

一种考虑时间因素的 COVID-19 SIR 模型, 包含无法检测到的感染者

黄雨, 张冉冉, 窦晟霆, 苏畅, 刘博, 何佳铭

(重庆邮电大学 通信与信息工程学院, 重庆 400065)

摘要: 本文对 COVID-19 进行了数学和数值分析。为了预测 COVID-19 的趋势, 我们提出了一个随时间变化的 SIR 模型, 该模型跟踪 t 时刻的传播率和恢复率。利用中国权威机构提供的数据, 我们发现一天的预测误差几乎小于 3%。根据我们的模型, 可以预测中国的转折点和确诊病例总数。为了分析无法检测到的感染对疾病传播的影响, 我们扩展了模型, 考虑了两类感染者: 可检测到的感染者和无法检测到的感染者。疫情是否爆发由一个 2×2 矩阵的谱半径来表征。如果 $R_0 > 1$, 则该矩阵的光谱半径大于 1, 存在疫情爆发。同时绘制了疫情爆发的相变图, 并显示在 2020 年 3 月 2 日, 有几个国家处于 COVID-19 疫情爆发的边缘。为了说明社会距离的有效性, 我们分析了配置随机网络中疾病传播的独立级联模型。展示了两种拉开社会距离的方法, 它们可以降低有效繁殖数 R_e 。

关键词: COVID-19、SARS-CoV-2、冠状病毒、时间依赖性 SIR 模型、检测不到的感染、群体免疫、超级传播者、独立级联、社会距离。

中图分类号: TN929

文献标志码: A

A Time-Dependent SIR Model for COVID-19

With Undetectable Infected Persons

Yu Huang, Ranran Zhang, Shengting Dou, Chang Su, Bo Liu, Jiaming He

(School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, P. R. China)

Abstract: In this paper, we conduct mathematical and numerical analyses for COVID-19. To predict the trend of COVID-19, we propose a time-dependent SIR model that tracks the transmission and recovering rate at time t . Using the data provided by China authority, we show our one-day prediction errors are almost less than 3%. The turning point and the total number of confirmed cases in China are predicted under our model. To analyze the impact of the undetectable infections on the spread of disease, we extend our model by considering two types of infected persons: detectable and undetectable infected persons. Whether there is an outbreak is characterized by the spectral radius of a 2×2 matrix. If $R_0 > 1$, then the spectral radius of that matrix is greater than 1, and there is an outbreak. We plot the phase transition diagram of an outbreak and show that there are several countries on the verge of COVID-19 outbreaks on Mar. 2, 2020. To illustrate the effectiveness of social distancing, we analyze the independent cascade model for disease propagation in a configuration random network. We show two approaches of social distancing that can lead to a reduction of the effective reproduction number R_e .

Keywords: COVID-19, SARS-CoV-2, Coronavirus, time-dependent SIR model, undetectable infection, herd immunity, superspreader, independent cascade, social distancing.

0 引言

2019 年 12 月初, 中国武汉首次确诊了冠状病毒感染者, 随后几周, 这种疾病在中国大陆和其他国家广泛传播, 引发了全球恐慌。该病毒被命名为“SARS-CoV-2”, 引发的疾病被命名为“2019 冠状病毒病 (简称“COVID-19”)”。根

据中国政府的官方声明, 截至 2020 年 3 月 2 日, 已有 80151 人感染该疾病, 2943 人死亡。为阻止病毒的传播, 中国和其他国家的政府采取了一些策略, 如城市封锁、交通停运、社区管理、社交距离和健康教育知识的宣传。

与严重急性呼吸系统综合症 (SARS) 和其他传染病不同, COVID-19 的一个问题特点是存在无

症状感染（症状非常轻微）。这些无症状感染者不知道自己的传染能力，因此在被发现之前就会让更多人感染[1]。在这种情况下，传播率会急剧上升。根据世界卫生组织最近的报告[2]，COVID-19 患者中只有 87.9% 发烧，67.7% 干咳。如果以体温作为检测 COVID-19 感染病例的手段，则有超过 10% 的感染者无法被检测出来。

由于最近疫情的发展，我们有兴趣为 COVID-19 解决以下重要问题：

Q1) 是否有可能遏制 COVID-19？全城封锁、交通管制、社区管理、健康教育知识宣传等常用措施能否有效遏制 COVID-19？

Q2) 如果 COVID-19 能够得到控制，疫情的高峰期是什么时候？又将在何时结束？

Q3) 无法检测到的感染对疾病传播有何影响？

Q4) 如果 COVID-19 无法控制，那么需要感染人群的比例是多少才能实现群体免疫？

Q5) 减少人际交往、取消群众集会等疏远社会的方法对控制 COVID-19 的效果如何？

对于 (Q1)，我们在本文中分析了中国的病例，旨在预测病毒的传播方式。具体而言，我们建议使用与时间相关的易感-感染-康复 (SIR) 模型来分析和预测感染人数和康复人数（包括死亡人数）。在传统的 SIR 模型中，有两个时间不变变量：传播率 β 和恢复率 γ 。传播率 β 表示每个人在单位时间内平均与随机选择的其他人有 β 次接触。传统的 SIR 模型忽略了 β 和 γ 的时变特性，过于简单，无法精确有效地预测疾病的趋势。因此，我们建议使用与时间相关的 SIR 模型，其中传播率 β 和康复率 γ 都是时间 t 的函数。我们的想法是使用机器学习方法跟踪传播率 $\beta(t)$ 和康复率 $\gamma(t)$ ，然后利用它们预测未来某一时间 t 的感染人数和康复人数。我们的随时间变化的 SIR 模型可以动态调整关键参数，如 $\beta(t)$ 和 $\gamma(t)$ ，以适应控制策略的变化，这与文献中现有 SIR 和 SEIR 模型（如 [3]、[4]、[5]、[6]、[7] 和 [8]）不同。例如，我们观察到，从我们的模型来看，全城封锁会大大降低传播率。大多数用于预测 COVID-19 的数据驱动和曲线拟合方法，如 [9]、[10] 和 [11]，似乎都能完美地跟踪数据；但它们缺乏对疾病传播的物理洞察力。此外，它们对 2020 年 2 月 12 日湖北省确诊病例定义的突然变化非常敏感。另一方面，我们的时变 SIR 模型可以检验中国政府的疫情控制政策，并提供合理的解释。利用中华人民共和国国家卫生健康委员会 (NHC)

提供的数据[12]，我们发现，除了 2020 年 2 月 12 日由于确诊病例定义的改变而无法预测外，确诊病例数的单日预测误差几乎小于 3%。

对于 (Q2)，基本繁殖数 R_0 定义为在感染期间，一个典型感染者在完全易感人群中恢复之前的预期额外感染（继发病例）数[13]，[14]，是检验疾病是否会暴发以及需要对多大比例的人群接种疫苗才能根除疾病的常用指标之一。事实上，在任何特定时间，不同比例的人群对任何特定疾病都具有免疫力。为此，有效繁殖数 $R_e(t)$ 被用来量化疾病在部分易感人群中的瞬时传播[14]。提前了解 $R_0(t)$ 和 $R_e(t)$ 有助于政府制定更准确的防疫政策。在经典的 SIR 模型中， R_0 简单地定义为 β/γ ，因为一个感染者平均需要 $1/\gamma$ 天才能康复，在此期间，他（平均）会接触到 β 个人。在我们的时变 SIR 模型中，时间 t 的有效繁殖数定义为 $\beta(t)/\gamma(t)$ ，用 $R_e(t)$ 表示。如果 $R_e(t) > 1$ ，疾病将以指数形式传播，并感染总人口的一部分。相反，疾病最终会被控制住。因此，通过观察 $R_e(t)$ 随时间的变化，甚至预测未来的 $R_e(t)$ ，我们可以检验某些流行病控制政策是否有效。根据中华人民共和国国家卫生健康委员会 (NHC) 提供的数据[12]，我们预测 2020 年 2 月 17 日为转折点（峰值），即有效繁殖数小于 1 的那一天。此外，在我们的（确定性）模型中，如果中国维持当前的传染病控制政策，疫情将在高峰期过后约 6 周结束。在这种情况下，根据我们的（确定性）模型预测，中国的确诊病例总数约为 80,000 例。

对于 (Q3)，我们将 SIR 模型扩展为两类感染者：可检测到的感染者（类型 1）和不可检测到的感染者（类型 2）。感染者属于 1 型（或 2 型）

的概率为 w_1 （或 w_2 ），其中 $w_1 + w_2 = 1$ 。1 型

（或 2 型）感染者的传播率为 β_1 （或 β_2 ），恢复率为 γ_1 （或 γ_2 ）。该模型的基本繁殖数为：

$$R_0 = w_1 \frac{\beta_1}{\gamma_1} + w_2 \frac{\beta_2}{\gamma_2}. \quad (1)$$

实际上，1 型感染者的传播率低于 2 型感染者（因为 1 型感染者可以被隔离）。在这种模式下，疾病是否可控取决于 2×2 矩阵的谱半径。如果该矩阵的频谱半径大于 1，则说明疾病爆发。反之，如果它小于 1，则不会爆发。一个有趣的结论是，如果 (1) 中的基本繁殖数 R_0 大于（或小于）1，则该矩阵的频谱半径大于（或小于）1。利用约翰霍普金斯大学 GitHub [16] 中 2020 年 1 月 22

