China Data Institute

Center for Geographic Analysis
Harvard University

社会地理计算联合研究中心
Geocomputation Center for Social Science

武漢大學 WUHAN UNIVERSITY

遥感信息工程学院 School of Remote Sensing and Information Engineering

新冠病毒数据分析系列公益讲座COVID-19 Data Analysis Webinars

第八讲 疫情多尺度时空扩散模式及多元流网络的相关性分析

主讲人：秦昆，武汉大学

点评人：柳林，辛辛那提大学

主持人：鲍曙明，中国数据研究所

合办单位：上海财经大学长三角与长江经济带研究院

北京时间 2020年6月19日上午9-10点

目 录

1 研究目标与背景

2 新冠疫情的多尺度时空扩散分析

3 多元流网络与新冠疫情的相关性分析

4 总结与讨论

研究目标与背景

1. 研究目标与背景

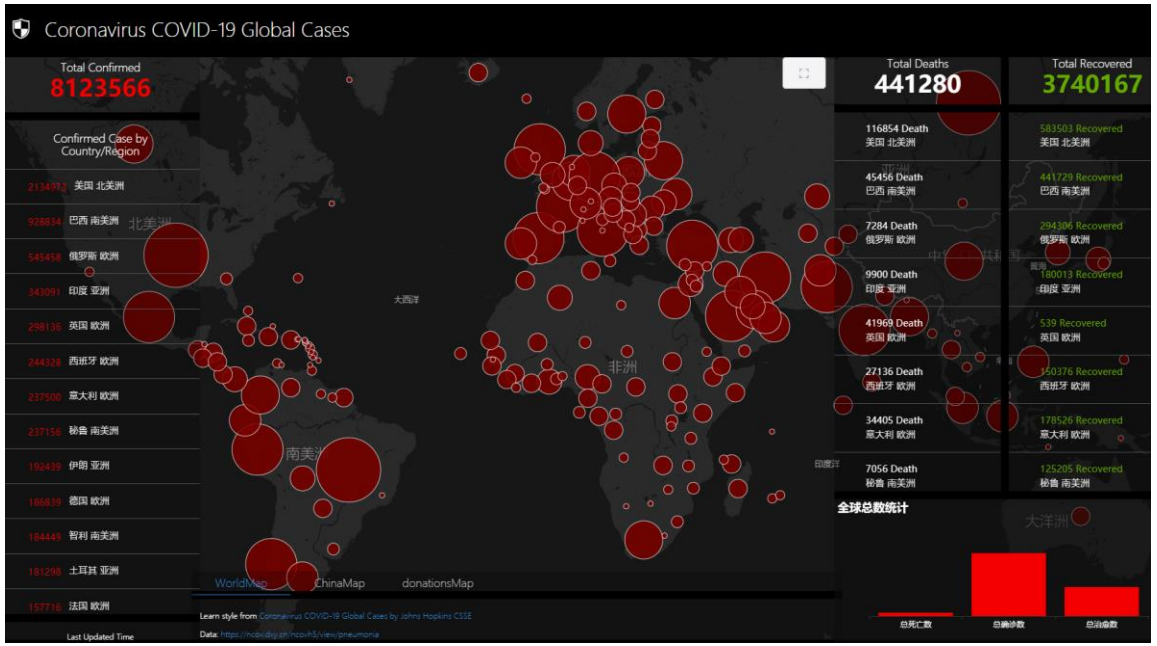
自2020年1月至今，“新冠”疫情已呈全球蔓延趋势，为人类带来了巨大的灾难。

如何充分利用时空大数据分析技术为疫情防控贡献一点力量？是地信人则无旁贷的责任。

1)新冠疫情扩散在不同时空尺度有什么模式和规律？

2)疫情扩散与人口流动情况有什么关联？

3)多元流网络演化与疫情扩散有什么相关性？



数据截止时间：2020年6月17日

1. 研究目标与背景

1)“流”数据（Stream Data）

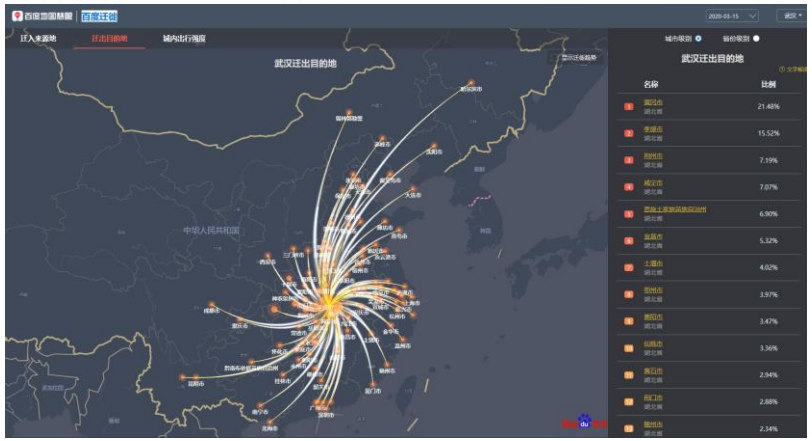
“流数据”：系统的部分部分之间、要素之间、以及系统与外部环境之间的物质、信息、能量等的移动和交换，而形成的各种流数据。

地理流：流的两端连接的是不同的空间位置。

特点：

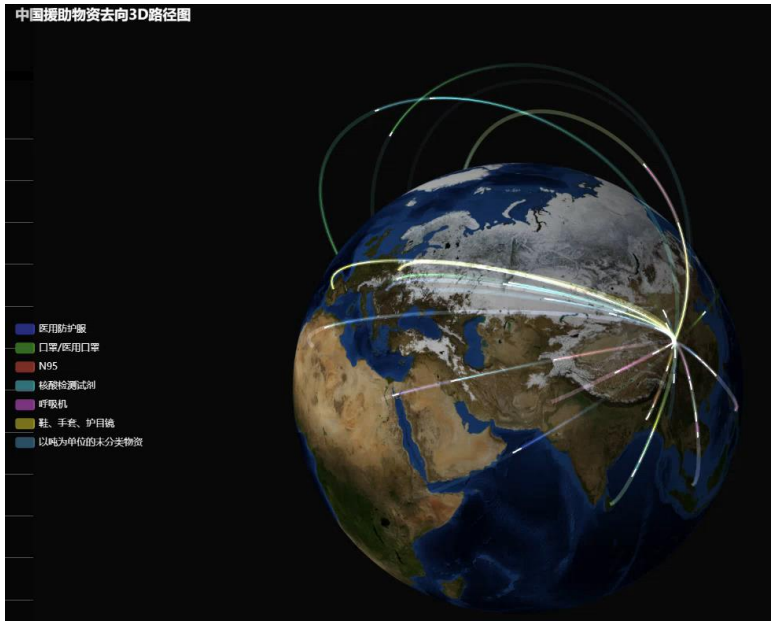
(1)有特定的起点、终点；

(2)有物质、信息、能量等的移动和交换。



百度迁徙流

航空流



中国援助物资流

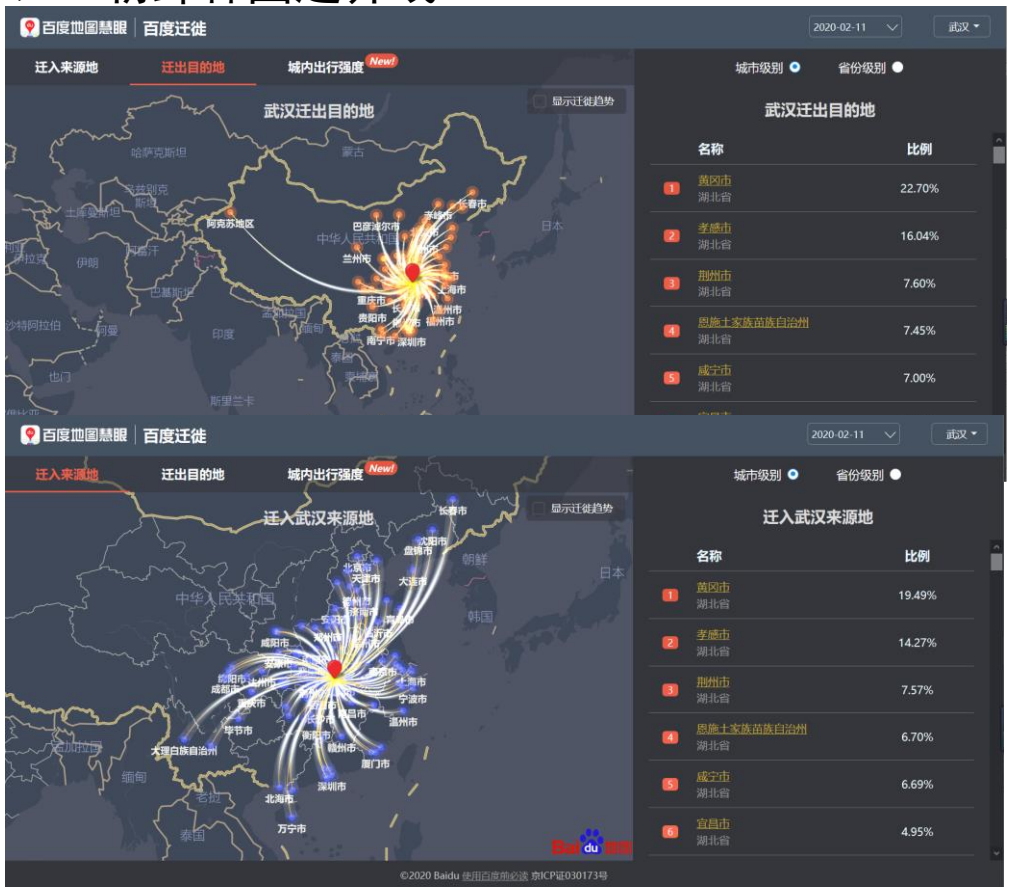
1. 研究目标与背景

2)时空邻近度：

时空邻近度：地理空间任意两匀质区域(含点)之间的时空邻近度，对给定的“流”，正比于二者之间的总流量，反比于从一端到达另一端的平均时间。

“邻国相望，鸡犬之声相闻，民至死不相往来”、“柏林墙”、“朝鲜韩国边界线”

李小文,曹春香,常超一. 地理学第一定律与时空邻近度的提出 [J]. 自然杂志, 2007,29(2): 69-71



1. 研究目标与背景

3)相关性与空间相关性

世界上万事万物都是相互关联的

Everything in the world are interrelated

(1)相关分析

相关分析： 两个或者两个以上变量之间在某种意义下所存在的规律，一个变量的数据发生变化，另外一个变量随之变化。

因果关系：在发现两个相关事物的基础上进一步探究哪一事物为因、哪一事物为果？

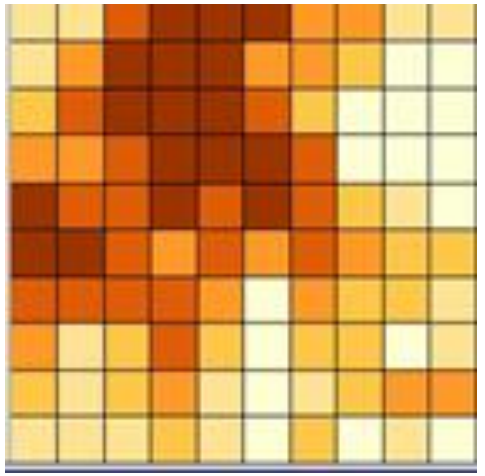
大数据时代，相关关系的探索更具重要价值。

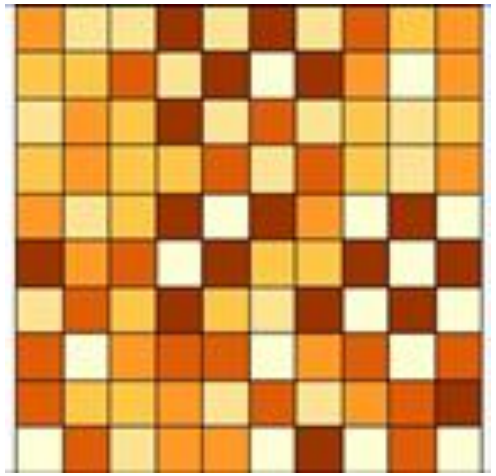
$$r_{xy}=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}$$

(2)空间相关性

地理学第一定律：任何事物都是空间相关的，距离近的事物比距离远的事物的空间相关性更大。

Tobler's First Law (TFL): Everything is related to every thing else, but near things are more related than distant things.



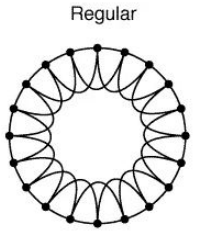


1. 研究目标与背景

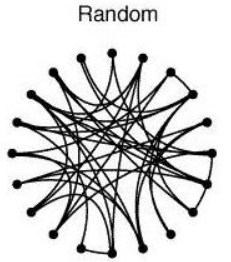
4)网络化挖掘

21世纪是复杂性和网络化的世纪

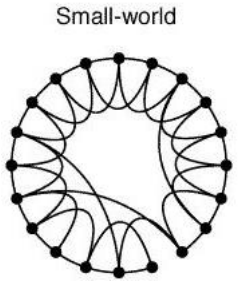
1）规则网络理论阶段：规则网络理论的发展得益于图论和拓扑学等应用数学的发展。



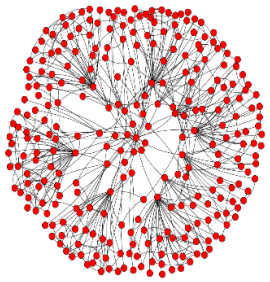
2）随机网络理论阶段：随机网络是指由N个节点构成的图中以概率P随机连接任意两个节点而形成的网络。

