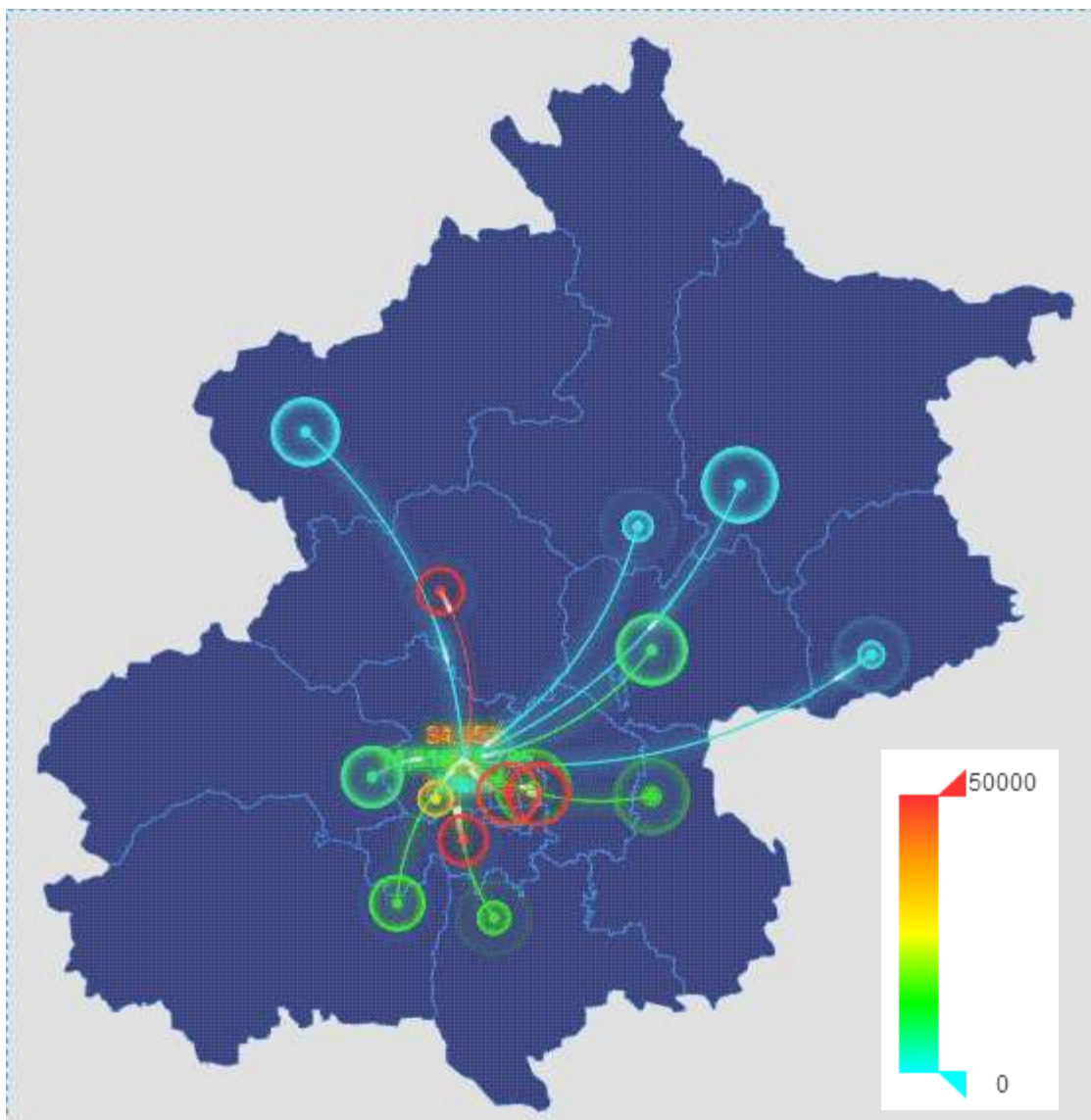
居住地	工作地	人数	占比
海淀	朝阳	72729	30.73%
海淀	西城	64071	27.07%
海淀	丰台	24513	10.36%
海淀	东城	22819	9.64%
海淀	昌平	22742	9.61%
海淀	石景山	12917	5.46%
海淀	大兴	5714	2.41%
海淀	顺义	3286	1.39%
海淀	通州	2486	1.05%
海淀	房山	2088	0.88%
海淀	门头沟	1357	0.57%
海淀	延庆	812	0.34%
海淀	怀柔	587	0.25%
海淀	平谷	292	0.12%
海淀	密云	276	0.12%
合计		236689	100.00%



