**2017年第四届中国可视化与可视分析大会**

**数据可视分析挑战赛-挑战1**

**（ChinaVis Data Challenge 2017 – mini challenge 1）**

**参 考 答 案**

（本答案仅供参考，鼓励参赛者发现本参考答案范围以外的任何有趣的故事）

# 一、挑战1.1

## 题目内容

伪基站常流动于人口密集的区域，以各种名义向一定范围内的手机发送垃圾短信，因此，了解掌握伪基站出行的时空模式，能够帮助执法人员尽早阻止和抓获不法分子，从而更好地维护社会秩序。然而仅仅从垃圾短信中很难确定其对应的伪基站，即无法确定来自同一台伪基站设备的垃圾短信，相同的垃圾短信有可能来自不同的伪基站，同一个伪基站可能不送不同的短信。请从宏观时空分析的角度出发，对垃圾短信数据进行可视分析，揭示伪基站的总体时空活动规律。（建议参赛者回答此题文字不多于1500字，图片不多于8张）

## 参考答案

本题可以有多种角度提供分析结果。在参考答案中，我们按照工作日和非工作日分别分析伪基站不法分子犯罪的时空特征：

1. 工作日伪基站时空特征的答题要点如下：
   1. 伪基站不法分子一般从早上6:00开始发送垃圾短信活动;
   2. 移动方向为从外环逐渐涌向城内，如图1（a）所示，早上6:00到7:00伪基站出现在外环，随后城内逐渐增多；
   3. 伪基站发送垃圾短信从数量上来说会出现两个高峰时段，早高峰时段为8:00到10:00，晚高峰时段为18:00到20:00，如图1（b）所示；
   4. 早晚高峰时垃圾短信主要集中在地铁、城市主干道路、环线上；
   5. 晚上22:00左右基本停止发送垃圾短信犯罪行为，如图1（c）所示。

