参赛队号：（由大赛组委会办公室填写）

2021年（第七届）全国大学生统计建模大赛

参赛学校：华中科技大学

论文题目：基于微博数据挖掘的突发事件灾害链建模

参赛队员：马世拓，张誉馨，肖渝楠

指导老师：王新宇，王然

目录

图表目录
图片

表格

基于微博数据挖掘的突发事件灾害链建模

摘要

为了从统计意义上描述**突发事件**与**灾害链**的时空分布规律以及引发因果关系，从微博上获取了海量突发事件数据进行**数据挖掘**。采用改进的**共现矩阵**与**复杂网络**理论进行建模能够从图论的角度展现灾害链所形成网络的拓扑结构，对灾害引发的因果关系与危害强度进行了一定的讨论。此外，从传播学的角度还可以分析不同种类的事件报道时间与爆发时间的差值来描述事件的重要程度。最终得到的统计学规律对于灾害的防控与避险有一定的参考价值。

关键词：突发事件，灾害链，数据挖掘，共现矩阵，复杂网络

Modeling of Emergency Disaster Chain Based on Micro-blog Data Mining

Abstract

In order to describe the **emergency**'s temporal and spatial distribution and its casual **disaster chain** from a perspective of statistics,we obtain a large amount of emergency data for **data mining** on Weibo. We adopt the improved **co-occurrence matrix** and **complex network** theory ,which can not only show the topological structure of the disaster chain's network from the graph theory Angle, but also discuss the causality and damage intensity caused by the disaster.From the perspective of communication, we can analyze the difference between the reporting time and the outbreak time of different kinds of events to describe the importance of events.The ultimate outcome of research is expected to make a contribution to the prevention and control of disasters.

Keywords: emergencies, disaster chain, data mining, co-occurrence matrix, complex network

引言

在现实生活中，灾害的出现往往并不是孤立的，而是有一定关联性的。那么为了描述灾害成灾的关联性，我们引入灾害链的概念对灾害事件进行建模分析。灾害链的建模对于各类突发事件的防控，有着重大意义。

自灾害链的概念初次被提出，就有一些学者试图从灾害引发的地质学机理或气象学原理层面进行推断，而从数据与统计方法研究灾害链则相对晚一些。早期的数据统计推断基于简单的共现矩阵理论和非常少量的数据，这些数据多发生在清末民国甚至古代唐宋时期，时间久远可信度低，而且事件之间的时间跨度大难以形成说服力。并且少量数据计算比较简便，但这一方法是否能在海量数据集上使用还未得到充分的验证。

随着信息化时代的到来，突发事件的记载不再依赖于史书等难以保证可靠性的文献，而是可以通过微博等网络平台进行数据的获取与处理。而大数据挖掘的手段，更能保证数据能够被充分地挖掘与利用。近年来，有关数据挖掘的研究热度居高不下，使用数据挖掘方法去提取出繁杂信息项中的模式是灾害链建模目前最有效的手段。

目前基于数据挖掘的灾害链模式数据挖掘主要依赖于靠以聚类算法为主要代表和以群体智能算法为主要代表的两大算法体系,但也仅局限于稍小的一个数据集,分析算法的时间复杂度和数据集空间的复杂度,这些算法在海量数据集上会造成很大的开销。另外随着自然语言处理等技术的成熟，我们能够从网络上的文本中提取出突发事件的一些特征。

为了将灾害链成灾的研究方法从原有的自然灾害扩充至包含社会突发事件在内的网络突发事件，同时利用海量数据挖掘的技术对灾害链进行时间和空间上联系更加密切的建模，我们考虑进行这样一个选题。

我们将从微博数据库中搜集大量数据，并过滤出进一步统计分析所需的数据。数据来源真实可靠，而且这一海量数据具有较高的价值，可以用于探索以往的一些研究方法在海量数据上的应用效果，并分析有没有一些新的发现。

这一选题将对社会上重大突发事件的防控提供一定参考意见，从而降低突发事件带来的损失。

国内外研究综述

重大灾害的相继产生并非相互独立的事件，而是存在一定关联的，若发生一种突发事件极易引起连锁效应。我们把灾害相继出现的模式，或者灾害在时空维度上的传播形成链式的有序结构称为灾害链。

一. 灾害链研究综述

1987 年我国地震学家郭增建首次提出灾害链的理论概念：灾害链就是一系列灾害相继发生的现象[1]。灾害链描述了自然界中各种灾害之间的联系，通常把地震后伴生的滑坡,泥石流等次生灾害称作第一类灾害链，而除了科学界公认的海气相互作用外的地气耦合则被称为第二类灾害链[2]。而灾害链除引发机制问题以外，风险评估问题在灾害链的风险预防中同样有着重要作用。王翔等人总结了灾害链风险评估的模式并对区域灾害链风险进行了系统分析[3]。

此外，国内外有一些学者将灾害链研究应用于特定事件种类例如洪涝灾害研究[4]、地震灾害研究[5]、电力系统灾害研究[6]等。此类研究可以在很大程度上指导类似事件的应对策略，同时减少经济损失，有着重要意义。

二. 复杂网络研究综述

在灾害链体系的研究过程中，基于共现矩阵与复杂网络的建模是一种简单而有效的研究方法。复杂网络是一种图论模型，具有相对较为复杂的拓扑结构[7]。复杂网络建模试图解决三个问题：第一，找出可以刻画网络拓扑结构和行为的统计特性，并且给出衡量这些统计特性的方法；第二，构建网络模型以便帮助我们理解这些统计特性背后的真正意义；第三，基于这些统计特性，研究网络中的行为与局部规则[8,9]。

对于复杂网络的研究，目前比较多地集中于网络的社团特性挖掘，尤其是基于演化聚类的社团发现[10]。近年来，随着深度学习技术的发展，越来越多的研究也采用了深度学习来进行复杂网络的建模[11]。

基于灾害成灾的历史数据进行复杂网络的建模，可以从各灾害形成的系统中建立图论模型，从而分析其拓扑结构并发现其特性[12,13]。共现矩阵与复杂网络理论模型较为简单容易实现，计算量比较小，但目前基于复杂网络的灾害链模型受限于历史数据，且研究灾害种类有一定局限性，所以需要在更庞大数据集的背景下对模型进行改进。

三. 群体智能算法研究综述

群体智能算法是启发式算法中一个重要的分支，也是计算机科学中一类非常重要的问题。群体智能算法试图模拟现实中群体生物的行为，对一些问题进行优化（如 TSP 问题等）[14]。

1992 年 Marco 从蚂蚁通过信息素浓度的决策行为获得灵感从而提出了蚁群算法，将所有可能走的路线设置为待优化问题的解空间，通过信息素矩阵反复迭代来保存最优信息[15]。但蚁群算法存在收敛速度慢、容易陷入局部最优等缺点。

1995 年 Kennedy 提出基于对鸟群行为研究的研究粒子群算法是目前为止应用最为广泛的群体智能算法：粒子通过当前位置和对于自己最优位置之间的距离以及与群体最好位置的距离迭代来获得最优解[16]。虽然这种方法有收敛速度快等优点，但仍存在精度低、易发散等缺点。

Yang 等人提出的萤火虫群算法根据相对亮度来决定萤火虫的移动方向，萤火虫会向最优位置集中，从而找到全局的最优个体值，不仅可以优化单峰函数也可以优化多峰函数，适用范围更广[17]。但这种方法只能在一定的前提下实现，所以还存在一定的局限性。

群体智能算法在灾害链建模中有重要作用。我们把灾害的爆发地点在二维平面上的横纵坐标与爆发时点，作为某一突发事件在时空维度下的特征，并将事件的类别作为点的标记。按照时间顺序，将地区相近的突发事件用有向边进行连接，我们就可以构建一个具有三维拓扑结构的图论模型。那么问题被抽象为：如果相似于网络中某一向量的向量在立体网络中多次出现，就可以认为该类空间向量是一种灾害关联向量，从而确定出该向量对应节点位置的成链规则[18]。胡明生、洪流等人基于改进萤火虫群算法对灾害网络进行了建模，分析了一些地区灾害链成灾的规律[19]。

此方法具有一定的新意与独特性，将灾害链转化为路径优化问题进行求解，但目前此方法尚未得到大规模应用，而且应用受限于数据量，当数据量过于庞大时计算量也会十分复杂。

四. 数据挖掘在灾害链研究中的应用综述

近年来，数据挖掘技术的热度只增不减，而将数据挖掘技术融入灾害链建模是很重要的一种方法。

一些聚类算法能够对突发事件的密集爆发区域进行聚类从而得到这些突发事件在时空维度上的特征。其中，基于密度的聚类算法的簇只需要考虑样本间密度，无需事先确定簇的个数。只需要提前设置好邻域半径，便可以描述样本与样本之间、聚类簇与聚类簇之间的关系[20,21]。

过去的一些研究也对时空数据的挖掘提供了一些可行的策略。Ray 等人提出概念迁移的概念，将时间序列进行切片，对每一个切片考虑其状态，然后考虑状态随时间推移的变化规律[22]。另外，在时空数据中，我们通过对 K 均值聚类算法的目标函数进行改进调整，得到了调和 K 均值聚类算法用于描述时空数据的属性特征[23]。Karine 等人考虑将时间序列元素的下标作为分析特征引入时空序列的分析流程中来[24]。对于时空聚类的高效算法，Vladimir 等提出了 AUTOCLUST+算法，对原有时空序列聚类的效率进行了更高程度上的优化[25]。这些研究都为我们分析突发事件的时空分布提供了很好的方法，但目前这些研究缺乏对整个时空域上聚类结果动态变化的效果评估。

有关网络上突发事件的传播数据挖掘目前主要是一些基础统计学方法。Chen 等人基于时空序列的统计学特征与文本情感识别方法，对美国 Harvey 飓风事件的影响与信息传播进行了挖掘[26]。Yao 等人发现转发行为表现出聚合特征，具有不同属性的用户在新浪微博上有特殊的转发习惯[27]。Zhong 等人从特征工程的角度，对比支持向量机、随机森林、CatBoost 和 LightGBM 四种算法，从文本以及用户行为等多个角度挖掘了突发事件在微博传播中的重要特征[28,29,30]。

这些研究都从数据挖掘的角度对网络上的突发事件进行了建模分析，但事件局限性比较强，未能包括各种突发事件，而且对灾害相互引发的机理与特性并未作出讨论。

五. 本文的主要工作

此前的一系列研究都具有其重要的前沿和先驱意义,它们为我们的研究提供了许多具有建设性的分析方法,以及一些在实践中具有重要的参考价值的结论。但此前的一些研究仍然存在着三个主要的限制:

1. 事件种类的局限性，仅仅考虑某一类特定突发事件尤其是自然灾害类事件，而对其他事件的联系并未进行讨论。

2. 事件关系的局限性，以往的研究大都考虑的是两个事件在一定时间段内爆发的相关关系，而没有考虑先后次序，即引入时间以后两个事件形成的因果关系其实是研究较少的。

3. 数据体量的局限性，以往的灾害链研究数据基于历史文献记录，不仅时间长周期长而且数据体量小，无法充分反映其在海量数据上的适用性。

鉴于前人的工作成果，我们在这项研究中主要的工作如下：

1. 收集海量微博数据，筛选出突发事件并对其进行统计意义上的分析与可视化，并结合密度聚类等手段发现其在时空分布上的统计规律。

2. 从共现矩阵与复杂网络的角度，利用图论建模的手段进行灾害链引发的进一步分析，考虑因果关系而非传统相关关系。

3. 从统计学角度，分析不同类别的突发事件爆发时间与发生时间之间的时间差，从而发现相应的传播意义的规律。

我们的工作成果从统计学意义上揭示了突发事件的发生规律,并对灾害链的引发因果关系进行了一定程度上的阐释,对于灾害发生的防控与避灾有一定借鉴意义。

变量符号与意义

表 1. 变量符号与定义

符号名称	含义
A,B	两种突发事件，可能相同也可能不同
O	突发事件组成的集合
M	突发事件两两组合形成的事件组
T	两件突发事件相继爆发的时间差
C_A	A 事件发生的总数
C_{AB}	A 事件和 B 事件爆发时间差在[-T,T]内的事件总数
$C_{AB}(t)$	A 事件在 B 事件爆发时间差等于 t 的事件总数
$C_{<A,B>}(t)$	A 事件在 B 事件之前爆发且爆发时间差等于 t 的事件总数
J_{AB}	对 AB 而言的 Jaccard 指数
$J_{AB}(t)$	对 AB 而言爆发时间差等于 t 的 Jaccard 指数
$J_{<A,B>}(t)$	对 A 作为条件 B 作为结果爆发时间差等于 t 的 Jaccard 指数
$CO(A,B)$	A，B 的共现率
$D(<A,B>)$	从事件 A 引发事件 B 的危害程度
$I(A\rightarrow B)$	同样代表从 A 引发 B 的改进共现率
α,β	主特征系数和次要特征系数
w_{ij}	节点 ij 之间边的权值
S_i	节点 i 的点权
g_{ik}	节点 i 和节点 k 之间的最短路径总数
$g_{ik}(v)$	节点 i 和节点 k 的最短路径中通过节点 v 的最短路径数量

$b(v)$	节点 v 的点介数
k_i	节点 i 的度数
m_i	直接相连到 i 的节点总数
n_{ijk}	i 与 j 之间和 i 与 k 之间的边的总数。
a_{jk}	若 jk 之间没有边则取 0；若 jk 之间仅有一条边则取 1；有两条边则取 2。
$\overline{x}_1$	样本平均数
μ_1	样本期望
S_1^2	样本方差
n_1	样本数量
t	检验统计量

突发事件的时空特征

一. 数据的获取

我们使用 Python3.8.2 下的 jupyter notebook 作为我们的编程工具开展实验。数据分析的实验环境为 CPU: Intel i7+GPU: NVIDIA GEFORCE GTX 1650，操作系统为 Windows 10，数据获取则使用大型服务器作为工具。