### 3.用于职住分离情况（海淀区）

海淀区白天用户通勤来源地

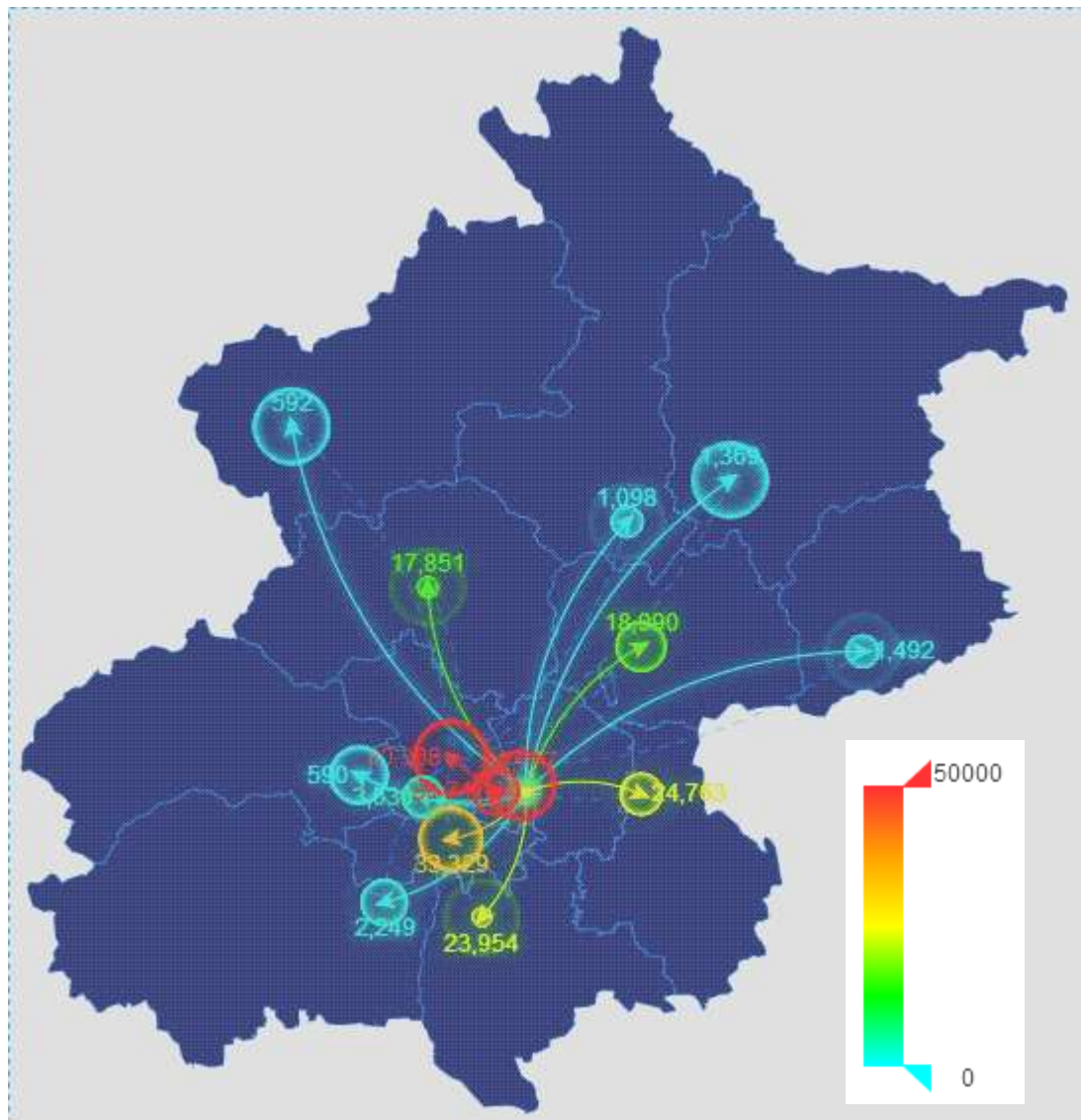
工作地	居住地	人数	占比
海淀	昌平	112379	27.66%
海淀	朝阳	80308	19.76%
海淀	丰台	65910	16.22%
海淀	西城	46642	11.48%
海淀	石景山	31652	7.79%
海淀	东城	14634	3.60%
海淀	大兴	14052	3.46%
海淀	通州	11853	2.92%
海淀	房山	11011	2.71%
海淀	顺义	7066	1.74%
海淀	门头沟	6795	1.67%
海淀	密云	1111	0.27%
海淀	平谷	1019	0.25%
海淀	怀柔	980	0.24%
海淀	延庆	940	0.23%
合计		406352	100.00%



