

Covid-19 tracking utilizing graph convolutional network

Choujun Zhan, Jianbin Li

JunhuiLu, ShuntaoZhang

Copyright © 2020 Li Jian Bin

Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License (the “License”). You may not use this file except in compliance with the License. You may obtain a copy of the License at <http://creativecommons.org/licenses/by-nc/3.0>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

First printing, Nov 2020

Contents

I

研究背景和回顾

1	研究背景	7
1.1	研究起源	7
1.2	问题描述	7
1.3	研究意义	7
2	现有的相关研究	9
2.1	Literature Review (浓缩总结后)	9
2.1.1	接触网络模型	9
2.1.2	基于接触网络的传播模型	9
2.1.3	防疫策略以及相关的评估标准	9
2.2	阅读笔记(详细)	9
2.2.1	传统的传播模型的文章阅读笔记	9
2.2.2	基于Agent-based传播模型	9
2.2.3	COVID-19病毒防控策略	9
2.3	针对前人的研究存在的问题提出的改进思路	12

II

数据介绍与动态网络构建

3	数据介绍	17
3.1	Gowalla Data	17
3.2	Foursquare_small Data	17
4	动态接触网络构建	19
4.1	区域数据划分	19
4.1.1	判断该点是否在New York城市内	19

III

基于动态网络的传播模型

IV

干预措施

5	邻接检测	27
6	区域检测	29
6.1	数据分析	29
6.1.1	个体不带口罩，政府执行区域检测隔离措施	29
7	基于历史轨迹的区域检测	31
8	基于图神经网络的区域检测推荐	33

V

实验结果

9	模型参数设置	37
10	结果分析	39

VI

Discussion and Conclusion

研究背景和回顾

1	研究背景	7
1.1	研究起源	
1.2	问题描述	
1.3	研究意义	
2	现有的相关研究	9
2.1	Literature Review（浓缩总结后）	
2.2	阅读笔记(详细)	
2.3	针对前人的研究存在的问题提出的改进思路	



1. 研究背景

1.1 研究起源

1.2 问题描述

1.3 研究意义



2. 现有的相关研究

2.1 Literature Review (浓缩总结后)

2.1.1 接触网络模型

2.1.2 基于接触网络的传播模型

2.1.3 防疫策略以及相关的评估标准

2.2 阅读笔记(详细)

2.2.1 传统的传播模型的文章阅读笔记

2.2.2 基于Agent-based传播模型

2.2.3 COVID-19病毒防控策略

COVID-19 病毒传播迅速，传染力极强，全球只有少量的国家能放置该病毒的传播，面对病毒的快速传播，不同的国家采取了不一样的检测措施，针对检测不同的检测对象可以分为三种方法[crozier2021put].

- Symptomatic testing : Symptomatic testing 检测策略仅对有症状的感染者进行检测，迅速对感染者追踪隔离，因此在检测率上往往有较高的准确率。目前使用该策略的国家有日本。
- asymptomatic testing : asymptomatic testing 是一种针对无症状感染者的检测策略，这类方法目前是多数国家普遍采取的检测措施，并且不同的国家会针对不同的人群进行检测，英国和奥地利专注于定期检测保护易感染人群，高风险地区。德国和比利时专注于对无症状感染者在5-7天进行每日检测，这类检测策略可以维持了成本和防疫力度的初步方法，对高风险地区进行检测可以大大提高检测命中率，但是，随着时间的推移以及感染人数的增加，该方法的可行性还有待证明。

- Mass testing: Mass testing大量的对社区的所有人群检测，或者对少量人口停止社区活动并进行重复性的检测。目前使用这类方法的有中国和越南，这类方法可以高效的控制病毒的传播，从而有效的抵御疫情，然而，这种方法成本极高，对于一些经济实力较弱的国家很难长期执行，一旦出现短时间内超级传播者经过多个地区进行全市甚至全省的检测，便需要极高的成本。

以上的三种方法在实际生活呈现出不同的状况。除了中国坚持Mass testing的方法使得整个国家仅有少数的个体感染外，其他的方法并没有。尽管Mass testing是真正意义上控制了感染人数的增长，然而这需要大量的劳动力成本和材料成本。因此，提高检测效率和降低成本扮演着控制病毒传播的关键地位，目前大量的研究针对经济和疫情的控制之间的平衡做出了相关的研究。