我们部署了基于 Scrapy 框架的爬虫来获取微博上突发事件的报道。收集的时间跨度从 2009 年到 2019 年一共十一年，囊括社会安全、自然灾害、事故灾难和公共卫生四大类事件，而这四类事件又可以进一步细化为刑事案件、地质灾害等一共 18 个小类。类别关系如图 1 所示：

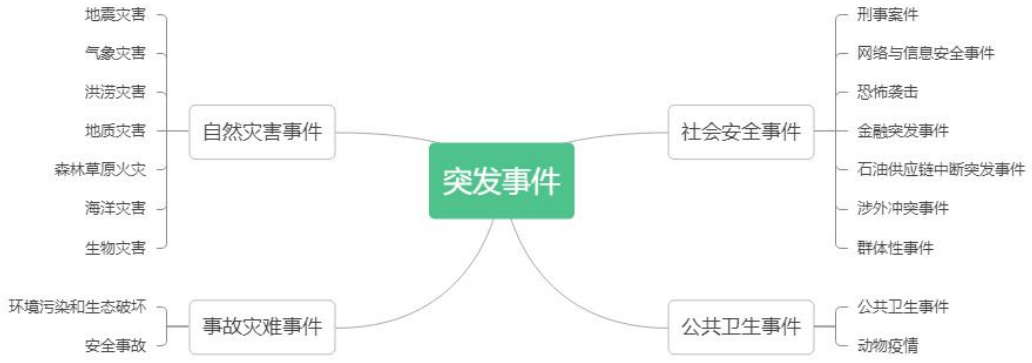


图 1. 事件关系的分类层级示意图

数据容量约 308G，采取传统的纯粹统计方法难以对这类海量数据进行分析处理。基于此，我们使用数据挖掘的流程策略对此进行分析。

二. 数据的预处理

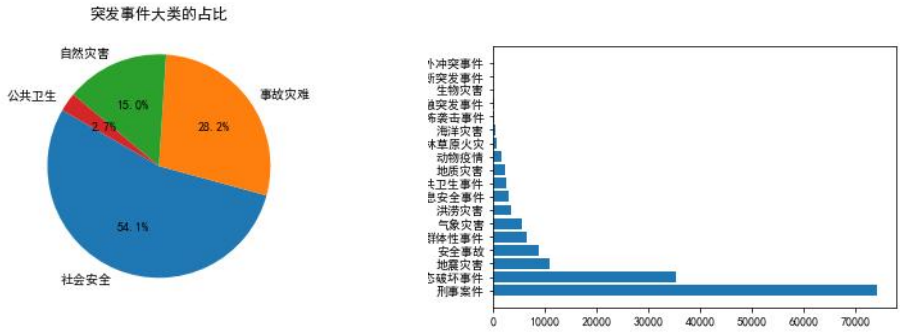
首先是数据的标签。我们基于 NLP 领域常用的 BERT 模型对文本进行了文本分类，对事件的类别进行标注。其次，我们从文本内容中基于正则表达式和关键词检索的方式提取出事件爆发的时间和具体地理位置，作为爆发的时空坐标。突发事件在时空三维欧几里得空间内的时空坐标是我们分析突发事件灾害链最重要的特征。

考虑到收集的数据中有很大一部分都不是突发事件，我们需要对非突发事件进行一个剔除。我们采取方式为根据爆发时点筛选，若该新闻不是突发事件则通常不会报道具体时间和地点。此外，有些非突发事件同样具有时点和地点，对于此类事件我们根据分类的分数进行剔除。经过这一步，我们的数据量大大减小，剩余数据条数约 83.3 万条。而在这 83.3 万条数据中又存在大量重复报道的信息，需要进行数据的去重。我们利用爆发时点重合性再进行筛选，保留报道时间最早的条目，最终筛选出数据约 15.6 万条。

三. 数据的统计特性与可视化

数据中四个大类的占比分别为：社会安全类类 84285 条，事故灾难类 44214 条，自然灾害类 23764 条，公共卫生类 4281 条。而各大类进行细化的条目数和

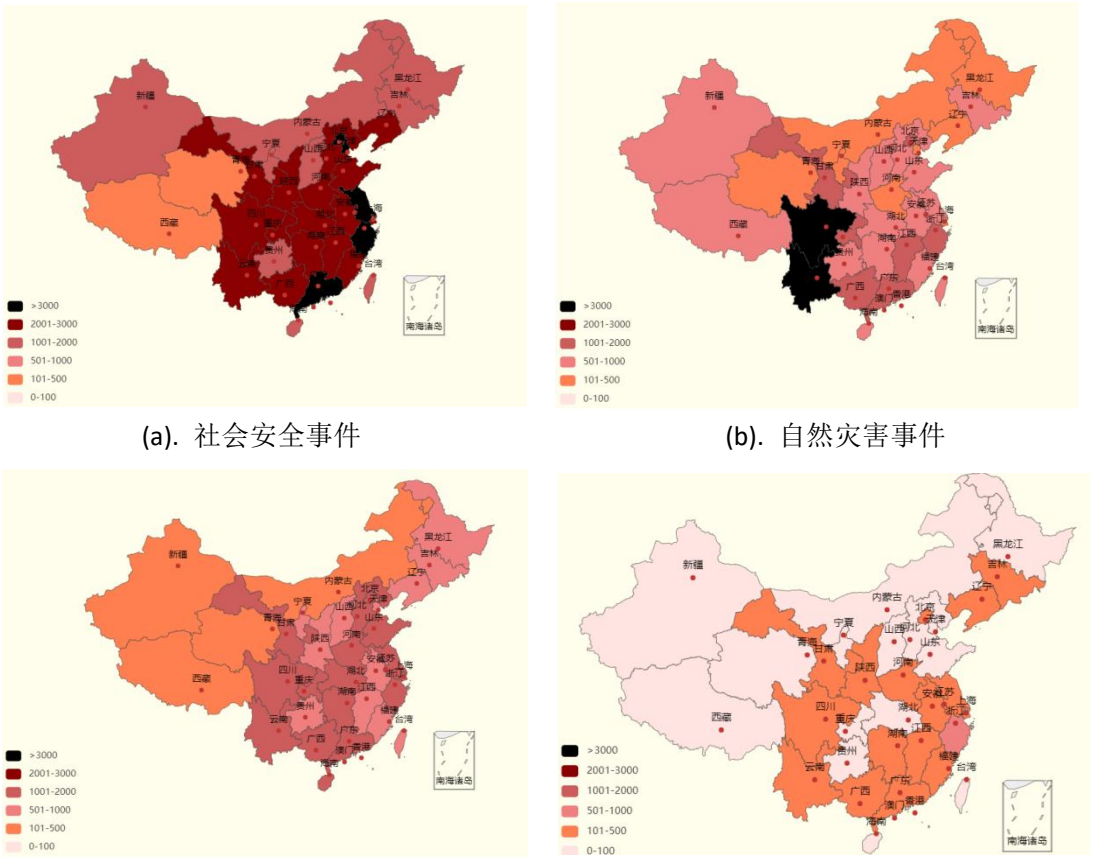
各类的占比大小如图 2 所示：



(a). 大类占比　　(b). 细化类统计

图 2. 突发事件分类占比示意图

为了研究四类突发事件在空间上的统计分布，我们将突发事件的爆发地点绘制在中国地图中，如图 3 所示：



(a). 社会安全事件　　(b). 自然灾害事件

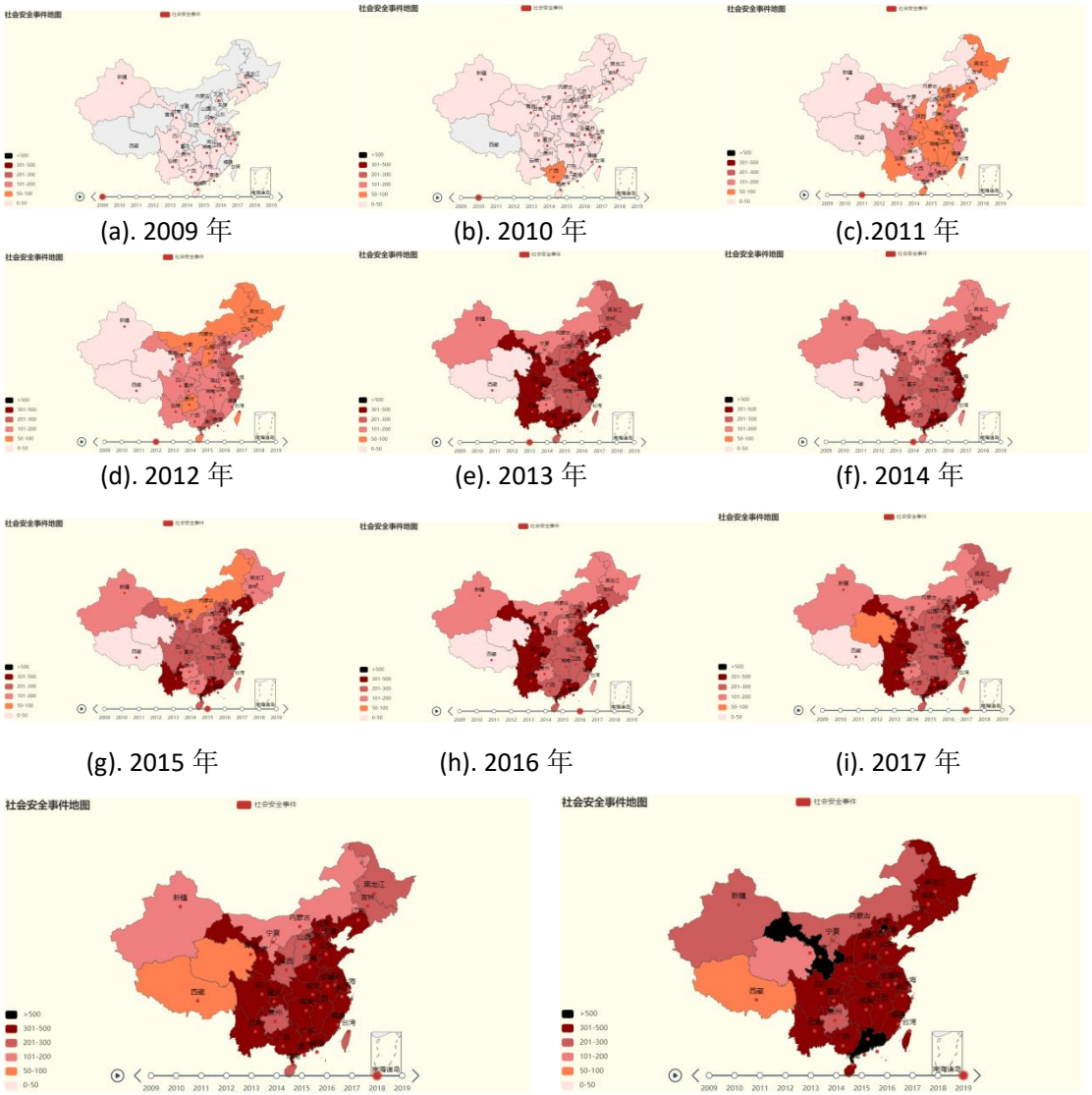
(c). 事故灾难事件　　(d). 公共卫生事件

图 3. 各类事件的爆发地点统计

（*注：图 3-图 5 的完整大图列在附录中）

图 3 中颜色越深表示爆发频率越大，可以看出：社会安全事件是各种突发事件中占比最高的，也是发生最频繁的，而社会安全事件则高发于我们眼中的“圣地”——京津地区，江浙沪地区和粤港澳地区。自然灾害事件多发于云南-四川一带，在横断山脉与青藏高原附近的特殊地质构造与生态环境确实决定了这一地区属于自然灾害的高发地。公共卫生事件爆发最少（在 2019 年新冠疫情还未开始爆发），但浙江一带为公共卫生事件高发地区。

除此以外，我们还根据时间推演的顺序将四类突发事件做了时间轴的切分。由于社会安全事件是四类突发事件中占比最多的一类，所以我们以社会安全事件为例，我们将带有时间推演的爆发地点统计图绘制在图 4 中：



(a). 2009 年　(b). 2010 年　(c).2011 年

(d). 2012 年　(e). 2013 年　(f). 2014 年

(g). 2015 年　(h). 2016 年　(i). 2017 年

(j). 2018 年　(k). 2019 年

图 4. 社会安全类事件随时间轴推移的分布

可以看到，社会安全类事件在华中地区、华北地区、华东地区和华南地区爆发较多，而且随着时间轴推移爆发密度越来越高。为了考察随着时间轴推移，我们还绘制了突发事件条数随着时间推移变化的曲线图如图 5 所示：