日至 2020 年 3 月 2 日的历史数据,我们将研究扩展到其他一些国家,包括日本、新加坡、韩国、意大利和伊朗。我们的数值结果表明,包括韩国、意大利和伊朗在内的几个国家都在渗透阈值曲线之上,它们即将在 2020 年 3 月 2 日爆发 COVID-19。与文献中的其他时间依赖性流行病模型(如著名的 tsiR 模型[17]、[18]和最近的《科学》文章[19])相比,我们的模型进一步考虑了无法检测到的感染者的影响(另见[19]中对我们的论文[20]预印版的评论)。Oliver Wyman(全球知名的管理咨询公司之一)在其 COVID-19 Pandemic Navigator Core Model [21]中将我们的时变 SIR 模型的预印本[20]作为两个主要参考文献之一。他们还在白皮书中纳入了政府应对行动和 Google COVID-19 社区流动性报告,用于分析无法检测到的感染者的影响。这显示了进一步扩展我们的模型以获得更准确预测的潜力。

英国首相鲍里斯-约翰逊(Boris Johnson)曾建议,感染 COVID-19 并痊愈的人数比例要足够高,以实现群体免疫。为了解决(Q4)中的问题,我们认为群体免疫与 SIR 模型中易感人群数量的减少相对应,在至少有 $1 - \frac{1}{R_0}$ 的人感染 COVID-19 并康复后,就可以实现群体免疫。

对于问题 5,我们考虑了两种常用的拉开社交距离的方法:(i) 允许每个人保持的人际交往不超过其正常交往的一小部分,以及(ii) 取消群众聚会。为了分析社交距离,我们必须考虑到社会网络(及其网络结构)。为此,我们考虑了疾病在随机网络中传播的独立级联(IC)模型,该网络基于度分布 $p_k, k=0,1,2,\dots$ 。IC 模型已被广泛用于研究病毒营销中的影响力最大化问题(例如,[22])。在 IC 模型中,一个受感染的节点可以以一定的传播概率将疾病传播给邻近的易感节点(通过边)。如此反复传播,我们就会得到一个长期包含受感染节点集合的子图。通过将 IC 模型中的传播概率与 SIR 模型中的传播率和恢复率联系起来,我们展示了社会距离的两种结果:(i) 对于允许每个人保持的人际接触不超过(平均)其正常接触的 a 部分的社会距离方法,有效繁殖数减少了 a^2 倍;(ii) 对于通过移除边数大于或等于 k_0 的节点来取消大规模聚集的社会距离方法,有效繁殖数减少了 $\sum_{k=0}^{K_0-2} k q_k / \sum_{k=0}^{\infty} k^2 q_k$ 倍,其中 q_k 是 p_k 的多余度分布。

本文接下来的内容安排如下:在第二节中,我们提出了随时间变化的 SIR 模型。然后,我们在第三节中将该模型扩展检测不到感染者的 SIR

模型。在第四节中,我们考虑了在一个由度分布指定的随机网络中疾病传播的独立级联模型。在第五部分,我们进行了几个数值实验来说明我们模型的有效性。在第六部分,我们提出了控制 COVID-19 的一些讨论和建议。本文在第七节中结束。

1 随时间变化的 SIR 模型

1.1 易感-感染-恢复(SIR)模式

在典型的传染病数学模型中,人们通常将病毒与宿主的相互作用和流行病的演变简化为几种基本的疾病状态。其中一个最简单的流行病模型被称为易感-感染-恢复(SIR)模型[15],包括三种状态:易感状态、感染状态和恢复状态。处于易感状态的个体是指在 t 时间尚未感染疾病的个体,但如果与感染疾病的人接触,则有可能被感染。感染状态指的是在 t 时间患有疾病的个体,可能会感染易感个体(如果他们相互接触)。痊愈状态指的是在 t 时间,患者已经痊愈或死亡,不再具有传染性。之所以将死亡人数计入康复状态,是因为从流行病学的角度来看,这基本上是一回事,无论康复还是死亡都不会对疾病的传播产生太大影响。因此,可以有效地将其从疾病的潜在宿主中排除[23]。用 $S(t)$ 、 $X(t)$ 和 $R(t)$ 表示时间 t 的易感人数、感染人数和康复人数。总结上述 SIR 模型,我们认为它与 COVID-19 疫情非常相似,本文将采用 SIR 模型作为基本模型。

在传统的 SIR 模型中,有两个时间不变变量:传播率 β 和恢复率 γ 。传播率 β 表示每个个体在单位时间内平均与随机选择的其他人有 β 次接触。传统的 SIR 模型忽略了 β 和 γ 的时变特性,这种假设过于简单,无法精确有效地预测疾病的趋势。因此,我们提出了随时间变化的 SIR 模型,其中传播率 β 和恢复率 γ 都是时间 t 的函数。

1.2 时变 SIR 模型的微分方程

在传统的 SIR 模型中, $S(t)$ 、 $X(t)$ 和 $R(t)$ 这三个变量受以下微分方程控制(参考书[15]):

$$\begin{aligned}\frac{dS(t)}{dt} &= \frac{-\beta S(t)X(t)}{n}, \\ \frac{dX(t)}{dt} &= \frac{\beta S(t)X(t)}{n} - \gamma X(t), \\ \frac{dR(t)}{dt} &= \gamma X(t).\end{aligned}$$

我们注意到

$$S(t) + X(t) + R(t) = n, \quad (2)$$

其中 n 为总人口。设 $\beta(t)$ 和 $\gamma(t)$ 为 t 时刻的传播率和恢复率。用 $\beta(t)$ 和 $\gamma(t)$ 代替上述微分方程中的 β 和 γ ，可得

$$\frac{dS(t)}{dt} = \frac{-\beta(t)S(t)X(t)}{n}, \quad (3)$$

$$\frac{dX(t)}{dt} = \frac{\beta(t)S(t)X(t)}{n} - \gamma(t)X(t), \quad (4)$$

$$\frac{dR(t)}{dt} = \gamma(t)X(t). \quad (5)$$

三个变量 $S(t)$ 、 $X(t)$ 和 $R(t)$ 仍然满足 (2)。

现在我们简要解释一下这三个等式的直观含义。如果我们假设总人口为 n ，那么随机选择的人处于易感状态的概率为 $S(t)/n$ 。因此，一个处于感染状态的人在单位时间内（平均）会接触 $\beta(t)S(t)/n$ 个处于易感状态的人，这意味着新感染者的数量为 $\beta(t)S(t)X(t)/n$ （因为在时间 t 时有 $X(t)$ 人处于感染状态）。相反，处于易感状态的人数将减少 $\beta(t)S(t)X(t)/n$ 。此外，由于处于感染状态的每个个体的康复率为 $\gamma(t)$ ，因此在时间 t 时 $\gamma(t)X(t)$ 人康复。在公式 (5) 中说明了时间在 t 时 $R(t)$ 的差异。由于 $S(t)$ 、 $X(t)$ 和 $R(t)$ 这三个变量仍然满足 (2)，我们可以得到

$$\frac{dX(t)}{dt} = -\left(\frac{dS(t)}{dt} + \frac{dR(t)}{dt}\right),$$

即从易感状态变为感染状态的人数减去从感染状态变为康复状态的人数（见 (4)）。

1.3 离散时间时间相关 SIR 模型

由于 COVID-19 数据按天更新[12]，我们将 (3)、(4) 和 (5) 中的微分方程修正为离散时差方程：

$$S(t+1) - S(t) = \frac{-\beta(t)S(t)X(t)}{n}, \quad (6)$$

$$X(t+1) - X(t) = \frac{\beta(t)S(t)X(t)}{n} - \gamma(t)X(t), \quad (7)$$

$$R(t+1) - R(t) = \gamma(t)X(t). \quad (8)$$

同样，三个变量 $S(t)$ 、 $X(t)$ 和 $R(t)$ 仍然满足 (2)。

在疾病传播初期，确诊病例数量很少，大部分人群处于易感状态。因此，在分析 COVID-19 的

初始阶段时，我们假设 $\{S(t) \approx n, t \geq 0\}$ ，并进一步将 (7) 简化如下：

$$X(t+1) - X(t) = \beta(t)X(t) - \gamma(t)X(t). \quad (9)$$

根据上述差分方程，我们很容易得出每天的 $\beta(t)$ 和 $\gamma(t)$ 。根据 (8)，我们可以得出

$$\gamma(t) = \frac{R(t+1) - R(t)}{X(t)}. \quad (10)$$

在 (9) 中代入 (8) 可以得出

$$\beta(t) = \frac{[X(t+1) - X(t)] + [R(t+1) - R(t)]}{X(t)}. \quad (11)$$

根据某一时期的历史数据 $\{X(t), R(t), 0 \leq t \leq T-1\}$ ，我们可以使用公式 (10) 和 (11) 得到相应的 $\{\beta(t), \gamma(t), 0 \leq t \leq T-2\}$ 。利用上述信息，我们可以使用机器学习方法预测随时间变化的传播率和恢复率。