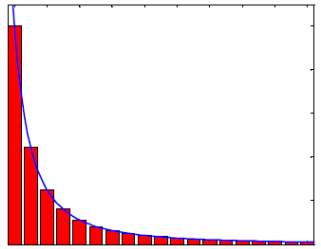


3) 复杂网络理论阶段：复杂网络：具有不同于规则网络和随机网络的特性的复杂网络，如小世界特性、无标度特性、层次特性、自相似特性、自组织特性等。



小世界理论(六度分割理论)：你和任何一个陌生人之间所间隔的人不会超过6个。



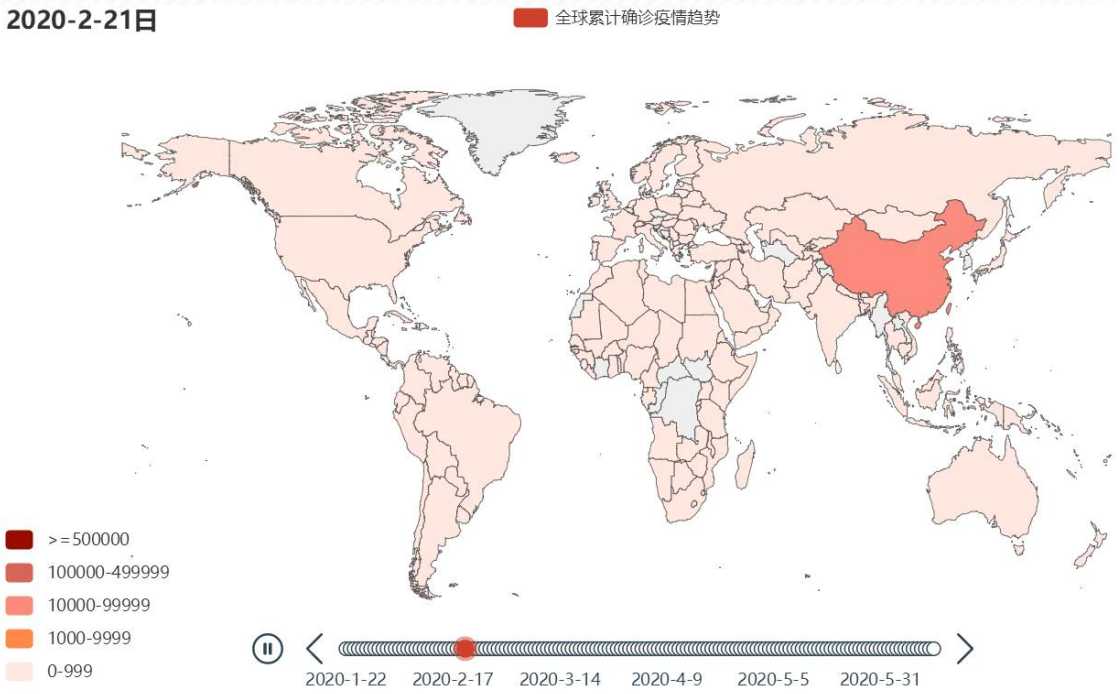


无标度特性（scale-free model）：幂律分布、二八定律。

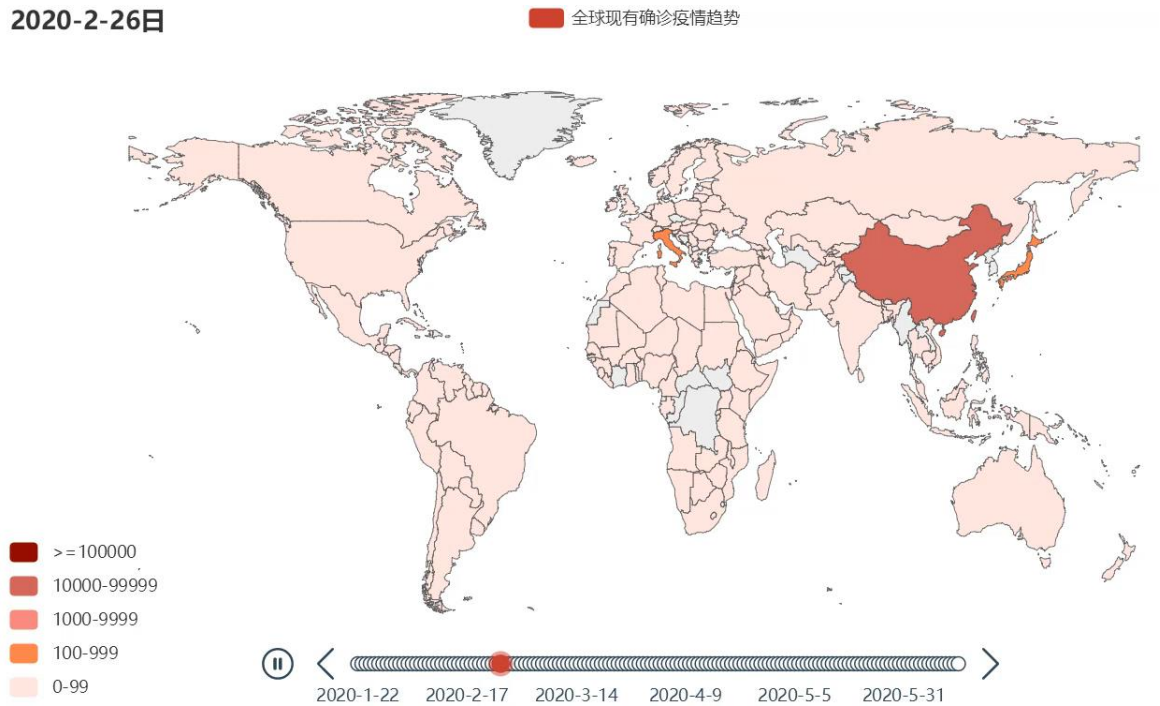
新冠疫情的多尺度时空扩散分析

2. 新冠疫情的多尺度时空扩散分析

2.1 全球尺度疫情时空扩展分析

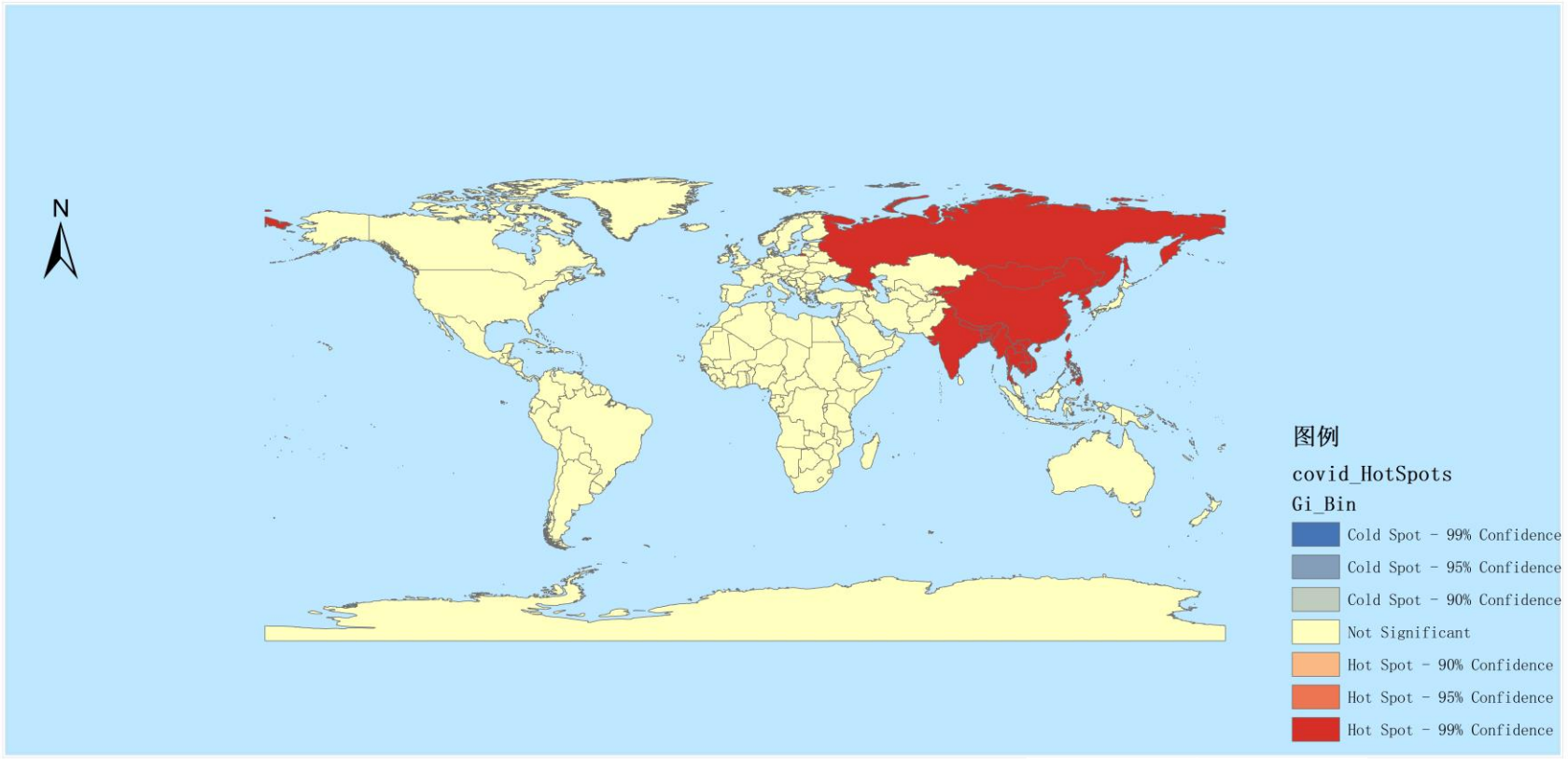


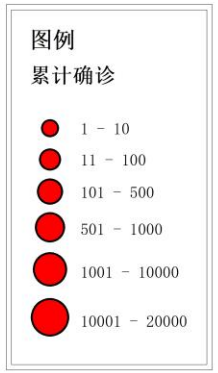
全球累计确诊疫情时空扩展

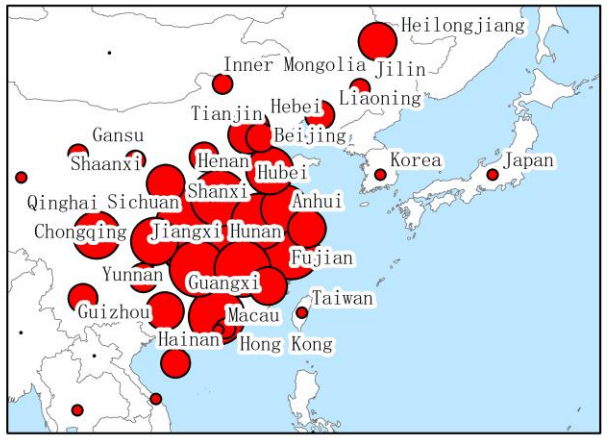


全球现有确诊疫情时空扩展

全球累计确诊冷热点分布 2.12



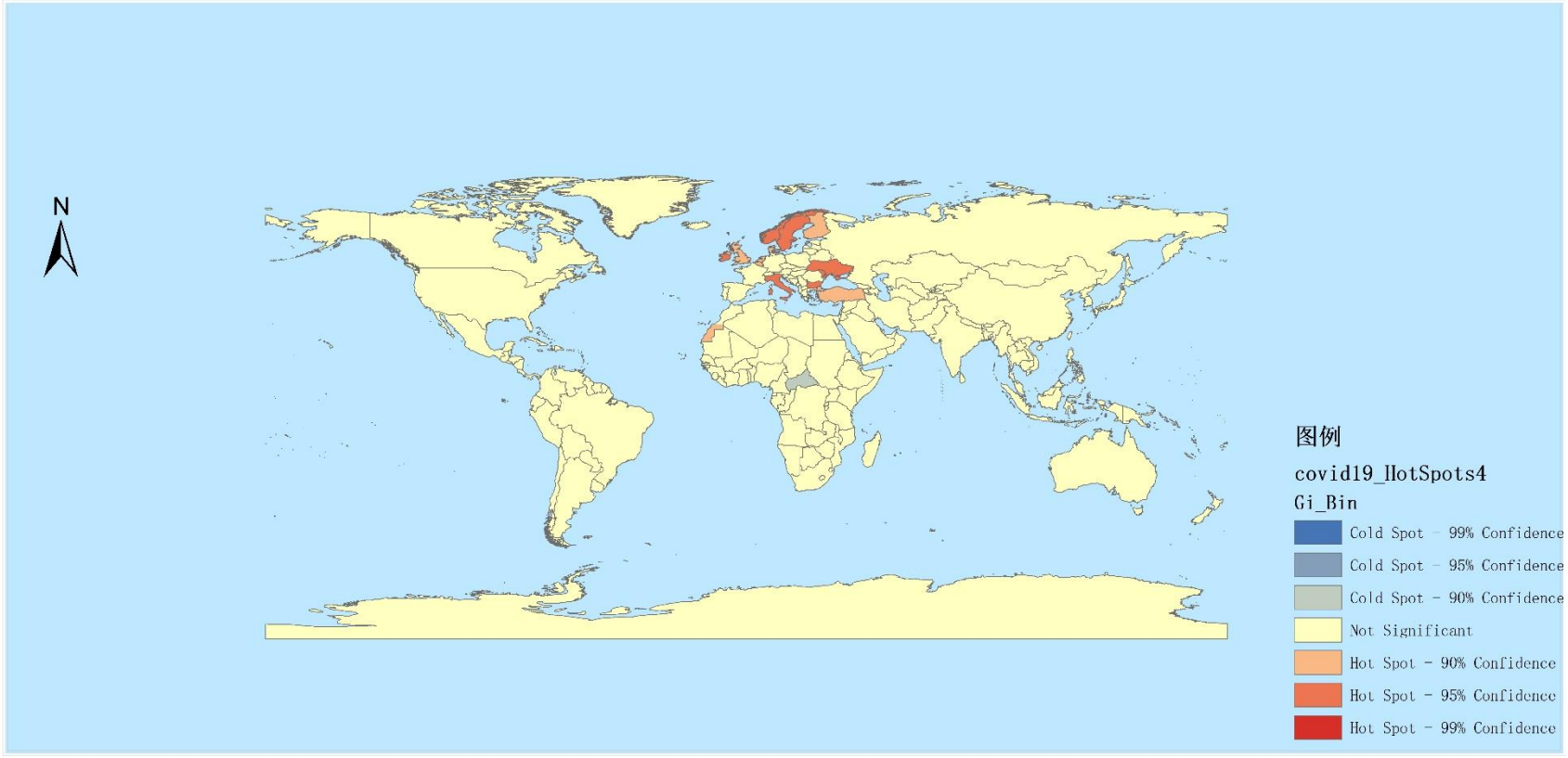


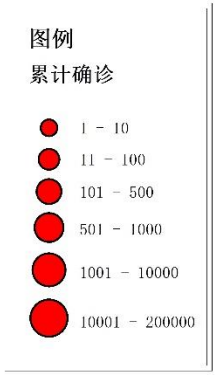


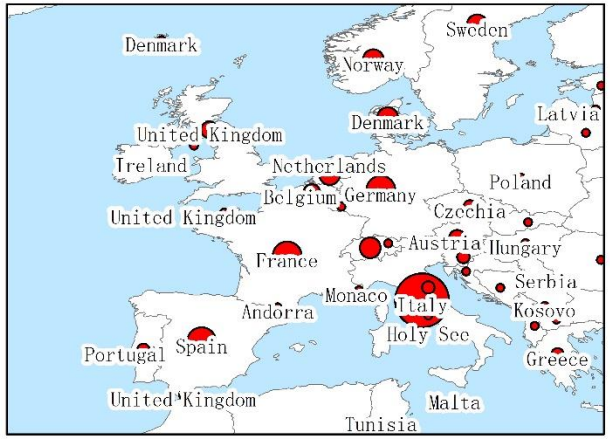
国家	累计确诊
China	44759
Singapore	50
Thailand	33
Japan	28
Korea, South	28

- 二月中旬疫情热点集中在亚洲

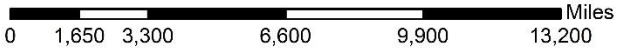
全球累计确诊冷热点分布 3.12





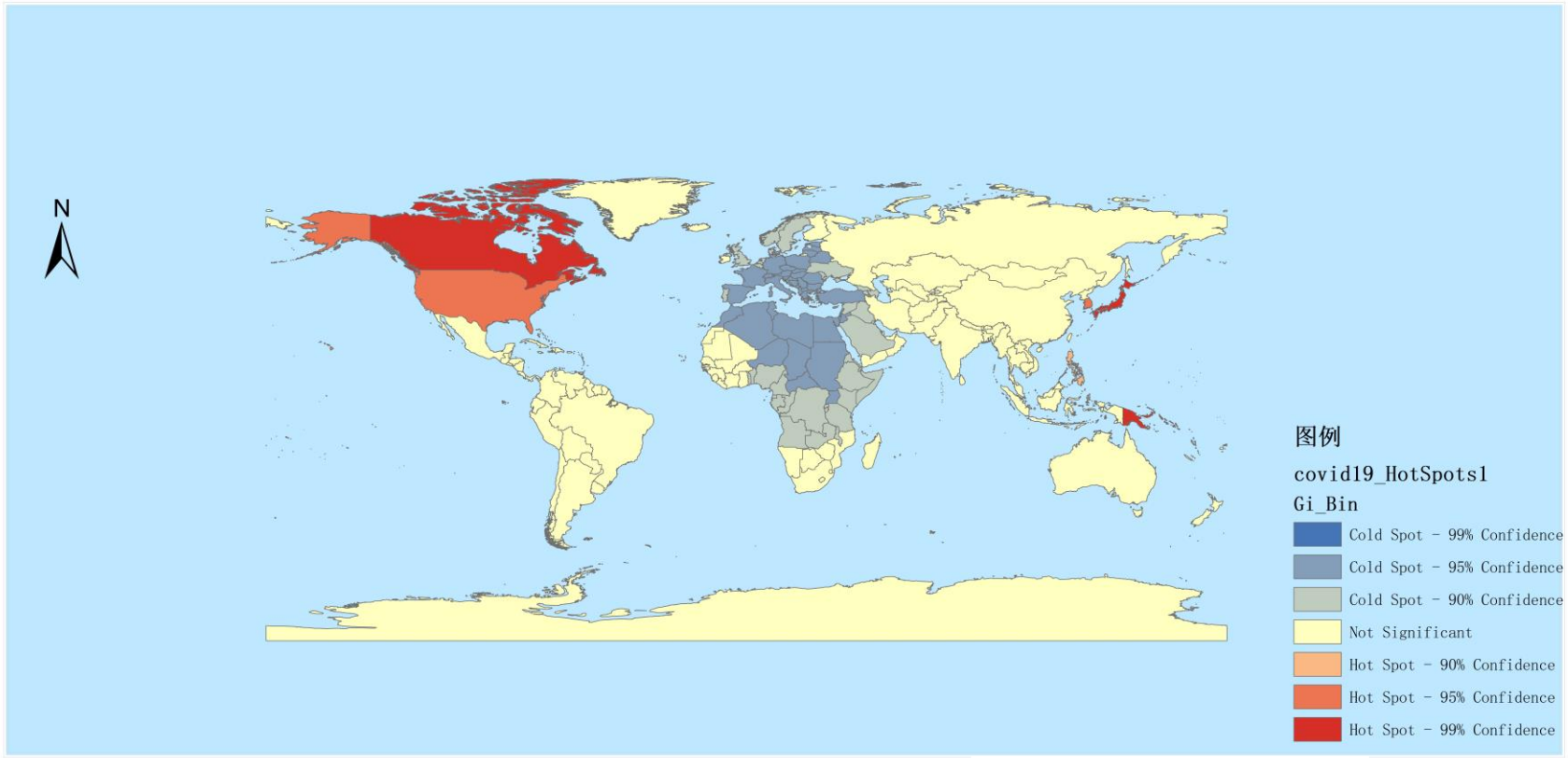


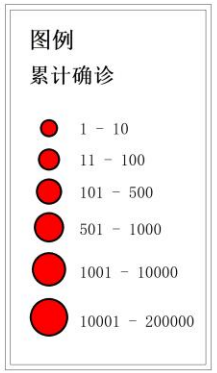
国家	累计确诊
Italy	15113
Iran	10075
France	2293
Spain	2277
Germany	2078

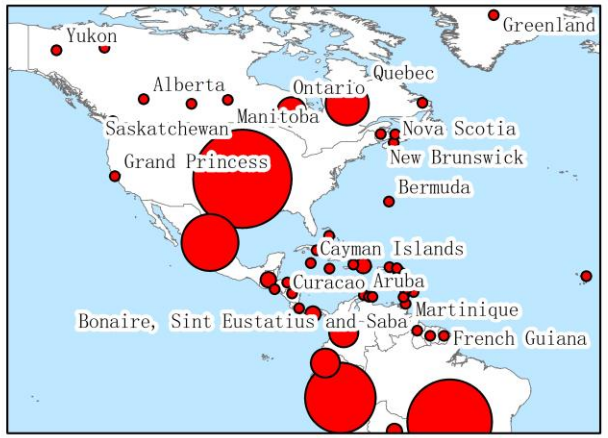


- 三月中旬疫情热点转移到欧洲

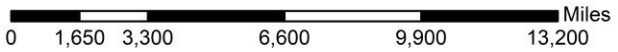
全球累计确诊冷热点分布 6.10







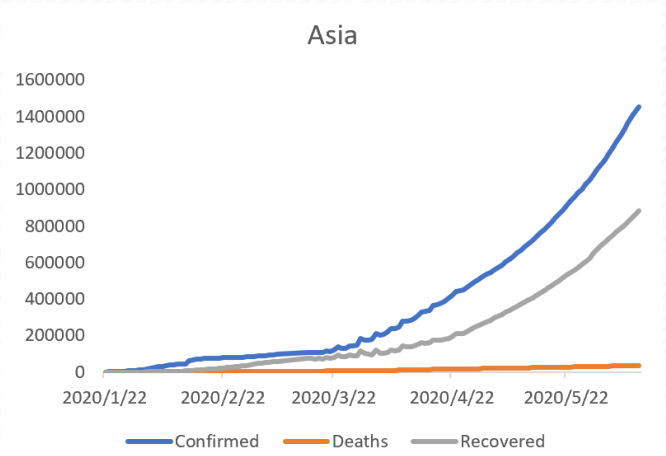
国家	累计确诊
United States	2000702
Brazil	772416
Russia	493023
United Kingdom	291588
India	286605

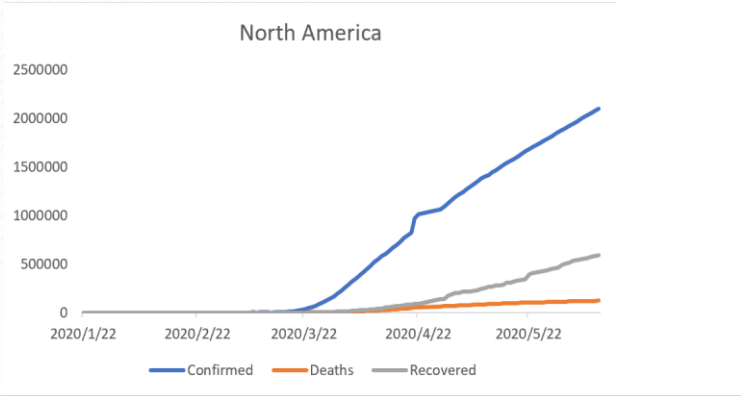


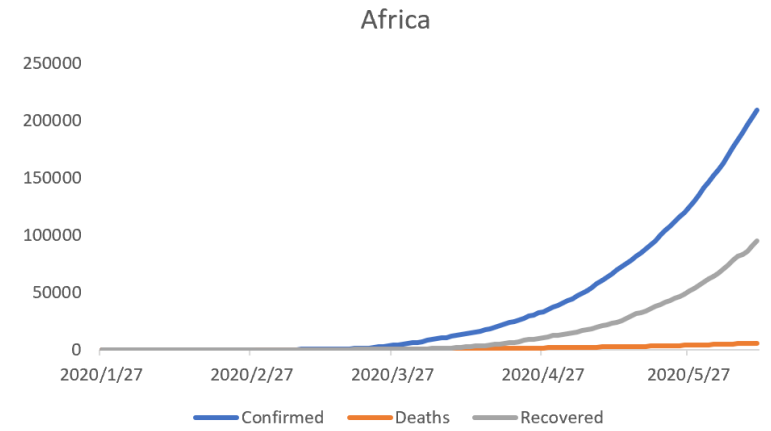
- 六月中旬疫情热点转移到美洲

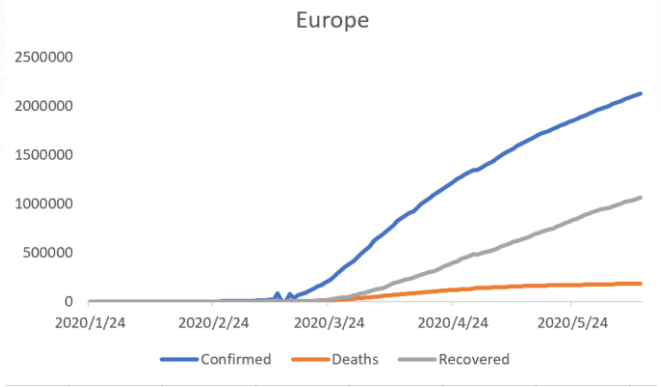
2. 新冠疫情的多尺度时空扩散分析

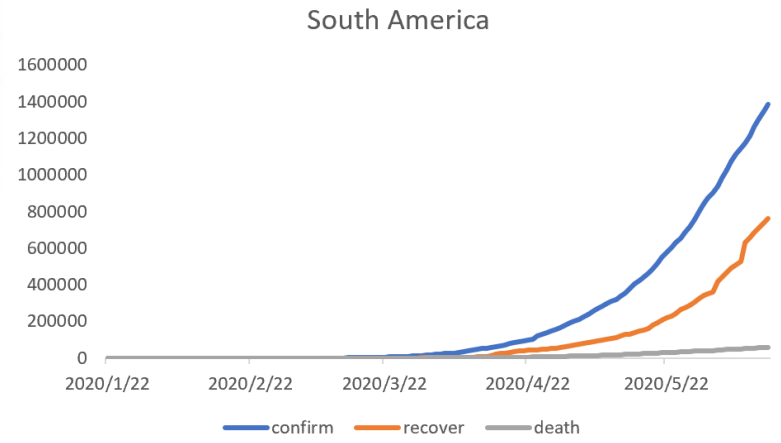
2.2 洲级尺度疫情时空扩展分析(6大洲的累计确诊、死亡、治愈数时序分析)

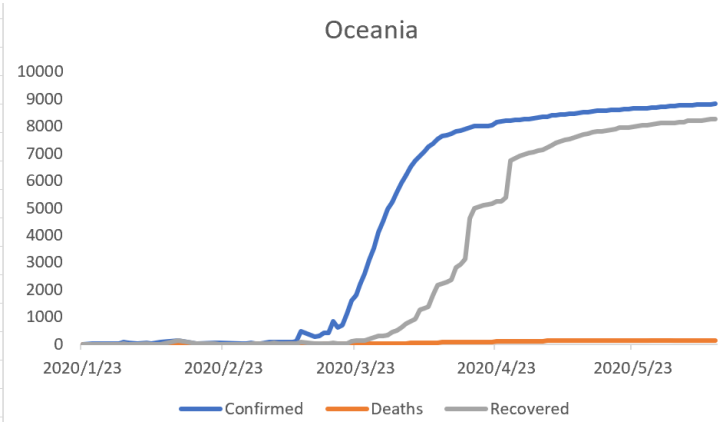










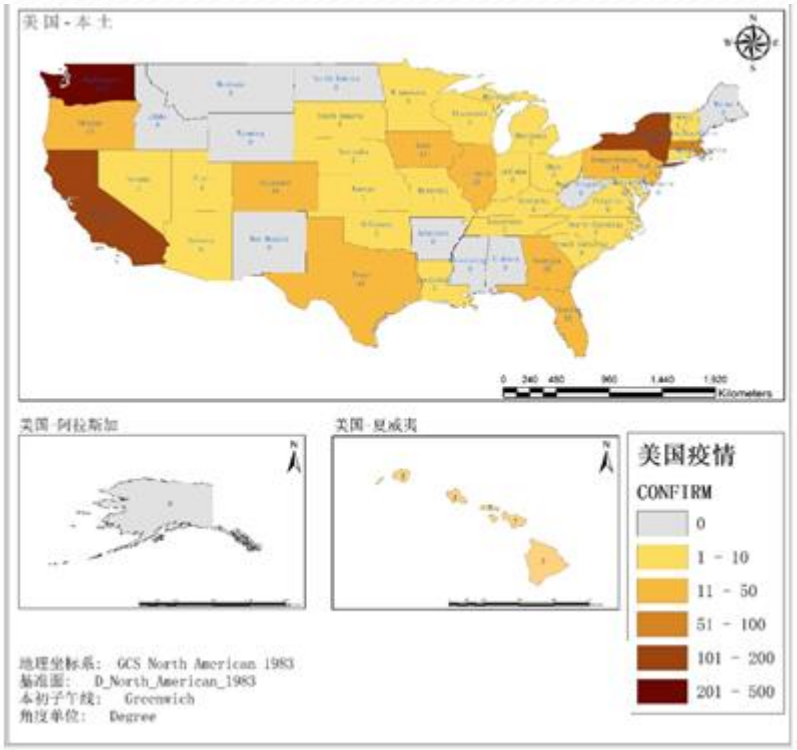


亚洲从二月份开始疫情爆发，现在仍处于上升趋势；
欧洲从三月份开始疫情爆发，目前增速放缓。
北美洲从三月中下旬疫情逐渐变得严重，南美洲从四月份开始呈现爆发式增长。
大洋洲疫情总体较轻。