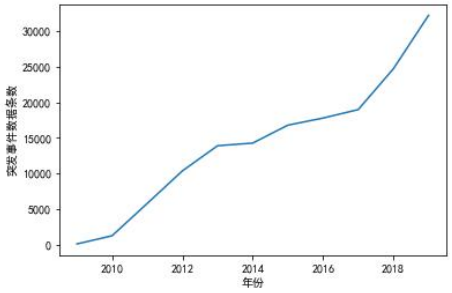


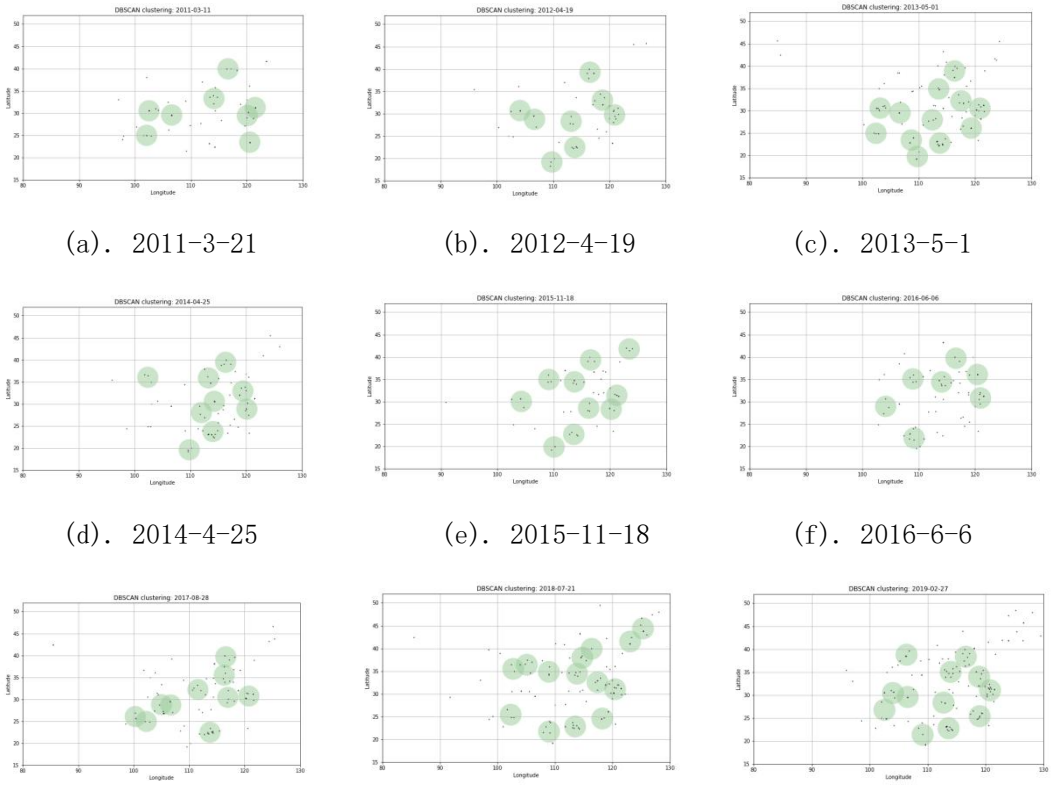
图 5. 突发事件信息条数随时间推移变化的曲线图

从图 5 中可以看到，即使进行了数据去重，每一年的突发事件总量仍然是指数式上升。2009 年突发事件不超过 1000 条，但很快在接下来的十年内迅速上升。这反映了信息化时代突发事件能够被更快更多地报导，从而让用户及时准确掌握第一手信息。

四. 对爆发点的聚类分析

另外，为了发掘突发事件爆发密集区域随时间变化的推移演化，我们还尝试利用 DBSCAN 聚类算法对突发事件的爆发地点进行聚类，试图发掘爆发最密集的点的演化规律。

我们将邻域半径设置为 100km（经验数值，可以调整），即：相邻两个聚类簇之间的联通半径不超过 100 km。考虑到爆发具有一定的时间性，我们对数据按照爆发时间进行切片分析。结果如图 6 所示：



(a). 2011-3-21　(b). 2012-4-19　(c). 2013-5-1

(d). 2014-4-25　(e). 2015-11-18　(f). 2016-6-6

(g). 2017-8-28　(h). 2018-7-21　(i). 2019-2-27

图 6. 部分时间切面的 DBSCAN 聚类效果

（*注：横纵坐标分别表示各地点的经纬度）

为了更好地描述高密度区域，我们将每个聚类簇内样本的坐标取平均值作为样本中心点，描述在一定范围内的高密度区域演化规律。

从聚类结果中可以发现几条规律：

1. 随着时间的推移，爆发密度集中区域的数量也在随之上升。

2. 爆发区域虽然一直在变换，但也多集中于华东、华中、华南、华北和四川地区，其中广东几乎是雷打不动，江浙沪、京津冀地区也是常客。

3. 突发事件在夏季爆发会比在冬季爆发更多。

突发事件的演化特征

一. 基于共现矩阵的灾害链建模

共现矩阵是灾害链建模中一类非常有效的模型。可以采取其基本思想对微博数据进行初步的统计与建模。我们记所有的突发事件类别形成了事件集合 O，那么突发事件的引发关系就可以描述为事件形成的二元组向量，即类似于 $M=\{<A,B>|A,B\in O\}$ 的形式。在这一向量中，A 作为引发事件，B 作为被引发事件，是有一定先后顺序的。

传统的共现矩阵建模中，每一项的含义是在两个突发事件的时间差位于时间区间[-T,T]内（这里我们取 T 为 90 日），仅仅是强调共现关系而没有考虑先后顺序。事实上，灾害链建模考虑的问题并非单纯的“共现”形成的相关关系，而是由时间顺序连结形成的“因果关系”。所以，如果把共现矩阵看作图的关联矩阵，那么这个图论图是一个有向图而非无向图。

我们将这一共现矩阵的部分展示如表 2 所示：

表 2. 突发事件统计的共现矩阵部分展示

事件种类	刑事案件	环境污染和生态破坏	地震灾害	安全事故	群体性事件
刑事案件	6080188	2707872	716468	532018	386081
环境污染和生态破坏事件	2824174	1382881	337225	257599	189836
地震灾害	932107	406118	166686	84044	60279
安全事故	564188	258102	67670	57786	38260
群体性事件	432758	198432	49264	40087	34007
气象灾害	436031	191531	54504	37845	28941
洪涝灾害	236803	98801	30913	21400	15694
网络与信息安全事件	162946	77620	17657	16483	12563
公共卫生事件	270880	136234	38527	25831	19379
地质灾害	194777	78818	32803	16859	12546
动物疫情	85220	41219	10792	8832	6252
森林草原火灾	37236	16428	5704	3403	2487
海洋灾害	24663	11381	3188	2394	1826
恐怖袭击事件	11028	5227	1306	1051	852
金融突发事件	9193	4264	1178	956	667
生物灾害	5424	2308	588	499	331
石油供应中断突发事件	495	258	76	85	40
涉外冲突事件	295	168	44	42	27
事件种类	**气象灾害**	**洪涝灾害**	**网络与信息安全事件**	**公共卫生事件**	**地质灾害**
刑事案件	414853	246229	158625	252901	163112

环境污染和生态破坏事件	198683	110438	80821	143671	70567
地震灾害	75841	46511	23110	38565	40653
安全事故	41490	23082	17243	22841	15552
群体性事件	31957	18521	13740	18622	11731
气象灾害	41180	23711	12526	16238	15327
洪涝灾害	23312	19141	7298	5633	11146
网络与信息安全事件	11691	7708	6717	5071	4404
公共卫生事件	21123	9468	6168	28200	6890
地质灾害	18323	13119	5461	5668	12098
动物疫情	6169	3388	2831	4224	2403
森林草原火灾	2943	2029	1224	2073	1520
海洋灾害	1959	988	809	802	743
恐怖袭击事件	818	467	394	446	288
金融突发事件	692	466	305	310	325
生物灾害	540	471	184	124	227
石油供应中断突发事件	102	24	24	8	20
涉外冲突事件	21	12	17	10	17

事件种类	**动物疫情**	**森林草原火灾**	**海洋灾害**	**恐怖袭击**	**金融突发事件**
刑事案件	83804	29001	19513	10768	9613
环境污染和生态破坏事件	41911	13407	9506	5642	4991
地震灾害	12700	6681	3928	1749	1489
安全事故	8962	2955	2283	1119	1048
群体性事件	6900	2076	1612	873	795
气象灾害	6610	1940	1854	803	654
洪涝灾害	3933	795	952	476	390
网络与信息安全事件	3065	843	803	402	381
公共卫生事件	3115	1003	783	445	493
地质灾害	2701	866	882	327	292
动物疫情	2081	344	428	248	214
森林草原火灾	569	774	175	88	68
海洋灾害	407	131	276	46	48
恐怖袭击事件	197	102	59	32	36
金融突发事件	172	55	42	16	27
生物灾害	93	17	14	7	7
石油供应中断突发事件	8	3	10	1	3
涉外冲突事件	7	2	1	2	3

（*注：表格中每一项表示该行对应的突发事件在前，该列对应的突发事件在后，形成的时间差不超过 90 日的事件组数量。）

那么我们其实可以发现，对于这一矩阵的某一个元素，该行和该列的和是不相等的。关于这一点，我们可以理解为，若对该元素的某一行进行求和，则是该行代表的事件 A 作为先验条件引发事件数目；而对该列进行求和，则是对该列代表的事件 B 作为结果，即由其他所有事件引发 B 的数目。即使是对于同一种事件，它作为原因和作为结果，所引发或被引发的事件数目其实是不一样的。

为更直观描述引发的现象，这里引入共现率矩阵。描述共现率的指标有很多，其中，Jaccard 指数是一类计算共现率十分优良的指标[31]。Jaccard 指数定义如下：

$$J_{AB}=\frac{C_{AB}}{C_A+C_B-C_{AB}}(0\leqslant J_{AB}\leqslant 1) \quad (1)$$

为了研究灾害链的时间性和因果性，我们对原始的 Jaccard 指数稍作改动：

第一点改动是引入时间因素。将爆发时间差再按照 1-90 日一共 90 个时间差值进行分切，将每个事件组$M=\{<A,B>|A,B\in O\}$中的事件再按时间差进行分堆，并重新计算 J_{AB} 的数值：

$$CO(A,B)=\sum_{t=1}^{90}\frac{J_{AB}(t)}{t} \quad (2)$$

$$J_{AB}(t)=\frac{C_{AB}(t)}{C_A+C_B-C_{AB}(t)} \quad (3)$$

其中，$C_{AB}(t)$表示时间差等于 t 的事件组数量。那么在(2)-(3)中，我们引入了时间成分进行修正。引入时间因素的原因在于我们需要充分考虑到两事件若相隔太久远则因果关系越弱，所以关联性应该是与时间长度呈负相关的关系。

第二点改动是引入先后顺序。由于在我们研究的共现矩阵中每一项带有先后顺序性，所以我们对（3）进行进一步改进：

$$J_{<A,B>}(t)=\frac{C_{<A,B>}(t)}{C_{<A,:>}+C_{<:,B>}+C_{<A,B>}(t)} \quad (4)$$

在(4)中，$C_{\langle A,B\rangle}(t)$规定了事件的先后顺序是从 A 引发 B，而记号 $C_{\langle A,:\rangle}$表示由 A 引发的所有事件总数。这样，我们从原有的无向图 Jaccard 系数计算变换到有向图的 Jaccard 系数计算。

计算得到的的 Jaccard 系数非常小，为了放大它们之间的差异，我们对数据进行 Z-score 标准化：

$$J_i=\frac{J_i-\min(J_i)}{\min(J_i)-\max(J_i)} \quad (5)$$

我们将标准化后的共现率矩阵形成的热力图以及对共现率矩阵再度归一化以后的马尔可夫矩阵（这一变化可以清晰看到由一个事件引发另一事件的概率）热力图呈现在图 7 中：

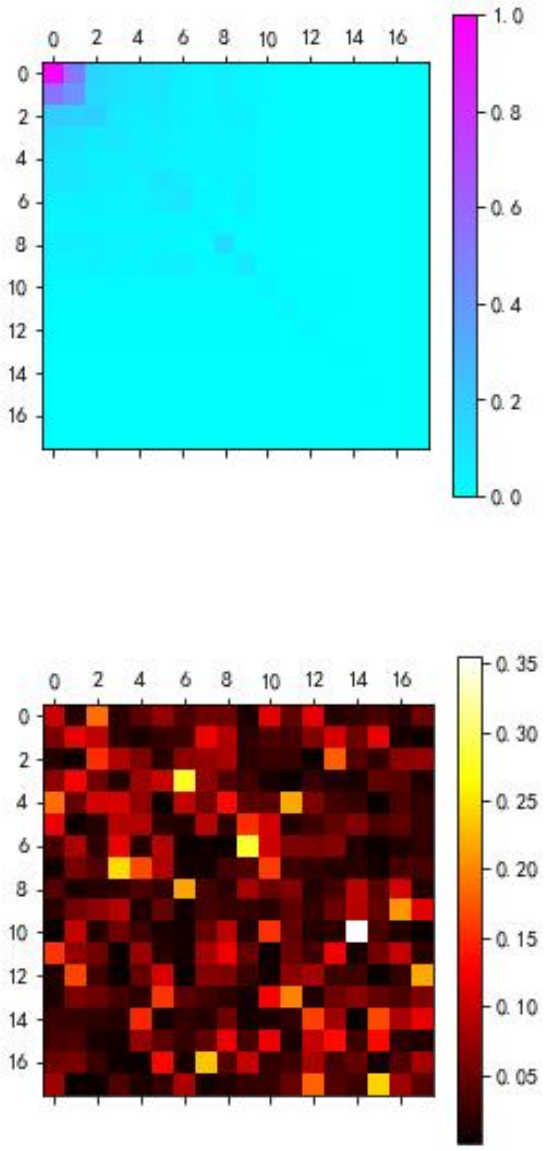


图 7. 经 Z-score 标准化以后的共现率矩阵热力图和马尔可夫矩阵热力图

可以看到：除了刑事案件和环境污染与生态破坏之间的共现率较高以外其它都比较低，大部分甚至都接近于 0。这说明有些事件因果性是很弱的。

当一类突发事件爆发后，除了可能直接引起下一类突发事件以外，还可能引起次生事件。为了描述一类事件引发的次生灾害的严重程度，我们定义事件 A 与其引发事件 B 的危害程度如式(6)所示：

$$D(<A,B>)=\alpha I(A\rightarrow B)+\beta\sum_{c\in O}I(B\rightarrow c) \quad (6)$$