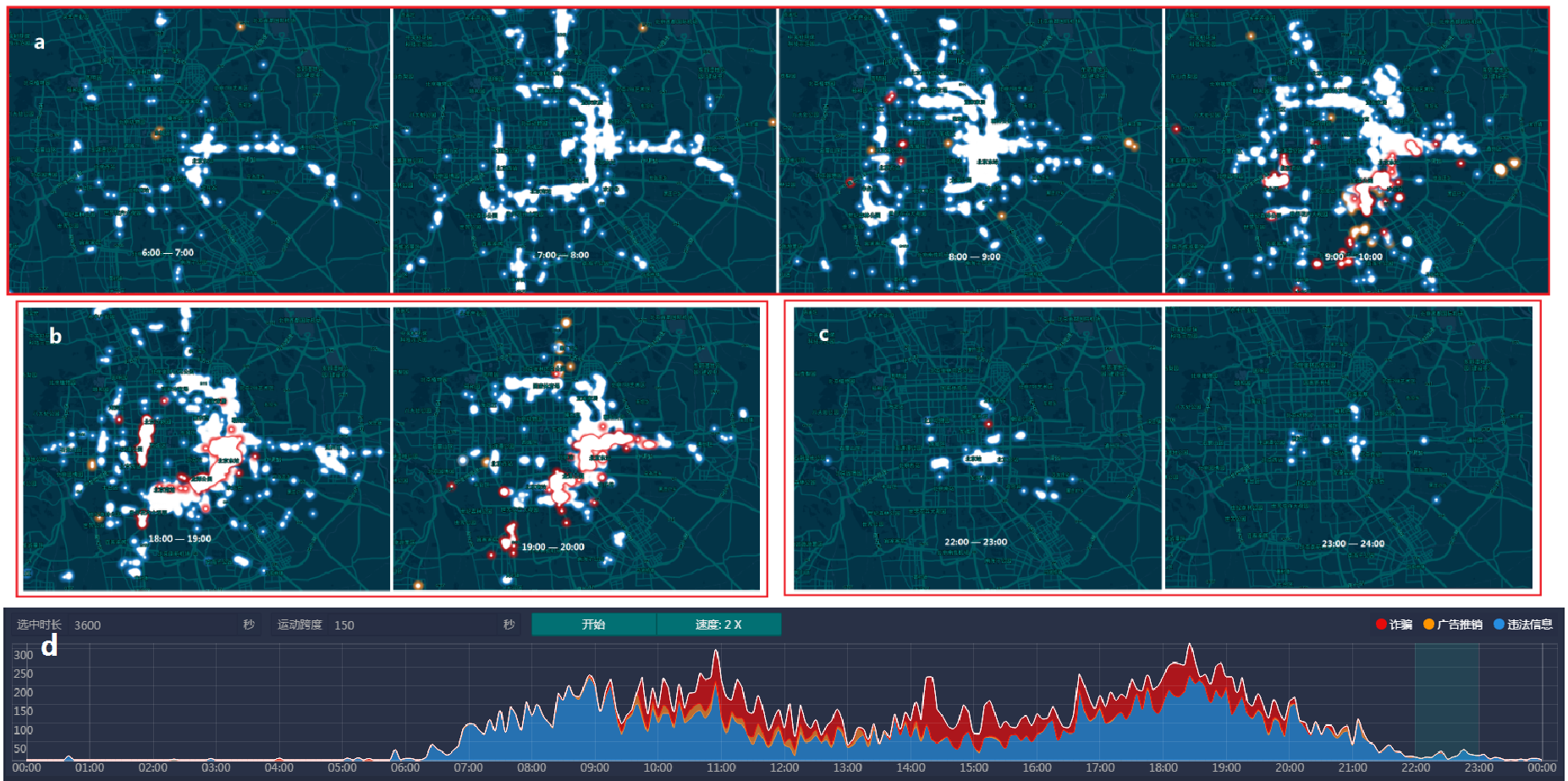


图1 典型工作日伪基站的时空分布示例，图中选取日期为4月11号: (a)早上从外环逐渐涌向城内,并出现早高峰; (b)18:00-20:00为晚高峰时间段；(c)晚上22:00之后伪基站逐渐暂停发送垃圾短信；(d)垃圾短信分布时间趋势图.

1. 非工作日伪基站的时空特征答题要点如下：
   1. 周末伪基站不法分子发送垃圾短信的早高峰消失，如图2（a）所示；
   2. 周末垃圾短信数量整体上少于工作日短信数量；
   3. 晚上22:00点之后依然还在发送垃圾短信，且深夜短信常出现在城东地区，如图2（b）所示。