1.4 通过岭回归跟踪传输速率 $\beta(t)$ 和恢复速率 $\gamma(t)$

在本小节中，我们通过线性系统中常用的有限脉冲响应（FIR）滤波器来跟踪和预测 $\beta(t)$ 和 $\gamma(t)$ 。用 $\beta(t)$ 和 $\gamma(t)$ 表示预测的传输速率和恢复速率。根据 FIR 滤波器，它们的预测结果如下：

$$\begin{aligned} \hat{\beta}(t) &= a_1\beta(t-1) + a_2\beta(t-2) + \dots + a_J\beta(t-J) + a_0 \\ &= \sum_{j=1}^J a_j\beta(t-j) + a_0, \end{aligned} \quad (12)$$

$$\begin{aligned} \hat{\gamma}(t) &= b_1\gamma(t-1) + b_2\gamma(t-2) + \dots + b_K\gamma(t-K) + b_0 \\ &= \sum_{k=1}^K b_k\gamma(t-k) + b_0, \end{aligned} \quad (13)$$

其中， J 和 K 是两个 FIR 滤波器的阶数（ $0 < J, K < T-2$ ）， $a_j, j=0,1,\dots,J$ 和 $b_k, k=0,1,\dots,K$ 是这两个 FIR 滤波器的脉冲响应系数）。

有几种广泛使用的机器学习方法可用于估计 FIR 滤波器的脉冲响应系数，如普通最小二乘法（OLS）、正则化最小二乘法（即岭回归）和偏最小二乘法（PLS）[24]。在本文中，我们选择岭回归作为估计方法，它可以解决以下优化问题：

$$\min_{a_j} \sum_{t=J}^{T-2} (\beta(t) - \hat{\beta}(t))^2 + \alpha_1 \sum_{j=0}^J a_j^2, \quad (14)$$

$$\min_{b_k} \sum_{t=K}^{T-2} (\gamma(t) - \hat{\gamma}(t))^2 + \alpha_2 \sum_{k=0}^K b_k^2, \quad (15)$$

其中 α_1 和 α_2 为正则化参数。

1.5 跟踪时间相关 SIR 模型的感染人数 $X(t)$ 和康复人数 $R(t)$

在本小节中,我们将展示如何在随时间变化的 SIR 模型中使用两个 FIR 滤波器来跟踪和预测感染人数和康复人数。给定一段历史数据 $\{X(t), R(t), 0 \leq t \leq T-1\}$, 我们首先通过(10)和(11)得出 $\{\beta(t), \gamma(t), 0 \leq t \leq T-2\}$ 。然后,我们求解岭回归(目标函数为(14)和(15),约束条件为(12)和(13))来学习 FIR 滤波器的系数,即 $a_j, j=0,1,\dots,J$ 和 $b_k, k=0,1,\dots,K$ 。学习到这些系数后,我们就可以通过(12)和(13)中训练好的岭回归预测 $t=T-1$ 时刻的 $\beta(t)$ 和 $\gamma(t)$ 。

要预测 $t=T$ 时的 $X(t)$ 和 $R(t)$, 我们只需用(8)和(9)中的 $\beta(t)$ 和 $\gamma(t)$ 替换 $\beta(t)$ 和 $\gamma(t)$ 。由此得出:

$$\hat{X}(T) = (1 + \hat{\beta}(T-1) - \hat{\gamma}(T-1))X(T-1), \quad (16)$$

$$\hat{R}(T) = R(T-1) + \hat{\gamma}(T-1)X(T-1). \quad (17)$$

为了预测 $t > T$ 时的 $X(t)$ 和 $R(t)$, 我们利用(12)和(13)估算 $\beta(t)$ 和 $\gamma(t)$ 。与(16)和(17)类似,我们预测 $X(t)$ 和 $R(t)$ 如下:

$$\hat{X}(t+1) = (1 + \hat{\beta}(t) - \hat{\gamma}(t))\hat{X}(t), t \geq T, \quad (18)$$

$$\hat{R}(t+1) = \hat{R}(t) + \hat{\gamma}(t)\hat{X}(t), t \geq T. \quad (19)$$

算法 1 概述了我们的跟踪/预测方法的详细步骤。

算法 1: 跟踪离散时间时变 SIR 模型

输入: $\{X(t), R(t), 0 \leq t \leq T-1\}$, 正则化参数 α_1 和 α_2 , FIR 滤波器的阶数 J 和 K , 预测窗口 W

输出: $\{\beta(t), \gamma(t), 0 \leq t \leq T-2\}$, $\{\beta(t), \gamma(t), t \geq T-1\}$, 以及 $\{X(t), R(t), t \geq T\}$ 。

1. 分别使用公式(11)和(12), 计算出 $\{\beta(t), \gamma(t), 0 \leq t \leq T-2\}$

2. 使用(14)和(15)训练岭回归。

3. 分别通过(12)和(13)估计 $\beta(T-1)$ 和 $\gamma(T-1)$ 。

4. 分别使用(16)和(17)估计第二天感染人数 $X(T)$ 和康复人数 $R(T)$ 。

5. While $T \leq t \leq T+W$ do

6. 分别估计(12)和(13)中的 $\beta(t)$ 和 $\gamma(t)$ 。

7. 分别使用(18)和(19)预测 $X(t+1)$ 和 $R(t+1)$ 。

8. End while

我们注意到,这种确定性流行病模型是基于 $X(t)$ 和 $R(t)$ 的平均场近似值。这种近似是大数定律的结果。因此,当 $X(t)$ 和 $R(t)$ 相对较小时,平均值场近似可能没有预期的那么精确。在这种情况下,我们可能不得不采用随机流行病模型,如马尔可夫链。

2 检测不到感染者的 SIR 模型

根据世界卫生组织的最新报告[2], COVID-19 患者中只有 87.9% 发烧, 67.7% 干咳。这说明存在无症状感染。最近的研究[7]和[25]也指出了 COVID-19 无症状携带者的存在。这些人没有意识到自己的传染性,从而使更多的人受到感染。在这种情况下,传播率会急剧上升。如果没有大规模检测等方法来检测这些无症状感染者,这些人将无法被检测到。

本文中的无法检测到的感染者是指在现行防疫政策下尚未检测到或检测不到的感染者。为了将无法检测到的感染者考虑在内,我们在本节中提出了包含无法检测到的感染者的 SIR 模型。我们假设有两类感染者。可检测到(有明显症状)的感染者被归类为1型感染者,无症状且无法检测到的感染者被归类为2型感染者。对于一个感染者来说,其成为1型感染者的概率为 w_1 , 成为2型感染者的概率为 w_2 , 其中 $w_2 + w_1 = 1$ 。此外,这两类感染者根据是否接受治疗或隔离,其传播率和康复率也不同。我们将 $\beta_1(t)$ 和 $\gamma_1(t)$ 分别表示第1类感染者在第 t 时间的传播率和康复率; 同样, $\beta_2(t)$ 和

$\gamma_2(t)$ 分别表示第 2 类感染者在第 t 时间的传播率和康复率。

2.1 检测不到感染者的 SIR 模型的控制方程

现在我们来推导有两类感染者的 SIR 模型的控制方程。让 $X_1(t)$ 或 $(X_2(t))$ 为时间 t 时的第一类（或第二类）感染者人数。与前面小节中 (7)、(8) 的推导类似，我们假设 $\{S(t) \approx n, t \geq 0\}$ 在流行病的初始阶段，并将 $X(t)$ 分成两类感染者。我们有以下差分方程：

$$X_1(t+1) - X_1(t) = \beta_1 X_1(t) w_1 + \beta_2 X_2(t) w_1 - \gamma_1 X_1(t), \quad (20)$$

$$X_2(t+1) - X_2(t) = \beta_1 X_1(t) w_2 + \beta_2 X_2(t) w_2 - \gamma_2 X_2(t), \quad (21)$$

$$R(t+1) - R(t) = \gamma_1 X_1(t) + \gamma_2 X_2(t), \quad (22)$$

其中， β_1 、 β_2 、 γ_1 和 γ_2 为常数。值得注意的是，这些常量也可以是随时间变化的，正如我们在之前中所做的那样。不过，在本节中，我们将它们设为常量，以关注无法检测到的感染者的影响。将 (20) 和 (21) 以矩阵形式重写，可得到以下矩阵方程：

$$\begin{bmatrix} X_1(t+1) \\ X_2(t+1) \end{bmatrix} = \begin{bmatrix} 1 + \beta_1 w_1 - \gamma_1 & \beta_2 w_1 \\ \beta_1 w_2 & 1 + \beta_2 w_2 - \gamma_2 \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix},$$