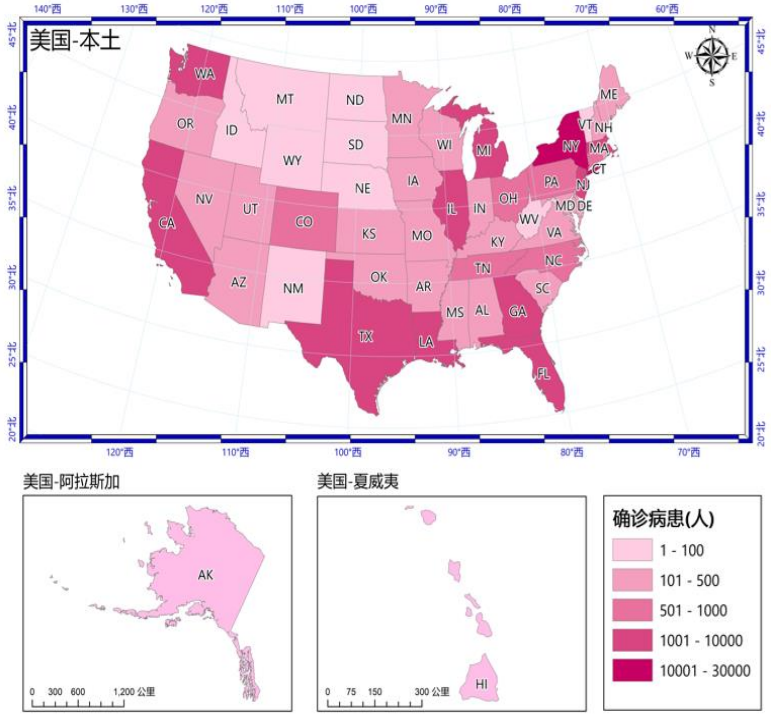
2. 新冠疫情的多尺度时空扩散分析

2.3 国家尺度的疫情时空扩散分析

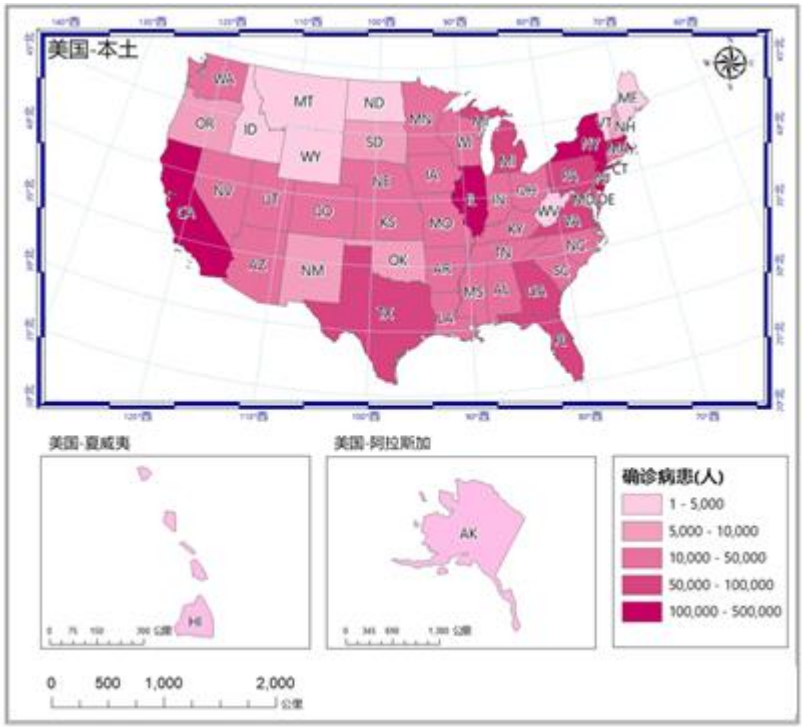
1）美国新冠疫情的时空扩展



2020年3月11日



2020年3月15日

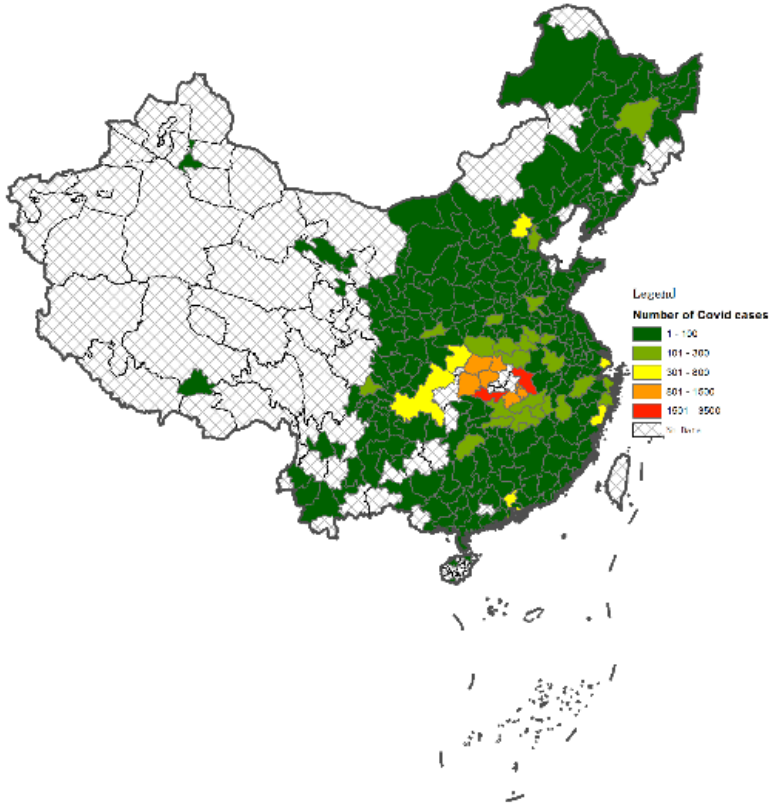


2020年6月14日

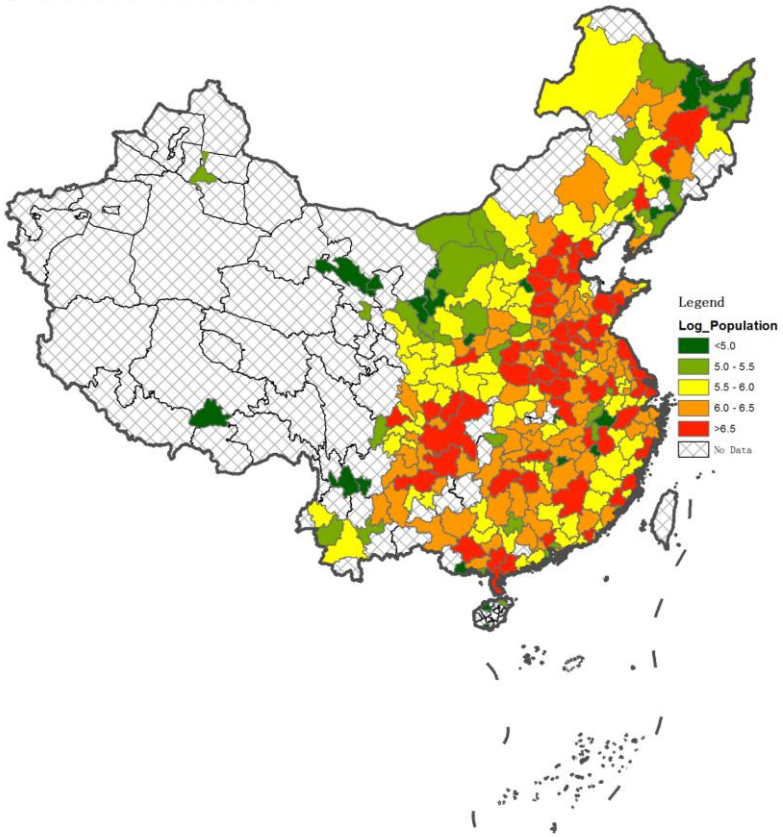
2. 新冠疫情的多尺度时空扩散分析

2.3 国家尺度的疫情时空扩散分析

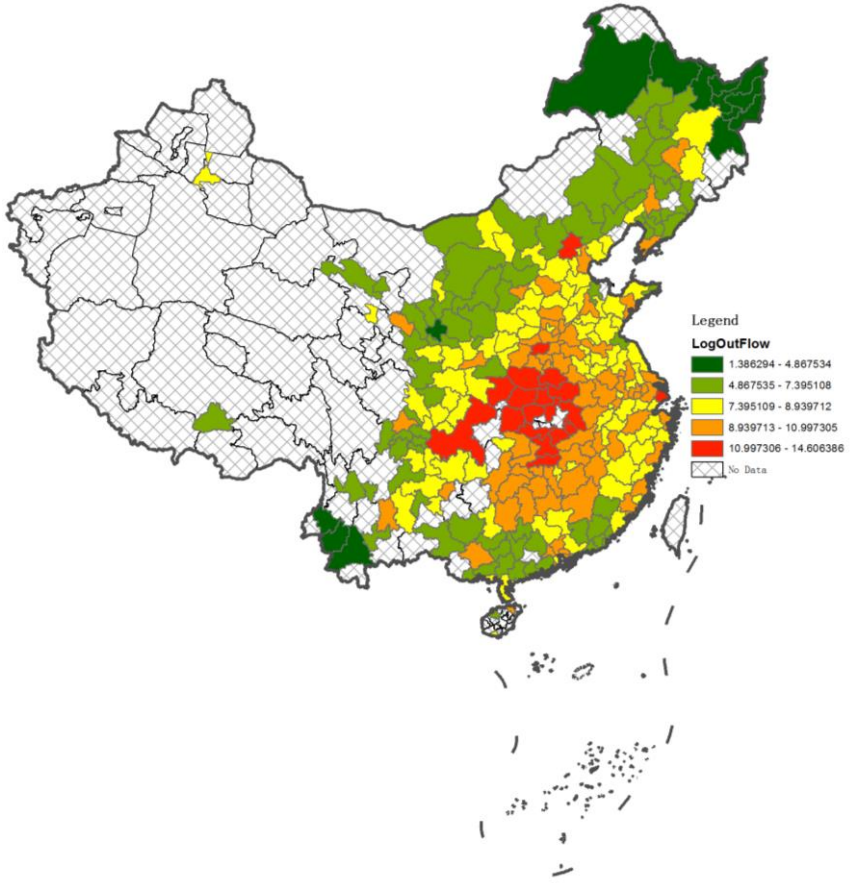
2）中国的新冠疫情及与人口流动的相关性分析



COVID-19人数分布(截止到4月8号)



城市（户籍）人口分布



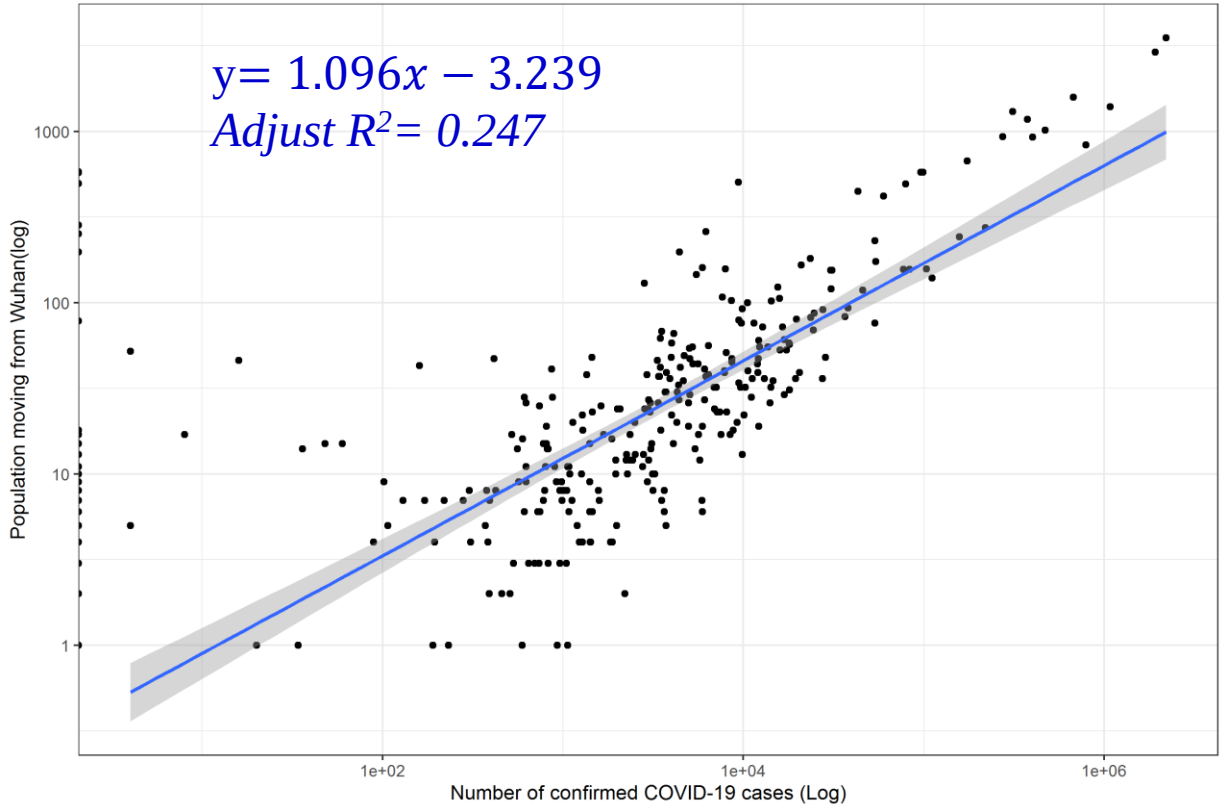
2020年1月1日-1月23日武汉市流出人口

数据来源：Jia J S et al. Population flow drives spatio-temporal distribution of COVID-19 in China [J]. Nature, 2020.

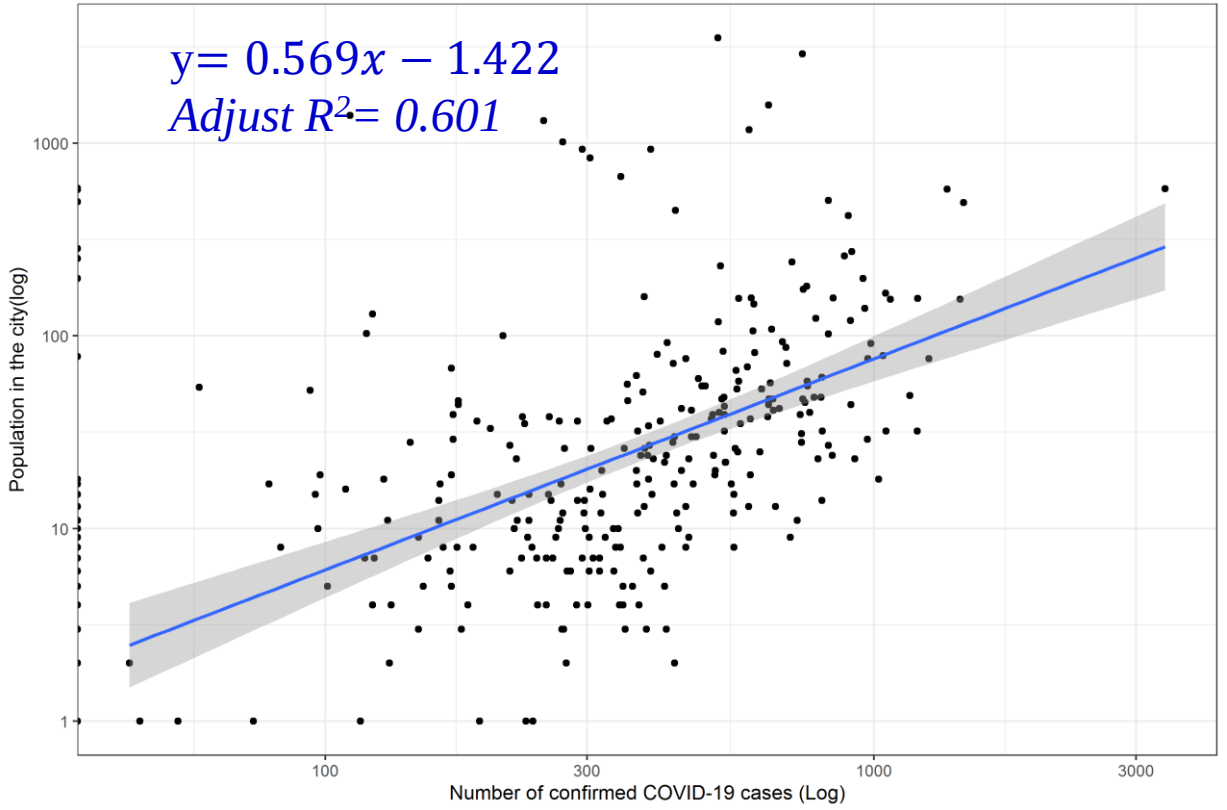
2. 新冠疫情的多尺度时空扩散模式

2.3 国家尺度的疫情时空扩散分析

2）中国的新冠疫情及与人口流动的相关性分析



COVID-19人数分布 *VS.* 人口分布(log)



COVID-19人数分布 *VS.* 武汉市流出人口分布(log)

2. 新冠疫情的多尺度时空扩散模式

地理加权回归分析技术

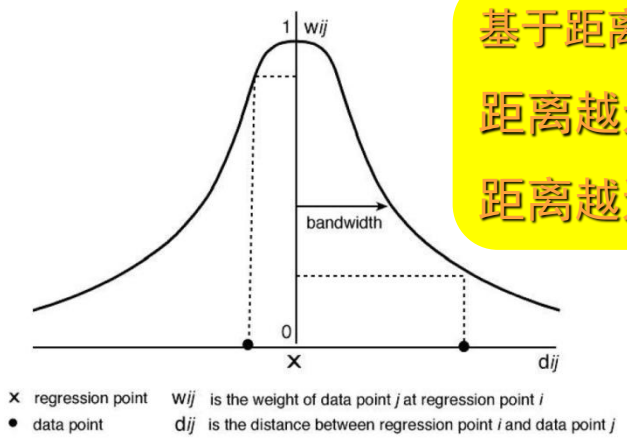
- 基本地理加权回归分析(GWR)模型

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1,n} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

- 解算算子（加权线性最小二乘）

$$\hat{\beta}_i = (\mathbf{X}^{\mathbf{t}}\mathbf{W}_{\mathbf{i}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{W}_i y$$

$$\mathbf{W_i} = \begin{matrix} w_{i1} & 0 & 0 & \dots & 0 \\ 0 & w_{i2} & 0 & \dots & 0 \\ 0 & & w_{i3} & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & w_{in} \end{matrix}$$