[[kolumbus2021effectiveness]] 提出了一个在多种高效干预措施下的SEIR的仿真模型，目的是权衡经济影响和健康之间的联系。在模型的模拟方面，作者模仿了现实情况，设置了一个E-R网络，每个节点的平均度为3.6(在没有任何干预下他引用别人的文章说是合理的),病人的平均潜伏期和平均感染期分别设置了6, 8天，而感染率 $q = 0.5$, 病发概率为 $p_{symp} = 1 - q^{1/8}$, 总节点数10000, 整个实验模型模拟了540天。在干预措施上作者在Lockdowns下按概率切断节点之间的边，也就是说，封锁程度0.3意味着存在连边的两个节点的边可能有30%的概率会切断，针对隔离和追踪措施的干预下，对感染者的直接接触者进行检测，只隔离检测呈阳性的人。此外，作者使用封锁程度来衡量劳动力的损失情况，即考虑到的只是封锁程度与劳动力存在线性关系。结果表明，检测和隔离的结合能有效的减少经济经济成本呢死亡率，但是大量的检测有显著的效果。跟踪结合适度的测试能力可以在没有封锁的情况下实现遏制。(copy) 跟踪和检测结合起来以降低死亡率和经济成本的政策极其有用，它们有可能在不施加任何社交距离限制的情况下遏制疫情。这突出了交易的棘手的社会问题——这些收益与病人的隐私相冲突，追踪不可避免地侵犯了病人的隐私。

[[eilersen2020cost]] 专注于衡量在有限的隔离与检测的情况下减缓成本收益同样地使用SEIR仿真模型进行研究，但作者不仅将个体分为30%Work,40%Family,15%Friends,15%Public,而且在状态的持续时间有所调整，令S,I,R的状态平均维持在时间3天，从E到I的平均时间为5天，其中有一半的时间维持着E的状态，后半段时间维持Pre-symptomatic(infectious)的过度状态。在个体的设置上是模仿了丹麦的人口分布，平均每个个体有2位亲人，10位工友，两组朋友每组有5位。每个个体每一轮(0.5day)具有一定概率抽到扮演着这4中角色的机会，总共5000个个体，最后感染率的设置是美国死亡率 + 23%。结果表明，图2.2.3是实验的结果，减少工作、社交接触、减少工作间的规模程度能有效的延缓并降低感染人数。接着，作者使用 1 step tracing and quarantine strategy(1STQ)策略，即

- close the workplaces of people who are tested positive for the disease.
- isolate their regular social contacts for a limited period
- keep symptomatic individuals in quarantine until they recover.

这种策略与overall Lockdown相比可以大大减少损失的工作日。如图2.2.3为采取封锁与未采取任何措施的结果，非常有意思的是，当Lockdown解除以后的30天开始检测隔离措施，隔离了大量的个体，最高甚至达到约 30%，而no testing 的感染人数尽管并没有隔离任由