其中， $w_2 = 1 - w_1$ 。设 \mathbf{A} 为上述系统方程的转移矩阵，即

$$\mathbf{A} = \begin{bmatrix} 1 + \beta_1 w_1 - \gamma_1 & \beta_2 w_1 \\ \beta_1 w_2 & 1 + \beta_2 w_2 - \gamma_2 \end{bmatrix}. \quad (23)$$

众所周知（从线性代数中），如果 \mathbf{A} 的谱半径（特征值的最大绝对值）小于 1，那么这样的系统就是稳定的。换句话说，当 t 变为无穷大时， $X_1(t+1)$ 和 $X_2(t+1)$ 将逐渐收敛为有限常数。在这种情况下，就不会爆发。相反，如果频谱半径大于 1，就会爆发疫情，受感染的人数将随时间 t 呈指数增长（以频谱半径的速率）。

此外， w_1 和 w_2 还可以进一步写成随时间变化的形式。如果假设 $\gamma_1 = \gamma_2 = \gamma$ ，根据 (20) 和 (21)，我们可以得出以下表达式

$$X(t+1) = (1 + \beta_1 w_1(t) + \beta_2(1 - w_1(t)) - \gamma) X(t). \quad (24)$$

根据 (24)，如果假设 β_1 、 β_2 和 γ 在一段时间内是常数，我们可以测量 w_1 随时间的变化，即

$$w_1(t) = \frac{\frac{X(t+1)}{X(t)} - \beta_2 + \gamma - 1}{\beta_1 - \beta_2}. \quad (25)$$

由于疫情的进展， w_1 与 w_2 的比值会随着政府的防疫政策而发生变化。通常，在新的防疫政策出台后，该政策会在一段时间内对 w_1 与 w_2 的比值产生明显的影响，因此 w_1 通常以阶跃函数的形式出现。

2.2 基本再生数

为了进一步研究此类系统的稳定条件，我们设定

$$R_0 = w_1 \frac{\beta_1}{\gamma_1} + w_2 \frac{\beta_2}{\gamma_2}. \quad (26)$$

注意， R_0 只是新感染者的基本再生数，因为一个感染者平均可以再感染 β_1 / γ_1 （如果是类型 1）或 β_2 / γ_2 （如果是类型 2）人，并且这种情况分别以 w_1 或 w_2 的概率发生。在下面的定理中，我们表明，如果 $R_0 < 1$ ，则不会爆发疫情；如果 $R_0 > 1$ ，则会爆发疫情。因此，式 (26) 中的 R_0 被称为该模型中疫情爆发的渗透阈值 [15]。

定理 1：如果 $R_0 < 1$ ，那么公式 (23) 中矩阵 \mathbf{A} 的谱半径小于 1，且不会发生疫情爆发。另一方面，如果 $R_0 > 1$ ，那么公式 (23) 中矩阵 \mathbf{A} 的谱半径大于 1，会发生疫情爆发。

证明：（定理 1）

首先，我们注意到 γ_1 和 γ_2 是恢复率，在离散时间设置中，它们不可能大于 1，即感染者至少需要一天时间才能康复。因此，矩阵 \mathbf{A} 是一个正矩阵（其所有元素均为正数）。由 **Perron-Frobenius** 定理可知，该矩阵谱半径是 2×2 矩阵中的较大特征值。

现在我们找到矩阵 \mathbf{A} 的较大特征值。设 \mathbf{I} 为 2×2 单位矩阵，

$$\tilde{\mathbf{A}} = \mathbf{A} - \mathbf{I}. \quad (27)$$

通过公式

$$\tilde{\mathbf{A}} = \begin{bmatrix} \beta_1 w_1 - \gamma_1 & \beta_2 w_1 \\ \beta_1 w_2 & \beta_2 w_2 - \gamma_2 \end{bmatrix}. \quad (28)$$

$$z_1 = \beta_1 w_1 - \gamma_1 + \beta_2 w_2 - \gamma_2, \quad (29)$$

$$z_2 = \beta_1 w_1 \gamma_2 + \beta_2 w_2 \gamma_1 - \gamma_1 \gamma_2. \quad (30)$$

很容易证明 \mathbf{A} 的两个特征值是

$$\lambda_1 = \frac{1}{2}(z_1 + \sqrt{z_1^2 + 4z_2}), \quad (31)$$

和

$$\lambda_2 = \frac{1}{2}(z_1 - \sqrt{z_1^2 + 4z_2}). \quad (32)$$

注意 $\lambda_1 \geq \lambda_2$ 。由 (27) 可知, 转移矩阵 \mathbf{A} 的较大特征值是 $1 + \lambda_1$ 。如果 $R_0 < 1$, 则我们知道 $z_2 < 0$, $w_1 \frac{\beta_1}{\gamma_1} < 1$, 且 $w_2 \frac{\beta_2}{\gamma_2} < 1$ 。因此, 从 (29) 可知 $z_1 < 0$ 。鉴于 (31), 我们得出结论

$$\lambda_1 < \frac{1}{2}(z_1 + |z_1|) = 0.$$

这表明 $1 + \lambda_1 < 1$, 且 \mathbf{A} 的谱半径小于 1。另一方面, 如果 $R_0 > 1$, 则 $z_2 > 0$, 并且我们从 (31) 可知

$$\lambda_1 > \frac{1}{2}(z_1 + |z_1|) \geq 0.$$

这表明 $1 + \lambda_1 > 1$, 且 \mathbf{A} 的谱半径大于 1。系统参数和相变之间的关系将在下面小节中展示。

2.3 群体免疫

如果有足够多的人对疾病免疫, 尤其是通过接种疫苗, 群体免疫就是抵御传染病传播的一种方式。英国首相鲍里斯·约翰逊曾提出过一个有趣的策略, 即有足够多的人感染了 COVID-19 并从疾病中康复, 以实现群体免疫。问题是, 需要感染多少人才能实现 COVID-19 的群体免疫。

为了解决这个问题, 我们注意到, 在 SIR 模型中, 群体免疫对应于易感人群数量的减少。在我们之前的分析中, 我们都假设每个人在早期阶段都易患 COVID-19, 因此 $S(t)/n \approx 1$ 。对于群体免疫的分析, 我们假设随机选择的人在时间 t 易感的概率为 h 。这相当于 $1-h$ 比例的个体对该疾病免疫。在这样的假设下, 我们有

$$\frac{S(t)}{n} \approx h. \quad (33)$$

根据 (7) 中 $X(t)$ 的差分方程, 我们可以重写 (20)–(21) 以推导出群体免疫的控制方程如下:

$$X_1(t+1) - X_1(t) = \beta_1 X_1(t) h w_1 + \beta_2 X_2(t) h w_1 - \gamma_1 X_1(t), \quad (34)$$

$$X_2(t+1) - X_2(t) = \beta_1 X_1(t) h w_2 + \beta_2 X_2(t) h w_2 - \gamma_2 X_2(t), \quad (35)$$

$$R(t+1) - R(t) = \gamma_1 X_1(t) + \gamma_2 X_2(t). \quad (36)$$

与 (20)–(22) 中的原始控制方程相比, 唯一的区别是类型 1 (类型 2) 的传播率从 β_1 变为 $\beta_1 h$ (类型 2) 从 β_2 变为 $\beta_2 h$ 。因此, 群体免疫有效地将传播率降低了 h 倍。作为定理 1 的直接结果, 我们有以下推论。

推论 2: 对于由我们的 SIR 模型建模的传染病, 如果该模型中有两种感染者, 且 (26) 中的 R_0 大于 1, 则在至少 $1-h^*$ 比例的个体感染并从传染病中康复后, 即可实现群体免疫, 其中

$$h^* = \frac{1}{R_0}.$$

有关可能的有限免疫力的影响及其对 COVID-19 群体免疫的影响的更多讨论, 请参阅最近的论文 [26]。作者在那里讨论了在没有疫苗的情况下达到 COVID-19 群体免疫阈值可能产生的后果。特别是, [26] 指出, 医疗资源的枯竭不仅会导致 COVID-19 死亡率上升, 还会导致全因死亡率上升。

3 网络中疾病传播的独立级联 (IC)