基于距离的权重计算：

距离越近，权重越高

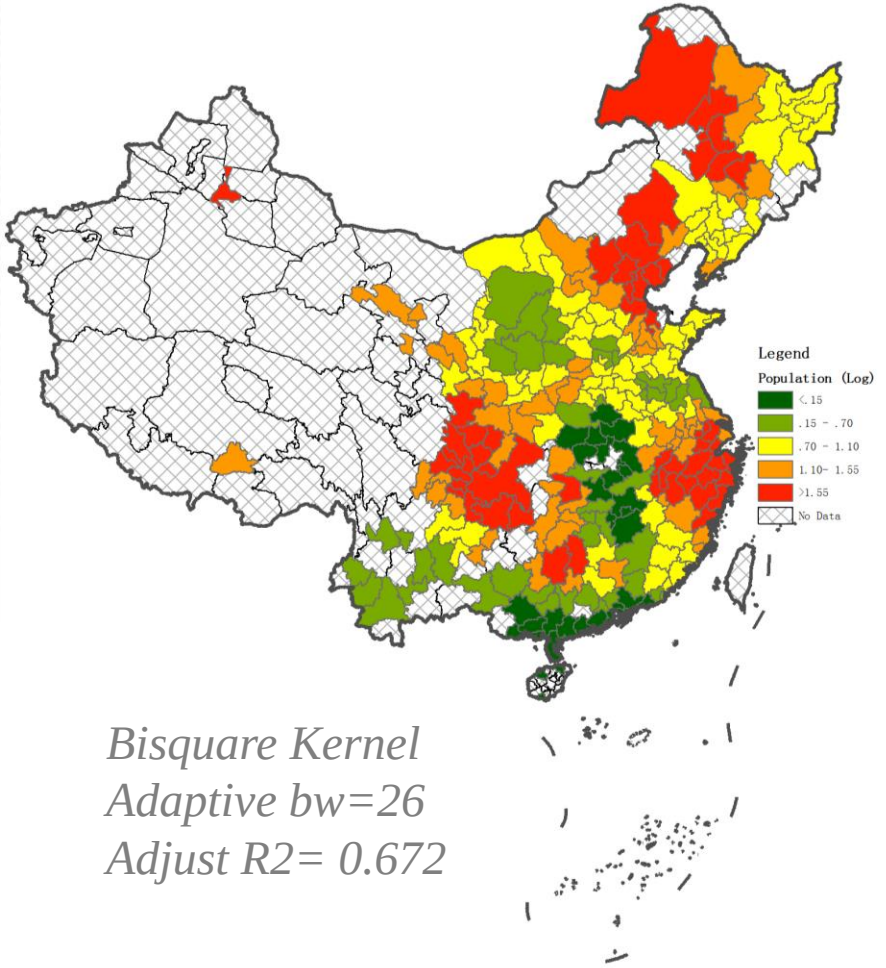
距离越远，权重越低

2. 新冠疫情的多尺度时空扩散模式

2.3 国家尺度的疫情时空扩散分析

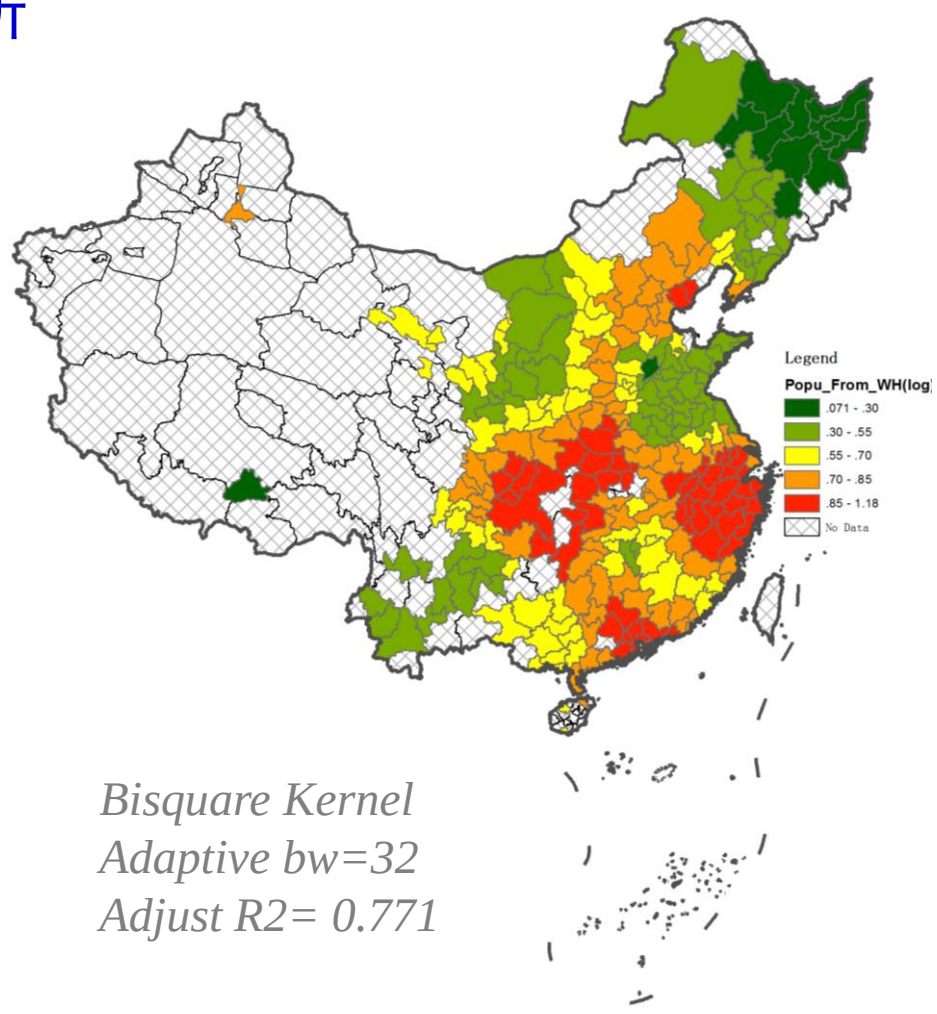
2）中国的新冠疫情及与人口流动的相关性分析

地理加权回归分析



Bisquare Kernel
Adaptive bw=26
Adjust R2= 0.672

COVID-19人数分布 *VS.* 人口分布(log)



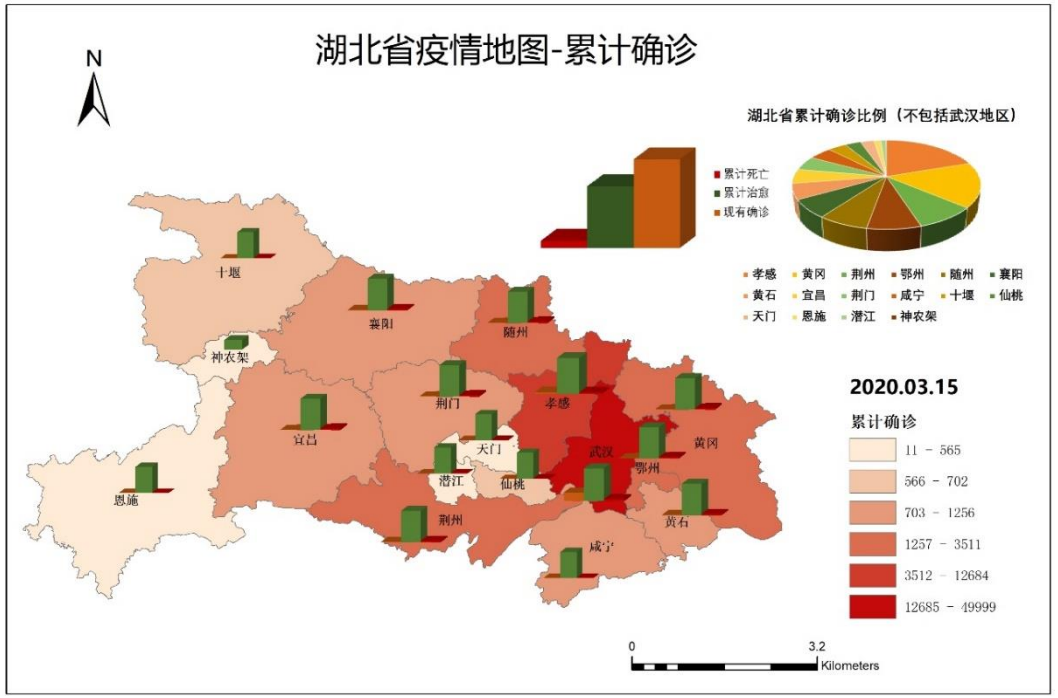
Bisquare Kernel
Adaptive bw=32
Adjust R2= 0.771

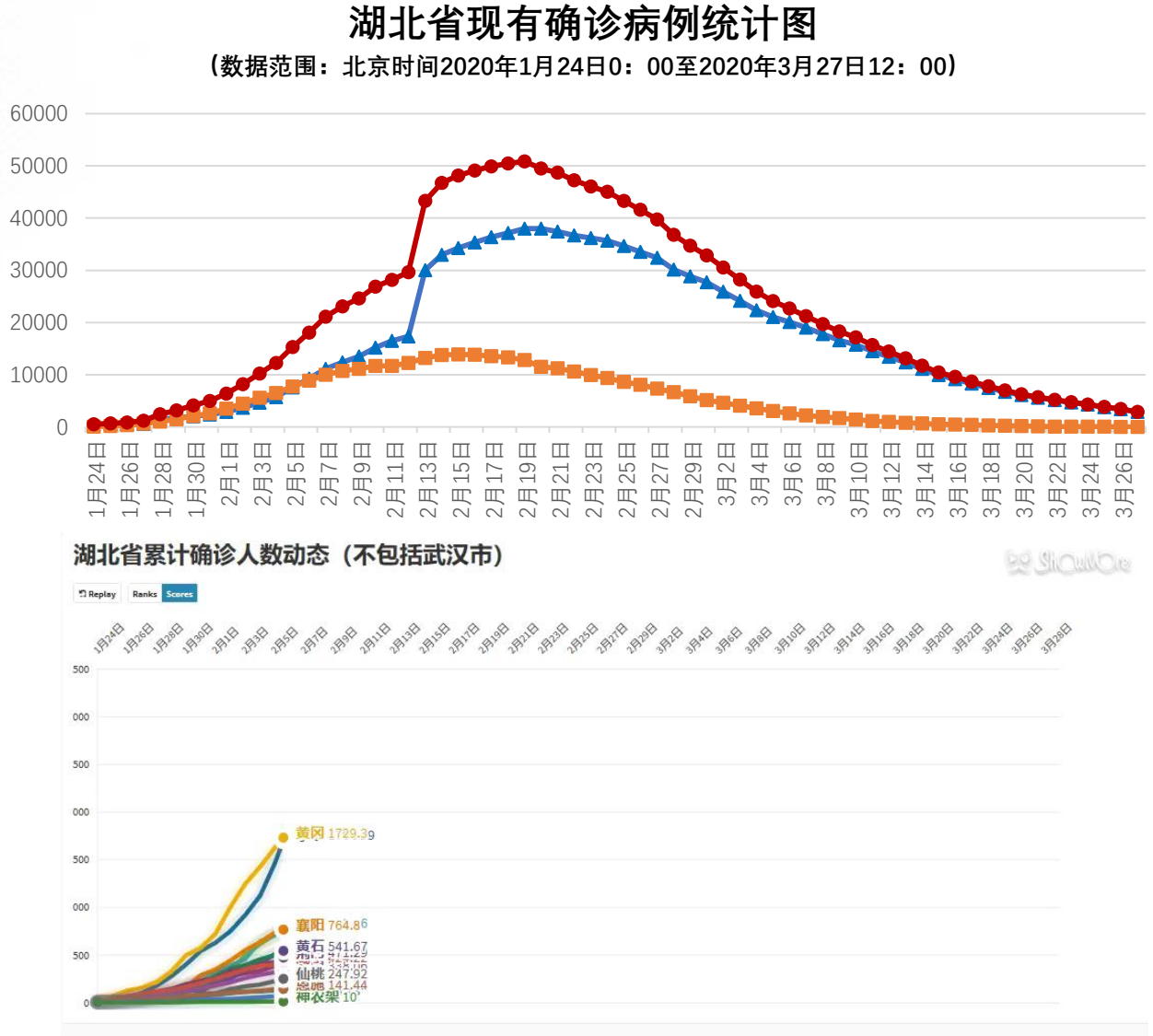
COVID-19人数分布 *VS.* 武汉市流出人口分布(log)

2. 新冠疫情的多尺度时空扩散模式

2.4 省级尺度的新冠疫情时空扩散模式分析

■ 湖北省新冠疫情的时空扩展模式

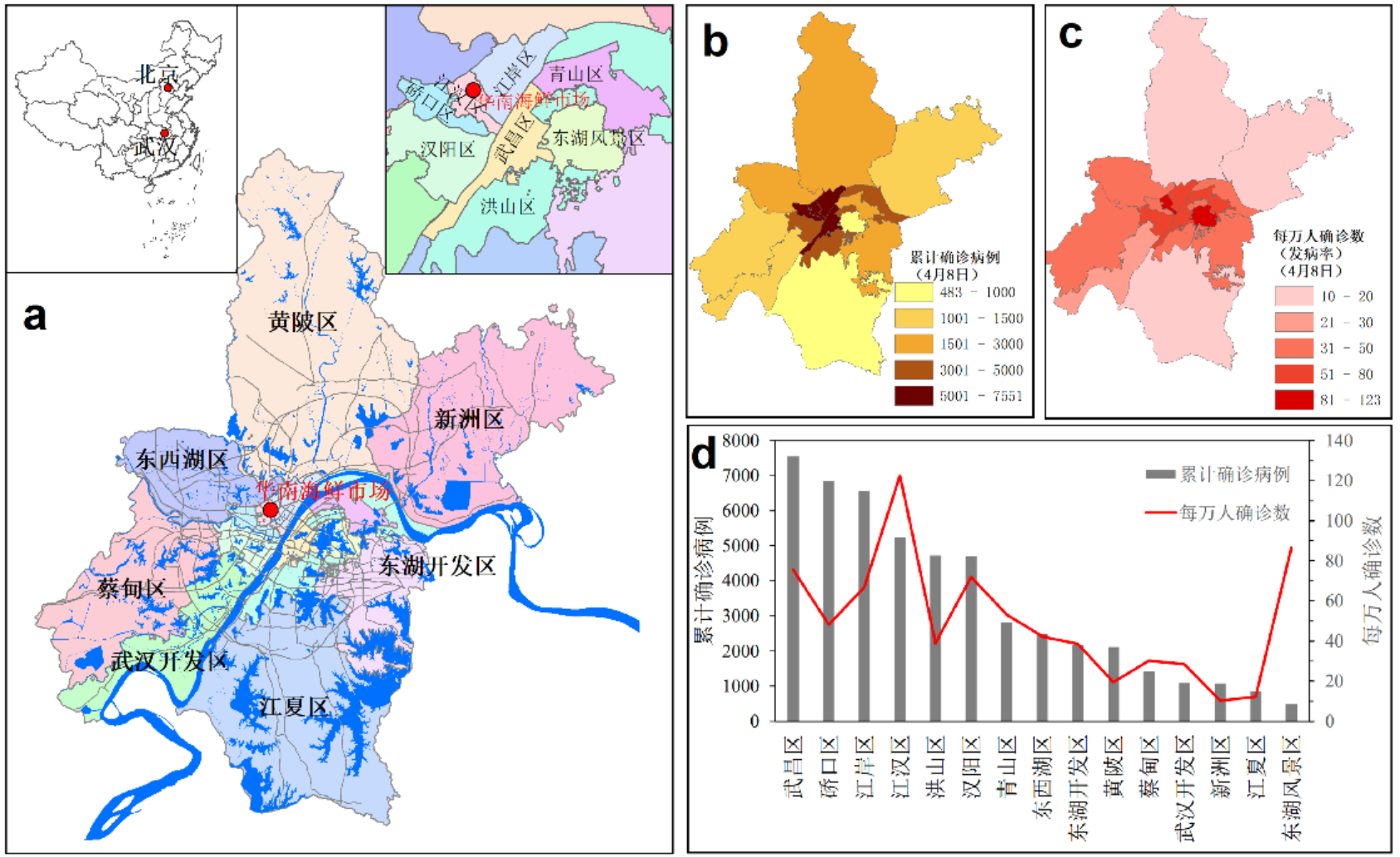




2020年1月底至二月中旬：湖北省的疫情呈爆发式增长，孝感、黄冈两市累计确诊人数仅次于武汉，

3月3日以后：湖北省除武汉外其他地区基本实现零新增。

2.4 武汉市新冠疫情时空扩散模式

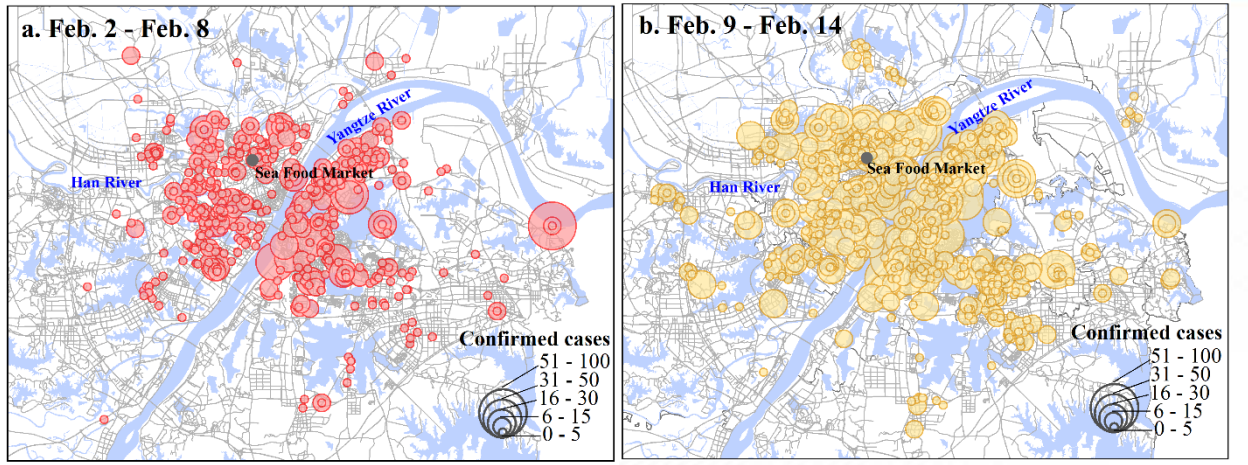


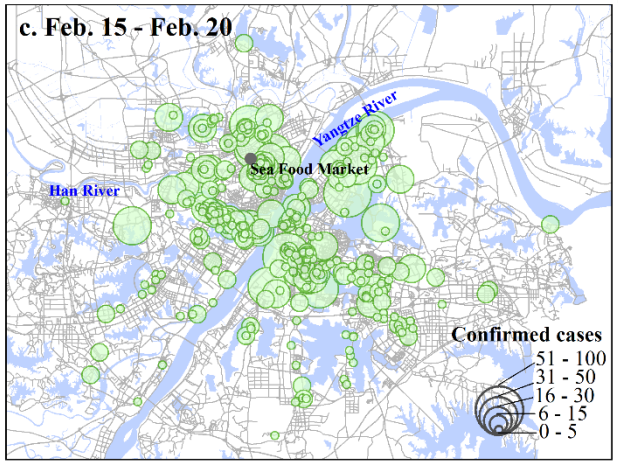
- 武汉市全市累计确诊50340例
- 武昌区确诊人数最多7551例
- 江汉区发病率最高122例/万人

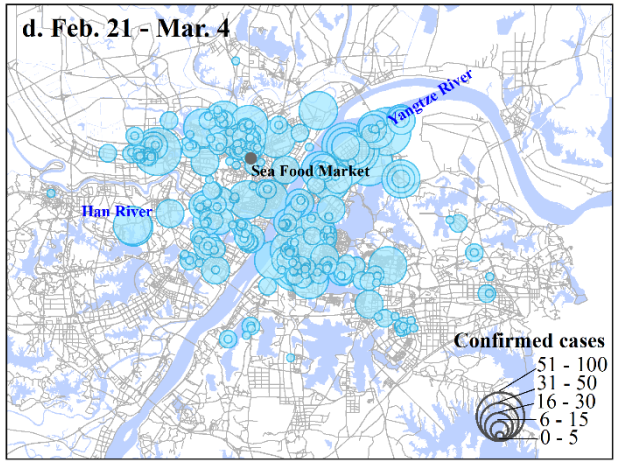
- 只公布了区县级确诊人数
- 研究武汉市新冠疫情时空扩散模式需要更精细的确诊病例数据

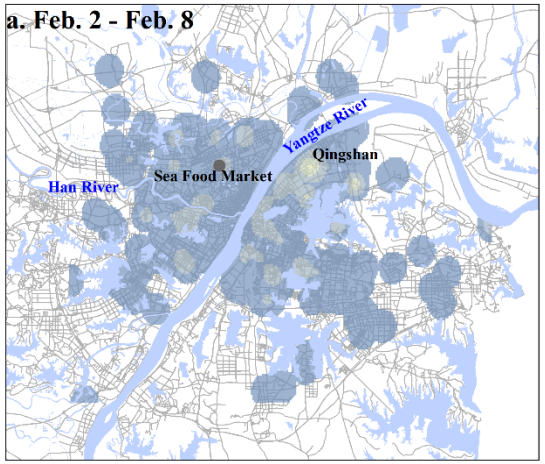
- 收集整理了武汉市疫情暴发时期，小区/社区疫情通告
- 将通告的文本信息结构化、空间化，得到小区尺度的确诊病例数据
- 2700+条数据，20000+确诊病例，覆盖武汉市130+街道