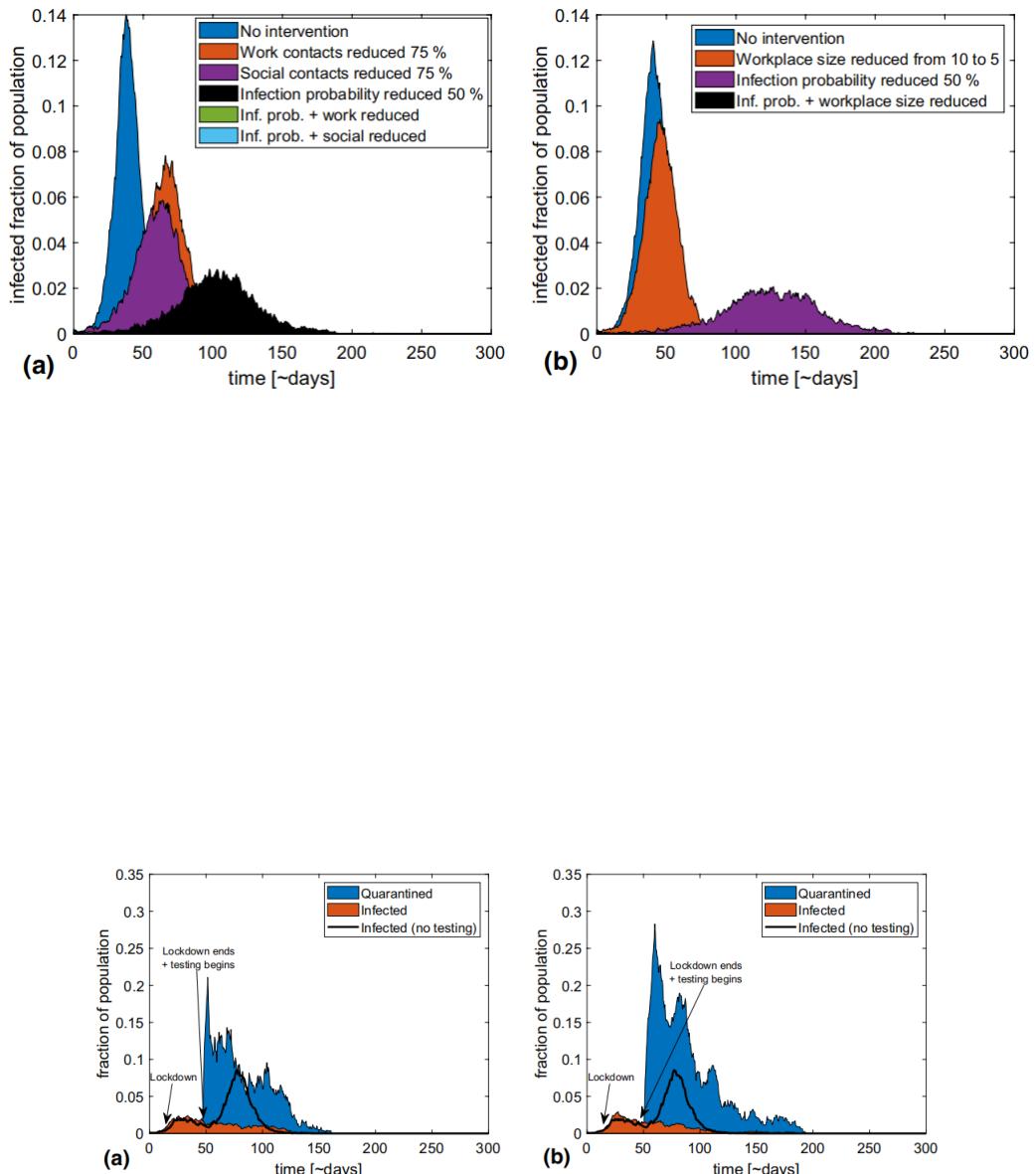


Figure 4. Various trajectories of the epidemic when combining a lockdown and a late-onset ISTQ strategy. (a) shows a possible course of an epidemic where restrictions in public and work life (by 75%) are implemented when 1% are infected and lifted after 30 days, being replaced by a testing and tracing regime with a testing probability of 20% per day and quarantine duration of 5 days. The black line shows the fraction of infected if no testing is implemented. We see that this level of testing and quarantine is sufficient to prevent a resurgence of the epidemic. (b) is similar, but here quarantine lasts 10 days. This is about as effective as (a) but costs a lot more in terms of number of people in quarantine.

病毒在相同的环境下传播，但是最高的感染人数不过10%， Quarantined 的节点是需要隔离5天的，无隔离措施的感染人数最高占约10%，并持续时间大概在50天左右，而采取隔离措施的情况却是隔离人数最高甚至接近30%，并持续约150天。我认为这样的隔离措施过于粗糙，假设病毒不会变异的情况下，尽管采取了这样的措施在50-100天的感染人数大大减少，隔离的人数和持续时间可以侧面反应成本，因此不难发现成本之大，也就是说尽管从结果上看采取1STQ的措施能大大减少损失的工作日，但是由于隔离的人数大大增加，损失了大量的劳动力。