# 3.用于职住分离情况（朝阳）

## 朝阳区夜晚用户白天通勤目的地

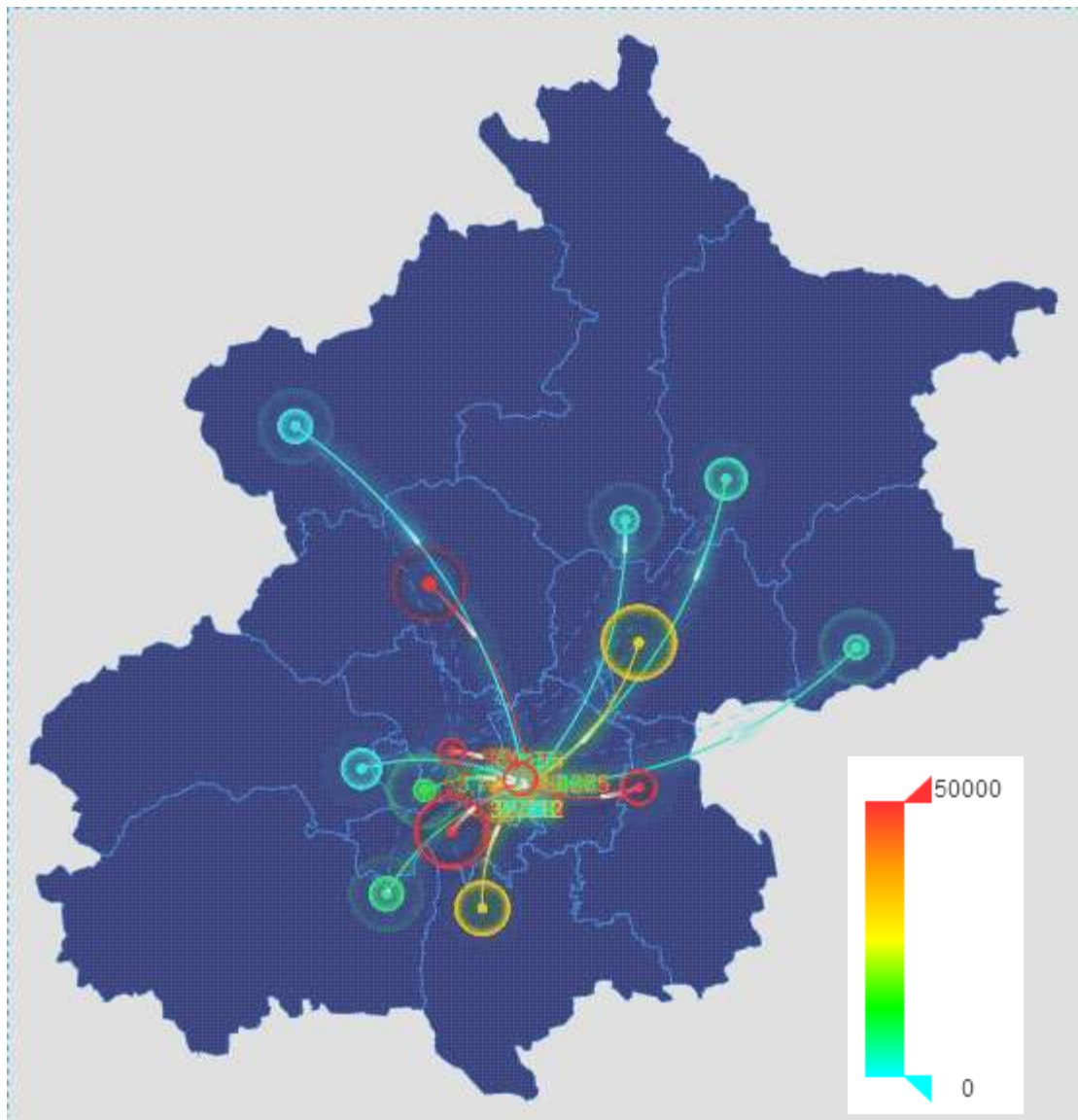
居住地	工作地	人数	占比
朝阳	东城	92677	26.21%
朝阳	海淀	80308	22.71%
朝阳	西城	50717	14.34%
朝阳	丰台	33329	9.43%
朝阳	通州	24763	7.00%
朝阳	大兴	23954	6.77%
朝阳	顺义	18990	5.37%
朝阳	昌平	17851	5.05%
朝阳	石景山	3630	1.03%
朝阳	房山	2249	0.64%
朝阳	平谷	1492	0.42%
朝阳	密云	1369	0.39%
朝阳	怀柔	1098	0.31%
朝阳	延庆	592	0.17%
朝阳	门头沟	590	0.17%
合计		353609	100.00%