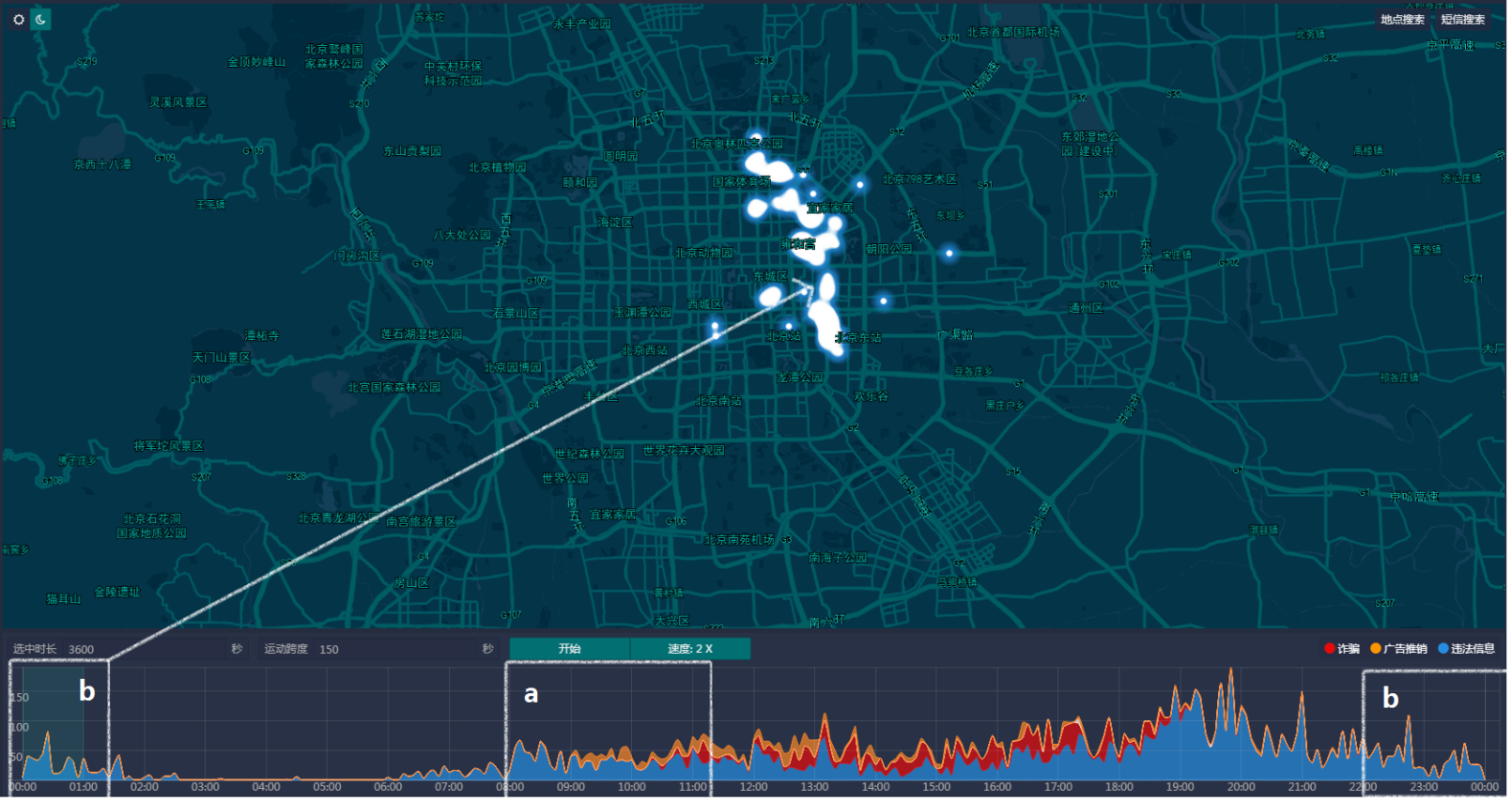


图2 周末伪基站的时空模式示例，图中选取日期为4月22号：（a)早高峰消失；（b)22:00后深夜依然有垃圾短信的出现，且常出现在城东.

1. 另外，不区分工作日和非工作日也可以找到伪基站的一般时空特征，举例如下：
   1. 垃圾短信常出现在人口密集、流量大的区域，如机场、火车站、地铁、景区、商业圈附近；
   2. 伪基站在某地区的活跃程度与该地经济水平具有一定的相关性，伪基站出现在城东（如朝阳区、东城区）的频率稍高于城西，如图3所示，选取不同日期不同时间段的伪基站热力分布图中都可以看到这个现象，城东人口密度大，经济发展相对较繁荣可能是导致这种现象的出现的原因之一；
   3. 从北京市行政区上进行对比，朝阳区垃圾短信分布最多。



图3 选取不同日期不同时段，从伪基站热力分布图中可以看出伪基站出现在城东的频率稍高于城西

1. 3月3号至3月16号两会期间，由于全城戒严、市内限行，安检力度加强，很大幅度限制了伪基站的活动，因此短信数量相对减少，3月16号之后的一个星期伪基站的活跃度则进入了快速增长期；
2. 4月2号至4号清明节假时市民在家休息或出门旅游，伪基站发送短信数明显减少，而且分布更加均匀。

# 二、挑战1.2

## 题目内容

不法分子通过设置伪基站设备能够发送不同类型的垃圾短信，请尝试对垃圾短信的具体内容进行分类，分类标准不限，例如：按垃圾短信类型可以分为广告、诈骗等等，按垃圾短信对人们的人生经济危害程度可以分为一般、严重等等。请尝试在问题1的基础上进一步分析伪基站发送不同类型垃圾短信的时空分布规律。（建议参赛者回答此题文字不多于1500字，图片不多于8张）