2.3 针对前人的研究存在的问题提出的改进思路

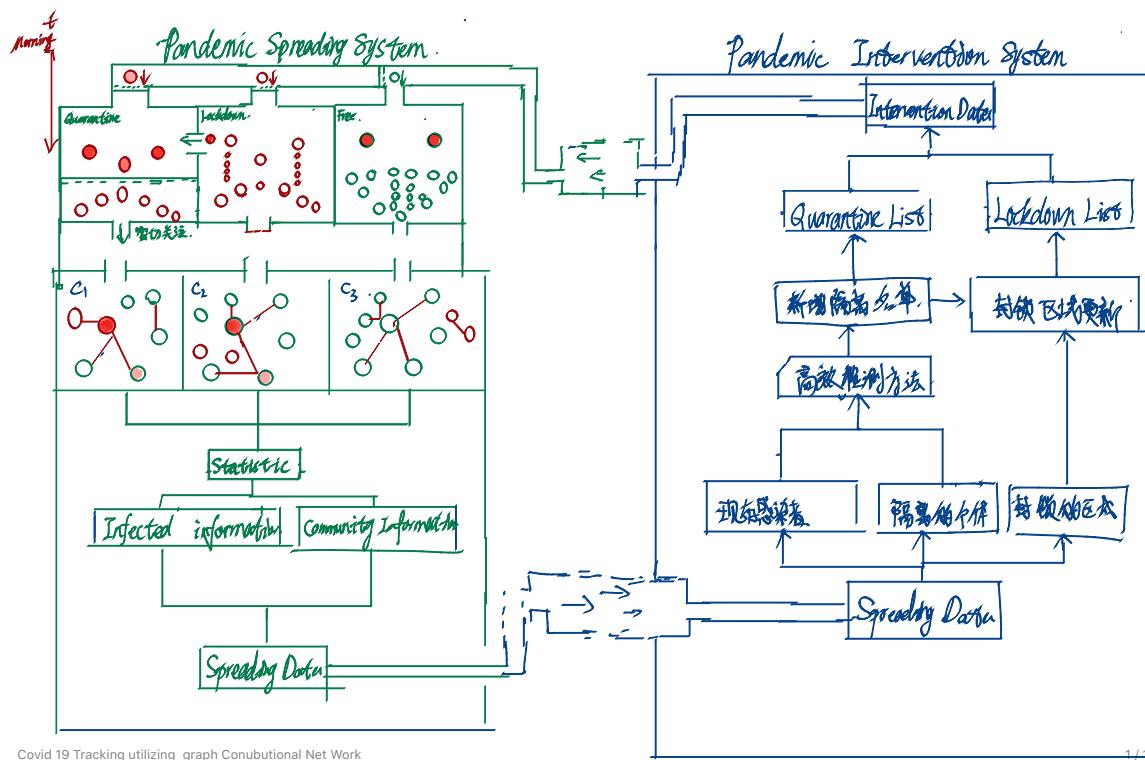


Figure 2.1: 整个系统包括病毒传播系统(Pandemic Spreading System,PSS)和病毒干预措施系统(Pandemic Intervention System,PIS)，在PSS中，红色的节点代表已确诊，红白的节点代表密切关注者，而绿白的节点代表易感人群和潜伏者。此外，流程图中设定了1个隔离区，1个封锁区，4个自由区作为示例。在隔离区，以虚实线分为必须隔离区和非必须隔离区，已确诊的节点隔离在必须隔离区，密切关注者则的行动与绿色的节点无差异，但必须采取戴口罩的措施，否则不能进入自由区。在自由区，节点可以根据偏好随意活动，只能跟同一区域内的节点接触。接着，每天统计感染者信息和所在的社区信息传输到PIS系统。收到了PSS传输的数据后，PIS根据现存的感染者和隔离的个体的状态信息，统计全部在自由区的有症状感染者所在的区域，对这些区域中所有的节点都进行检测并隔离，对于检测到为感染者的则安排到必须隔离区，而其他的节点则安排到非必须隔离区。接着，如果实行区域封锁，则将感染者所在区域进行全面封锁，除了已确诊的节点可以通往必须隔离区，封锁区域中的节点不能到其他的区域直到封锁期结束。