那么对于不同的主特征系数α和次要特征系数β取值，事件造成的次生灾害程度也有所不同。例如，对于洪涝灾害，若取α=0.8，β=0.2，它与其它引发事

件的危害程度如图 8 所示：

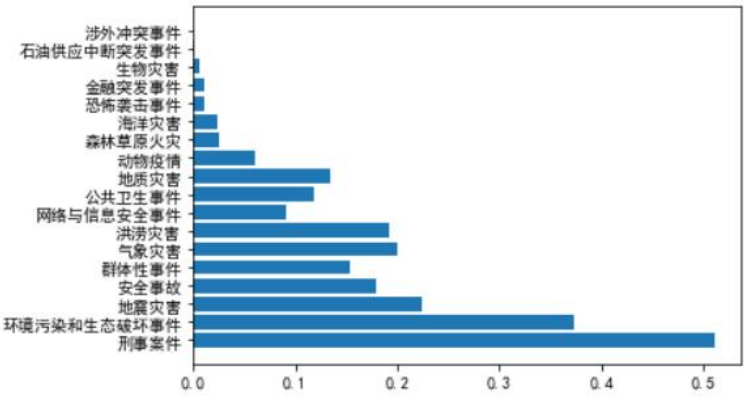


图 8. 洪涝灾害引发其他事件的危害程度

从图 8 中可以看到：刑事案件的危害是远高于其它危害的，其次是环境污染和生态破坏事件与地震灾害。涉外冲突、石油供应链中断等事件由于爆发次数极少所以危害暂时不明显。

若改变每一类突发事件的主特征系数和次要特征系数（控制α+β=1）可以发现，对于洪涝灾害引发环境污染和生态破坏时间的危害程度会如图 9 所示的线性变化：

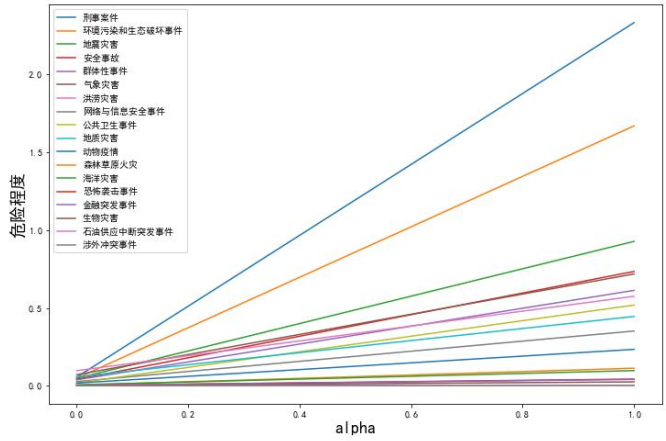


图 9. 洪涝灾害引发次生事件的危害程度变化

二. 基于复杂网络的进一步改进

基于复杂网络理论，我们将共现率矩阵视作一个图的关联矩阵，从而建立有向图模型，各边的权重为共现率矩阵对应位置的值。突发事件引发形成图的拓扑结构被描述在图 10 中：

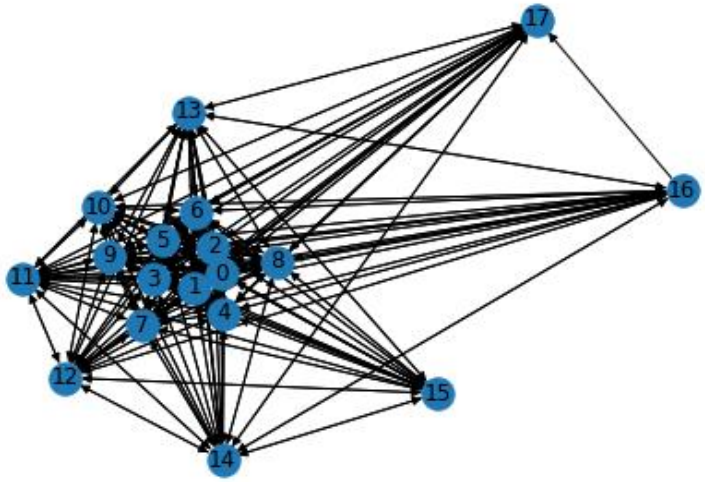


图 10. 各事件形成的共现图谱

(*注：图 10 中 0-17 号节点分别代表刑事案件、环境污染和生态破坏、地震灾害、安全事故、群体性事件、气象灾害、洪涝灾害、网络与信息安全时间、公共卫生事件、地质灾害、动物疫情、森林草原火灾、海洋灾害、恐怖袭击、金融突发事件、生物灾害、石油供应链中断、涉外冲突）

研究这一复杂网络的几个重要指标如下：

在图论中，研究网络的稳定性最简单的一个指标就是节点的度。我们将 0-17 号节点的度列在表 3 中：

表 3. 共现图谱中各节点的度数

事件	刑事案件	环境污染和生态破坏	地震灾害	安全事故	群体性事件	气象灾害
度	36	36	36	36	36	36
事件	洪涝灾害	网络与信息安全事件	公共卫生事件	地质灾害	动物疫情	森林草原火灾
度	36	36	36	36	36	36
事件	海洋灾害	恐怖袭击	金融突发事件	生物灾害	石油供应链中断	涉外冲突
度	36	36	36	32	31	31

由于共现图谱是一个带权图，我们认为，仅仅考虑节点的度是不够的，还需要充分考虑节点的权值。所以，我们用考虑边权值的点权描述这一节点的重要性。点权定义如(7)所示：

$$(7).\ S_i = \sum w_{ij}$$

除此以外，我们为了考虑到各节点的引发权值和被引发权值不同，所以还将点权分为引发权和被引发权，并列入表 4：

表 4. 共现图谱中各节点的点权

事件	刑事案件	环境污染和生态破坏	地震灾害	安全事故	群体性事件	气象灾害
引发权	2.266	1.647	1.02	0.737	0.622	0.692
被引发	2.333	1.669	0.928	0.734	0.613	0.718

权						
点权	4.599	3.316	1.948	1.471	1.235	1.41
事件	洪涝灾害	网络与信息安全事件	公共卫生事件	地质灾害	动物疫情	森林草原火灾
引发权	0.533	0.341	0.523	0.478	0.233	0.141
被引发权	0.575	0.352	0.518	0.445	0.234	0.113
点权	1.108	0.693	1.041	0.923	0.467	0.254
事件	海洋灾害	恐怖袭击	金融突发事件	生物灾害	石油供应链中断	涉外冲突
引发权	0.099	0.044	0.034	0.025	0.003	0.002
被引发权	0.097	0.043	0.04	0.025	0.002	0.002
点权	0.196	0.087	0.074	0.05	0.005	0.004

节点的点介数被定义为网络韧性的一种度量，是描述图的一种更加合理的方式能够合理地描述通过这一节点的最短路径数量占比情况，是衡量这一节点重要性的一个重要指标[32]。其计算公式为：

$$b(v)=\sum_{i\neq k}\frac{g_{ik}(v)}{g_{ik}} \quad (8)$$

我们将各突发事件的点介数列在表格 5 中：

表 5. 各节点的点介数

事件	刑事案件	环境污染和生态破坏	地震灾害	安全事故	群体性事件	气象灾害
点介数	34	34	34	34	38	34
事件	洪涝灾害	网络与信息安全事件	公共卫生事件	地质灾害	动物疫情	森林草原火灾
点介数	34	34	60	34	34	34
事件	海洋灾害	恐怖袭击	金融突发事件	生物灾害	石油供应链中断	涉外冲突
点介数	34	34	64	34	87.5	244.5

聚类系数是描述网络的稳定性的重要指标，其中，有向加权网络的聚类系数定义如下[33]：

$$C(i)=\frac{1}{m_i(m_i-1)}\sum_{j,k}\frac{k_i}{S_i}\frac{w_{ij}+w_{ji}+w_{ik}+w_{ki}}{n_{ijk}}a_{jk} \quad (9)$$

各节点的聚类系数如表 6 所示：

表 6. 各节点的聚类系数

事件	刑事案件	环境污染和生态破坏	地震灾害	安全事故	群体性事件	气象灾害
聚类系数	1.9998	1.9997	1.9996	1.9995	1.9994	1.9993
事件	洪涝灾害	网络与信息	公共卫生	地质灾害	动物疫情	森林草原

		安全事件	事件			火灾
聚类系数	1.9991	1.9987	1.9995	1.9992	1.9981	1.9973
事件	海洋灾害	恐怖袭击	金融突发事件	生物灾害	石油供应链中断	涉外冲突
聚类系数	1.997	1.9963	1.996	2.3926	2.5497	2.9494

我们可以发现：由于这一网络大部分节点都与其他所有节点直接相连且有双向边，所以大部分事件的聚类系数都在 2.0 左右。生物灾害、石油供应链中断和涉外冲突聚类系数较高，一方面是因为它们的度与其它节点并不相等，另一方面是它们连接的所有边权值都较小，通过了很多最短路径。

总体来说，由于这一图谱接近完全图，所以这一图论图会具备拓扑学意义上的高度稳定性。

三. 对爆发-报道时间差的统计分析

突发事件的爆发和被报道之间存在一个时间差。通过对时间差的统计分析能够从传播学的角度对突发事件在网络上的传播进行一定的分析。对于四类突发事件，事件类型不同往往被报道的时间与爆发时间的时间差也不同。所以，我们对时间差进行统计分析。我们将时间差大于 90 日的新闻视作 90 日，而将时间差小于 0 的事件视作报道有误而剔除。最后我们四类样本的均值如表 7 所示。除此以外，我们还对数据进行了基于 K-S 检验的正态性检验检验统计量如表 7 所示。

这四类突发事件的概率值都是非常接近 0，所以我们拒绝原假设，认为样本不服从正态分布。

表 7. 四类突发事件的时间差正态性检验

统计量	社会安全事件	自然灾害事件	事故灾难事件	公共卫生事件
均值/天	6.085	4.207	6.304	5.81
标准差/天	19.659	15.195	19.817	17.175
检验统计量	0.505	0.518	0.514	0.54
概率值	0	0	0	0

相比这些基本统计特征，更重要的是发现这些事件被报道时间差的一个差异。配对样本 t 检验用来揭示定量数据的对比关系，统计量定义为：

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \quad (10)$$

统计量之间的结果如表 8 所示。

表 8. 两样本进行双样本 t 检验结果

事件 1	事件 2	统计量	概率值	是否接受原假设
社会安全	自然灾害	13.627	2.98E-42	否
社会安全	事故灾难	-1.891	0.0586	是
社会安全	公共卫生	0.899	0.3686	是
自然灾害	事故灾难	-14.223	7.72E-46	否
自然灾害	公共卫生	-6.224	4.91E-10	否
事故灾难	公共卫生	1.575	0.1152	是

取显著性标准为 0.05，可以很明显地看到，自然灾害与其它三者之间都存在明显的差异，说明自然灾害爆发则需要更快更及时地报道，这一点也比较符合我们的常规认知。

另外，我们对十八个小类两两组合进行了双样本 t 检验，检验结果显示每两个小类都是有所差异的。而对它们的均值进行观察，诸如网络与信息安全事件、金融突发事件等不易引起他人关注的事件平均的延迟时间达到了 9.86 天、8.26 天。而对于地震灾害、气象灾害等延迟时间相对较低，分别为 3.56 天、4.12 天。当然，随着近年来网络与信息化时代的进一步发展，地震的灾害播报效率也在进一步提升。

四. 案例：以新型冠状肺炎病毒为例

2020 年初爆发的新型冠状肺炎病毒是一个典型的突发性公共卫生事件。根据上述方法建模，求解归一化的马尔可夫矩阵来得到概率，可以得到引发概率最高的五类事件分别为公共卫生事件、动物疫情、群体性事件、刑事案件和网络与信息安全事件。针对这些事件，虽然有很多内生性因素导致引发关系受到干扰，但大体上结论符合认知。

首先对于最高的公共卫生事件，自身的引发说明这一事件具有较强的时间与空间扩散效应，造成直接影响是严重的。世界卫生组织 3 月 11 日表示，新冠肺炎疫情的爆发已经构成一次全球性“大流行”。2020 年 6 月 7 日，国务院新闻办公室发布《抗击新冠肺炎疫情的中国行动》白皮书。10 月 8 日，中国同全球疫苗免疫联盟签署协议，正式加入“新冠肺炎疫苗实施计划”。截至 2020 年 12 月 31 日 24 时，据 31 个省（自治区、直辖市）和新疆生产建设兵团报告，现有确诊病例 370 例（其中重症病例 9 例），累计治愈出院病例 82067 例，累计死亡病例 4634 例，累计报告确诊病例 87071 例，现有疑似病例 1 例。累计追踪到密切接触者 905493 人，尚在医学观察的密切接触者 13584 人。

其次，动物疫情可能是受到内生性因素干扰，在这一次事件中并未过多发生。但也可能仅仅针对这一次事件没有发生，在其它公共卫生事件中仍然需要注意防范。

第三，群体性事件，这一点也可以解释。在武汉封城期间，有许多居民对封城举措表示不理解，也对相关部门的支持工作例如食物配送、上门量体温、全员核酸检测等工作上的一些疏忽表示不满，也有对疫情产生爆发的恐慌和不知所措等情绪进而产生过激风险感知，当负面情绪开始在时间维度或空间维度上聚集，就导致了网络上或者现实中的群体极化现象，进而发生群体性事件。这一过程也与媒体在传媒预警过程中的工作有关，在“新型肺炎”疫情的初期阶段，武汉当地媒体基本处于缺位失声状态，比如当地最主要的综合性市民报纸《楚天都市报》和《武汉晚报》，在 1 月 20 日之前，鲜少将疫情内容作为头版内容，即使内版报道，篇幅也比较少。因此，零碎的自媒体爆料内容与武汉当地媒体的缺位失声，让疫情未受到公众的重视，也贻误了信息公开和扩大社会影响的最佳时机。