## 参考答案

（一）垃圾短信短信分类可划分为三大类：广告推销类型、一般危害的违法信息类型和危害严重的诈骗类型，各类型又可细分为以下类型：

1. 广告类型

充值、房地产广告、旅游、教育培训、汽车交易、银行推广、金融服务小广告、电商网站推广、实体商店推广、电信运营商业务推广、其他网站及APP推广、其他等。

1. 违法类型

代开发票、办证刻章、复制电话卡、政治、色情服务、赌博、其他等。

1. 诈骗类型

打款诈骗、虚假中奖、虚假红包、身份冒充、其他等。

（二）不同类型的垃圾短信伪基站具有不同的时空模式，下面分类列举一些答题要点：

1. 广告类型垃圾短信时空特征答题要点：
   1. 节假日广告推销类型垃圾短信数量明显增多，如图4（a）选取日期为清明节假期4月3号时，垃圾短信时序统计图中表示广告类型垃圾短信的橙色增多；
   2. 广告推销类垃圾短信分布位置分散，如图4（b）中伪基站热力分布中橙色热力分布位置分散。

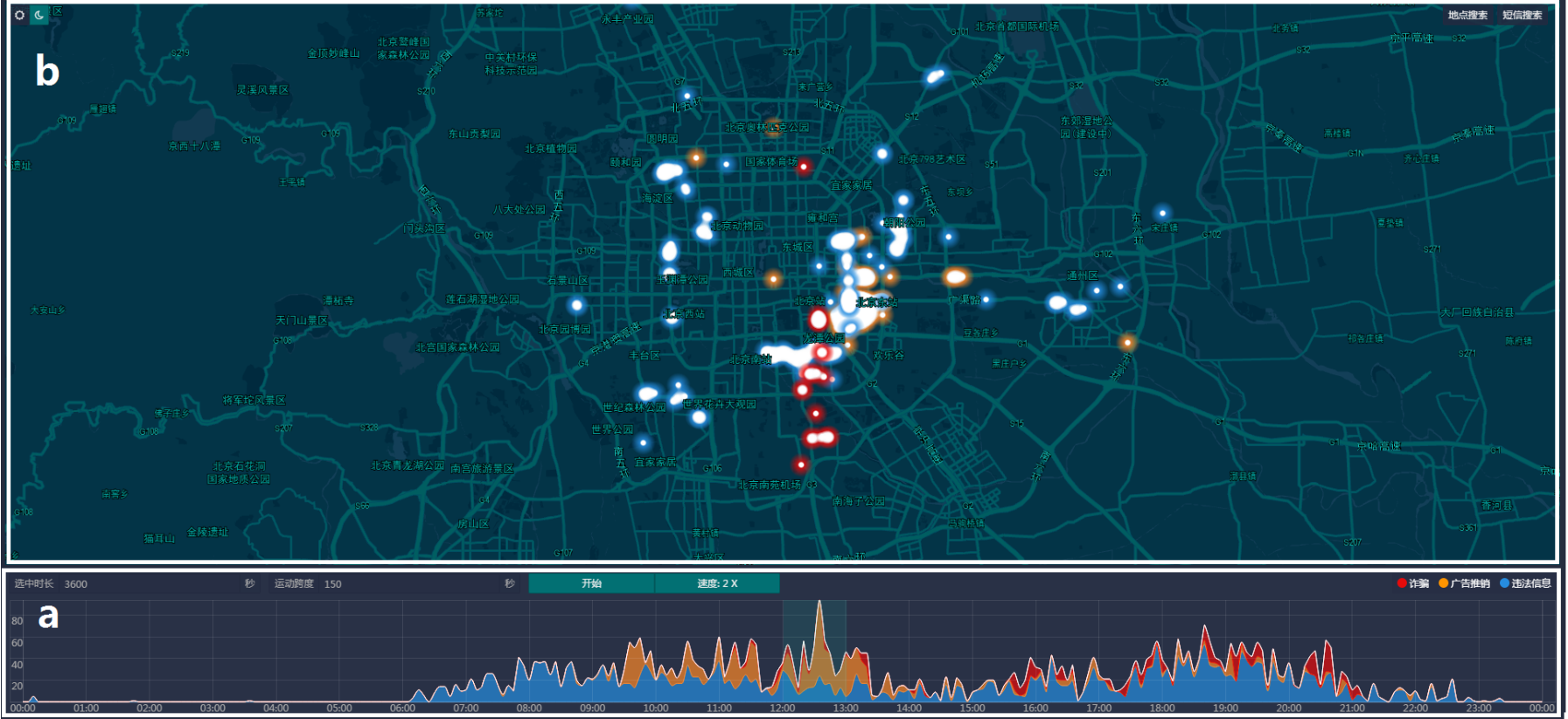