数据介绍与动态网络构建

3 数据介绍 17

- 3.1 Gowalla Data
- 3.2 Foursquare_small Data

4 动态接触网络构建 19

- 4.1 区域数据划分



3. 数据介绍

3.1 Gowalla Data

3.2 Foursquare_small Data

该数据集包括2012年4月12日至2013年2月16日从Foursquare收集的纽约市和东京的长期（约10个月）入住数据。其中文件名为dataset_TSMC2014_NYC.txt是纽约市的签入数据，一共227428行，文件名为dataset_TSMC2014_TKY.txt是东京的签入数据，一共573703行。地区偏移量的该地区（纽约市或东京），与UTC时间相差多少，单位为分钟，如-240，既该地区的当前时间等于UTC time的时间减去240分钟。

数据字段	description	数据类型
User ID	用户ID	int
Venue ID	场地ID	str
Venue category ID	场地类别ID	str
Venue category name	场地类别	str
Latitude	纬度	int
Longitude	经度	int
Timezone offset in minutes	时区偏移量（以分钟为单位）	int
UTC time	世界标准时间	str



4. 动态接触网络构建

4.1 区域数据划分

4.1.1 判断该点是否在New York城市内

在实际应用中，我们经常需要判断某个经纬度是否在某个区域内，而这个在程序中需要该区域的边界经纬度。

获取区域边界

我们可以从网站里获取经纬度：<https://nominatim.openstreetmap.org/ui/search.html>。在此网站内输入你想判断的区域名称，如，New York，点击Serach后再点击detail，复制该区域的OSM码。

Nový York (old_name:cs) Nuova York (old_name:it) NYC (short_name)	
Type	boundary:administrative
Last Updated	2021-06-14T13:01:23+00:00
Admin Level	5
Search Rank	10
Address Rank	16 (city / town / village)
Importance	0.8175766114518461
Coverage	Polygon
Centre Point (lat,lon)	40.7127281,-74.0060152
OSM	relation 75908
Place Id	259048015 (on this server)
Wikipedia	en:New_York_City
Calculated	
Computed	
Postcode	
Address Tags	剪切(Alt + A)

Figure 4.1: OSM 码

在此网站输入OSM码<http://polygons.openstreetmap.fr/index.py>，点击提交后就可以下载了。

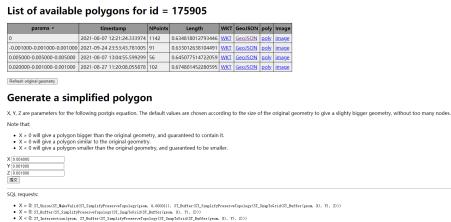


Figure 4.2: 下载界面

后续可以写一个方法，输入区域的名称后即可将该区域边界提取。

判断方法

由于New York城市是由多个点组成的一个不规则多边形，我们写了一个方法，用于判断某个经纬度坐标是否在New York这个城市里。此方法为射线法，步骤如下：

- 1) 从需要判断的点P向右引水平扫描线(即射线)。
- 2) 计算此射线与多边形的交点个数c。
- 3) 如果c为奇数，认为点P在多边形内；为偶数，则点P在多边形外。