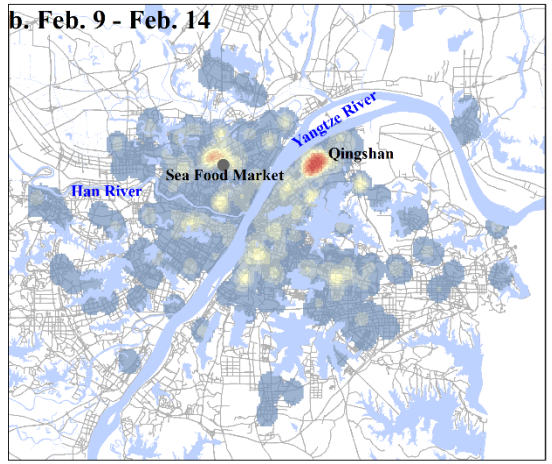
2.4 武汉市新冠疫情时空扩散模式

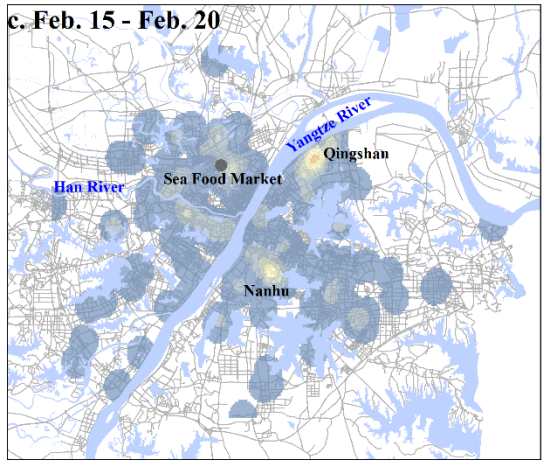


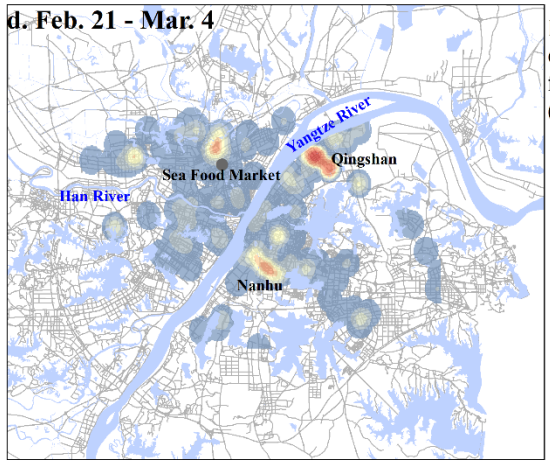


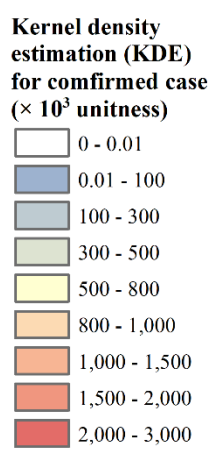


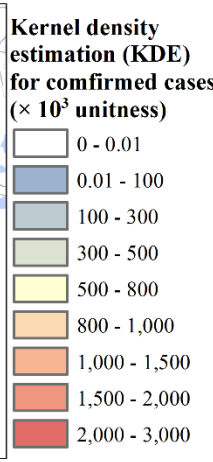








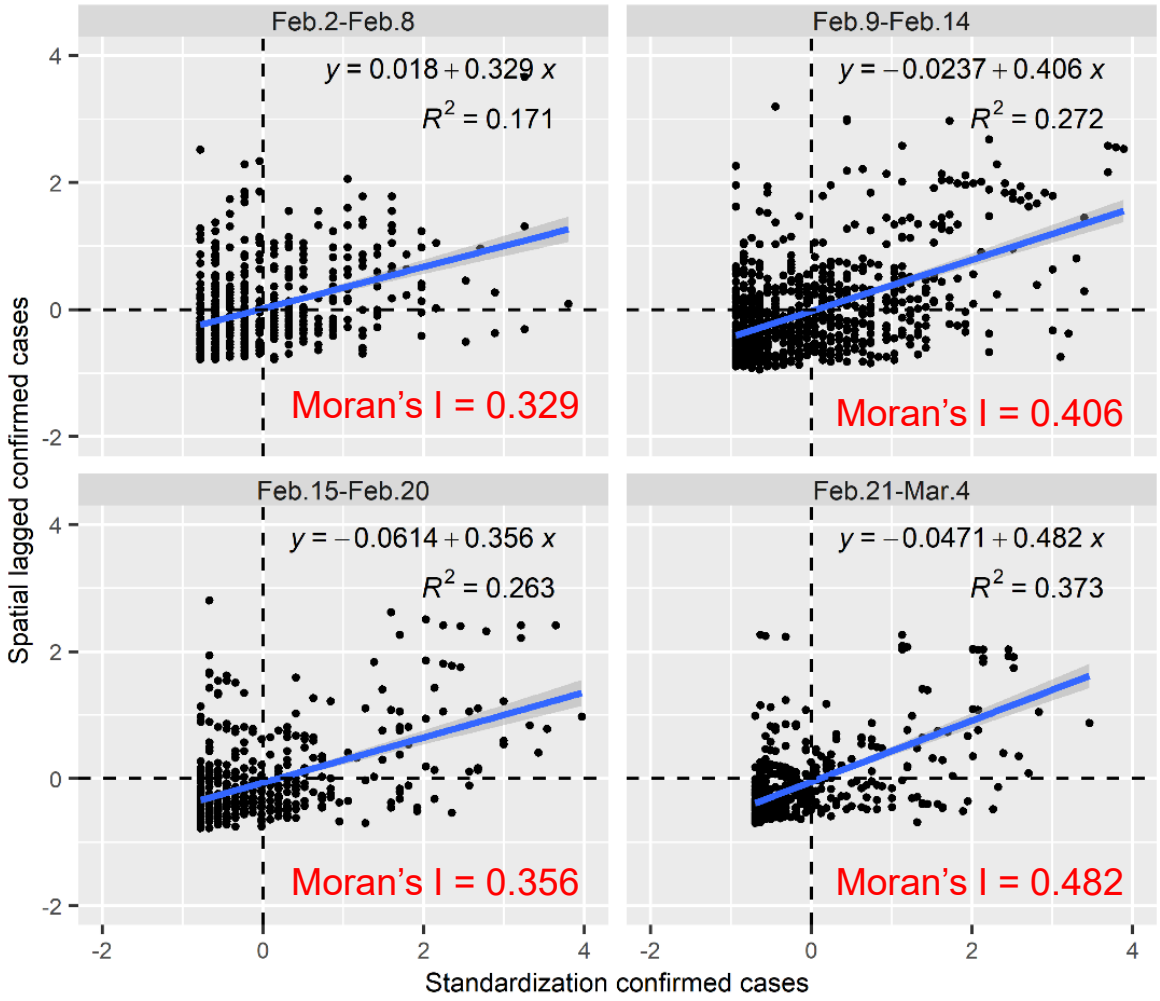




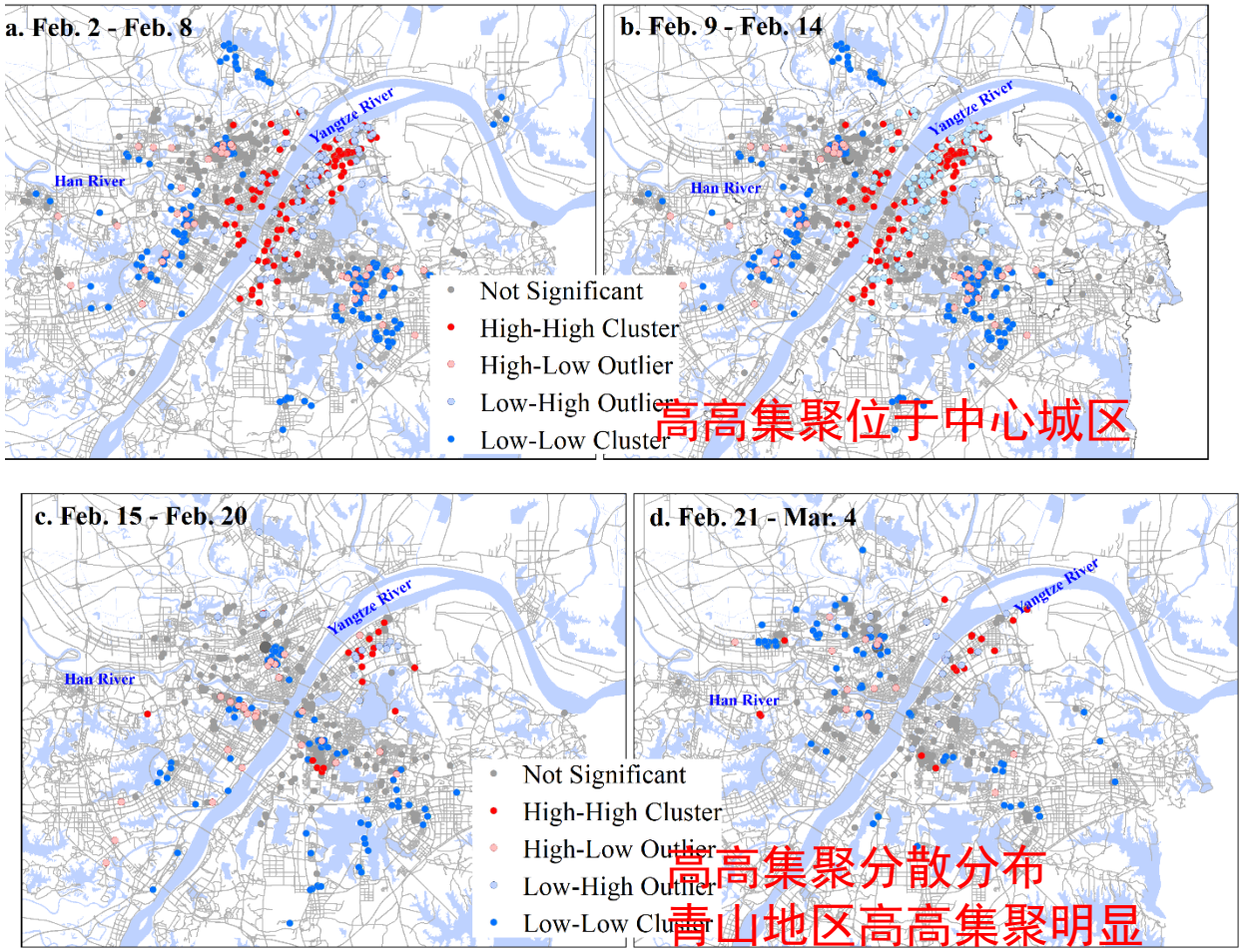
武汉市新冠肺炎确诊病例空间格局及核密度估计

- 2月2日-8日，武昌地区多个小区出现聚集性感染
- 2月9日-14日，武汉三日大会战（大排查），确诊病例空间分布比较均匀
- 2月15日之后，疫情散发趋势明显，但青山红钢城和南湖居住区仍是疫情暴发热点区域