图4 图中选取日期为4月3号清明节，橙色表示广告推销类垃圾短信，(a)从统计图中可以看出假期时广告推销类垃圾短信橙色区域明显增多;(b)并且分布位置分散.

1. 违法类型垃圾短信时空特征答题要点：
   1. 违法信息类垃圾短信数量最多，如图5所示，选取日期为4月12号，图中蓝色表示违法信息类垃圾短信，从5（a）统计图中可以看出违法信息类垃圾短信数量最多；
   2. 覆盖面积广，分布区域分散，如图5（b）地图热力分布中蓝色违法信息类覆盖面积最广，分布区域分散；
   3. 工作日时早晚上下班高峰期出现次数高于其它类型，如图5（c）早晚高峰时段违法信息类型垃圾短信数量高于其它类型。

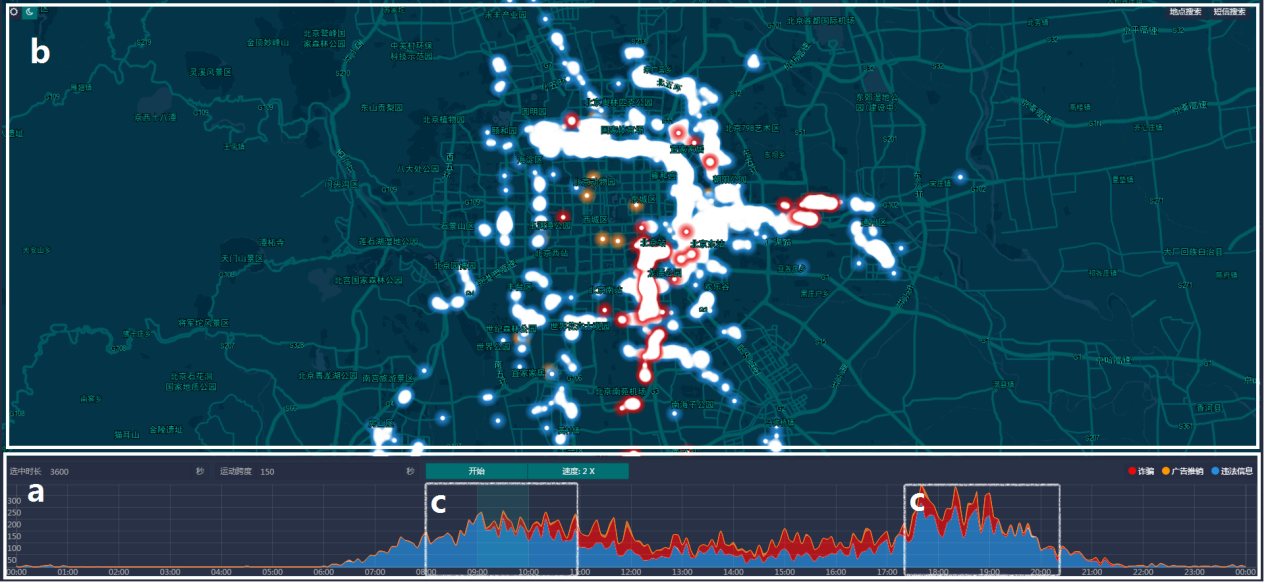


图5 图中选取日期为4月12号周三，其中蓝色表示违法信息类垃圾短信，(a)从统计图中可以看出违法信息类垃圾短信数量最多；(b)从地图热力分布中可以看出蓝色违法信息类覆盖面积最广，分布区域分散；(c)早晚高峰时段违法信息类型垃圾短信数量高于其它类型.