如下图(a),c为奇数，所以P点在多边形外，图(b), c为偶数，所以P点在多边形内。

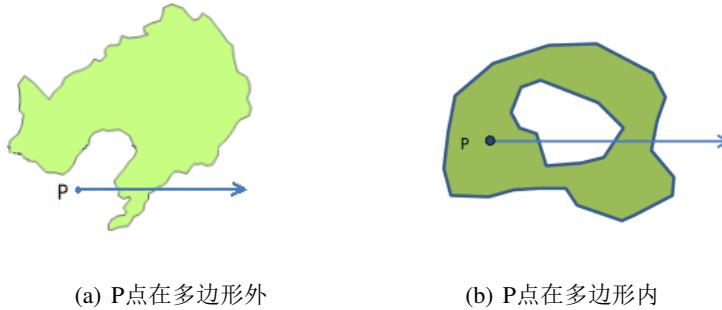


Figure 4.3: 射线法

在我们的数据集里，New York城市的边界是由很多个点（经纬度）组成的，我们可以把这些点连起来，就形成了一个多边形，再使用射线法去判断。

我们的将此方法转化为程序的主要思路为：

- 1) 我们规定，点P为需要判断的点， x_p, y_p 分别为P的经度和纬度。 $N = \{P_n = (x_n, y_n) | n \in [0, k]\}$ 为New York城市的边界经纬度集合，其中k为New York城市经纬度的点的数量。
- 2) 按顺序将New York城市的每两个点连起来（如 P_1, P_2 或 $P_2, P_3 \dots$ ），得到线段AB
- 3) 每次都用点P向右作射线，判断点P是否穿过线段AB
- 4) 如穿过，c加一，反之，c保持不变。
- 5) 直至后面将New York城市的所有点都判断后，再判断c的奇偶性，如c为奇数，则说明点P在多边形内，反之，点P在多边形外。

由于射线法本身存在缺陷，且在转化成代码时也会有相应的问题，我们在使用射线法前先

进行了初步的判断以及排除一些情况：

- 1) 判断点P是否属于N，如是，则说明P在New York城市内。
- 2) 判断点P是否属于线段AB上，如是，则说明P在New York城市内。（伪代码是否要给出来？）
- 3) 点P正好位于线段AB的下端点，则在射线法里跳过此线段的判断。（怎么描述能更清楚？）
- 4) 线段在射线的上方、线段在射线的下方、线段在射线的左方，则在射线法里跳过此线段的判断。

公式如下：

$$isInRegion(N, P) = \begin{cases} in, & 2 \nmid \sum_{n=1}^k M(P_n, P_{n+1}, P) \\ out, & 2 \parallel \sum_{n=1}^k M(P_n, P_{n+1}, P) \end{cases}$$

其中，k为New York城市经纬度的点的数量， $M(P_n, P_{n+1}, P)$ 的作用为计算每一条线段与P是否有交点，公式在此给出：

$$M(P_n, P_{n+1}, P) = \begin{cases} 0, & x_{meet} < x_p \\ 1, & x_{meet} > x_p \end{cases}$$

$$x_{meet} = x_n - \frac{x_n - x_{n+1}}{y_n - y_{n+1}}(y_n - y_p)$$

其中， x_{meet} 为P点的射线与线段AB的交点的横坐标，其原理为三角形的相似性，原公式如下：

$$\frac{x_n - x_{n+1}}{y_n - y_{n+1}} = \frac{x_n - x_{meet}}{y_n - y_p}$$

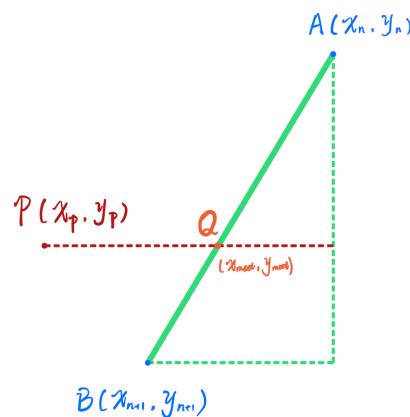


Figure 4.4: similar triangles

在上面的公式里， $isInRegion(N, P)$ 中，如果 $\sum_{n=1}^k M(P_n, P_{n+1}, P)$ 能被2整除，说明c为偶数，所以我们通过射线法可以知道点P在该区域内，反之，则在该区域外。

干预措施



5	邻接检测	27
6	区域检测	29
6.1	数据分析	
7	基于历史轨迹的区域检测	31
8	基于图神经网络的区域检测推荐	33



5. 邻接检测



6. 区域检测

6.1 数据分析

6.1.1 个体不带口罩，政府执行区域检测隔离措施

在个体不带口罩，政府执行区域检测隔离措施的情况下，我们对各个状态(SEAIR)进行分析，画出了下列图片，其中，Y轴为某状态当前人数/总人数，X轴为iteration，即迭代的次数。