2. 新冠疫情的多尺度时空扩散模式



全局空间自相关分析



局部空间自相关分析

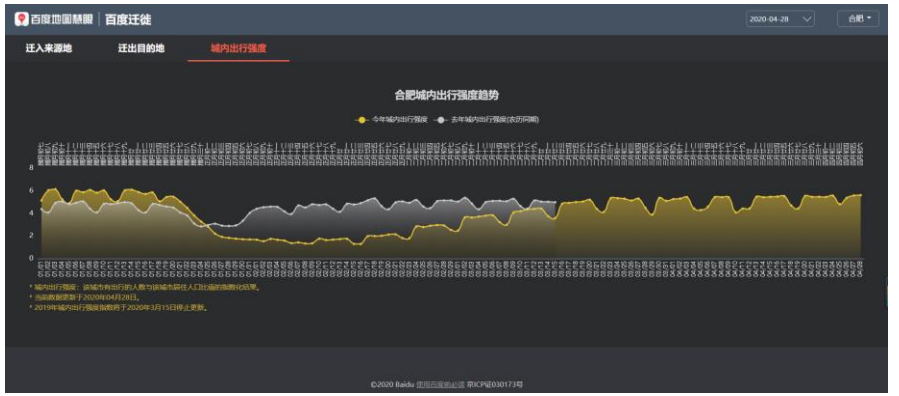
多元流网络网络构建及疫情相关性分析

3. 多元流网络构建及疫情相关性分析

3.1 百度迁徙人流网络构建及疫情相关性分析

1）百度迁徙人流网络构建方法

百度迁徙数据：百度迁徙大数据利用基于地理位置的服务技术，全程、动态、即时、直观地展示了全国人口迁徙的轨迹和特征。



热门迁入地比例：迁入到某城市的人口与全国迁入总人口的比值

热门迁出地比例：从某城市迁出的人口与全国迁出总人口的比值

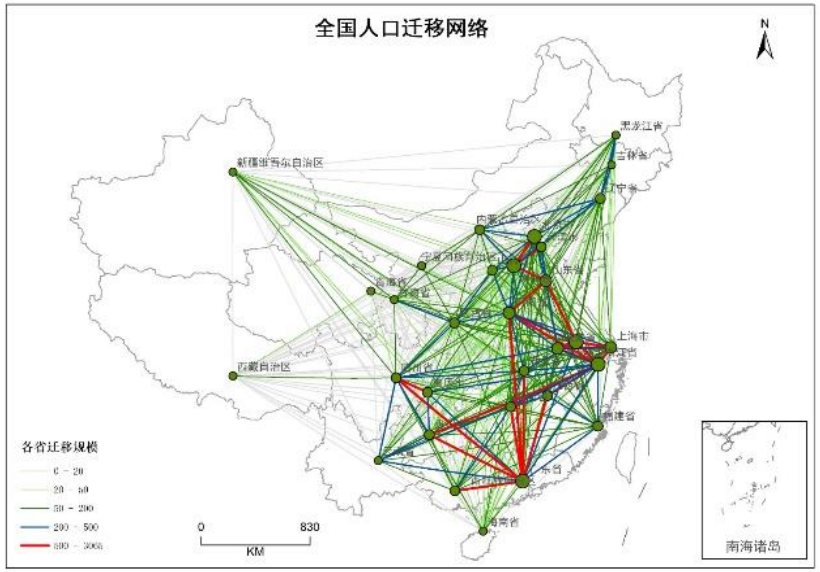
迁移规模指数：反映迁入或迁出人口规模。

城市出行强度：该城市有出行的人数与该城市居住人口比值的指数化结果。

3. 多元流网络构建及疫情相关性分析

3.1 百度迁徙人流网络构建及疫情相关性分析

1）百度迁徙人流网络构建方法



给网络节点赋予权重值，对构建好的网络进行可视化。

网络密度：网络中各个成员之间联系的紧密度，成员之间的联系越多，该网络的密度越大。

出入度：在无向图中指网络中相邻的点的个数。

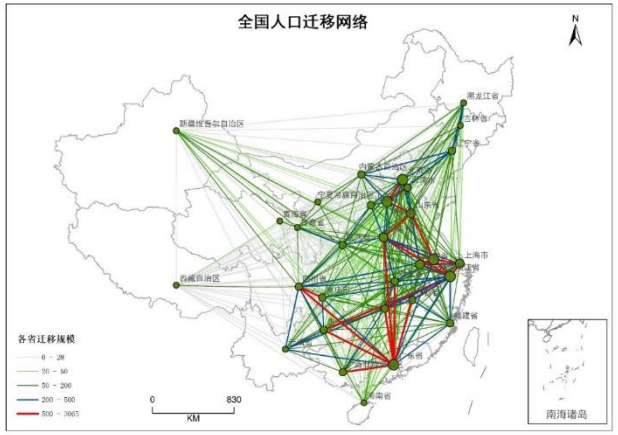
“出入强度”：指在网络中节点迁出规模与迁入规模的强度，即规模累加。

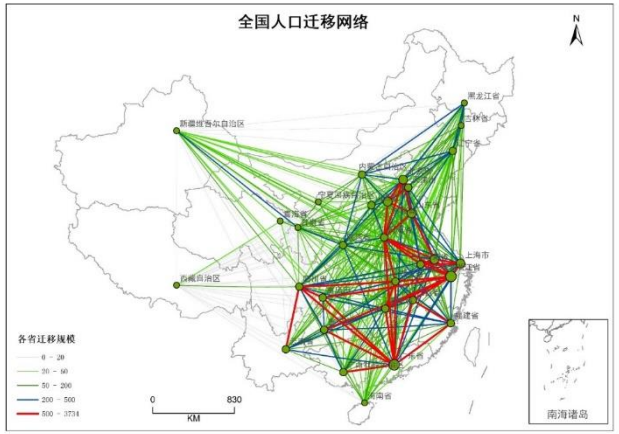
核心度：研究网络中哪些节点处于核心地位，哪些节点处于边缘地位。

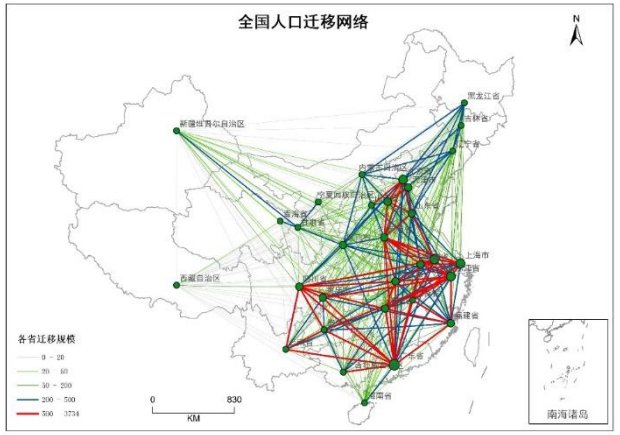
3. 多元流网络构建及疫情相关性分析

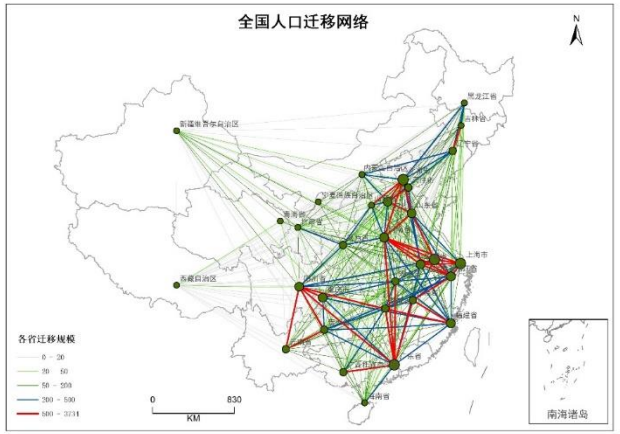
3.1 百度迁徙人流网络构建及疫情相关性分析

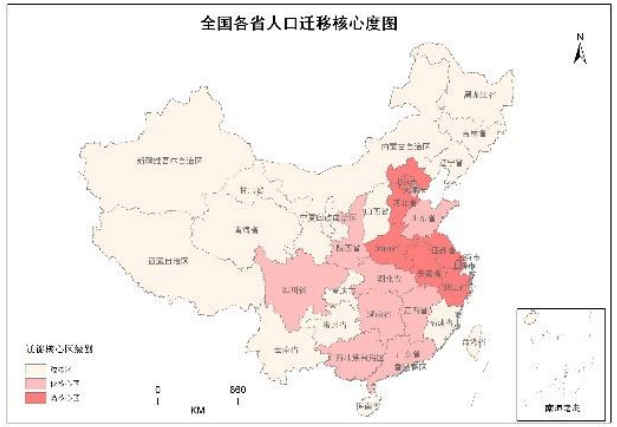
2）百度迁徙网络与疫情的相关性分析



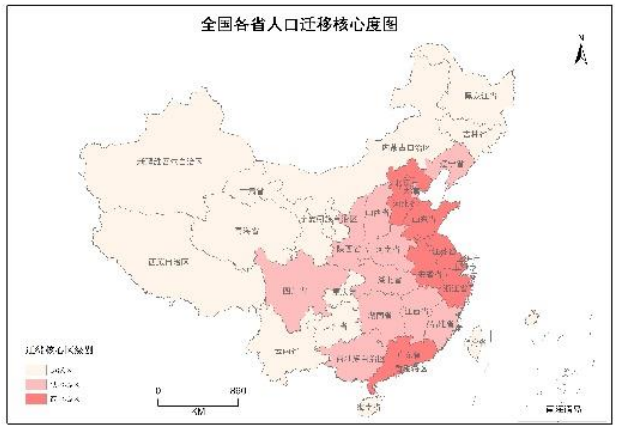




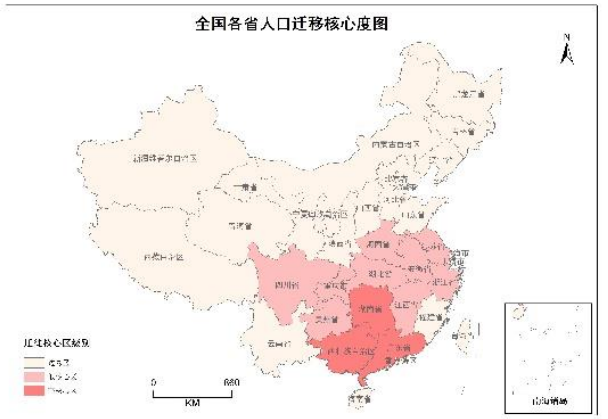




1.1日-1.7日



1.8日-1.14日



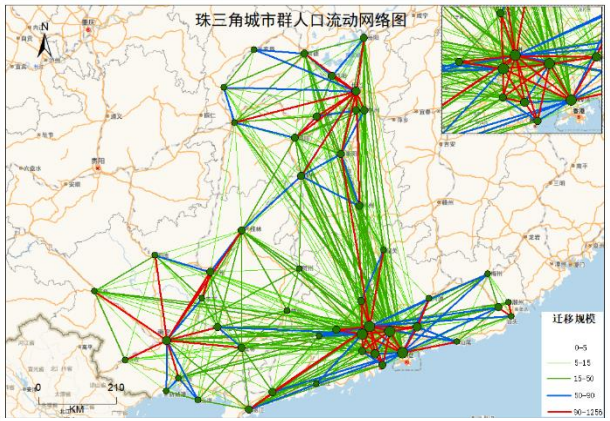
1.15日-1.21日

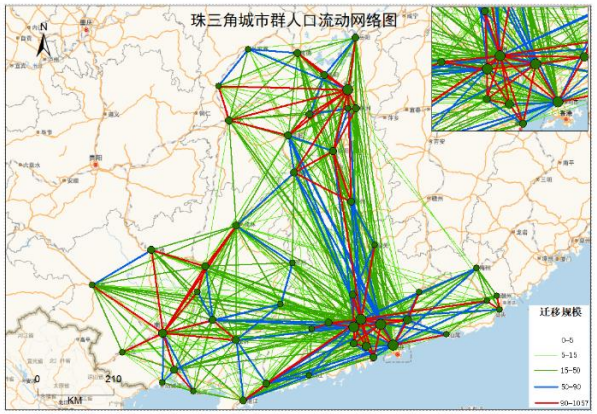
1.22日-1.28日

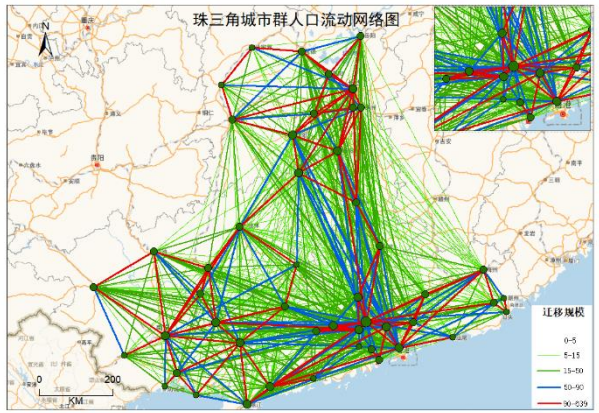
3. 多元流网络构建及疫情相关性分析

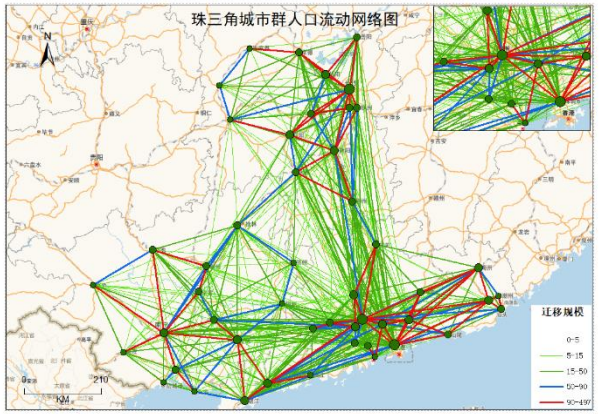
3.1 百度迁徙人流网络构建及疫情相关性分析

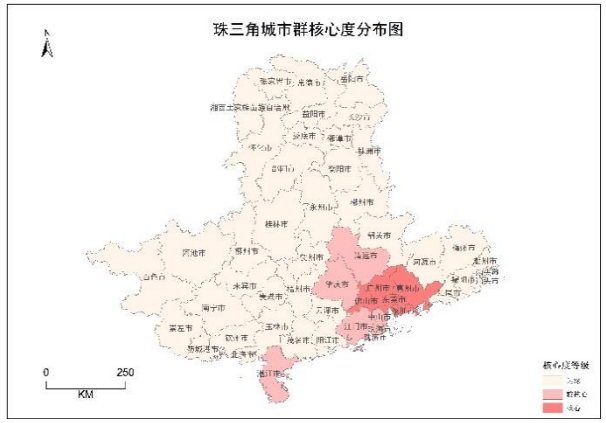
2）百度迁徙网络与疫情的相关性分析



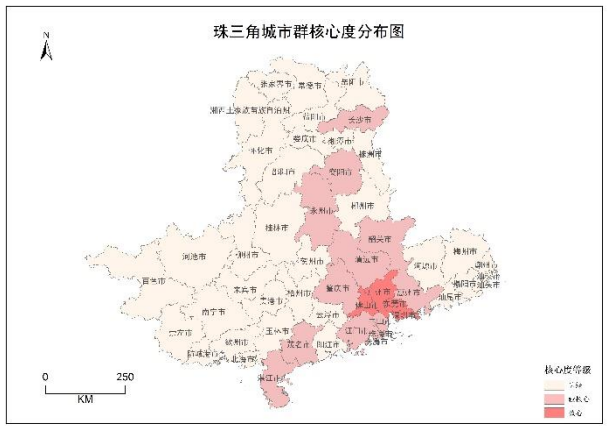




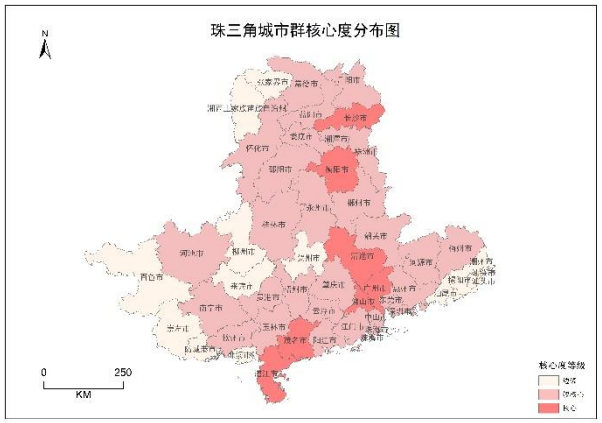




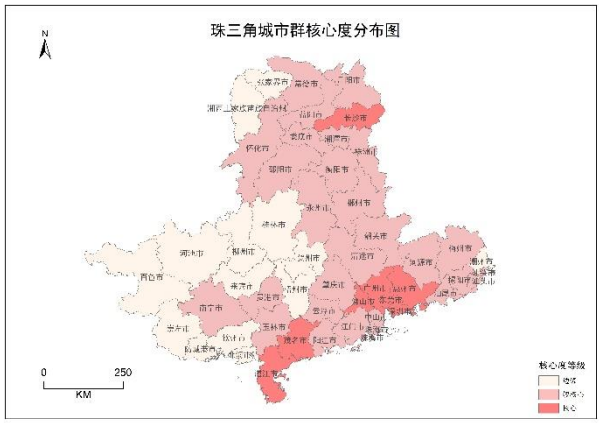
1.1日-1.7日



1.8日-1.14日



1.15日-1.21日

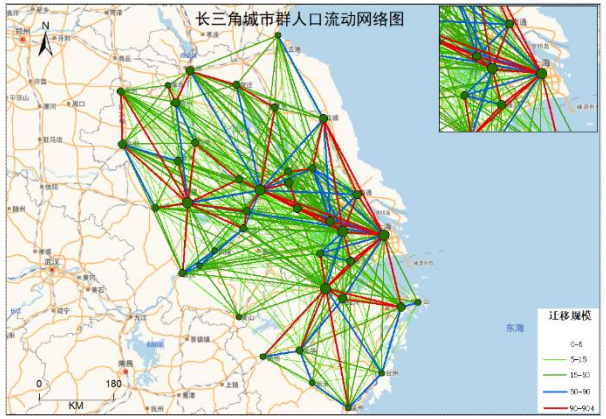


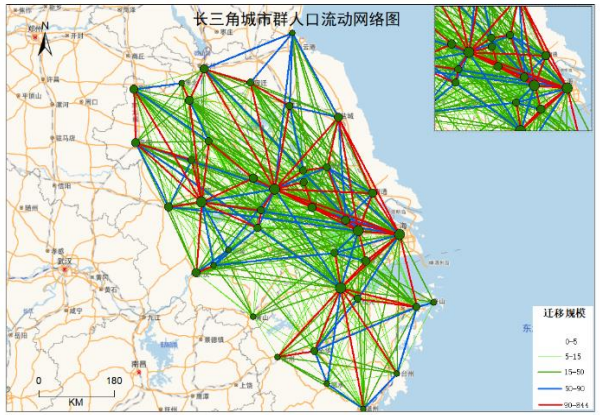
1.22日-1.28日

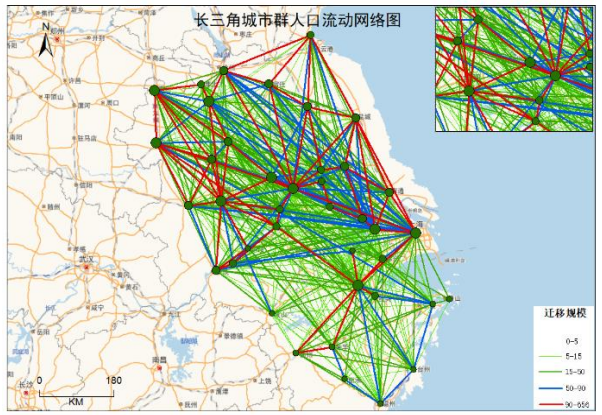
3. 多元流网络构建及疫情相关性分析

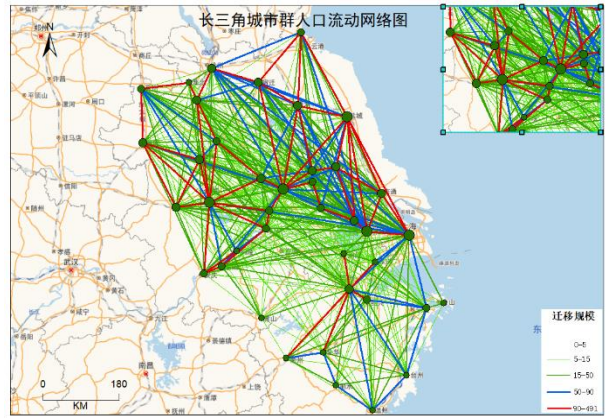
3.1 百度迁徙人流网络构建及疫情相关性分析

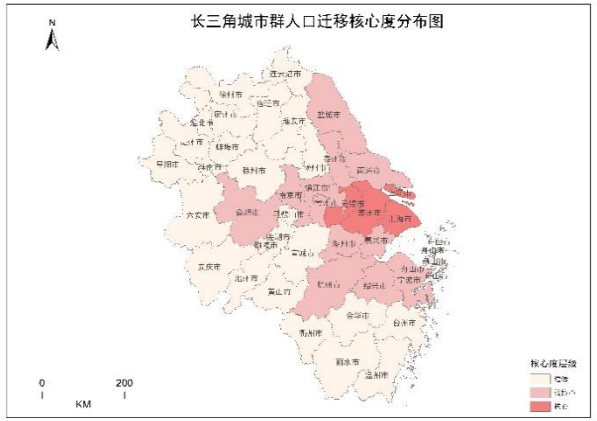
2）百度迁徙网络与疫情的相关性分析

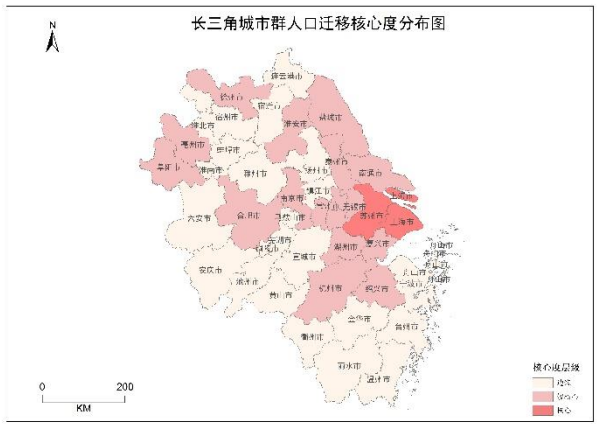


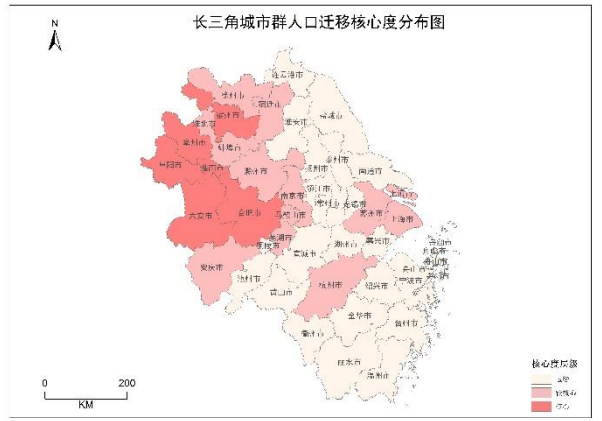


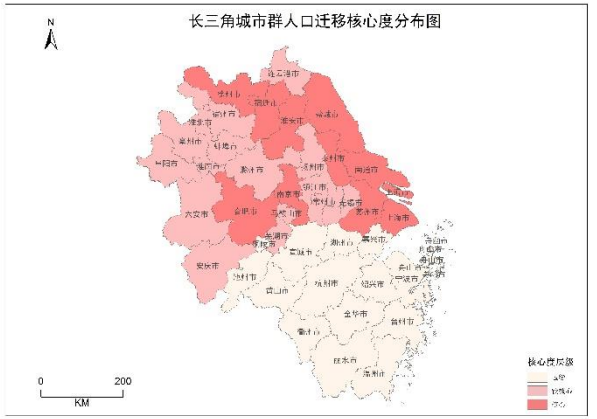










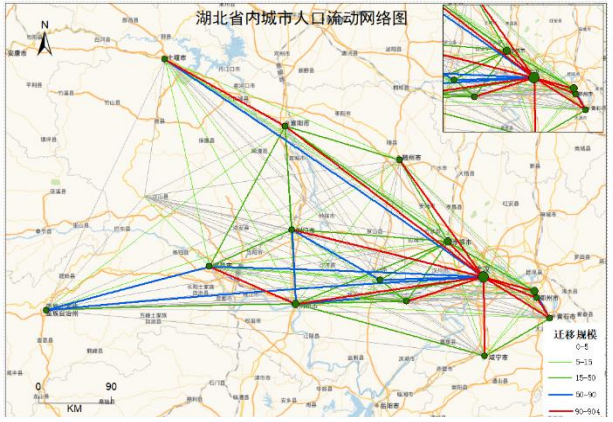


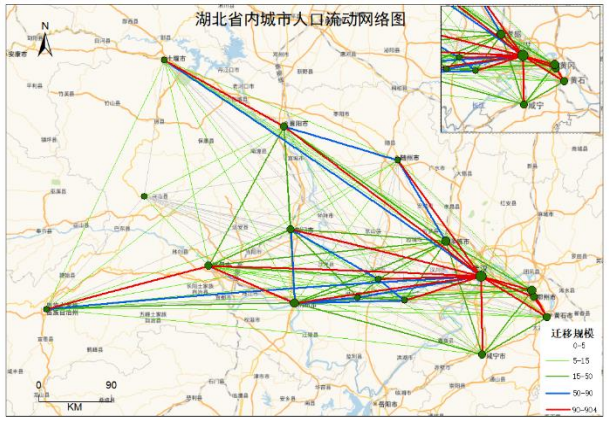
1.1日-1.7日　　1.8日-1.14日　　1.15日-1.21日　　1.22日-1.28日

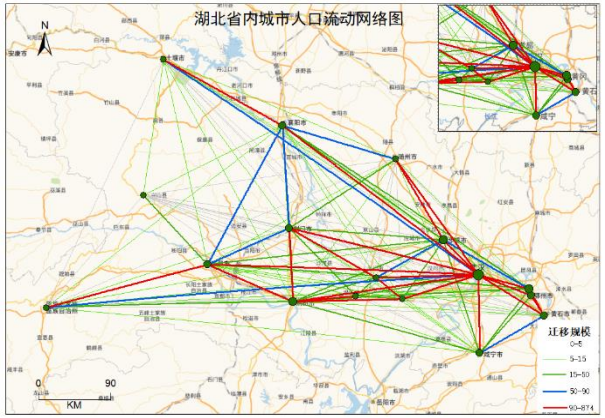
3. 多元流网络构建及疫情相关性分析

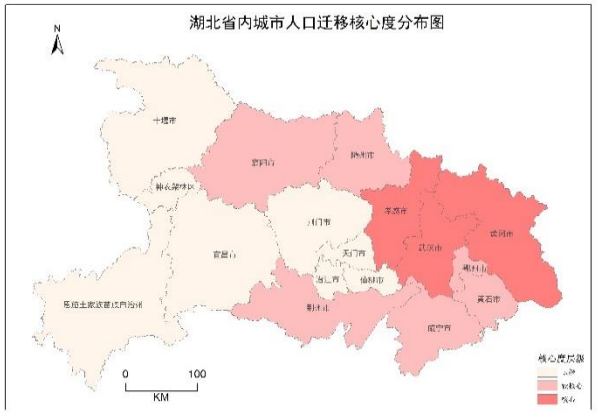
3.1 百度迁徙人流网络构建及疫情相关性分析

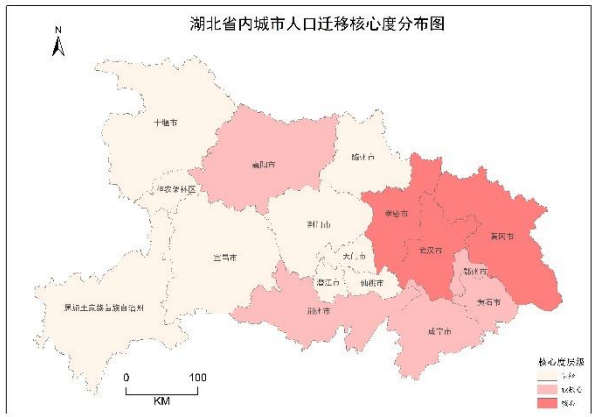
2）百度迁徙网络与疫情的相关性分析

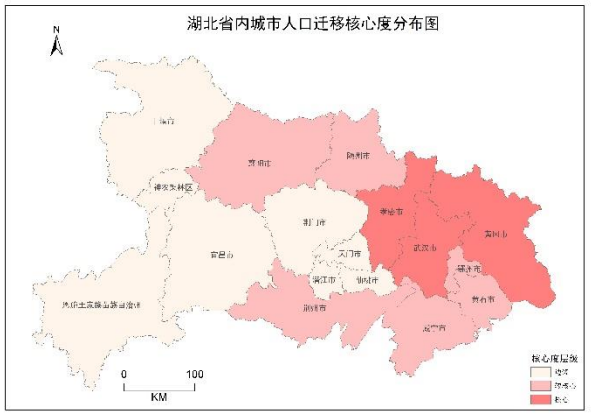


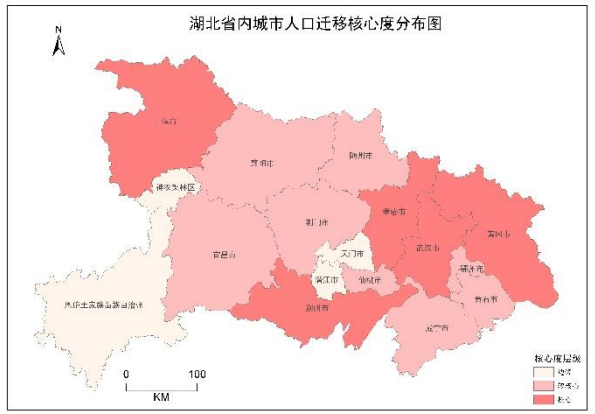












1.1日-1.7日　　1.8日-1.14日　　1.15日-1.21日　　1.22日-1.28日

湖北省内各城市中迁移规模是以武汉为核心，向周围扩散，在武汉封城前的一段时间里，迁移规模并未发生明显改变，其中，孝感、黄冈、黄石、鄂州等城市也是人口迁移规模较大的区域。