第四，刑事案件，这一事件结果的产生是由于受到数据量失衡和内生因素干扰导致，不具备一般性的借鉴意义，决策者在进行判断时不应将其作为特定的演化模式。

第五，网络与信息安全事件，这一点也符合公众认知。在封城期间，由于媒体预警缺位，官方对疫情信息的不准确把握导致公民在疫情信息的传播中产生一

定偏差。严重情况下就导致谣言在网络上扩散，由于公民的过激风险感知与恐慌情绪容易相信谣言进而引发一系列后续事件，这一类事件可以视作网络与信息安全事件。包括期间出现的冒充政府工作人员进行诈骗等安全事件进一步证实了网络安全事件的管控与预防必要性。那么在这一过程中，媒体最应该做的就是严防预警失声和缺位，快速反应及时传递准确信息和辟谣，科学防疫，传递正确措施，安抚民众情绪，让公民抱有正能量，相信政府，配合政府工作。

对结果的讨论与结论

一. 时空的分布规律

（1）从时空维度上来讲，突发事件多发生在中原以及东南部和京津冀地区，而西北、东北等地相对较少。一方面是由于人口因素，人口相对密度小，冲突不容易发生。其次，边境地区多为少数民族，这也从侧面反映了国家在对少数民族地区的管理中采取了相当正确的方法，使这些边境地区社会安定人民幸福。

（2）随着时间的推移，突发事件的爆发密度越来越高。信息化时代的到来，使得更多的突发事件能够被及时报道，所以虽然看起来不稳定的因素似乎在增加，但实际上并非突发事件真的在指数式增长，从另一个角度也可以看到媒体传播在灾害防控过程中的重要作用。

（3）一个非常有趣的现象是，社会安全事件多发于我们认为最富庶的北上广等地区。出现这一现象的原因也是由于人口密度一旦密集就容易发生冲突从而造成一定程度的社会安全事件。那么一个地区的 GDP 和社会安全事件的爆发风险是否确实存在一定的联系，还有待进一步研究论证。

（4）夏季突发事件的爆发频次往往多于冬季。这一点也很好理解。例如，暴雨洪涝灾害多发于夏季，而它又容易引发其他灾害比如山体滑坡，甚至水体污染导致公共卫生事件等。

二. 状态的转移规律

（1）通过共现率矩阵求解得到的最大几项概率都是刑事案件与环境污染与生态灾害之间的转化，这是一个比较新颖的发现。不排除由于样本类别不平衡导致结果有一定偏差，但是这一结果在一定程度上也从统计学的角度反映了环境污染防治的另一个重要意义。

出现这一现象的原因是由于事件的互通性。对一件事情处理不当会诱发更多的事情，导致社会的动荡与不安。这就要求我们必须要在事前进行多重准备与事发后的立刻行动，这样才能尽可能减少天灾人祸带来的损失。

（2）事件并非独立存在，不同的事件之间存在一定的关联，对一个突发事件的防控需要从多个角度来采取措施。同时，一个事件的爆发可能引起多米诺效应。所以要重视突发事件的及时处理，从而避免诱发其他事故。

对状态转移规律的研究让我们更直观地看到事件的关联，从而对不同类型突发事件有更全面的认识。通过大数据的手段研究实际问题，从多个角度去看待并解决问题，对不同事件提出看法与建议是一项很有价值的工作。

三. 传播时间的统计规律

（1）从突发事件的传播时间来看，只有自然灾害类事件的爆发-报道时间差显著差异于其它事件，这是由于自然灾害事件的特殊性。自然灾害类时间传播快危害大，能够很容易引发一系列严重的连锁反应，所以报道会比较快。以往的自然灾害在网络的传播容易造成长期延迟，而目前的自然灾害报道时间大都能够控制在 24 小时甚至 12 小时以内。

（2）而对大类突发事件进一步划分，我们的 18 小类的突发事件中任两项都是有差异的。其中，诸如网络与信息安全事件等不易引起他人关注的事件平均的延迟时间较长。而对于地震灾害等传播快危害大的突发事件，以及涉外冲突这类

特殊事件，延迟时间相对较低。

出现这一现象的原因是由于事件的影响力和人们对该事件的敏感程度不同。从事件发生到被报道的时间可以反映外部世界对事件的重视程度。例如，自然灾害等突如其来的灾难不仅危害大而且后续的连锁事件很多所以能够在短时间内被关注并报道；再像公共卫生事件会对大范围内的人造成危害，因此人们对该类事件反应比较迅速；但像社会安全事件在短时间内只是损害一部分人的利益，所以从发生到被报道会经过一段时间的延迟。此外，传播时间也会随着不断发达的网络而逐渐变短，人们获取外界信息的速度不断加快。对于传播时间的研究可以更好地了解人们的关注和需求，充分发挥我们这次工作的价值。

四. 总结与展望

客观物质世界中存在着客观规律，探索突发事件发生的规律有助于我们从根本上大致把握导致突发事件发生的原因，从而能尽可能地剔除祸根，防范于未然，增强我们对突发事件的预防能力，同时也减轻突发事件的危害程度。导致灾害发生的因果链条是复杂、隐晦的，并不容易轻易被发现和理清，且存在偶然因素，干扰我们对因果关系链条上组成成分的判断。但是我们能做的，就是利用大数据分析，尽可能地扩大研究对象和样例，利用统计方面相关技术，找到突发事件中一些现象的恒常联系，对这些现象和事件加以特别地关注和回避，以避免相同的灾难再次发生，减少灾难对人类社会产生的消极影响。

致谢

在本文编写的过程中，首先我非常感谢我的导师王然老师对我平日里的悉心指导，另外数学与统计学院的王新宇老师也提供了很多有价值的参考意见。张誉馨同学、和肖渝楠同学在这一工作中也付出了很多努力。

另外，我还要感谢同实验室的贾彬学长在数据的处理方面提供了很大程度的帮助，还有钟盛涛学长、李逸尧学长在平时学习里提供了很多技术指导，在此一并感谢。

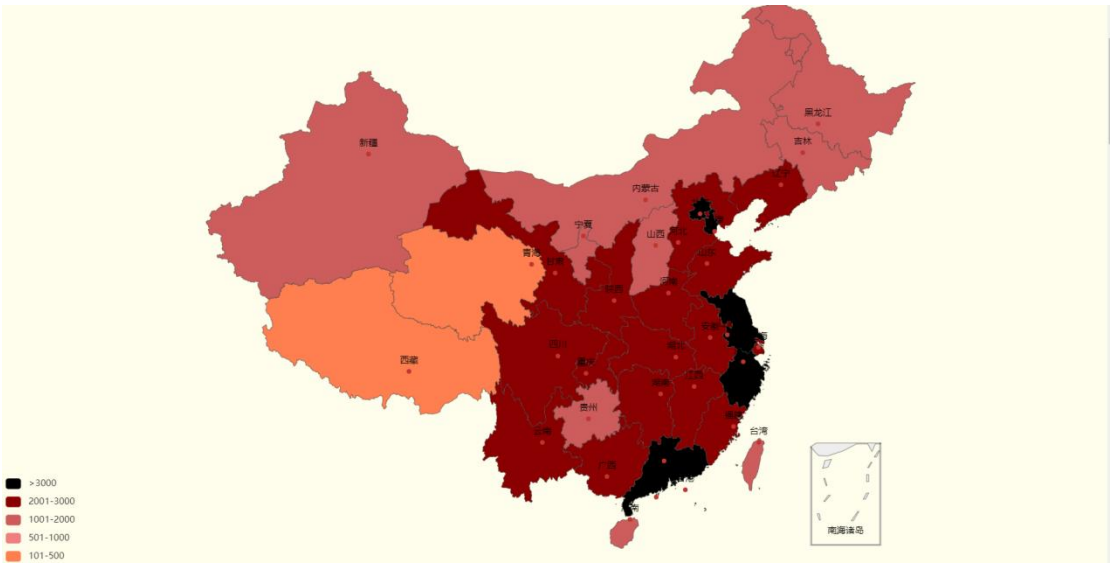
参考文献

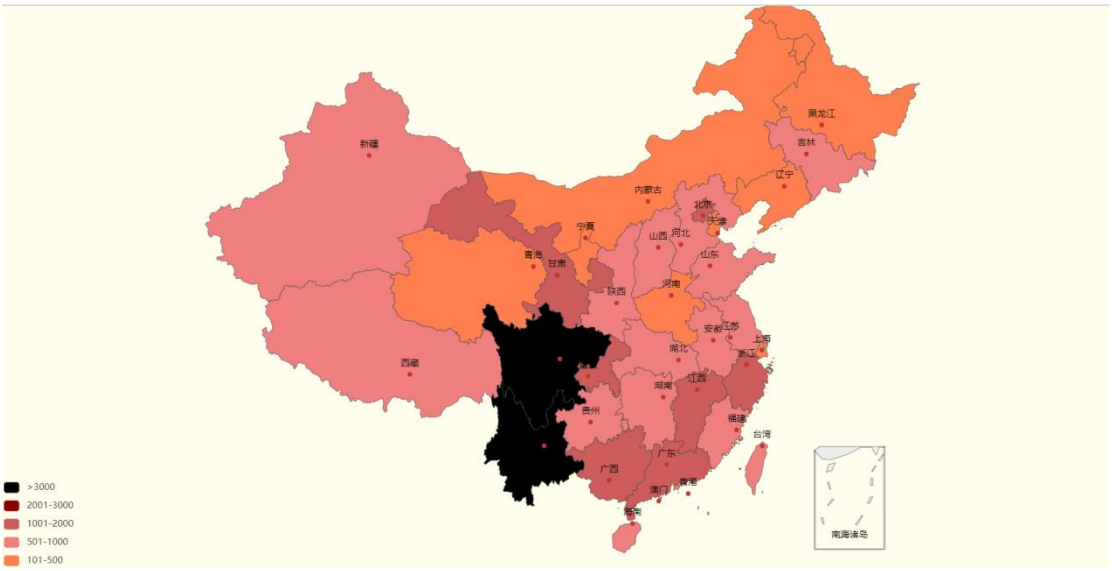
[1] 郭增建, 秦保燕. 灾害物理学简论[J]. 灾害学, 1987(2):30-38.
[2] 高建国. 中国灾害链研究[C]// 2007 中美灾害防御研讨会. 2007.
[3] 王翔. 区域灾害链风险评估研究[D].大连理工大学,2011.
[4] 刘永志,唐雯雯,张文婷,张行南,牛帅.基于灾害链的洪涝灾害风险分析综述[J].水资源保护,2021,37(01):20-27.
[5] 余世舟, 张令心, 赵振东,等. 地震灾害链概率分析及断链减灾方法[J]. 土木工程学报, 2010(S1):479-483.
[6] 孙宝军. 内蒙古电力系统自然灾害链分析[J]. 灾害学, 2020, v.35;No.139(04):10-14+49.
[7] Kurant M , Thiran P . Complex Networks[J]. all publications, 2005.
[8] Newman M . The structure and function of complex networks[J]. Siam Review, 2003.
[9] 许超. 复杂网络研究综述[J]. 卷宗, 2018, 008(033):242-244.
[10] 李辉，陈福才，张建朋，吴铮，李邵梅，黄瑞阳.复杂网络中的社团发现算法综述[J/OL].计算机应用研究:1-9[2021-5-23].https://doi.org/10.19734/j.issn.1001-3695.2020.06.0211.
[11] 汪黎明. 基于深度强化学习的复杂网络关键节点识别[D].安徽财经大学,2020.
[12] 胡明生, 贾志娟, 雷利利,等. 基于共现分析的历史自然灾害关联研究[J]. 计算机工程与设计, 2013, 34(006):2015-2019.
[13] 胡明生, 贾志娟, 雷利利,等. 基于复杂网络的灾害关联建模与分析[J]. 计算机应用研究, 2013(08):2315-2318.
[14] Mansour O . Group Intelligence: A Distributed Cognition Perspective[C]// Intelligent Networking and Collaborative Systems, 2009. INCOS '09. International Conference on. IEEE, 2009.
[15] 段海滨, 王道波, 朱家强,等. 蚁群算法理论及应用研究的进展[J]. 控制与决策, 2004(12):1321-1326.
[16] 倪庆剑, 邢汉承, 张志政,等. 粒子群优化算法研究进展[J]. 模式识别与人工智能, 2007.
[17] 吴斌, 崔志勇, 倪卫红. 具有混合群智能行为的萤火虫群优化算法研究[J]. 计算机科学, 2012, 39(005):198-200.
[18] 胡明生, 贾志娟, 刘思,等. 基于蚁群优化的历史灾害关联分析方法[J]. 计算机应用与软件, 2012.
[19] 胡明生, 贾志娟, 吉晓宇,等. 基于改进萤火虫群的区域灾害链挖掘方法[J]. 计算机应用与软件, 2012, 000(011):29-31.
[20] 王桂红. 基于密度的聚类算法研究[J]. 泉州师范学院学报, 2009(02):44-49.
[21] XIA LuNing, JING JiWu, 夏鲁宁,等. SA-DBSCAN:A self-adaptive density-based clustering algorithmSA-DBSCAN:一种自适应基于密度聚类算法[J]. 中国科学院大学学报, 2009, 26(4):530-538.