1. 诈骗类型垃圾短信时空特征答题要点：
   1. 诈骗类型垃圾短信对社会的危害最高，伪基站发送诈骗类型垃圾短信的频次最低；
   2. 诈骗类垃圾短信常出现在学校、火车站、机场、旅游景区等人口密集区域，如图6所示。

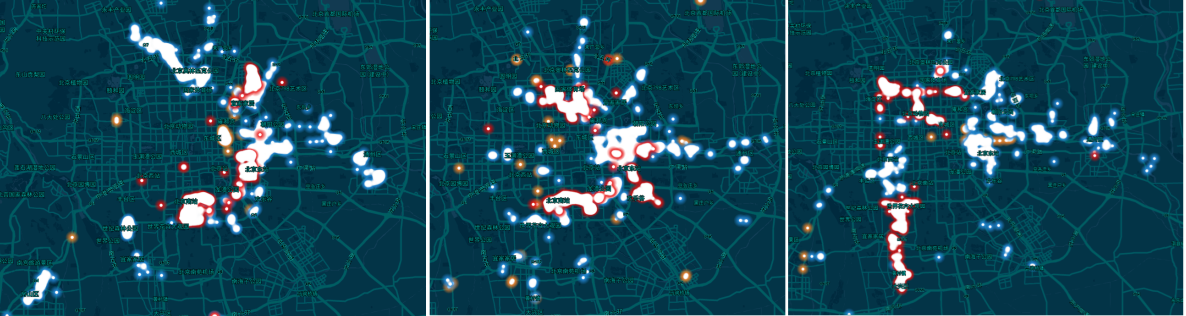


图6 图中选取多个日期不同时段进行分析，图中红色表示诈骗类型垃圾短信，从伪基站热力分布图中可以看出诈骗类型垃圾短信常出现在学校、火车站、机场、景区等人口密集区域，如图中北京南站、北京首都国际机场、海淀区高校附近.

1. 各种类型垃圾短信除了上述三大分类的总体时空特征外，一些具体细分类型的垃圾短信也会有不同的时空特征。

下面以违法类型中色情类垃圾短信举例：

* 1. 深夜凌晨色情短信出现频次高于其它时段；
  2. 周末深夜色情短信数量高于工作日；
  3. 常出现在东三环附近，如图7所示，东三环三里屯、工体、双井附近酒吧、夜店较多可能是产生这种情况的原因之一。