### 3.用于职住分离情况（朝阳）

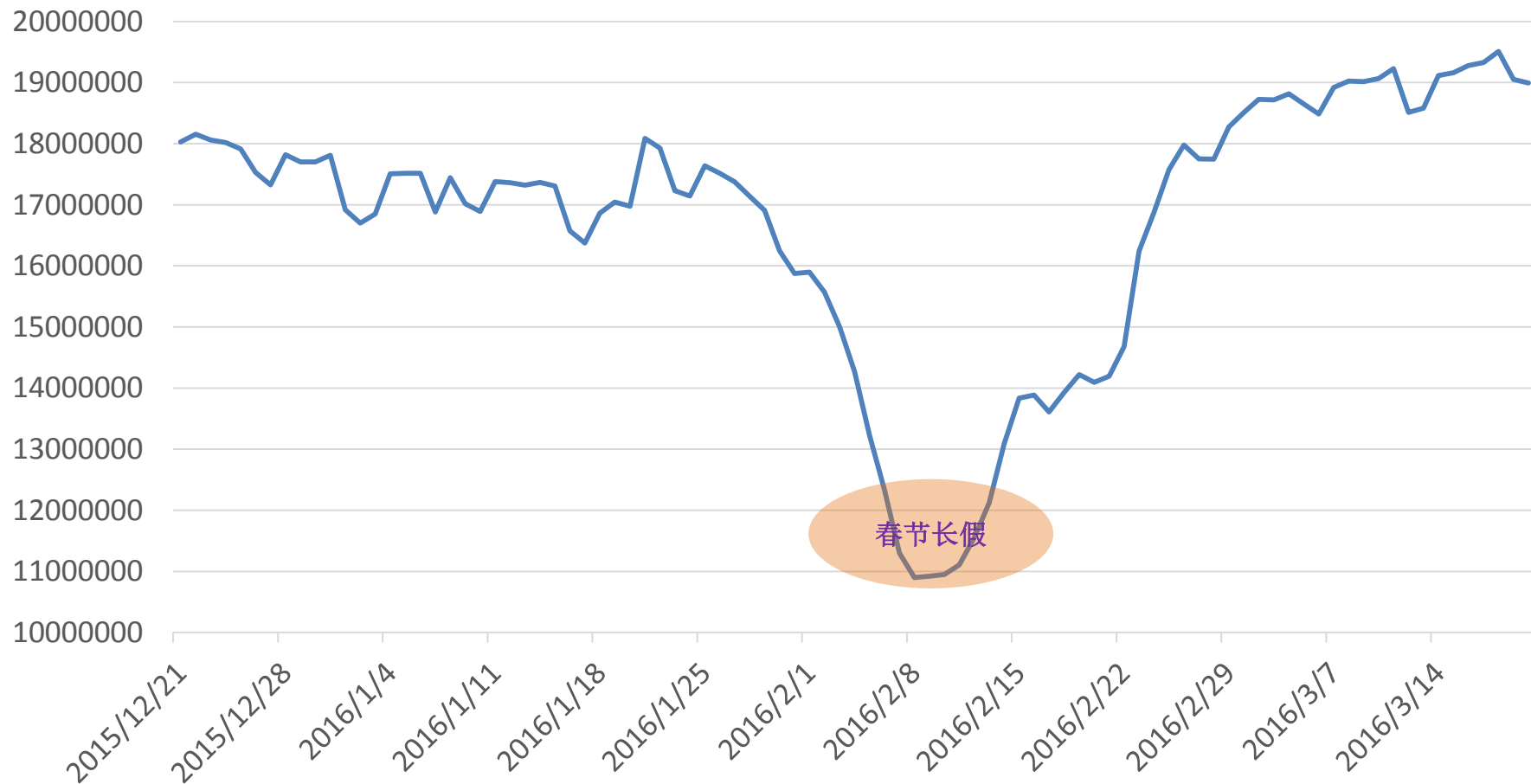
朝阳区白天用户通勤来源地

工作地	居住地	人数	占比
朝阳	通州	97144	18.46%
朝阳	昌平	92328	17.54%
朝阳	海淀	72729	13.82%
朝阳	丰台	68712	13.06%
朝阳	东城	61434	11.67%
朝阳	西城	40275	7.65%
朝阳	大兴	33155	6.30%
朝阳	顺义	32212	6.12%
朝阳	石景山	9436	1.79%
朝阳	房山	6188	1.18%
朝阳	密云	3798	0.72%
朝阳	平谷	3727	0.71%
朝阳	怀柔	2759	0.52%
朝阳	门头沟	1866	0.35%
朝阳	延庆	530	0.10%
合计		526293	100.00%