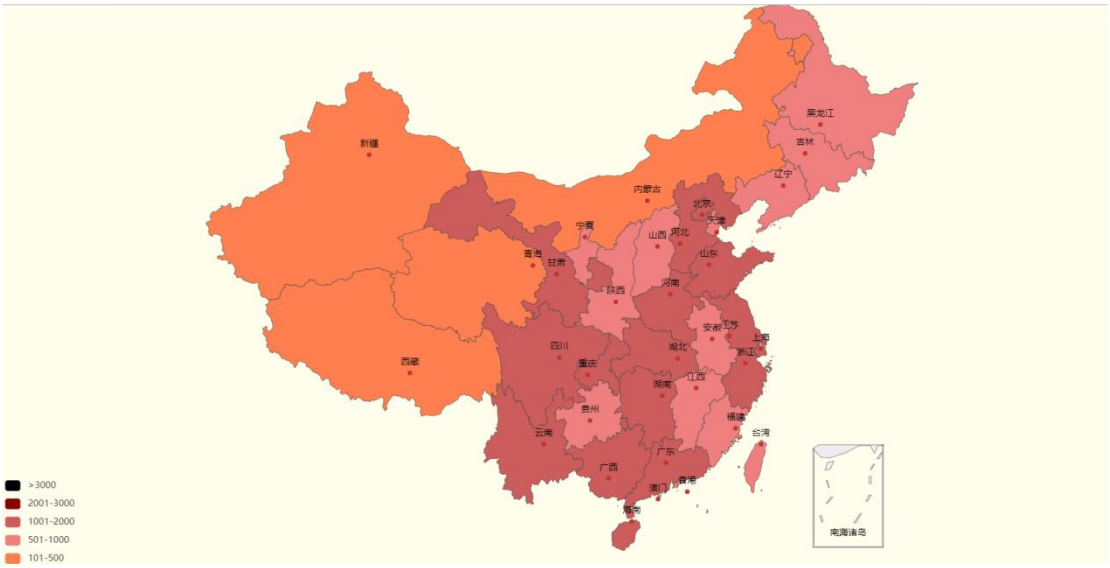
[22] Hickey R J， Black M M . Refined Time Stamps for Concept Drift Detection During Mining for Classification Rules[J]. Springer-Verlag, 2000.
[23] Zhang B， Hsu M， Dayal U . K-Harmonic Means - A Spatial Clustering Algorithm with Boosting[M]. Springer Berlin Heidelberg, 2001.
[24] Roddick J F， Hornsby K . [Lecture Notes in Computer Science] Temporal, Spatial, and Spatio-Temporal Data Mining Volume 2007 || Join Indices as a Tool for Spatial Data Mining[J]. 2001, 10.1007/3-540-45244-3(Chapter 9):105-116.
[25] Estivill-Castro V， Lee I . AUTOCLUST+: Automatic Clustering of Point-Data Sets in the Presence of Obstacles[C]// Temporal, Spatial, & Spatio-temporal Data Mining, First International Workshop Tsdm Lyon, France, September 12, Revised Papers. DBLP, 2001.
[26] Chen S， Mao J， Li G，et al. Uncovering Sentiment and Retweet Patterns of Disaster-related Tweets from a Spatiotemporal Perspective—A Case Study of Hurricane Harvey[J]. Telematics and Informatics, 2019, 47:101326.
[27] Yao,W., Jiao, P., Wang, W.and Sun, Y.,“Understanding human reposting patterns on Sina Weibo from a global perspective”,Physica A: Statical Mechanics and its Applications 518, 374-383(2019)
[28] Prokhorenkova L， Gusev G， Vorobev A，et al. CatBoost: unbiased boosting with categorical features[J]. 2017.
[29] Meng Q . LightGBM: A Highly Efficient Gradient Boosting Decision Tree. 2018.
[30] Shengtao Zhong,Rui Sheng,Ran Wang,Yiyao Li. Prediction of the propagation effect of emergencies microblog[P]. International Symposium on Multispectral Image Processing and Pattern Recognition,2020.
[31] 王铎.基于关联度的突发事件网络模型研究[D].大连理工大学：2010
[32] 钱珺, 王朝坤, 郭高扬. 基于社区的动态网络节点介数中心度更新算法[J]. 软件学报, 2018, 029(003):853-868.
[33] 马梦珂,倪静.基于度值和聚类系数的跨单元调度问题优化[J/OL].计算机应用研
究:1-7[2021-05-27].https://doi.org/10.19734/j.issn.1001-3695.2021.01.0024.

附录

1. 图 3-图 5 的高清图片

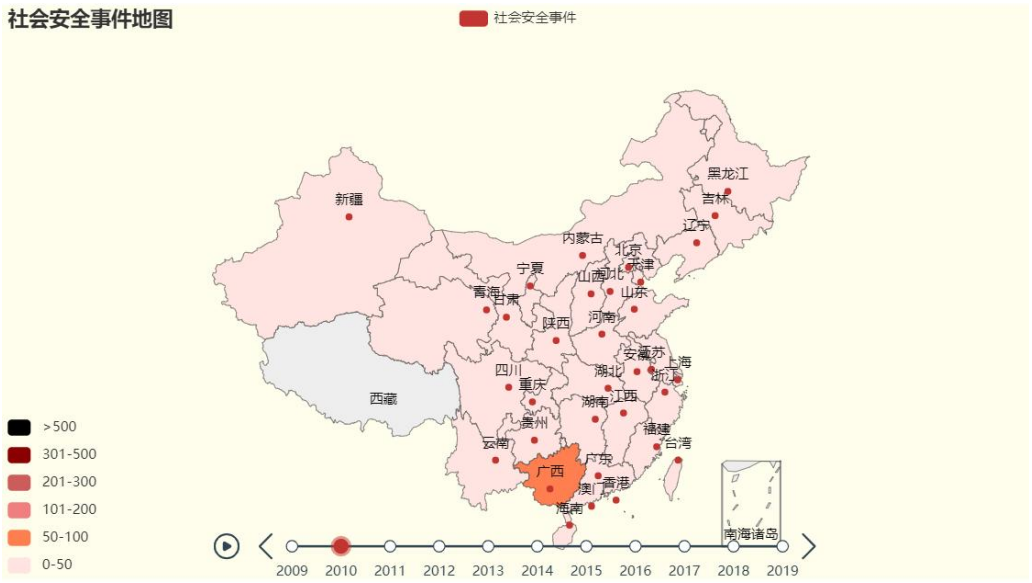


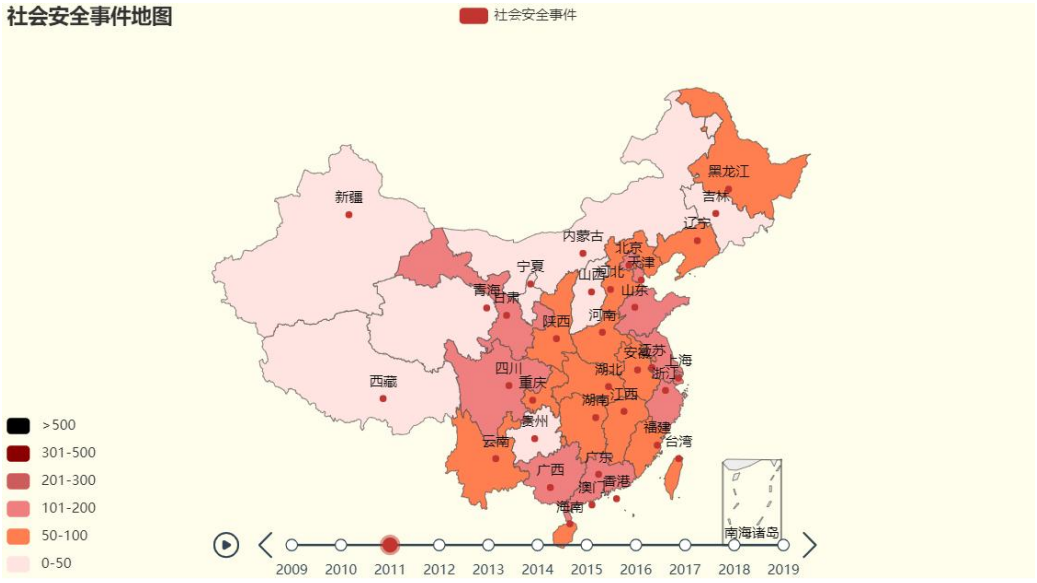


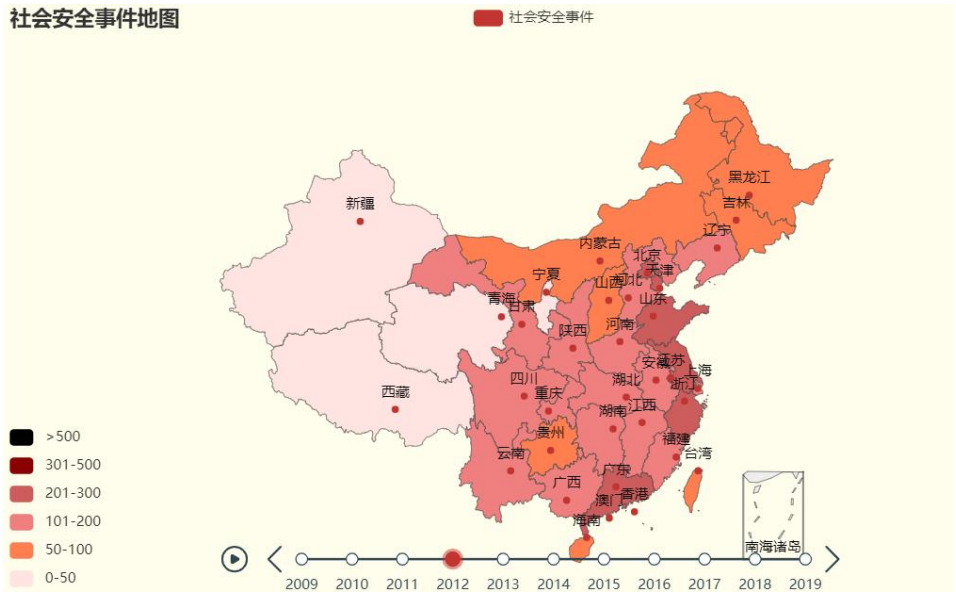

>3000
2001-3000
1001-2000
501-1000
101-500
南海诸岛

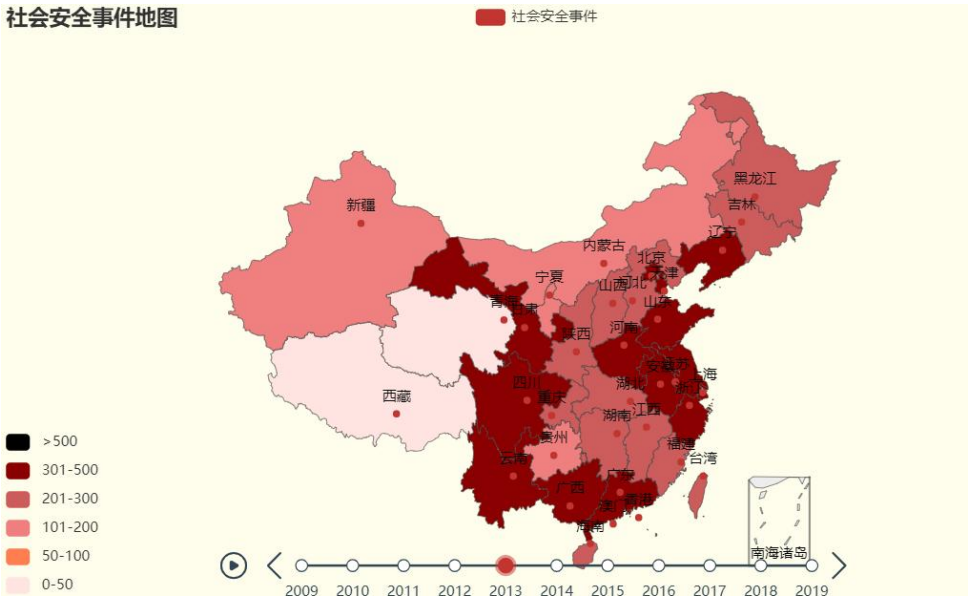

>3000
2001-3000
1001-2000
501-1000
101-500
南海诸岛

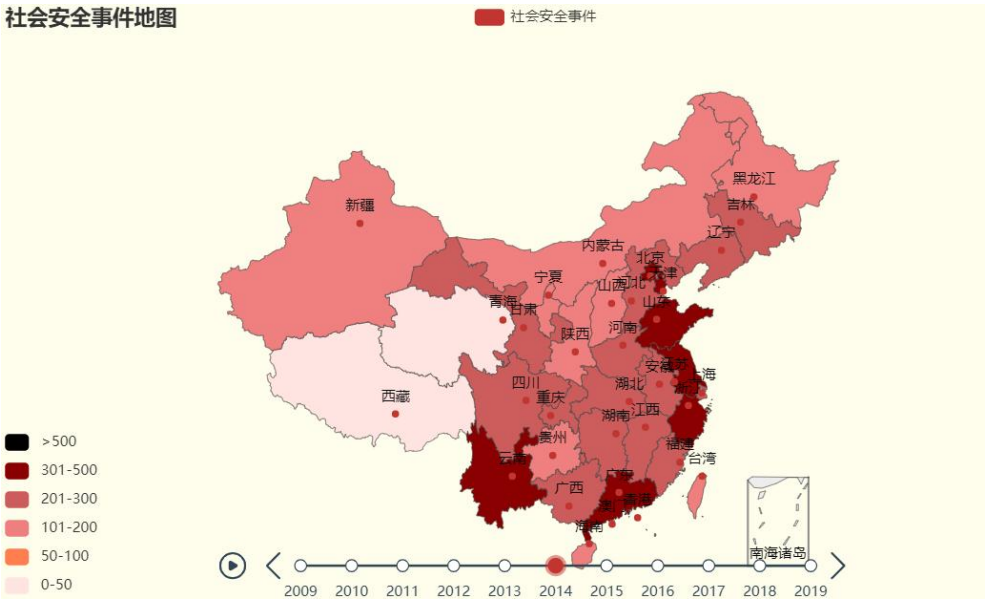

社会安全事件地图
社会安全事件
>500
301-500
201-300
101-200
50-100
0-50
南海诸岛
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