模型

我们在上一节的分析没有考虑社交网络的结构如何影响疾病的传播。还有其他广泛使用的政策, 例如社交距离, 无法通过第三节中具有不可检测感染者的 SIR 模型进行建模。为了考虑网络结构, 在本节中, 我们考虑了疾病传播的独立级联 (IC) 模型。Kempe、Kleinberg 和 Tardos 之前在 [22] 中研究 IC 模型, 用于解决病毒式营销中的影响力最大化问题。在 IC 模型中, 有一个由图 $G=(V, E)$ 建模的社交网络, 其中 V 是节点集, E 是边集。受感染的节点可以以一定的传播概率将疾病传播到相邻的易感节点 (通过边)。由于我们的模型中有两种类型的感染者, 我们用 ϕ_1 (分别为 ϕ_2) 表示类型 1 (分别为类型 2) 感染节点将疾病传播给感染节点的 (直接) 邻居的传播概率。一旦相邻节点被感染, 它就变成类型 1 (类型 2) 感染节点, 概率为 w_1 (分别为 w_2), 并且可以继续将疾病传播给其邻居。继续传播, 我们便会形成 G 的一个子图, 从长远来看, 该子图包含感染节点集。将这样的子图称为感染子图。一个有趣的问题是, 如何控制疾病的传播, 以便即使节点总数非常大, 感染子图中的节点总数也能保持较少。

3.1 配置模型中的感染树

对于大量人群来说，通常很难获得确切的网络结构，即网络 G 的邻接矩阵。但是，我们可以了解网络的一些特征，特别是节点的度分布。配置模型（参见书籍[15]）是一类由节点度分布指定的随机网络。在这种随机网络中，随机选择的节点的度为 k ，概率为 p_k 。节点的边随机连接到其他节点的边。由于边连接是随机的，如果沿着受感染节点的边将疾病传播到此类网络中的其他节点，则受感染的子图看起来是一棵树（概率很高）。树假设是配置模型最重要的特性之一。配置模型的另一个关键属性是超度分布。在连接到该节点的边上找到度为 $k+1$ 的节点的概率为

$$q_k = \frac{(k+1)p_{k+1}}{\sum_{\ell=0}^{\infty} (\ell+1)p_{\ell+1}}. \quad (37)$$

因此，除去到达节点的边，仍有 k 条边可以传播疾病。这也是 q_k 被称为超度分布的原因。请注意，超度分布 q_k 通常不同于度分布 p_k 。当 p_k 为泊松度分布时，它们是相同的。在这种情况下，配置模型简化为著名的 Erdős-Rényi 随机图。

由于感染子图在配置模型中是一棵树，我们想知道感染树的大小是否有限。如果感染节点的感染树的大小有限且概率为 1，则我们说没有爆发。让 u_1 （或 u_2 ）表示类型 1（或类型 2）节点的感染树的大小通过其特定邻居之一为有限的概率。然后，

$$u_1 = 1 - \phi_1 + \phi_1 \sum_{k=0}^{\infty} (w_1 q_k u_1^k + w_2 q_k u_2^k), \quad (38)$$

$$u_2 = 1 - \phi_2 + \phi_2 \sum_{k=0}^{\infty} (w_1 q_k u_1^k + w_2 q_k u_2^k). \quad (39)$$

为了直观地了解 (38) 的内涵，我们注意到邻居要么被感染，要么未被感染。它未被感染的概率为 $1 - \phi_1$ 。另一方面，它被感染的概率为 ϕ_1 。然后，以概率 w_1 （分别为 w_2 ），被感染的邻居属于类型 1（代表类型 2）。此外，以概率 q_k ，邻居有额外的 k 条边来传播疾病。根据树假设，如果被感染的邻居属于类型 1（代表类型 2），则这 k 条边都具有有限感染树的概率为 u_1^k （分别为 u_2^k ）。(39) 中的等式来自类似的论证。

定义

$$g_1(z) = \sum_{k=0}^{\infty} q_k z^k \quad (40)$$

为超度分布的矩生成函数。然后我们可以将 (38) 和 (39) 简化如下：

$$u_1 = 1 - \phi_1 + \phi_1 w_1 g_1(u_1) + \phi_1 w_2 g_1(u_2), \quad (41)$$

$$u_2 = 1 - \phi_2 + \phi_2 w_1 g_1(u_1) + \phi_2 w_2 g_1(u_2). \quad (42)$$

从式 (41) 和式 (42) 我们可以解出 u_1 和 u_2 从

$$u_1^{(0)} = u_2^{(0)} = 0 \text{ 开始并且}$$

$$u_1^{(m+1)} = 1 - \phi_1 + \phi_1 w_1 g_1(u_1^{(m)}) + \phi_1 w_2 g_1(u_2^{(m)}), \quad (43)$$

$$u_2^{(m+1)} = 1 - \phi_2 + \phi_2 w_1 g_1(u_1^{(m)}) + \phi_2 w_2 g_1(u_2^{(m)}). \quad (44)$$

很容易得出 $u_1^{(m+1)} \geq u_1^{(m)}$ 和 $u_2^{(m+1)} \geq u_2^{(m)}$ 。因

此，它们收敛于 (41) 和 (42) 的固定解 u_1^* 和 u_2^*

3.2 与之前的 SIR 模型连接

现在我们通过指定传播概率 ϕ_1 和 ϕ_2 在第三章中显示与 SIR 模型的连接。

假设随机选择的边的一端是 1 型节点。那么根据第三章的 SIR 模型，该 1 型节点平均感染

β_1 / γ_1 人。由于平均过剩度为 $\sum_{k=0}^{\infty} k q_k = g_1'(1)$ ，因

此被该类型 1 节点感染的邻居的平均数量为

$\phi_1 g_1'(1)$ 。为了使 1 型节点感染与第三节 SIR 模型

中相同的平均节点数，我们有

$$\phi_1 = \frac{\beta_1 / \gamma_1}{g_1'(1)}. \quad (45)$$

对于 ϕ_2 ，我们有

$$\phi_2 = \frac{\beta_2 / \gamma_2}{g_1'(1)}. \quad (46)$$

利用式 (45) 和式 (46) 中规定的传播概率 ϕ_1 和

ϕ_2 , 我们得到如下的稳定性结果。

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \geq \begin{bmatrix} w_1\beta_1/\gamma_1 & w_2\beta_1/\gamma_1 \\ w_1\beta_2/\gamma_2 & w_2\beta_2/\gamma_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} 1-\beta_1/\gamma_1 \\ 1-\beta_2/\gamma_2 \end{bmatrix}. \quad (51)$$

定理 3: 对于组态模型构造的随机网络中的 IC

代入

模型(疾病传播), 设传播概率 ϕ_1 和 ϕ_2 分别在式 (45) 和式 (46) 中指定。那么受感染树的大小是有限的, 概率为 1 如果

$$\mathbf{B} = \begin{bmatrix} w_1\beta_1/\gamma_1 & w_2\beta_1/\gamma_1 \\ w_1\beta_2/\gamma_2 & w_2\beta_2/\gamma_2 \end{bmatrix}, \quad (52)$$

$$\mathbf{z} = \begin{bmatrix} 1-\beta_1/\gamma_1 \\ 1-\beta_2/\gamma_2 \end{bmatrix}. \quad (53)$$

$$R_0 = w_1 \frac{\beta_1}{\gamma_1} + w_2 \frac{\beta_2}{\gamma_2} < 1.$$

可以把 (51) 写成下面的矩阵形式:

$$\mathbf{u} \geq \mathbf{B}\mathbf{u} + \mathbf{z}. \quad (54)$$

在这种情况下, 没有爆发。

很容易看出 \mathbf{B} 的两个特征值是

证明: (定理 3)

$$\tilde{\lambda}_1 = 0, \text{ 和 } \tilde{\lambda}_2 = w_1 \frac{\beta_1}{\gamma_1} + w_2 \frac{\beta_2}{\gamma_2} = R_0.$$

让 $\mathbf{u} = (u_1, u_2)^T$ 和 $\mathbf{e} = (1, 1)^T$ 。足以证明, 当

$R_0 < 1$ 时, $\mathbf{u} = \mathbf{e}$ 是式 (41) 和式 (42) 方程组的唯一解。我们用反证法来证明这一点。假设 (41) 和 (42) 存在一个解当 $R_0 < 1$ 时 $u_1 < 1$ 或 $u_2 < 1$

并且, 特征值 R_0 对应的特征向量为

$$\mathbf{v} = \left(\frac{\beta_1}{\gamma_1}, \frac{\beta_2}{\gamma_2} \right)^T.$$

由于 (40) 中的矩生成函数是一个凸函数, 我

递归展开 (54) m 次得到

们从 $g_1(u_1)$ 和 $g_1(u_2)$ 的一阶泰勒展开中得到

$$g_1(u_1) \geq g_1(1) + (u_1 - 1)g_1'(1), \quad (47)$$

$$g_1(u_2) \geq g_1(1) + (u_2 - 1)g_1'(1). \quad (48)$$

请注意 $g_1(1) = \sum_{k=0}^{\infty} q_k = 1$ 。将 (47) 和 (48) 替换

$$\begin{aligned} \mathbf{u} &\geq \mathbf{B}^{m+1}\mathbf{u} + (\mathbf{I} + \mathbf{B} + \dots + \mathbf{B}^m)\mathbf{z} \\ &= \mathbf{B}^{m+1}\mathbf{u} + \mathbf{e} + R_0^m\mathbf{v}. \end{aligned} \quad (55)$$