## 4.用于实时用户动态统计

图2： 移动手机在网用户日监测情况





## 4.用于实时用户动态统计

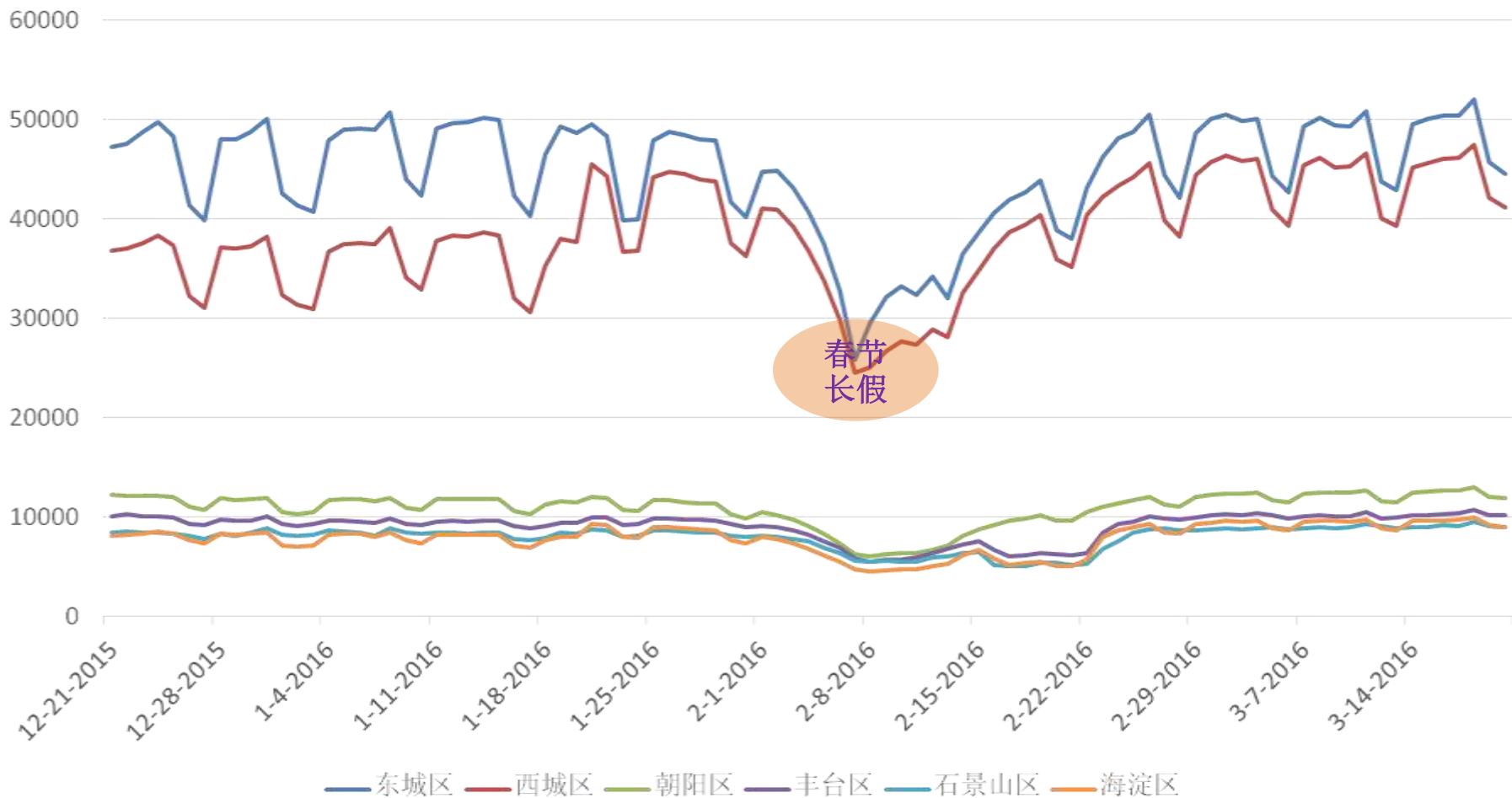
表1：春节长假期间城六区手机用户密度变化情况

单位：部/平方公里

行政区	春节期间 ( 2016.2.7-2.13 )	12月21日-3月20日 ( 剔除春节长假 )	比值
东城区	31218	44204	71%
西城区	26834	38509	70%
朝阳区	9673	16102	60%
丰台区	8935	13311	67%
石景山区	9729	13382	73%
海淀区	9029	14528	62%