3. 多元流网络构建及疫情相关性分析

3.1　百度迁徙人流网络构建及疫情相关性分析

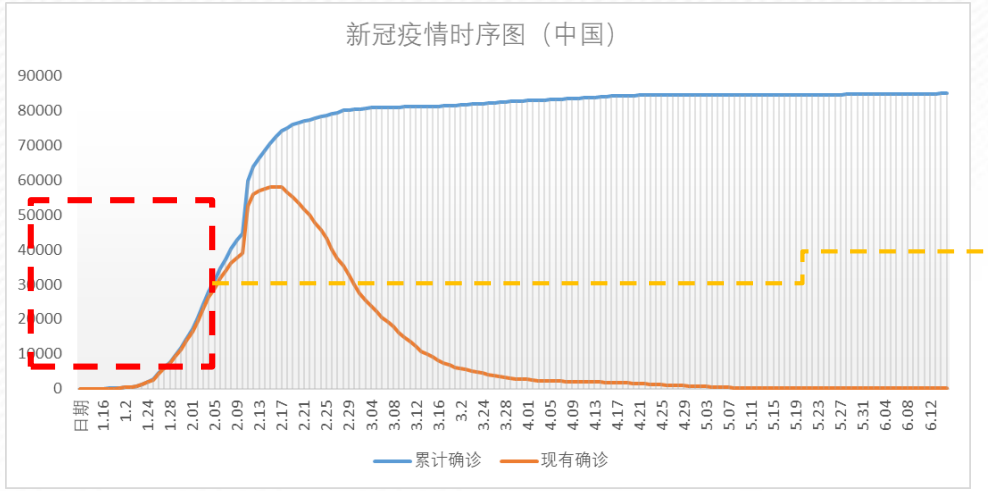
2）百度迁徙网络与疫情的相关性分析

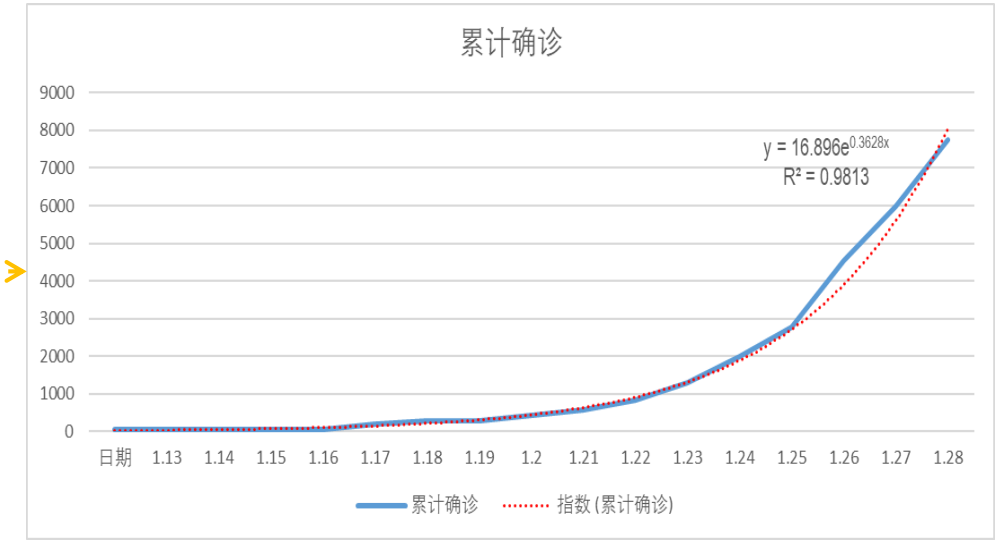
武汉市迁入湖北省内各市的人口流动强度与各地区累计确诊病例相关性随时间增强，最后稳定在0.92左右

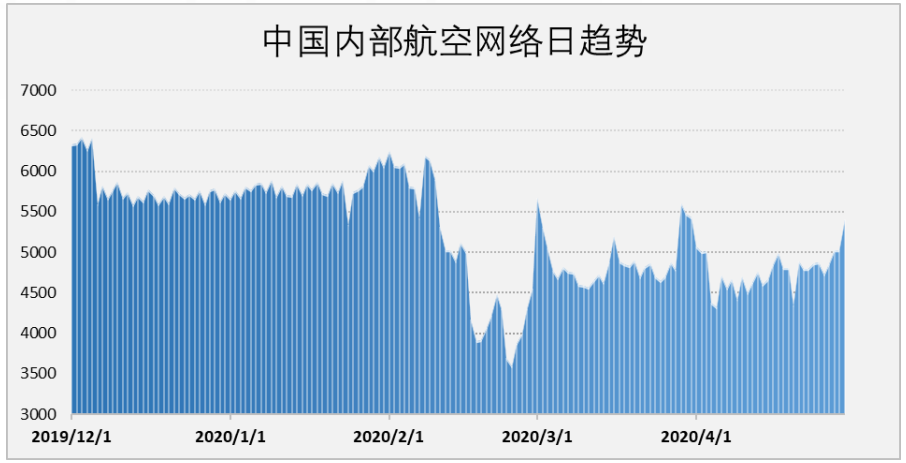
3. 多元流网络构建及疫情相关性分析

3.2 航空流网络构建及疫情相关性分析

1)中国航空网络与疫情相关性分析







一月末到二月初：中国的新冠疫情确诊人数呈指数级上升。

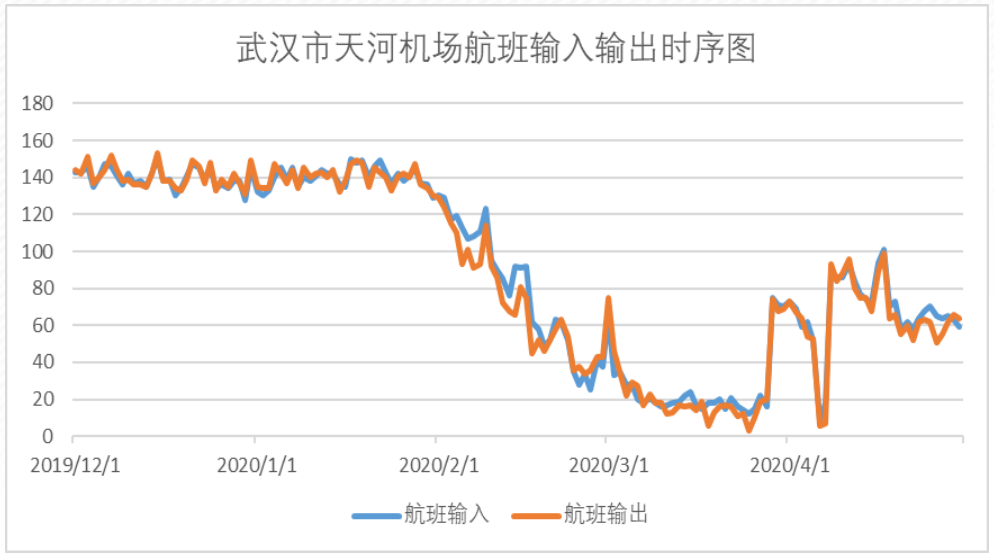
二月份：航班数呈现断崖式下跌，三月份逐步回升。

航班数的波动点时间迟滞于武汉封城以及现有确诊数达到极大值的时间数日。

数据源来自varflight http://www.variflight.com/

3.2 航空流网络构建及疫情相关性分析

1)中国航空网络与疫情相关性分析



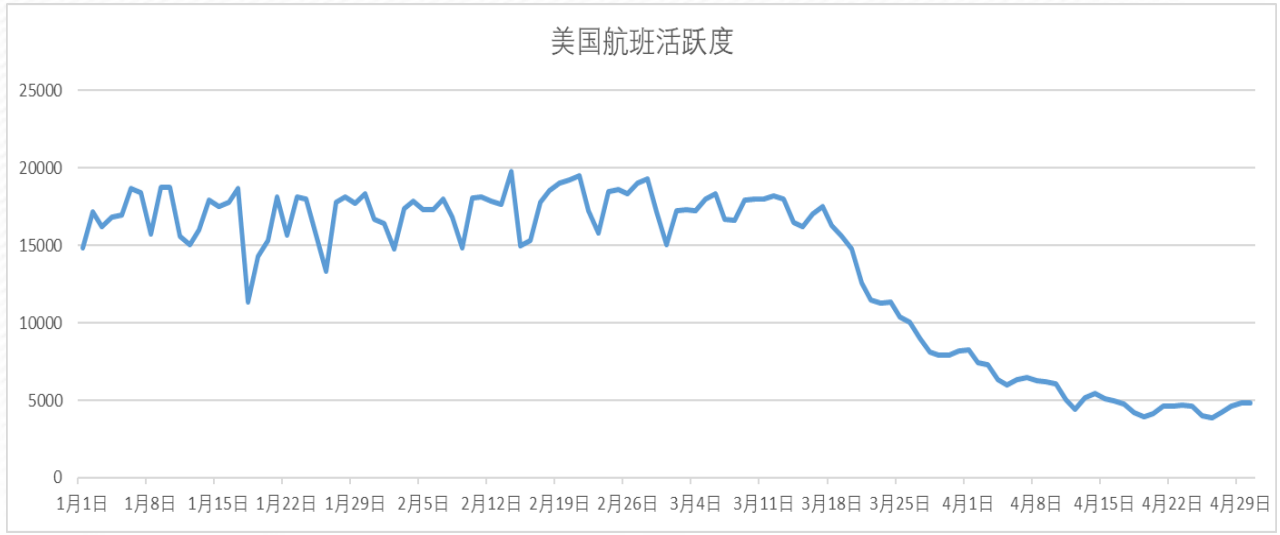
和全国航班的总体情况相比，武汉市的航班活跃度在一月底就开始有下滑趋势，与武汉封城的时间点大致相当。
同比于全国在三月上旬即有恢复迹象，武汉市的航空流在三月下旬恢复。

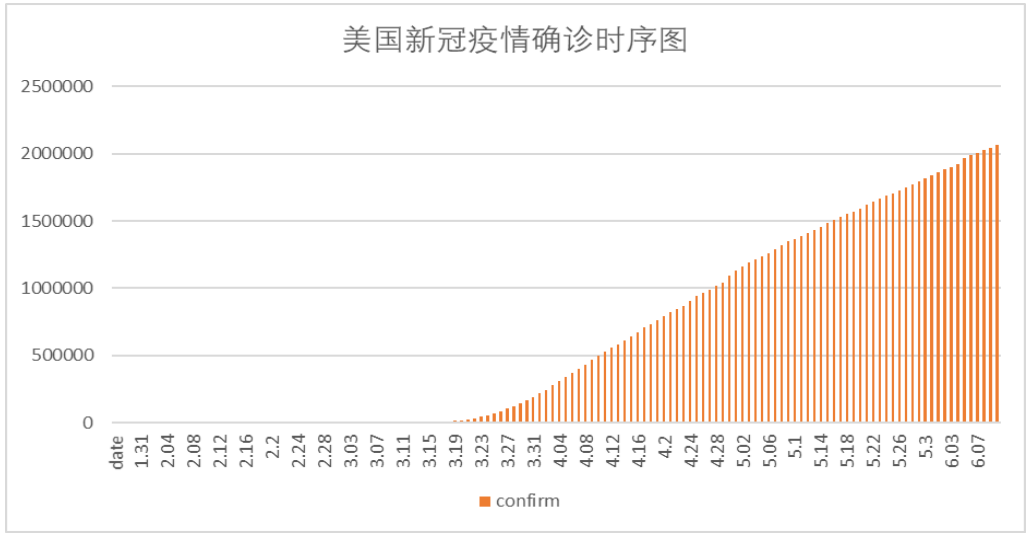
	封城前10天	封城后10天	2月份	3月份	4月份
相关系数 r	0.121	0.754	-0.918	0.170	0.311
显著性 p	0.740	0.012	0.000	0.360	0.094

数据源来自varflight http://www.variflight.com/

3.2 航空流网络构建及疫情相关性分析

2)美国航空网络与疫情相关性分析





	二月份	三月份	四月份
相关系数 r	0.286	-0.899	-0.854
显著性 p	0.133	0.000	0.000

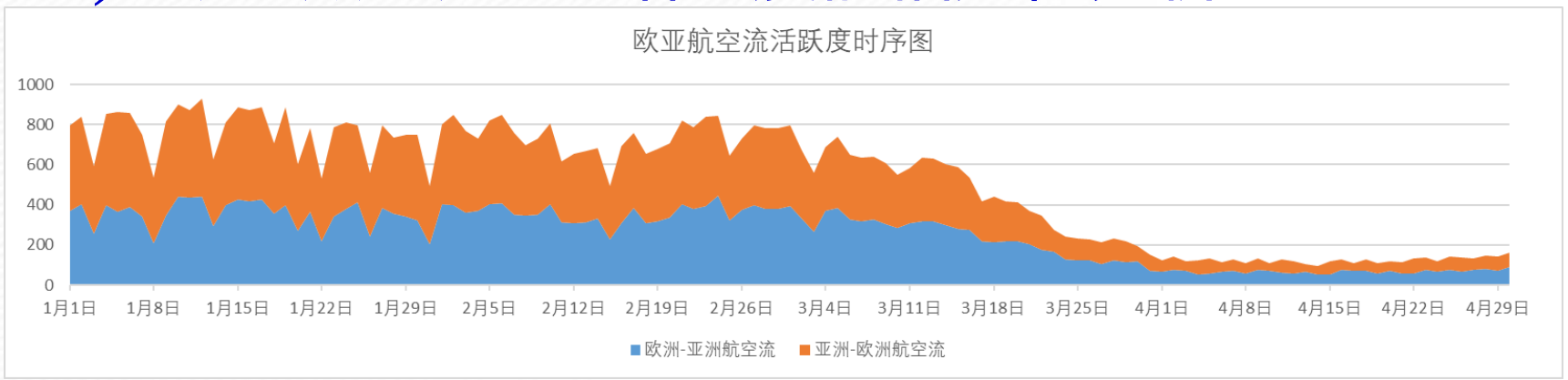
美国新冠疫情累计确诊和航班活跃度的相关系数分析

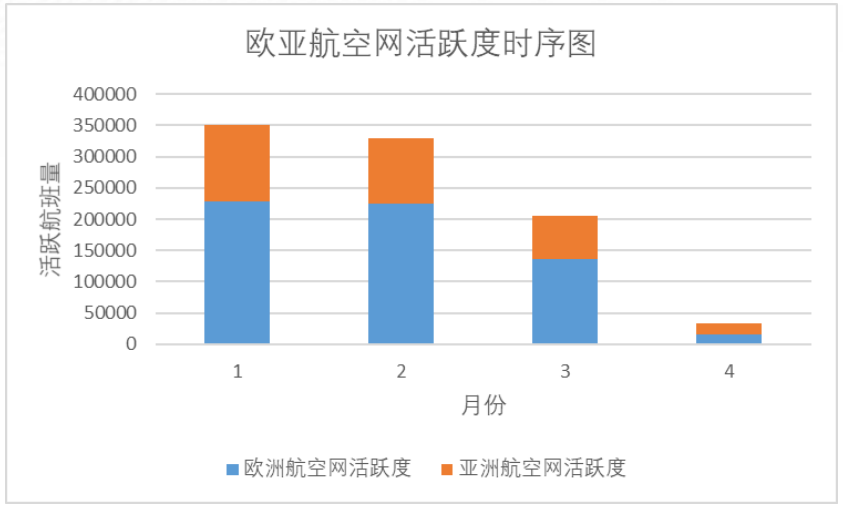
美国疫情在三月份和四月份的时候对于航空网络有较大影响

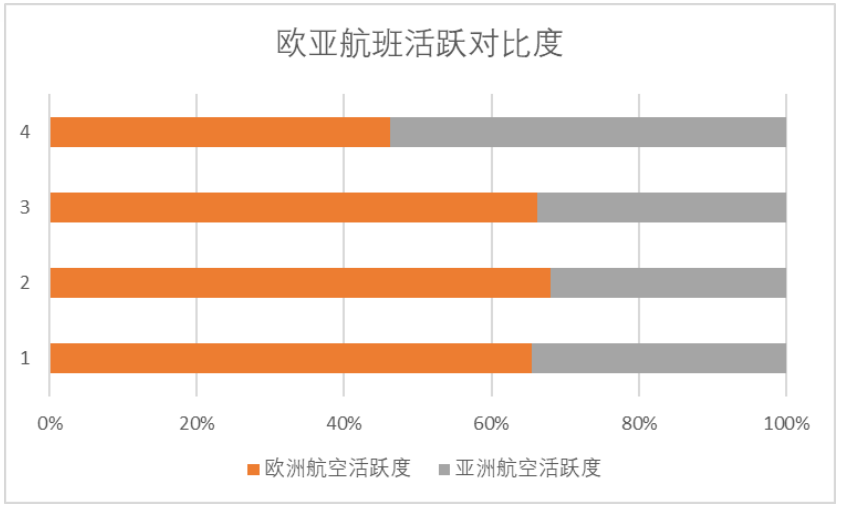
数据源来自opensky https://opensky-network.org/

3.2 航空流网络构建及疫情相关性分析

3)亚洲-欧洲航空网络与疫情相关性分析







疫情爆发后，欧亚航班在三月份的时候开始出现下跌，之后一直处于低迷状态。

亚洲航班的衰落状况比欧洲航班更轻。

到了四月份，亚洲航班的活跃占比超过了欧洲。

数据源来自opensky https://opensky-network.org/

3. 多元流网络构建及疫情相关性分析

3.3 高铁网络构建及疫情相关性分析

1）数据介绍

高铁详细客运量数据并不公开，难以获取实际人口流量。

因此我们采取替代方案，使用高铁班次数据展开研究，并用停运公告验证。

数据从12306官网爬取，详细记录了高铁沿途经停轨迹。

	A	B	C	D	E	F
1	train_No	start_St	arrive_St	serial_No	start_Time	arrive_Time
2	5l000000G241	上海虹桥	南京南	1	9:00	10:00
3	5l000000G241	南京南	济南西	2	10:00	11:59
4	5l000000G241	济南西	北京南	3	11:59	13:28
5	5600000G2011	杭州东	湖州	1	7:48	8:09
6	5600000G2011	湖州	南京南	2	8:09	8:56
7	5600000G2011	南京南	济南西	3	8:56	10:55
8	5600000G2011	济南西	北京南	4	10:55	12:20
9	5l00000G221B	上海虹桥	南京南	1	19:00	20:00
10	5l00000G221B	南京南	北京南	2	20:00	23:18
11	5i00000G2401	合肥南	蚌埠南	1	11:42	12:32
12	5i00000G2401	蚌埠南	济南西	2	12:32	13:59
13	5i00000G2401	济南西	北京南	3	13:59	15:28
14	4f00000G2605	西安北	郑州东	1	9:18	11:08
15	4f00000G2605	郑州东	石家庄	2	11:08	12:39