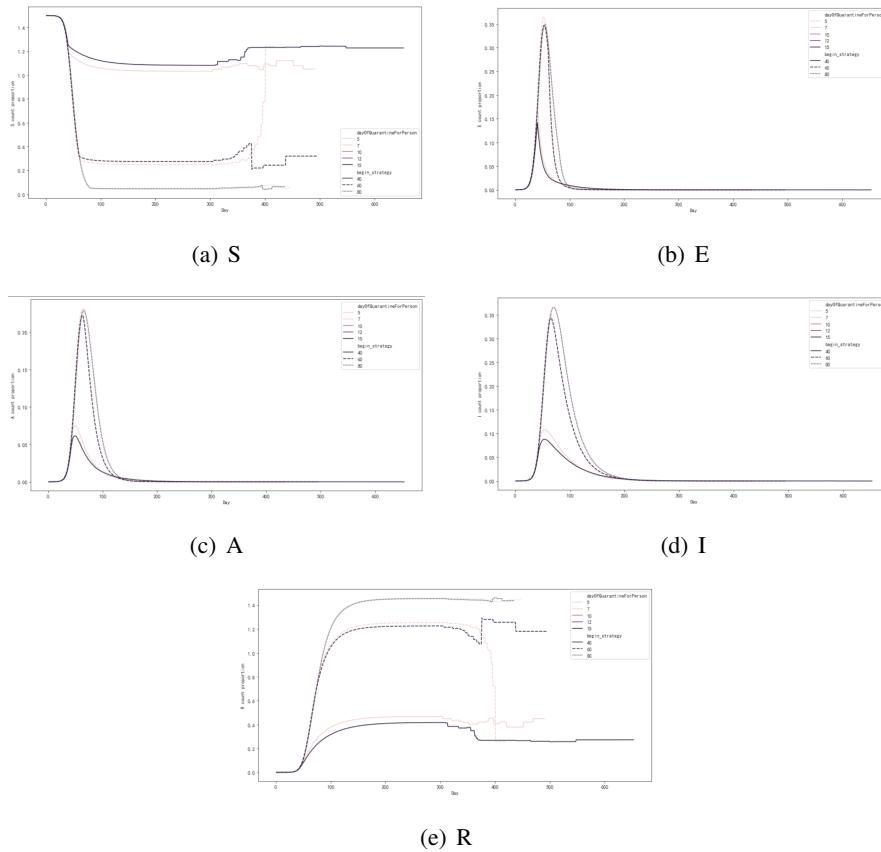


Figure 6.1: 各个状态人数占总人数

从图中可以看出，处于S状态和R状态图可以明显看到在iteration>300时，曲线的变化不像以往缓慢，S占比开始上升，而R开始下降，而在iteration>400时，S和R都开始出现垂直上升或下降或者暂缓不变的现象。于是，我们将S和R的占比画在同一幅图里，以观察他们的之间的关系。

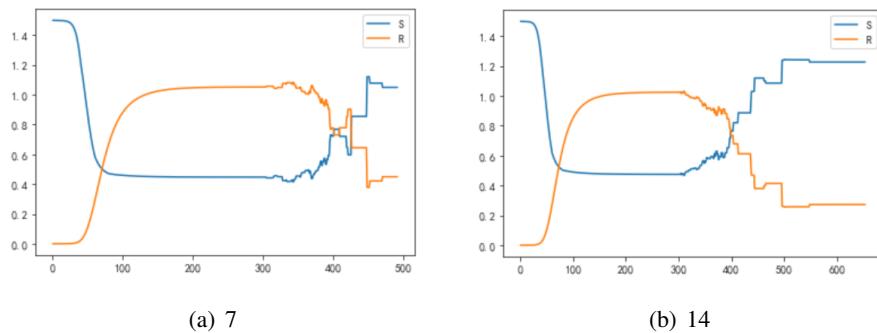


Figure 6.2: 处于S状态和R状态占比分析图

可以看到，S状态与R状态在dayOfQuarantineForPerson=7或dayOfQuarantineForPerson=14时都表现出对称性，目前未发现出现此现象的原因。



7. 基于历史轨迹的区域检测



8. 基于图神经网络的区域检测推荐

V

实验结果

9 模型参数设置 37

10 结果分析 39



9. 模型参数设置



10. 结果分析