# 4.用于实时用户动态统计

图3： 城六区移动手机在网用户密度



## 4.用于实时用户动态统计

表2： 各区平日与节假日手机用户密度对比

单位：部/平方公里

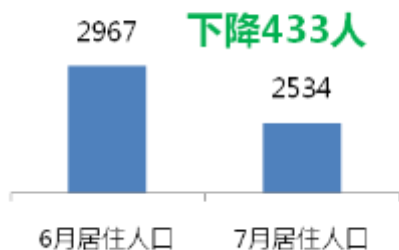
行政区	全月	平时	假日（周六日）	假日/平时
东城	47299	49221	43027	87%
西城	43315	45042	39476	88%
海淀	16804	17324	15646	90%
朝阳	17894	18287	17022	93%
丰台	14612	14766	14273	97%
大兴	6236	6280	6142	98%
石景山	14664	14698	14589	99%
顺义	4689	4692	4673	100%
通州	6209	6195	6247	101%
昌平	5574	5526	5680	103%
怀柔	2923	2891	3018	104%
房山	2865	2814	2977	106%
密云	1268	1241	1315	106%
门头沟	3154	3078	3306	107%
平谷	2521	2446	2677	109%
延庆	1594	1539	1691	110%

## 5. 用于整治拆迁效果跟踪监测

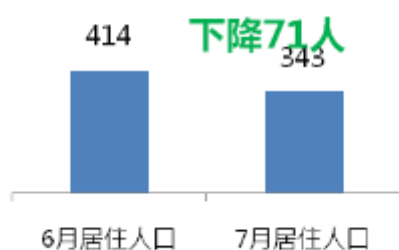
表3

监测重点区域	区县	7月人口数量	6月人口数量	增减数量
天坛区域简易楼	东城区	2534	2967	-433
建外街道永安里社区	朝阳区	343	414	-71
西北旺镇西玉河村	海淀区	1206	1074	132
新发地村	丰台区	14355	15637	-1282

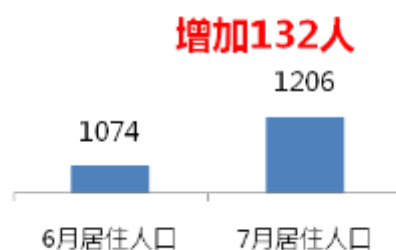
天坛区域简易楼  
(东城区)



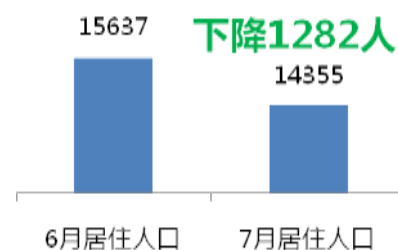
建外街道永安里  
(朝阳区)



西北旺镇西玉河村  
(海淀区)



新发地村  
(丰台区)