为 (41) 和 (42), 我们得到

$$u_1 \geq 1 + \phi_1 g_1'(1)(w_1 u_1 + w_2 u_2 - 1), \quad (49)$$

$$u_2 \geq 1 + \phi_2 g_1'(1)(w_1 u_1 + w_2 u_2 - 1). \quad (50)$$

把这两个方程写成矩阵形式会得到

由于 $R_0 < 1$, $\mathbf{B}^{m+1}\mathbf{u}$ 和 $R_0^m\mathbf{v}$ 都收敛于 0 向量 m 趋于无穷。让 m 趋于无穷在 (55) 得到 $\mathbf{u} \geq \mathbf{e}$ 。这与 $u_1 < 1$ 或 $u_2 < 1$ 的假设相矛盾。

3.3 社交距离

保持社交距离是减缓传染病传播的有效方法。保持社交距离的一种常见方法是允许每个人将其人际接触(平均)保持在正常接触的一小部分 α (例如, 参见[27], [28])。在我们的 IC 模型中, 这对应于每个节点随机断开其一条边的概率为 $1 - \alpha$ 。

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \geq \begin{bmatrix} \phi_1 g_1'(1)w_1 & \phi_1 g_1'(1)w_2 \\ \phi_2 g_1'(1)w_1 & \phi_2 g_1'(1)w_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} 1 - \phi_1 g_1'(1) \\ 1 - \phi_2 g_1'(1) \end{bmatrix}.$$

这可以通过使用式 (45) 和式 (46) 进一步简化。因此, 我们有

与前一小节中一样, 我们让 $u_1(u_2)$ 为类型为 1 的受感染树的大小(例如: 类型 2)节点通过它的一个特定的邻居是有限的。因此,

$$u_1 = 1 - \alpha^2 \phi_1 + \alpha^2 \phi_1 \sum_{k=0}^{\infty} (w_1 q_k u_1^k + w_2 q_k u_2^k), \quad (56)$$

$$u_2 = 1 - \alpha^2 \phi_2 + \alpha^2 \phi_2 \sum_{k=0}^{\infty} (w_1 q_k u_1^k + w_2 q_k u_2^k). \quad (57)$$

根据 (56)，感染节点的相邻节点只有在以下情况下才会被感染：(i) 连接这两个节点的边缘未被移除(概率为 α^2)，以及(ii) 疾病通过边缘传播

(传播概率为 ϕ_1)。发生的概率是 $\alpha^2 \phi_1$ 。那么概率是 w_1 (或者 w_2)，受感染的邻居是类型 1 (或者类型 2)。同样，以概率 q_k ，邻居有额外的 k 条边来传播疾病。根据树的假设，这 k 条边都有有限棵受感染的树的概率是 u_1^k 或 u_2^k 。如果受感染的邻居是类型 1 (或者类型 2)，则 (57) 中的方程可以从类似的论证推导出来。

与式 (38) 和式 (39) 中的两个方程相比，我们得出结论，这种社会距离方法将传播概率 ϕ_1 和 ϕ_2 分别降低到 $\alpha^2 \phi_1$ 和 $\alpha^2 \phi_2$ 。作为定理 3 的直接结果，我们有以下推论。

推论 4: 假设社交距离方法允许每个人的际接触(平均)保持在正常接触的一小部分 α 。对于配置模型构造的随机网络中的 IC 模型(疾病传播)，感染树的大小是有限的，其概率为 1 如果

$$\alpha^2 R_0 < 1. \quad (58)$$

则在这种情况下，没有爆发。

另一种常用的保持社交距离的方法是取消大型聚会。这样做的目的是为了消除人际交往频繁的“超级传播者”的影响。为此，我们考虑一个疾病控制参数 k_0 ，并在我们的 IC 模型中删除大于或等于 k_0 的边数的节点。与式 (38) 和式 (39) 的推导类似，我们有

$$u_1 = 1 - \phi_1 + \phi_1 \sum_{k=k_0-1}^{\infty} q_k \quad (59)$$

$$+ \phi_1 \sum_{k=0}^{k_0-2} q_k (w_1 u_1^k + w_2 u_2^k),$$

$$u_2 = 1 - \phi_2 + \phi_2 \sum_{k=k_0-1}^{\infty} q_k \quad (60)$$

$$+ \phi_2 \sum_{k=0}^{k_0-2} q_k (w_1 u_1^k + w_2 u_2^k).$$

看到 (59)，我们注意到，如果 (i) 疾病没有通过边缘传播(概率为 $1 - \phi_1$)，(ii) 疾病通过边缘传播并

且邻近节点被移除(概率为 $\phi_1 \sum_{k=k_0-1}^{\infty} q_k$)，则 1 型感

染者仅沿边缘感染有限数量的人，(iii) 疾病通过边缘传播，相邻节点仅感染有限数量的人(概率为

$\phi_1 \sum_{k=0}^{k_0-2} q_k (w_1 u_1^k + w_2 u_2^k)$)。 (60) 的论据也是类似的。

类似于定理 3 的稳定性结果，对于取消大规模集会的社交距离方法，我们有以下稳定性结果。

定理 5: 考虑一种社交距离方法，它通过移除边缘数大于或等于 k_0 的节点来取消大规模聚集。对于配置模型构造的随机网络中的 IC 模型(疾病传播)，假设传播概率 ϕ_1 和 ϕ_2 分别在式 (45) 和式 (46) 中指定。那么受感染树的大小是有限的，概率为 1 如果

$$\left(\frac{\sum_{k=0}^{k_0-2} k q_k}{\sum_{k=0}^{\infty} k q_k} \right) R_0 < 1. \quad (61)$$

则在这种情况下，没有爆发。

证明: (定理 5)

在定理 3 的证明中，我们让 $\mathbf{u} = (u_1, u_2)^T$ 和

$\mathbf{e} = (1, 1)^T$ 。只要满足式 (61) 中的不等式，就足以

证明 $\mathbf{u} = \mathbf{e}$ 是式 (59) 和式 (60) 中方程组的唯一解。

我们用反证法来证明这一点。假设 (59) 和 (60) 存在一个解, 当 (61) 中的不等式满足时, $u_1 < 1$ 或 $u_2 < 1$ 。

因为 $f(u) = u^k$ 是 $u_1 > 0$ 和 $k > 0$ 的凸函数, 所

以我们有 $u^k \geq 1 + (u-1)k$ 。由 (59) 和式 (60) 可知

$$u_1 \geq 1 - \phi_1 + \phi_1 \sum_{k=k_0-1}^{\infty} q_k + \phi_1 \sum_{k=0}^{k_0-2} q_k + \phi_1 \sum_{k=0}^{k_0-2} k q_k (w_1 u_1 + w_2 u_2 - 1), \quad (62)$$

$$u_2 \geq 1 - \phi_2 + \phi_2 \sum_{k=k_0-1}^{\infty} q_k + \phi_2 \sum_{k=0}^{k_0-2} q_k + \phi_2 \sum_{k=0}^{k_0-2} k q_k (w_1 u_1 + w_2 u_2 - 1). \quad (63)$$

将这两个方程写成矩阵形式, 用式 (45) 和式 (46) 得到

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \geq \left(\sum_{k=0}^{k_0-2} k q_k \right) \begin{bmatrix} \phi_1 w_1 & \phi_1 w_2 \\ \phi_2 w_1 & \phi_2 w_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} 1 - \phi_1 \left(\sum_{k=0}^{k_0-2} k q_k \right) \\ 1 - \phi_2 \left(\sum_{k=0}^{k_0-2} k q_k \right) \end{bmatrix} \quad (64)$$

$$= \frac{\left(\sum_{k=0}^{k_0-2} k q_k \right)}{g_1'(1)} \mathbf{B} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} 1 - \frac{\left(\sum_{k=0}^{k_0-2} k q_k \right)}{g_1'(1)} \frac{\beta_1}{\gamma_1} \\ 1 - \frac{\left(\sum_{k=0}^{k_0-2} k q_k \right)}{g_1'(1)} \frac{\beta_2}{\gamma_2} \end{bmatrix},$$

其中 \mathbf{B} 是 (52) 中的矩阵。注意 $g_1'(1) = \sum_{k=0}^{\infty} k q_k$ 只是

期望的超额度。根据与定理 3 相同的论证, 可以很容易地证明, 当 (61) 中的不等式满足时 $u_1 \geq 1$ 和 $u_2 \geq 1$ 。这与 $u_1 < 1$ 或 $u_2 < 1$ 的假设相矛盾。