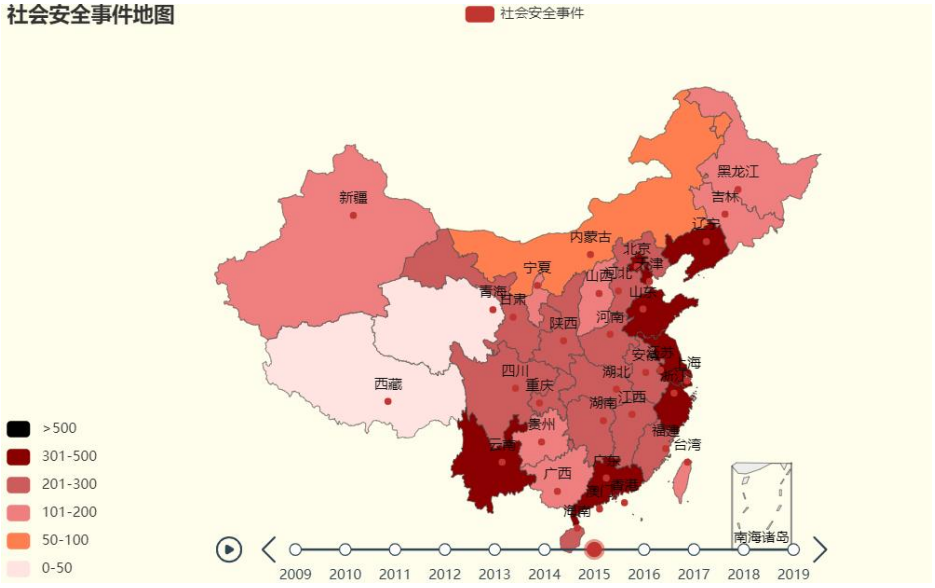

社会安全事件地图
社会安全事件
新疆
西藏
青海
甘肃
宁夏
内蒙古
黑龙江
吉林
辽宁
北京
天津
河北
山西
山东
陕西
河南
安徽
江苏
上海
浙江
湖北
四川
重庆
湖南
江西
贵州
福建
台湾
云南
广西
广东
澳门
香港
海南
南海诸岛
>500
301-500
201-300
101-200
50-100
0-50
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

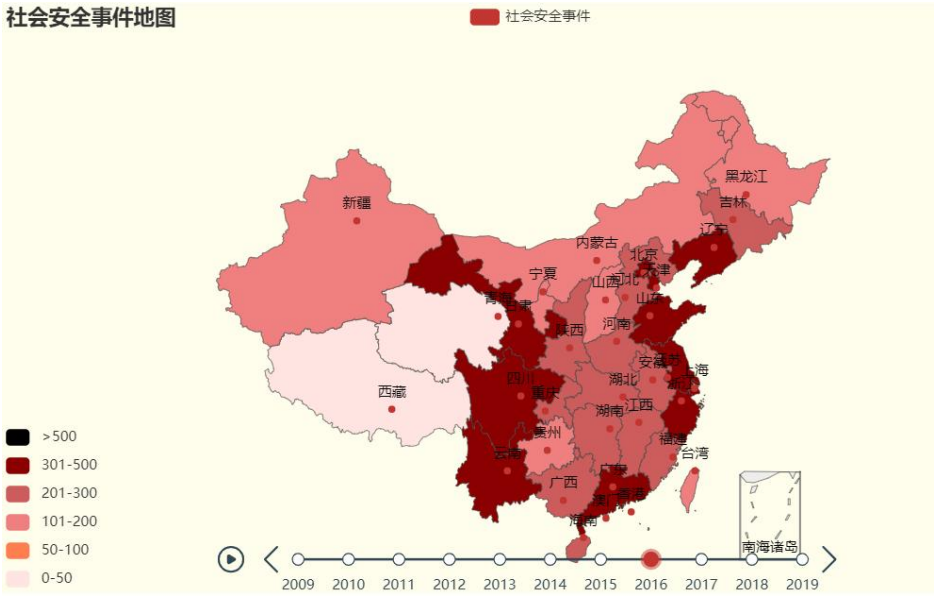

社会安全事件地图
社会安全事件
新疆
西藏
青海
甘肃
宁夏
内蒙古
黑龙江
吉林
辽宁
北京
天津
河北
山西
山东
陕西
河南
安徽
江苏
上海
浙江
湖北
四川
重庆
湖南
江西
贵州
福建
台湾
云南
广西
广东
澳门
香港
海南
南海诸岛
>500
301-500
201-300
101-200
50-100
0-50
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

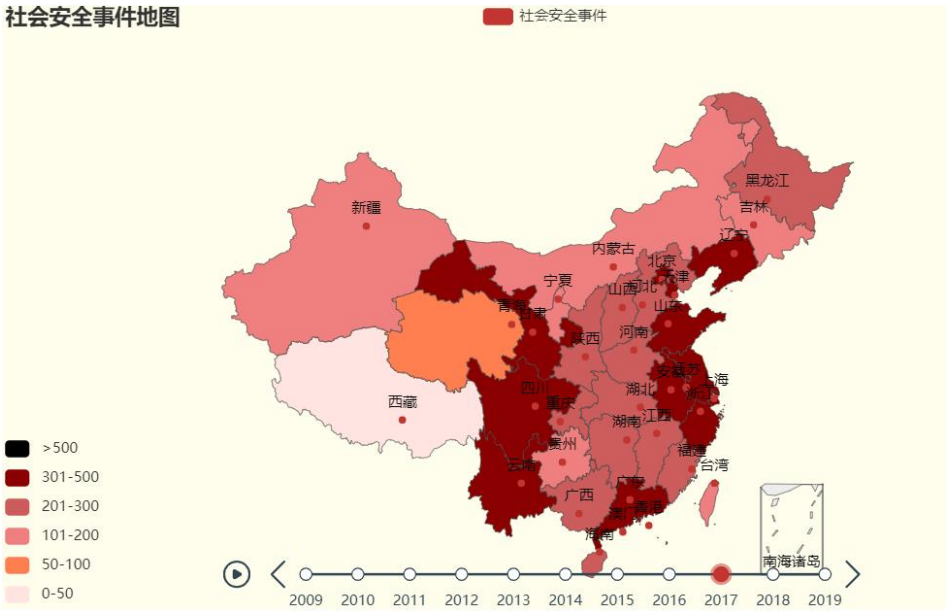

社会安全事件地图
社会安全事件
新疆
西藏
青海
甘肃
宁夏
内蒙古
黑龙江
吉林
辽宁
北京
天津
河北
山西
山东
陕西
河南
安徽
江苏
上海
浙江
湖北
四川
重庆
湖南
江西
贵州
福建
台湾
云南
广西
广东
澳门
香港
海南
南海诸岛
>500
301-500
201-300
101-200
50-100
0-50
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

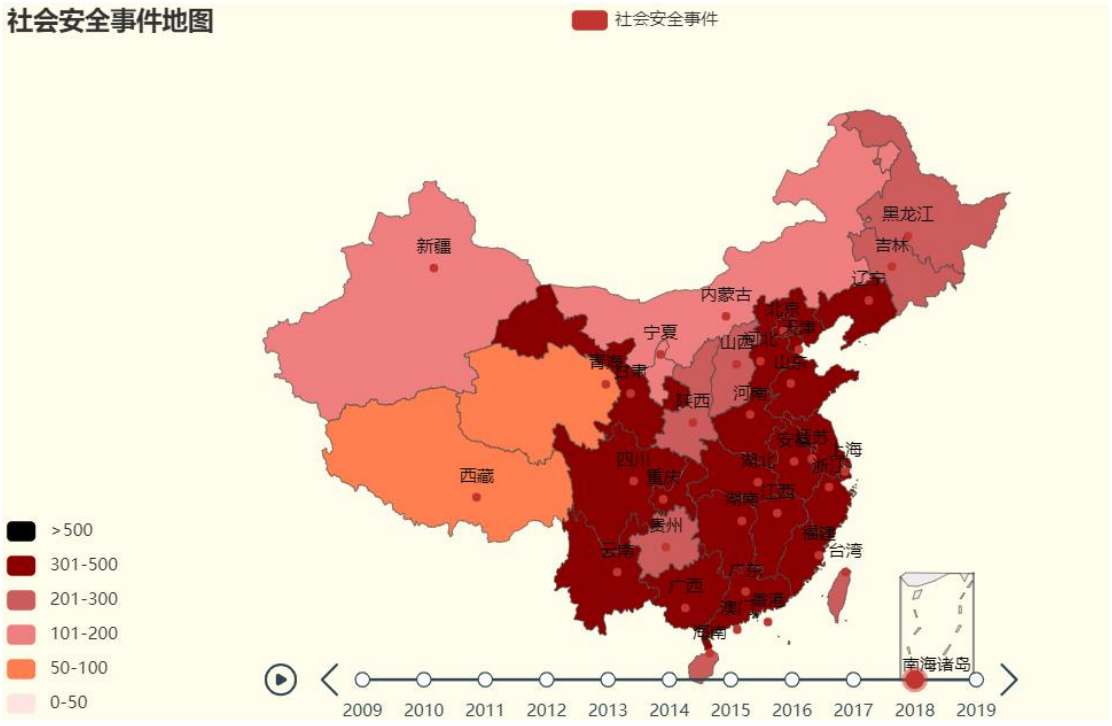

社会安全事件地图
社会安全事件
新疆
西藏
青海
甘肃
宁夏
内蒙古
黑龙江
吉林
辽宁
北京
天津
河北
山西
山东
河南
陕西
四川
重庆
湖北
安徽
江苏
上海
浙江
湖南
江西
贵州
云南
广西
广东
福建
台湾
香港
澳门
海南
南海诸岛
>500
301-500
201-300
101-200
50-100
0-50
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

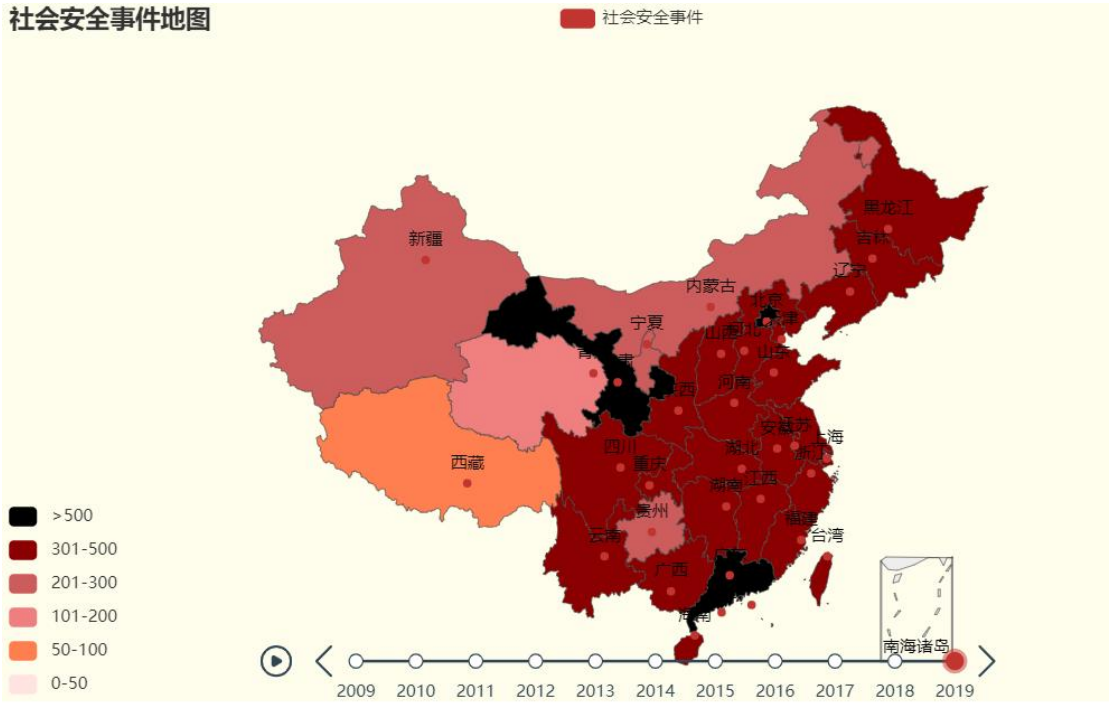

社会安全事件地图
社会安全事件
新疆
西藏
青海
甘肃
宁夏
内蒙古
黑龙江
吉林
辽宁
北京
天津
河北
山西
山东
河南
陕西
四川
重庆
湖北
安徽
江苏
上海
浙江
湖南
江西
贵州
云南
广西
广东
福建
台湾
香港
澳门
海南
南海诸岛
>500
301-500
201-300
101-200
50-100
0-50
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

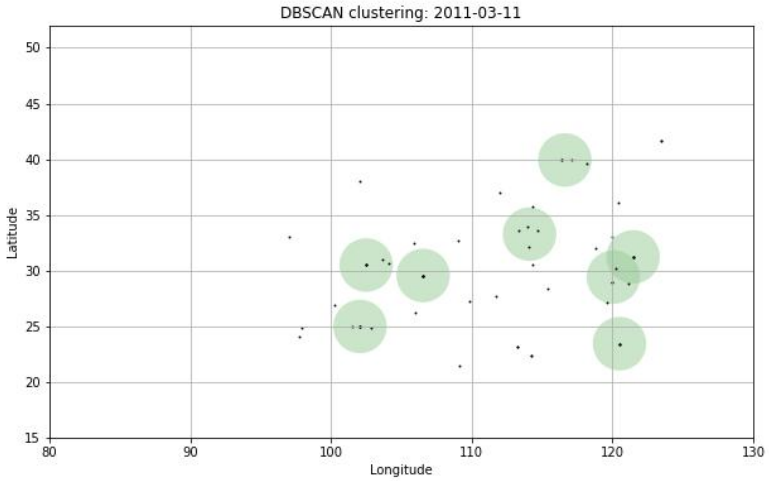

社会安全事件地图
社会安全事件
新疆
西藏
青海
甘肃
宁夏
内蒙古
黑龙江
吉林
辽宁
北京
天津
河北
山西
山东
河南
陕西
四川
重庆
湖北
安徽
江苏
上海
浙江
湖南
江西
贵州
云南
广西
广东
福建
台湾
香港
澳门
海南
南海诸岛
>500
301-500
201-300
101-200
50-100
0-50
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