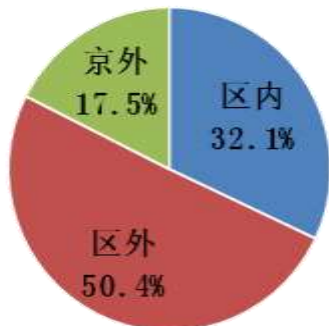


# 5. 用于整治拆迁效果跟踪监测

图4

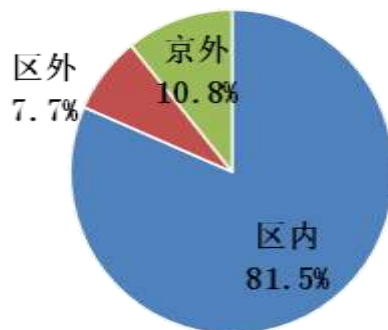
天坛区域简易楼（东城）

腾退人口流向



建外街道永安里（朝阳）

腾退人口流向



西北旺镇西玉河村（海淀）

腾退人口流向



新发地村（丰台）

腾退人口流向



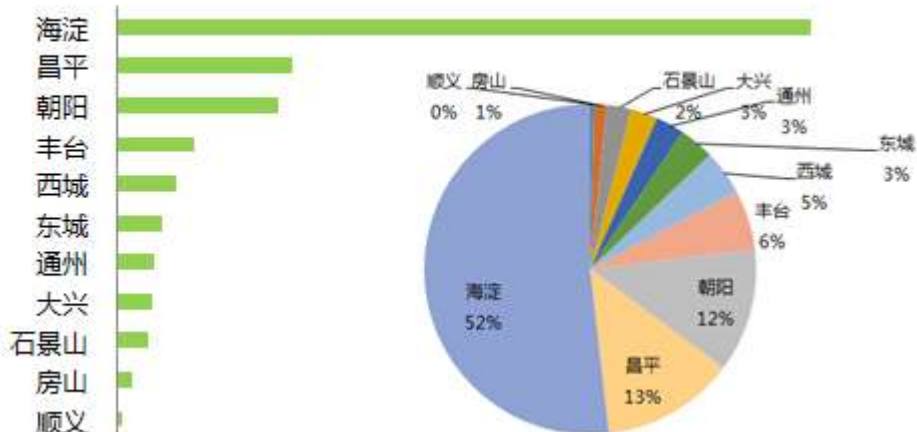
## 6. 用于人口属性行为分析

利用手机实名制相关信息，以及公安等部门的共享数据，逐步实现对手机持卡用户的身份属性分析，如性别、年龄、京籍、非京籍、行业职业等。目前，结合经普地理信息系统，对人员聚集热点区域、大型写字楼等建筑，分析其中白天用户的行业职业属性、居住地、通勤半径等。

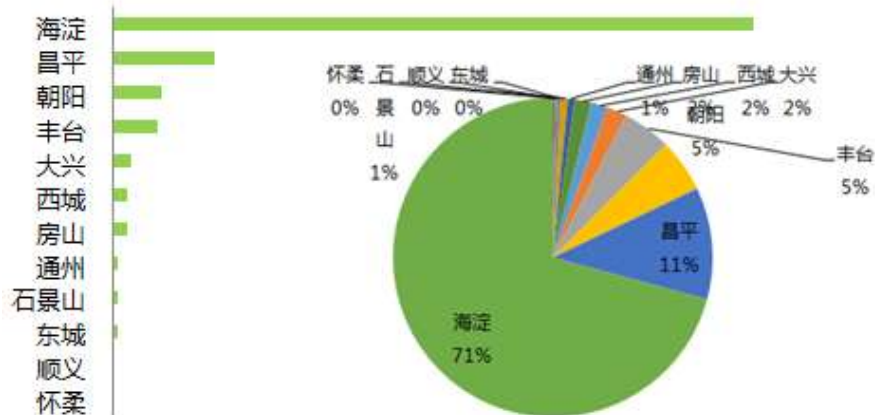
居住地	中关村理想大厦	中关村金融中心	清华科技大厦	鼎好电子商城
昌平	65	67	163	47
朝阳	60	32	61	30人以下
大兴	30人以下	30人以下	30人以下	30人以下
东城	30人以下	30人以下	30人以下	30人以下
房山	30人以下	30人以下	30人以下	30人以下
丰台	30人以下	30人以下	30人以下	30人以下
海淀	258	419	434	282
怀柔	0	30人以下	0	0
门头沟	0	0	0	30人以下
密云	0	0	0	0
平谷	0	0	30人以下	0
石景山	30人以下	30人以下	30人以下	30人以下
顺义	30人以下	30人以下	30人以下	30人以下
通州	30人以下	30人以下	30人以下	30人以下
西城	30人以下	30人以下	30人以下	30人以下
延庆	0	0	0	30人以下
白天用户	863	821	1130	579