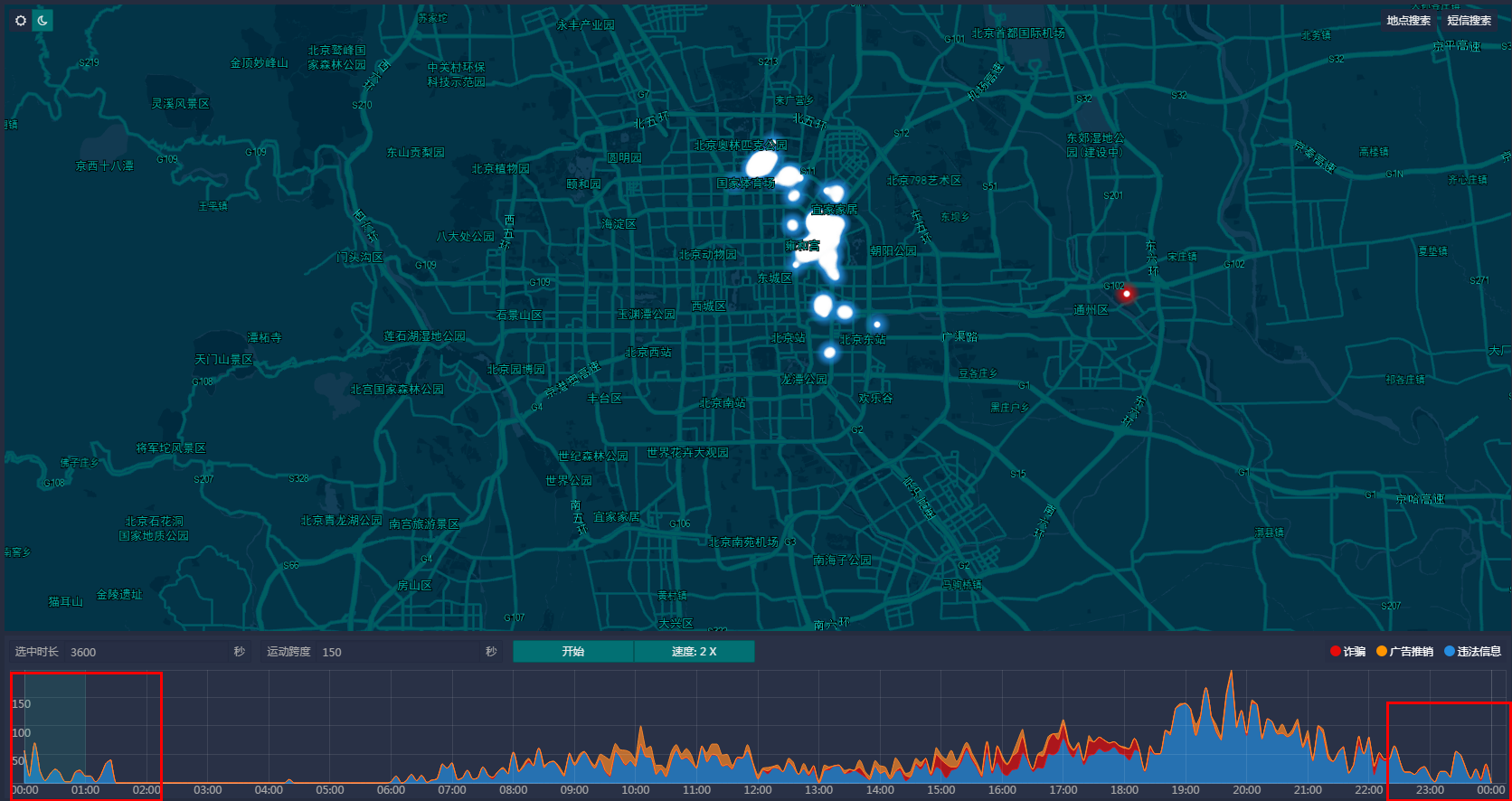


图7 选取日期为4月23号周日，周末时色情类垃圾短信数量高于工作日，深夜凌晨时色情类短信较多，且常出现在东三环附近

# 三、挑战1.3

## 题目内容

伪基站不仅破坏正常电信秩序，危害公共安全，扰乱市场秩序，而且严重损害群众财产权益，侵犯公民个人隐私，社会危害严重。据《人民网》统计，每年通过“伪基站”设备发送诈骗、赌博、推销、中奖等短信近千亿条，伪基站已成为社会一大公害。请结合以上两题中得到的伪基站行为模式，向执法人员提出打击整治伪基站的有效建议和方案，并结合数据分析结果进行说明。（建议参赛者回答此题文字不多于1000字，图片不多于5张）

## 参考答案

通过可视分析的方法对伪基站垃圾短信的时空模式和具体内容进行分析，可以发现伪基站高发时段、区域和时空规律，执法人员可以在掌握这些规律的基础上合理协调人力物力资源对伪基站不法分子进行有效打击，大大提高打击效率。

首先对伪基站垃圾短信进行分类，评估不同类型短信对社会和人们的危害程度，针对不同类型垃圾短信的时空模式，采取不同的打击措施，列举如下：

1. 针对诈骗类型伪基站不法分子的犯罪规律，如常出现在学校、车站、机场、景区等人口密集区域，建议执法人员在这些区域增加伪基站检测设备。
2. 一般违法信息垃圾短信的工作模式是早上从外环涌向主城区，因此执法人员可以在早上进入主城区的主要关卡加大打击伪基站的人力物力资源。
3. 考虑到城东东三环附近常出现色情类垃圾短信，传播不良信息，严重危害人们的身心健康，建议执法人员在附近相关区域定点设置伪基站检测设备，严厉打击不法分子。

（本题开放性程度较大，请评委酌情给分）