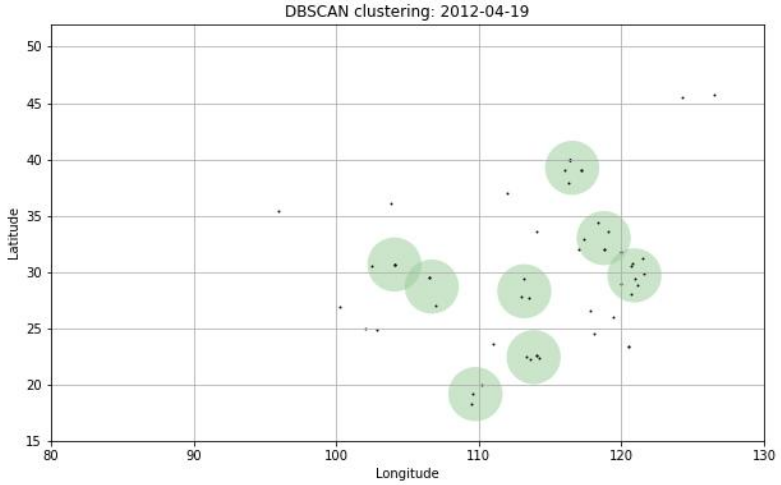

社会安全事件地图
社会安全事件
新疆
黑龙江
吉林
辽宁
内蒙古
北京
天津
宁夏
山西
河北
山东
青海
陕西
河南
安徽
上海
四川
湖北
西藏
重庆
湖南
江西
贵州
福建
台湾
云南
广西
广东
香港
海南
南海诸岛
>500
301-500
201-300
101-200
50-100
0-50
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

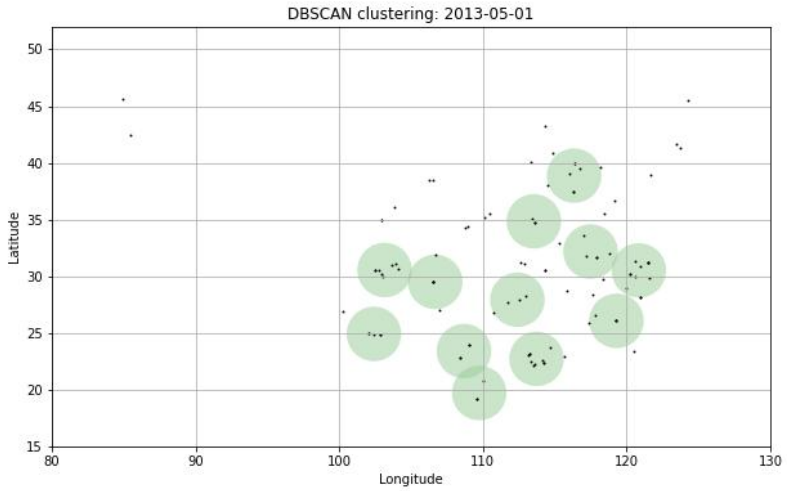

社会安全事件地图
社会安全事件
新疆
黑龙江
吉林
辽宁
内蒙古
北京
天津
宁夏
山西
河北
山东
青海
陕西
河南
安徽
上海
四川
湖北
西藏
重庆
湖南
江西
贵州
福建
台湾
云南
广西
广东
香港
海南
南海诸岛
>500
301-500
201-300
101-200
50-100
0-50
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

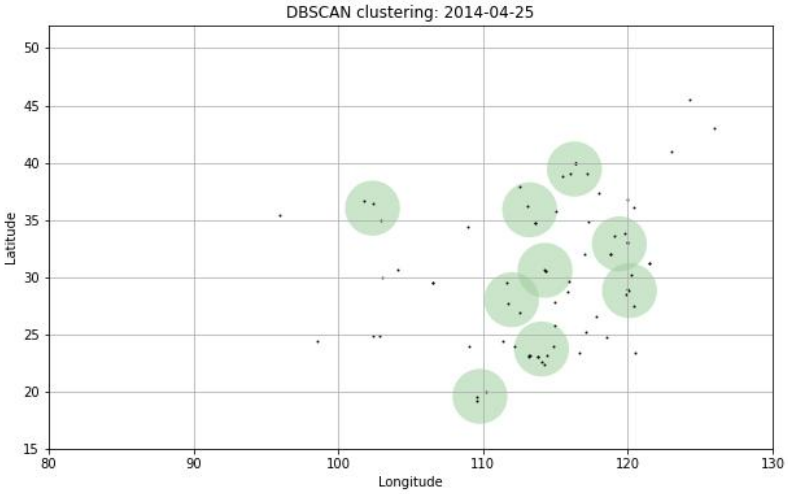

社会安全事件地图
社会安全事件
新疆
黑龙江
吉林
辽宁
内蒙古
北京
天津
宁夏
山西
青海
甘肃
山东
陕西
河南
安徽
江苏
上海
四川
湖北
浙江
西藏
重庆
湖南
江西
贵州
福建
台湾
云南
广东
广西
澳门
香港
南海诸岛
>500
301-500
201-300
101-200
50-100
0-50
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

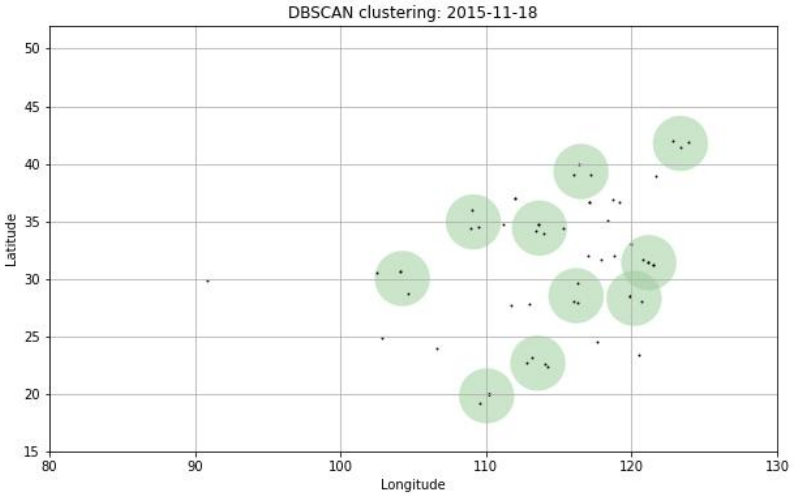

社会安全事件地图
社会安全事件
新疆
黑龙江
吉林
辽宁
内蒙古
北京
天津
宁夏
山西
青海
甘肃
山东
陕西
河南
安徽
江苏
上海
四川
湖北
浙江
西藏
重庆
湖南
江西
贵州
福建
台湾
云南
广东
广西
南海诸岛
>500
301-500
201-300
101-200
50-100
0-50
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

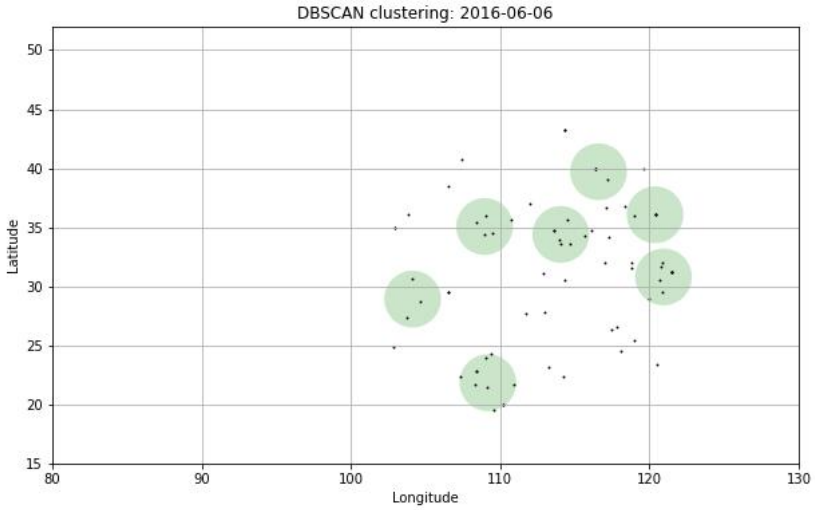

DBSCAN clustering: 2011-03-11
Latitude
Longitude
50
45
40
35
30
25
20
15
80
90
100
110
120
130

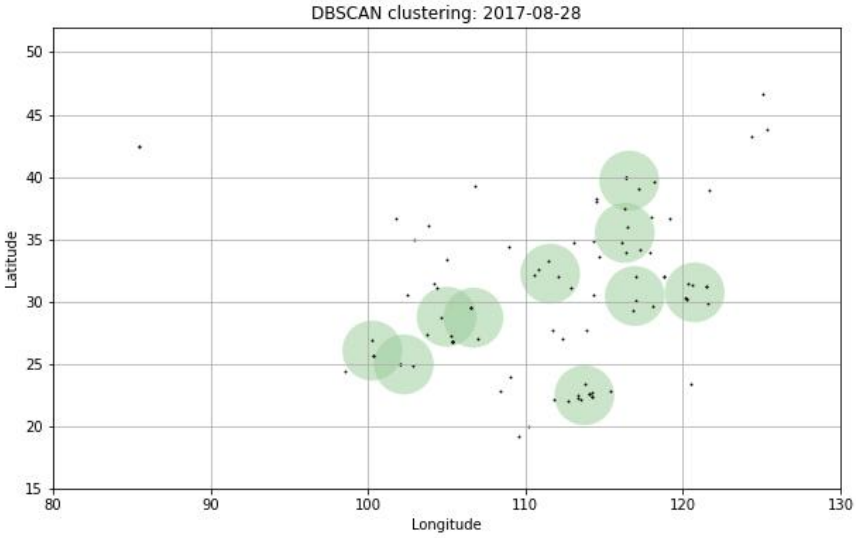

DBSCAN clustering: 2012-04-19
Latitude
Longitude
50
45
40
35
30
25
20
15
80
90
100
110
120
130

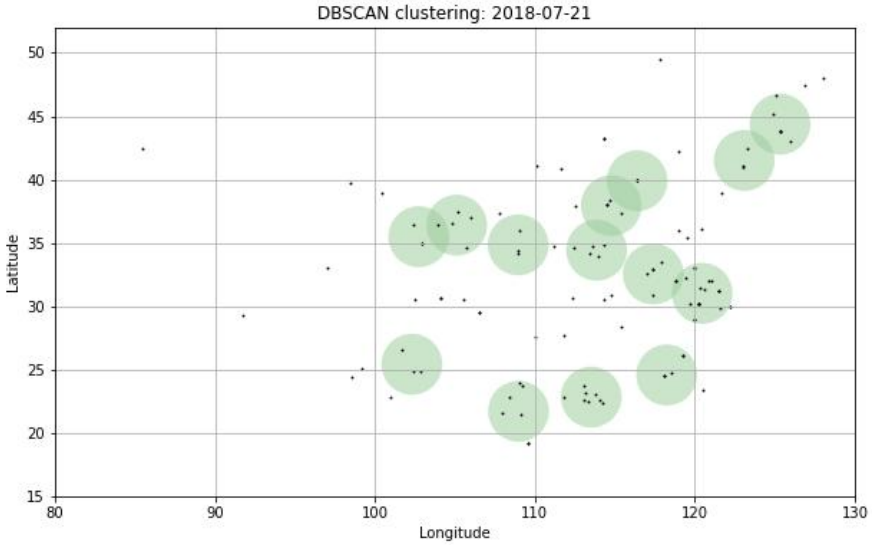

DBSCAN clustering: 2013-05-01
Latitude
Longitude
50
45
40
35
30
25
20
15
80
90
100
110
120
130

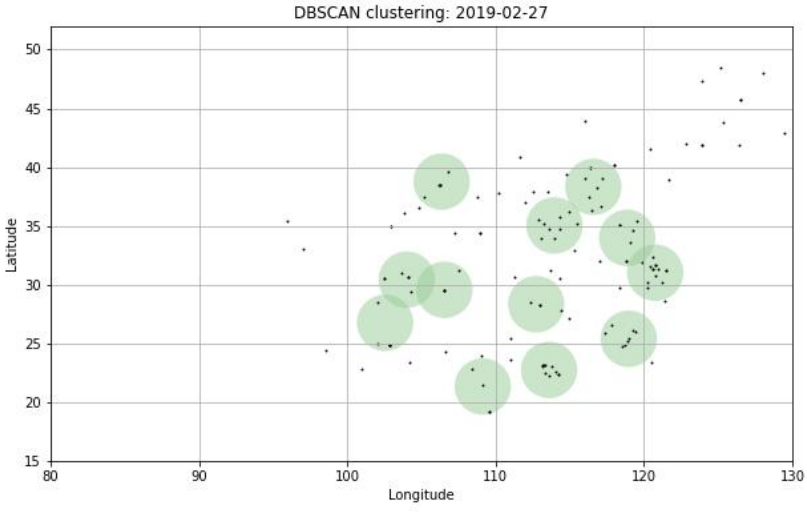

DBSCAN clustering: 2014-04-25
Latitude
Longitude
50
45
40
35
30
25
20
15
80
90
100
110
120
130


DBSCAN clustering: 2015-11-18
Latitude
Longitude
50
45
40
35
30
25
20
15
80
90
100
110
120
130


DBSCAN clustering: 2016-06-06
Latitude
Longitude
50
45
40
35
30
25
20
15
80
90
100
110
120
130


DBSCAN clustering: 2017-08-28
Latitude
Longitude
15
20
25
30
35
40
45
50
80
90
100
110
120
130


DBSCAN clustering: 2018-07-21
Latitude
Longitude
15
20
25
30
35
40
45
50
80
90
100
110
120
130


DBSCAN clustering: 2019-02-27
Latitude
Longitude
15
20
25
30
35
40
45
50
80
90
100
110
120
130