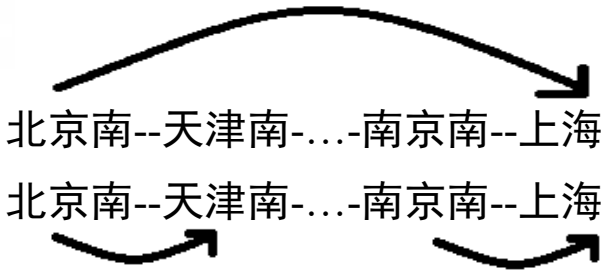
3.3 高铁网络构建及疫情相关性分析

2）高铁网络构建

高铁网络的边有两种构建方法：

- 仅提取始发站和终点站作为有向连接关系
- 提取所有沿途经停各站作为有向连接关系

高铁网络的节点为高铁站所在城市。

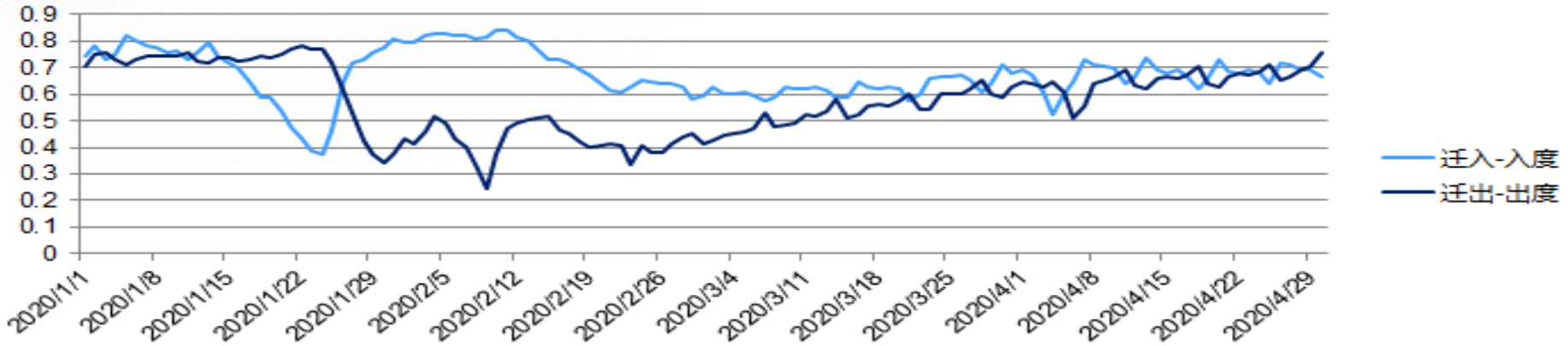


3. 多元流网络构建及疫情相关性分析

3.3 高铁网络构建及疫情相关性分析

3）高铁网络的代表性

验证高铁网络是否可以代表人口流动，将高铁网络节点的出、入度与百度迁徙的城市迁出、迁入规模指数进行相关分析：



整体上，高铁网络节点度和百度迁徙规模指数相关性在0.6左右，属于中等-强相关

- 迁入规模和入度相关性更强，仅在春节/武汉封城节点处相关性下降
- 迁出规模和出度相关性在整个春节/疫情假期中都较低，但是前后较高

高铁网络具有一定的人口流动代表性，但是精确程度有限

3. 多元流网络构建及疫情相关性分析

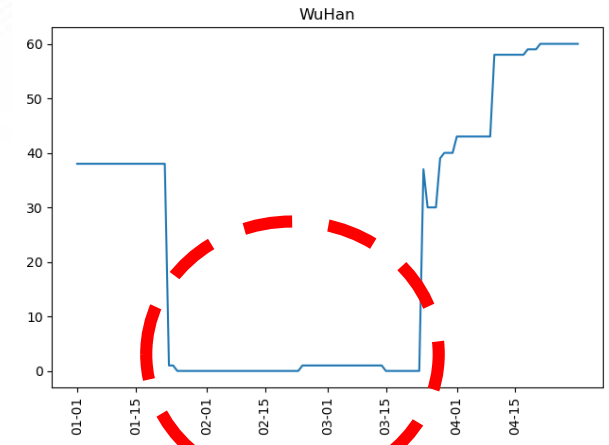
3.3 高铁网络构建及疫情相关性分析

4）疫情相关分析

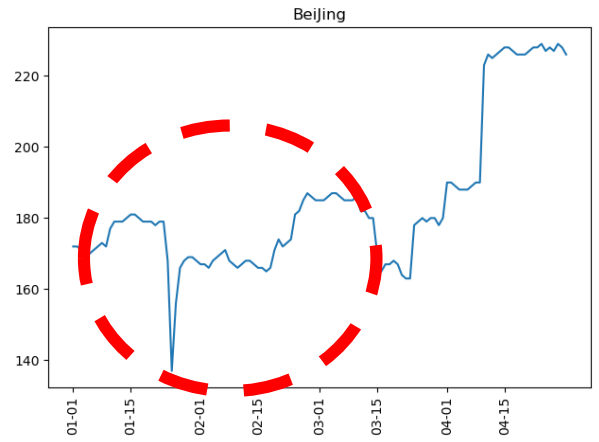
高铁网络从两个方面反应了疫情：

- 从节点角度，节点测度是否受到疫情影响而发生变化？
- 从网络角度，网络的结构和测度是否受到疫情影响而发生变化？

节点入度

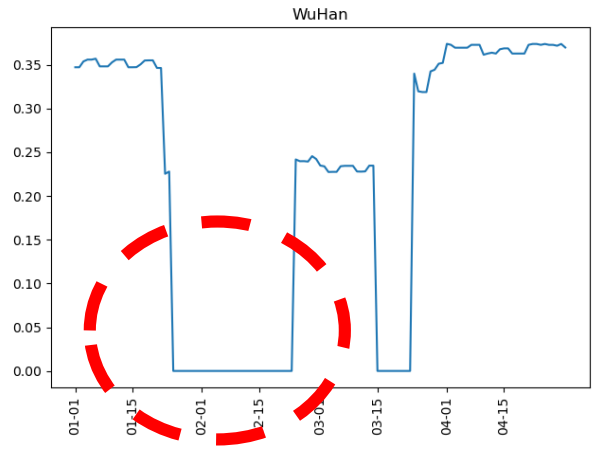


武汉

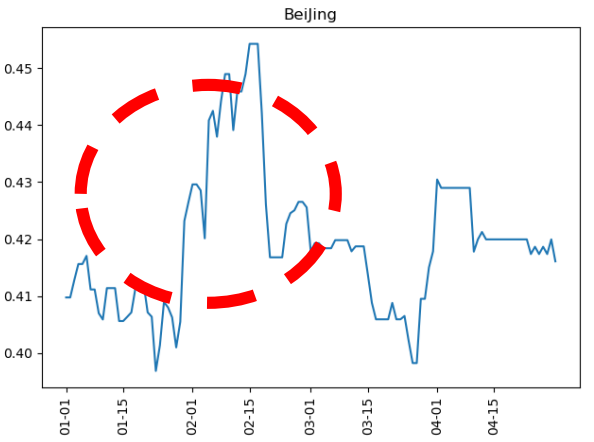


北京

节点中心性



武汉

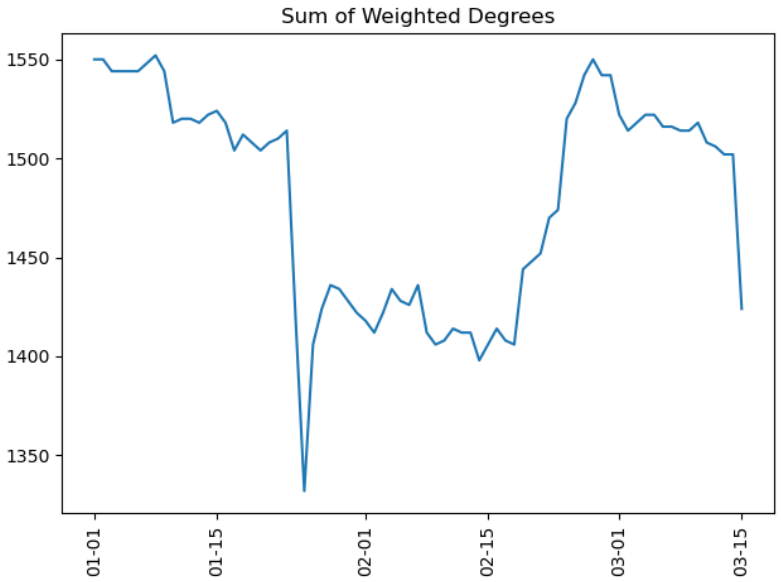


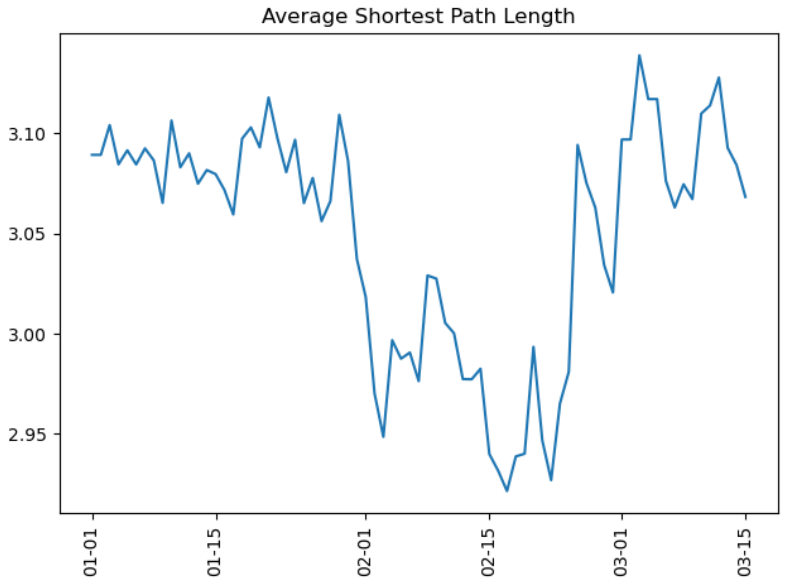
北京

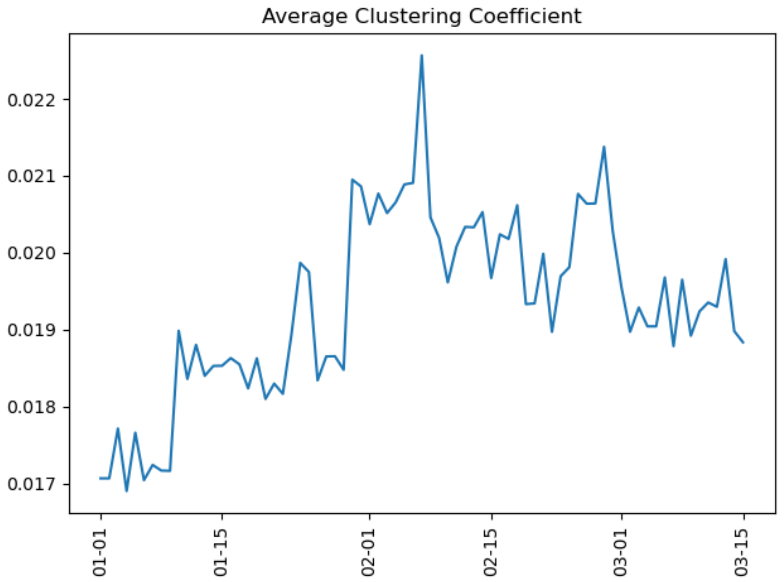
3. 多元流网络构建及疫情相关性分析

3.3 高铁网络构建及疫情相关性分析

4）疫情相关分析







分析疫情期间高铁网络总加权度、平均最短路径和平均聚集系数随着时间的变化

疫情期间高铁网络的规模减小而聚集程度增加。

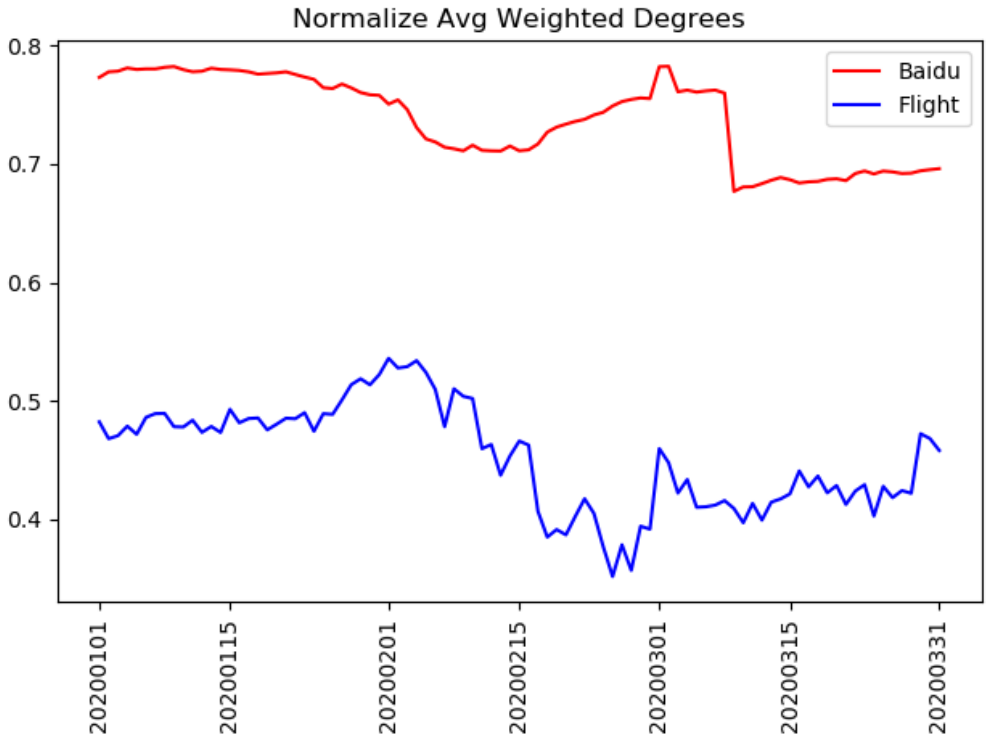
3. 多元流网络构建及疫情相关性分析

3.4 多元流网络的构建

1）单元网络的相关性分析

将航空网络和百度迁徙网络节点的加权度进行归一化显示，探究其随疫情变化所反映的特征，我们发现：

- 百度迁徙网络反映的是大基数的人口迁徙，疫情爆发/封城期人流量显著降低，网络加权度减小，其后疫情减缓，有一定程度的回升；
- 航空网络反映了部分的迁徙，疫情爆发/封城期更多人选择航空出行，因此网络加权度增加，而疫情峰值过去后便大幅降低。

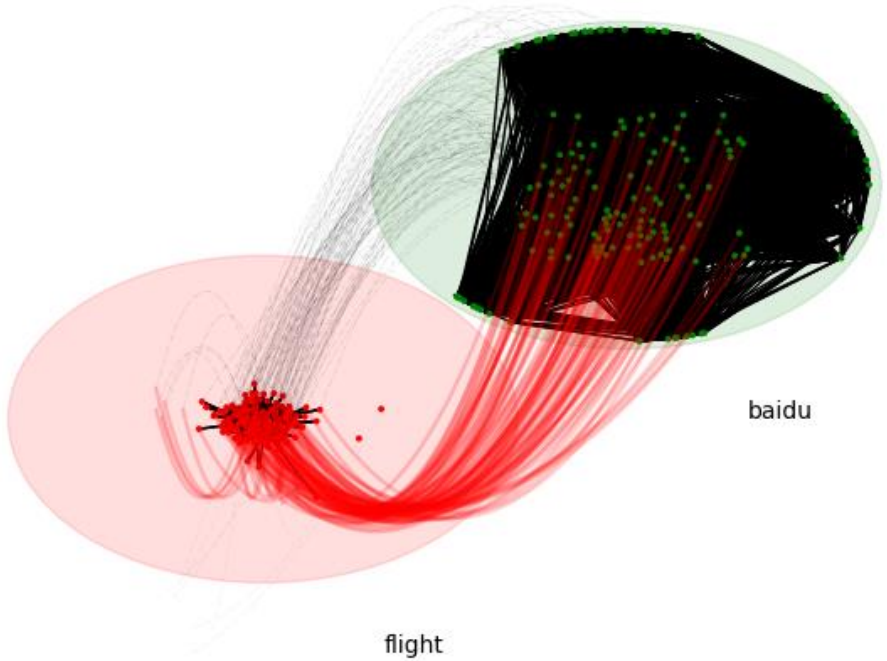


3. 多元流网络构建及疫情相关性分析

3.4　多元流网络的构建

2）迁徙多元流网络的构建

将航空网络和百度迁徙网络中的节点交集作为节点集，层与层之间代表同一个城市的节点之间进行连接，由此构造multiplex模型的城际迁徙多元流网络。



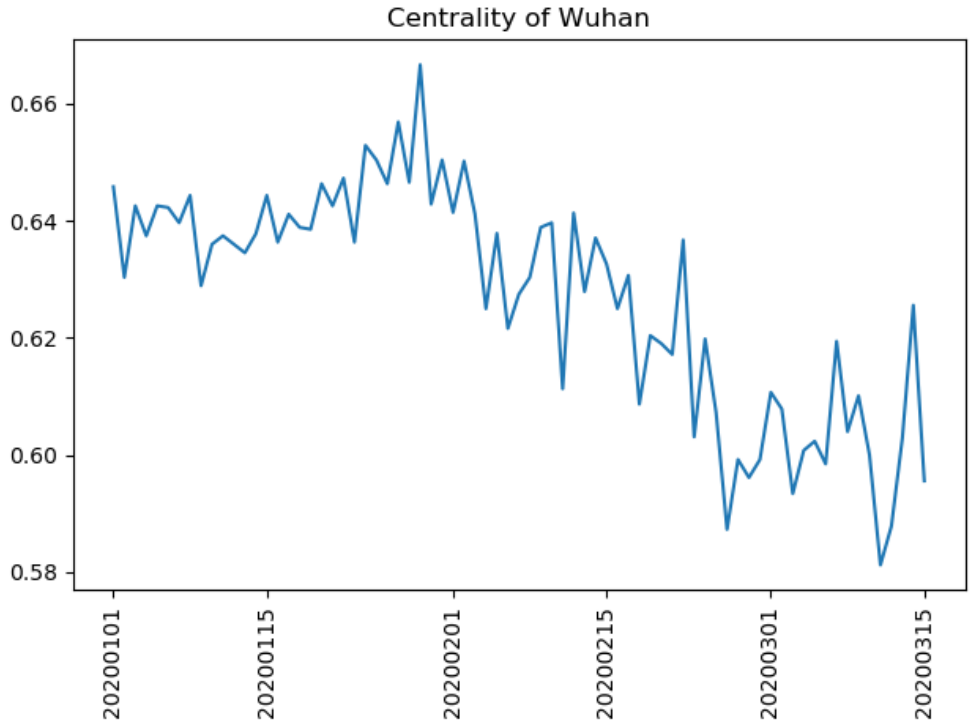
3. 多元流网络构建及疫情相关性分析

3.4 多元流网络的构建

3）迁徙多元流网络分析

使用多层网络的聚合网络进行分析。聚合网络是将各层按照一定的方法聚合成一个单层网络，用以代表multiplex模型的多层网络。

研究聚合网络中武汉市的中心性变化，
发现疫情封城之前武汉中心性达到高峰，
而封城后中心性急速下降，
这反映了实际武汉在疫情前后，
全国迁徙多元流网络中地位的变化。

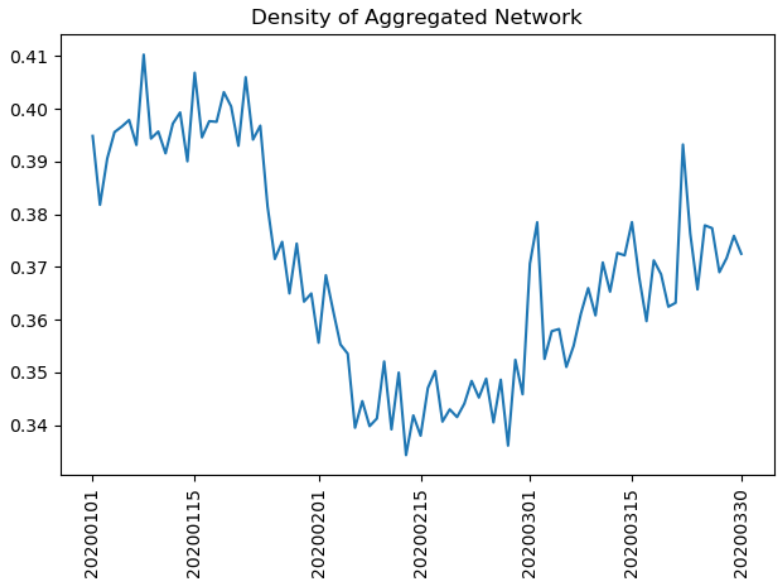


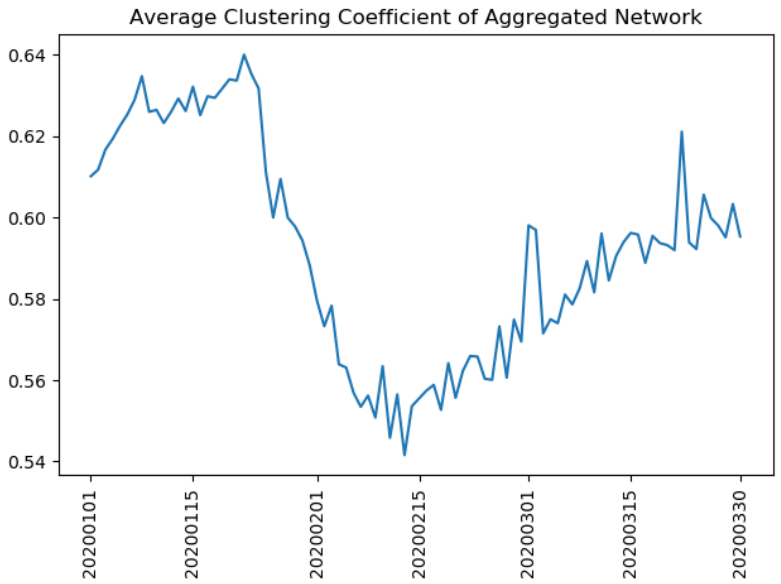
3. 多元流网络构建及疫情相关性分析

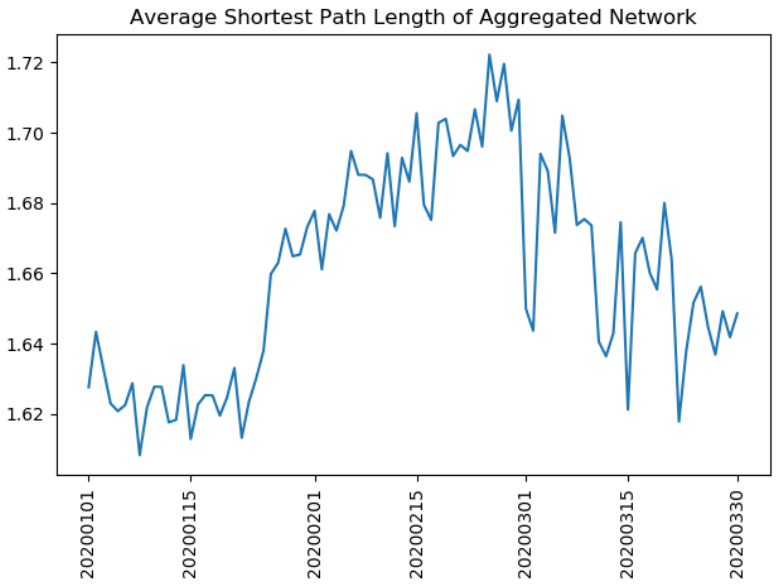
3.4 多元流网络的构建

3）迁徙多元流网络分析

研究聚合网络整体结构测度的变化，发现疫情期间聚合网络的密度、平均聚集系数均下降，而平均最短路径增加，说明疫情期间全国人口迁移多元流网络变得稀疏，而疫情之后则逐渐恢复稠密。







4

总结与讨论

4. 总结与讨论

- “流”数据分析
- 网络化挖掘
- 通过阻断“流”、防控疫情扩散

深入分析新冠疫情的扩散规律、科学防控，
我们一定能够战胜新冠疫情！

致 谢！

感谢合作团队！

合作团队：

鲍曙明、胡涛、
卢宾宾、许刚、余洋、陈江平、贾涛、
王晓悦、张炅焱、张凯、喻雪松、罗萍、熊淑萍、
王一喆、郭文杰、王文蕊、高昇、侯博文、陈斯璐
等。

欢迎加入疫情研究群与注册系列讲座！

加入疫情研究群

Covid-19研究六群

该二维码7天内(6月23日前)有效，重新进入将更新

在线系列讲座注册