# 6. 用于人口属性行为分析

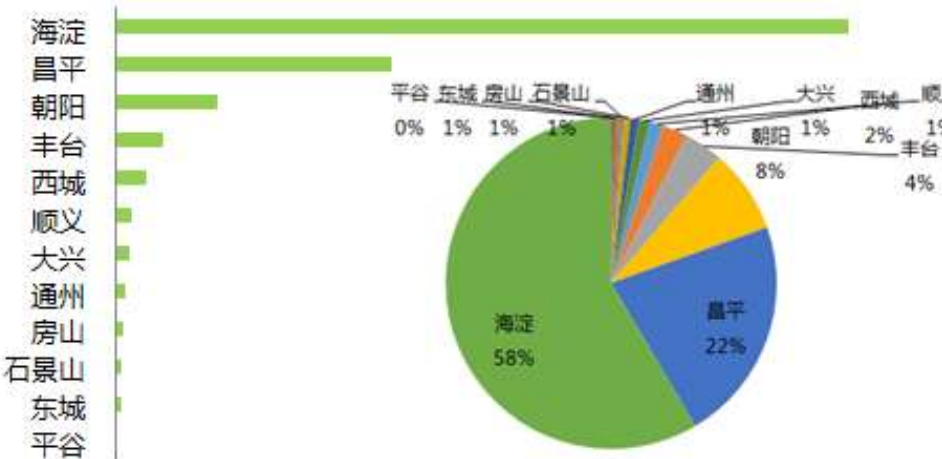
## 中关村理想大厦



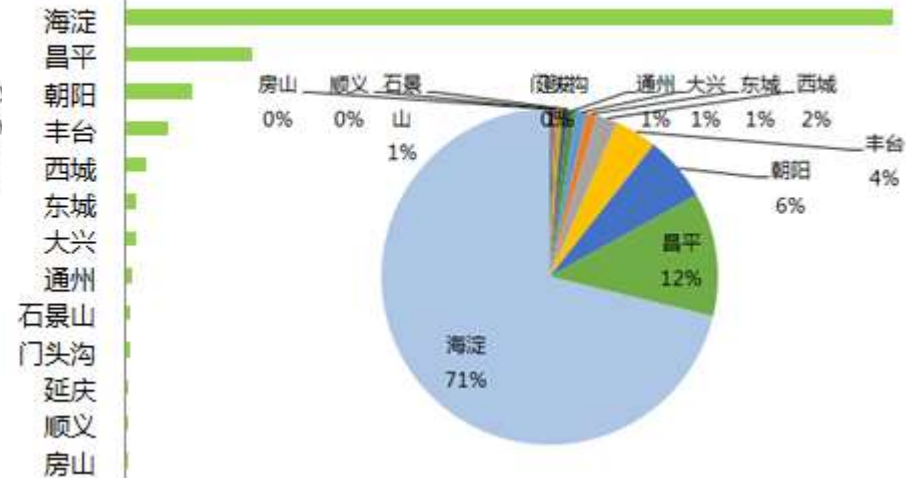
## 中关村金融中心



## 清华科技大厦



## 鼎好电子商城



## 五、问题与思考

关于使用移动通信数据测算人口，我们还在进一步的分析和研究之中。

1.手机稳定用户数量与人口数既有联系也有区别：在使用手机的用户中，现役军人、外籍和港澳台人口以及居住半年以下的外来人口，均不计入统计调查的常住人口范畴；部分老人、儿童虽然计入常住人口，但不持有手机；加上重号率、不同的运营商（联通、电信）等因素，人口的测算非常复杂。



## 五、问题与思考

随着手机实名制的逐步推开，部门数据的共享运用，利用用户资料对人口身份信息加以识别，以区分军人、外籍和港澳台人士等，得到更加准确、动态的人口数据；

针对部分老人和儿童，可以开展专项调查，摸清非手机持卡用户的数量及比重情况，用来补充手机监测测算人口数据的缺口；

## 五、问题与思考

2. 积极与联通、电信两家运营商沟通，以获得全市手机用户全面的位置信息和停留时长等，对全市常住人口数据进行测算。

3. 我们正在与公司一起开展轨迹拟合研究，以推算出较为准确的重号率，构建用户活动特征DNA，包括主要通话位置、活动轨迹、活动规律等；通过“DNA”比对，将重合度高的，默认为一机多卡、或多机多卡用户，对其进行合并剔重。

# 结束语

2014年10月国家统计局时任局长马建堂曾在“大数据和官方统计”国际会议上讲到未来的中国统计,一方面要运用信息技术、云计算等新技术采集、处理、公布传统数据,另一方面基于大数据、云计算的广泛应用,打造更科学、更坚实的政府统计数据来源的“第二轨”。

我们今天所做的基于移动通信大数据的人口动态监测工作只是刚刚起步,取得了初步的成果。虽然不能替代人口抽样调查,但毕竟是我们将来采集数据的一种新的方法,也是未来政府统计数据的另一个来源渠道。

# 结束语

大数据的意义在于改变人们思考世界的方式。通过对大数据在人口统计中的探索应用，我们体会到：信息化高速发展的今天，我们不能永远停留在旧有的数据采集模式中，我们要开拓创新，不断研究探索如何将新的数据生产方式与科学的统计制度结合运用，才能真正解放统计生产力，提高数据生产效率。

朋友们：大数据时代已经到来，让我们张开双臂去拥抱吧!

**介绍结束**

**谢谢大家！**