不幸的是, 在 (61) 中很难得到 k_0 的显式表达式来防止爆发。为此, 我们将在下一节中使用数值计算。

4. 数值结果

4.1 数据库

在本节中, 我们将使用第二节中的时间相关 SIR 模型和第三节中未检测到感染者的 SIR 模型对 COVID-19 趋势进行分析和预测。为了分析和预测 COVID-19, 我们从中华人民共和国国家卫生健康委员会 (NHC) 每日疫情通报中收集数据集 [12]。国家卫生健康委公布截至前一天 24 时的数据。我们收集 2020 年 1 月 15 日至 2020 年 3 月 2 日的确诊病例数、康复人数和死亡人数作为我们的数据集。确诊病例定义为实时逆转录聚合酶链反应 (rRT-PCR) 结果阳性的个体。值得注意的是, 湖北省自 2020 年 2 月 12 日起将确诊病例定义放宽为临床特征, 其他省份仍沿用原有定义。

4.2 参数设置

对于我们的时间相关 SIR 模型, 我们将用于预测 $\beta(t)$ 和 $\gamma(t)$ 的 FIR 滤波器的阶数设置为 3, 即 $J = K = 3$ 。模型的停止条件设置为 $X(t) \leq 0$ 。由于 2020 年 1 月 27 日之前的感染人数太少, 趋势不明显 (可能含有噪声), 我们只使用 2020 年 1 月 27 日之后的数据作为我们预测 $\beta(t)$ 和 $\gamma(t)$ 的训练数据。

我们使用 **scikit-learn** 库 [29] (Python 3 的第三方库) 来计算脊回归。预测 $\beta(t)$ 和 $\gamma(t)$ 的正则化参数分别设置为 0.03 和 10^{-6} 。由于传输速率 $\beta(t)$ 是非负的, 如果它小于 0, 我们将其设置为 0。然后, 我们使用算法 1 来预测 COVID-19 的趋势。

我们注意到我们的预测模型产生良好结果的局限性:

1. 我们的模型是一个确定性的流行病模型。它是基于 $X(t)$ 和 $R(t)$ 的平均场近似。这种近似是大数定律的结果。因此, 当 $X(t)$ 和 $R(t)$ 相对较小时, 平均场近似可能不像预期的那样准确。在这种情况下, 人们可能不得不求助于随机流行病模型, 比如马尔可夫链。我们将把它作为我们未来的工作。

2. 数据在开始时可能非常嘈杂。因此, 选择

一个好的起始日来训练岭回归进行预测是非常重要的。

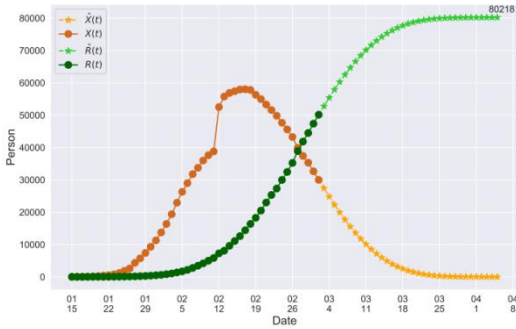


图 1. COVID-19 时变 SIR 模型的时间演化用深橙色(或绿色)标记的圆形实心曲线表示实际感染人数 $X(t)$ (康复人员 $R(t)$)，浅橙色(或绿色)的星号虚线表示预测感染人数 $\hat{X}(t)$ (恢复 $R(t)$ 人数)

3. 到 2020 年 3 月 2 日，一些国家的复苏速度可能还没有显示出明显的趋势。因此，很难预测这些国家的复苏速度。对此的一种补救办法是使用 1/30 天的固定康复率，这是医疗专业人员估计的康复/死亡的中位数。另一种方法是等待复苏速度的明显趋势。造成其治愈率低、治愈率趋势不清的主要原因是缺乏特效药和医疗资源短缺。有关这个问题的更多讨论，请参阅第五节。

4.3 时变 SIR 模型的时间演化

在图 1 中，我们展示了感染人数和康复人数的时间演变。实线为截至 2020 年 3 月 2 日的真实历史数据，虚线为我们对未来的预测结果。预测结果表明，如果中国政府继续采取封锁城市、停课等控制措施，新冠肺炎疫情将在 6 周内结束，确诊人数将达到 8 万人左右。

在图 2 中，我们显示了来自真实历史数据的测量值 $\beta(t)$ 和 $\gamma(t)$ 。我们可以看到 $\beta(t)$ 急剧下降， $\gamma(t)$ 略有上升。这是中国政府通过封锁城市和禁止交通等措施抑制传播速度 $\beta(t)$ 的直接结果。另一方面，由于缺乏有效的 COVID-19 药物和疫苗，恢复速度 $\gamma(t)$ 增长相对缓慢。此外，2020 年 2 月 12 日确诊病例的定义发生了变化，2020 年 2 月 11 日的相关数据没有参考价值。我们用灰色虚线曲线标记 $\beta(t)$ 和 $\gamma(t)$ 的这些数据点。

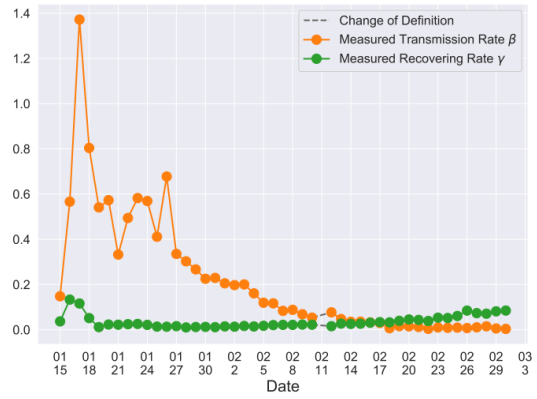


图 2. 监测 2020 年 1 月 15 日至 2 月 19 日新冠肺炎传播率

$\beta(t)$ 和恢复率 $\gamma(t)$ 。两条曲线分别根据式(11)和式(10)测量。

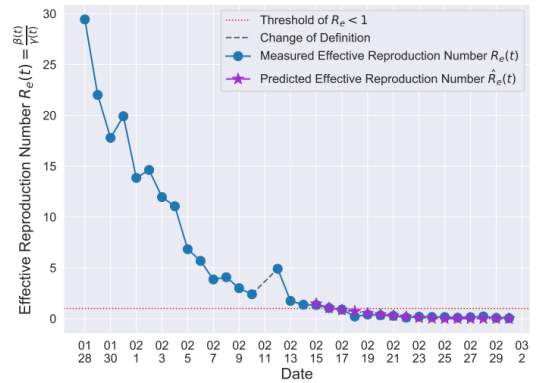


图 3. 中国 COVID-19 时变 SIR 模型的有效复制数 $R_e(t)$ 。

蓝色圆标实线为 2020 年 1 月 27 日至 2020 年 2 月 20 日给定数据的 $R_e(t)$ ，紫色星号虚线为 2020 年 1 月 27 日至 2020 年 2 月 15 日

数据的预测 $R_e(t)$ ，红色虚线为有效繁殖数的渗透阈值 1。

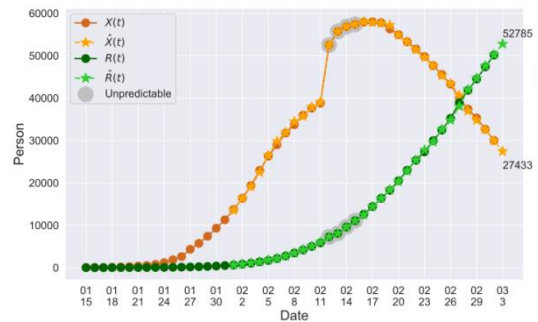


图 4. 一天内感染和康复人数的预测。由于确诊病例定义的改变而不可预测的点被标记为灰色。用深橙色(或绿色)标记的圆形实心曲线表示实际感染人数 $X(t)$ (康复人员 $R(t)$)，浅橙色(或绿色)星号虚线为预测感染人数 $\hat{X}(t)$ (恢复的人 $R(t)$)。

在流行病模型中, 一个关键问题是疾病能否得到控制, 流行病是否会结束, 或者是否会出现感染总人口 n 的一定比例的大流行。为了回答这个问题, 一个常用的度量是基本感染数 R_0 , 它被定义为感染者在完全易感人群中恢复之前的平均额外感染数[13]。在经典 SIR 模型中, R_0 简单地表示为 β / γ , 因为一个感染者 (平均) 需要 $1/\gamma$ 天才能康复, 在这段时间内, 他将 (平均) 接触 β 个人。在我们的时变 SIR 模型中, t 时刻的有效繁殖数, 用 $R_e(t)$ 表示, 定义为 $\beta(t) / \gamma(t)$ 。如果 $R_e(t) > 1$, 疾病将呈指数传播, 感染总人口 n 的一定比例。反之, 疾病最终将得到控制。因此, 通过观察 $R_e(t) > 1$ 随时间的变化, 甚至预测未来 $R_e(t) > 1$, 我们可以检验某些疫情控制政策是否有效

在图 3 中, 我们显示了测量的有效感染数 $R_e(t)$ 和预测的有效感染数 $R_e(t)$ 。蓝色圆圈标记的实线曲线是测量的 $R_e(t)$, 紫色星号标记的虚线曲线是预测的 $R_e(t)$ (从 2020 年 2 月 15 日开始)。很明显, $R_e(t)$ 自 2020 年 1 月 28 日以来急剧下降, 这意味着控制政策在中国起作用。更重要的是, 它表明转折点是 2020 年 2 月 17 日 $R_e(t) < 1$ 。在 2020 年 2 月 17 日之后的几天里, $X(t)$ 将呈指数级下降, 这将导致中国疫情的结束。我们的模型精确预测 $R_e(t)$ 将在 2020 年 2 月 17 日 (2020 年 2 月 14 日) 提前 3 天小于 1。结果表明, 该模型对 $\beta(t)$ 和 $\gamma(t)$ 的特征跟踪非常有效。

4.4 一天预测

为了显示我们模型的精度, 我们在图 4 中展示了第二天的预测结果 (一天预测)。它包含每天的预测感染人数 $X(t)$ (橙色星号虚线)、预测康复人数 $R(t)$ (绿色星号虚线) 以及实际感染人数和康复人数 (深橙色和深绿色圆圈标记的实线)。2020 年 2 月 12 日, 由于确诊病例定义的变化, 难以预测的日子。标记为灰色。预测曲线与实测曲线 (从真实历史数据中得到) 非常接近。在该图中, 我们还注释了预测感染人数 $X(t)=27,433$, 预测获救人数 $R(t)=52,785$ 在 2020 年 3 月 3 日。

我们在图 5 中进一步检查我们的预测准确性。除 2020 年 2 月 1 日、2 月 3 日、2 月 5 日预测的

恢复人数 $R(t)$ 外, 错误率均在 $\pm 3\%$ 以内。灰色虚线表示由于确诊病例定义的改变而不可预测的点。然而, 从 2020 年 2 月 16 日之后的预测结果来看, 我们发现我们的模型仍然可以准确地跟踪 $\beta(t)$ 和 $\gamma(t)$, 并且克服了定义变化的影响。

4.5 其他几个国家的基本再生产数字

除了中国的数据集, 我们还从 [16] 中的数据集测量了 2020 年 3 月 31 日几个国家的有效再现数 $R_e(t)$ 。这显示在表 1 的最后一列中。由于这些国家的累计恢复人数。

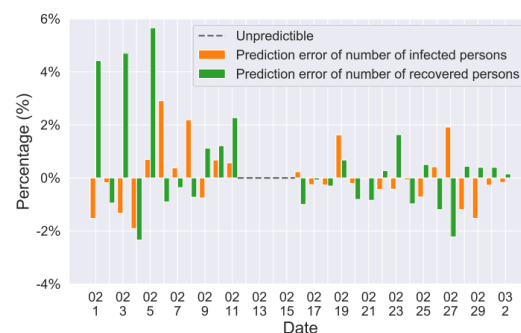


图 5. 一天内预测感染和康复人数的误差。灰色虚线表示因 2020 年 2 月 12 日确诊病例定义发生变化而导致的不可预测点。

据是有噪声的, 我们也显示了在平均恢复时间 $1/\gamma$ 的各种假设下的估计 $R_e(t)$ 。包括美国、英国、法国、伊朗和西班牙在内的五个国家的 $R_e(t)$ 值非常高。另一方面, 意大利政府于 2020 年 3 月 10 日宣布封锁并禁止人们聚集, 似乎正在控制疾病的传播。此外, 德国和大韩民国都有能力控制这种疾病的传播。

4.6 I1 型感染者的影响

在本小节中, 我们将展示无法检测到的 (2 型) 感染者如何影响流行病。特别是, 我们感兴趣的是解决无法检测到的感染者 (2 型) 的存在是否会导致爆发的问题。

为了进行数值研究, 我们需要在式 (20)–式 (22) 的差分方程组中固定一些变量。对于传输速率 (或康复率) 1 型感染者 β_1 (或 γ_1), 我们将其设置为测量值 $\beta(t)=0.00383$ (或 $\gamma_1=0.08493$) 在 2020 年 3 月 1 日在中国。其理由是, 2020 年 3 月 1 日以后, 中国发现了 1 型感染者, 并对其进行了治疗和隔离。此外, 由于没有针对 COVID-19 的药物,

我们可以假设这两种类型的感染者具有相同的康复率，即 $\gamma_2 = \gamma_1$ 。鉴于式 (20)–式 (22) 中的系统方程，仍然存在两个自由变量 w_2 和 β_2 ，其中 w_2 为感染者为 2 型的概率， β_2 为 2 型的传播率。

在图 6 中，我们说明了 w_2 和 β_2 如何影响 COVID-19 的爆发。这种图称为 [15] 中的相变图。图 6 中的黑色曲线为 (23) 中过渡矩阵 A 的谱半径等于 1 时的曲线。该曲线代表 COVID-19 的渗透阈值。如果 w_2 和 β_2 落在黑色曲线之上 (在橙色区域)，那么就会爆发。相反，如果 w_2 和 β_2 低于黑色曲线 (在黄色区域)，则不会发生疫情。如图 6 所示，我们想指出检测感染者的重要性。只要能够实际发现 90% 以上的感染者，并对其进行适当隔离和治疗，即使 2 型感染者 (即 β_2) 的传播率高达 0.7，也有可能遏制疾病的传播。另一方面，在检出率不高的情况下，抑制 2 型感染者的传播率也能有效控制疾病。例如，戴口罩和洗手可以是有效的防疫机制，以减少 β_2 。

Country	Estimated $R_e(t)$ when the average time to recover $1/\gamma$ is					$R_e(t)$ on Mar. 31, 2020
	14 Days	21 Days	28 Days	35 Days	42 Days	
United States of America	2.13	3.20	4.26	5.33	6.39	12.59
The United Kingdom	1.89	2.83	3.77	4.72	5.66	8.90
France	2.58	3.86	5.15	6.44	7.73	4.76
Iran	1.25	1.88	2.51	3.14	3.76	4.51
Spain	1.63	2.45	3.27	4.08	4.90	3.47
Italy	0.79	1.19	1.59	1.98	2.38	3.08
Germany	1.49	2.24	2.98	3.73	4.48	2.80
Republic of Korea	0.44	0.66	0.88	1.11	1.33	1.68

在接下来的实验中，我们将研究扩展到其他国家，包括日本、新加坡、韩国、意大利和伊朗。我们从 Johns Hopkins University 的 GitHub 中收集了 2020 年 1 月 22 日至 2020 年 3 月 2 日的历史数据 [16]。对于这些国家，在确诊病例数量迅

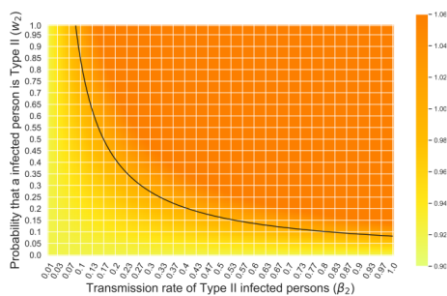


图 6. 爆发相对于 β_2 和 w_2 的相变图。黑色曲线为渗透阈值。橙色区域表示疾病将爆发，而黄色区域表示疾病已得到控制。

速增加的初始阶段，传播率 $\beta(t)$ 和恢复率 $\gamma(t)$ (根据第 2 章中与时间相关的 SIR 模型测量) 可视为 β_2 和 γ_2 。这是因为在这段时间内，没有任何防疫干预，基本上没有发现所有的感染者。有趣的是，不同的国家可能有不同的 β_2 和 γ_2 。另一方面，根据世卫组织的报告，只有 87.9% 的 COVID-19 病例出现发烧 [2]。如果以体温作为检测手段，那么新冠肺炎的检出率只有 87%。为此，我们设置 $w_1 = 87.9\%$ 。

我们利用之前研究中为中国指定的 β_1 和 γ_1 ，用两个变量 β_2 和 γ_2 绘制了图 7 中的相变图。同样，黑色曲线是 (23) 中过渡矩阵 A 的谱半径等于 1 时的曲线。这条曲线代表了 COVID-19 疫情的渗透阈值。如果 β_2 和 γ_2 落在黑色曲线之上 (在橙色区域)，那么就会爆发。相反，如果 β_2 和 γ_2 低于黑色曲线 (在黄色区域)，则不会发生疫情。日本、新加坡、韩国、意大利、伊朗等确诊病例较多的国家如图 7 所示。从图 7 中，我们观察到新加坡和日本都低

于渗透阈值。但日本更接近渗透门槛。相反，韩国和意大利都超过了渗透阈值，在 2020 年 3 月 2 日处于潜在爆发的边缘。这两个国家必须紧急实施防疫政策。毫不奇怪，2020 年 3 月 10 日，意大

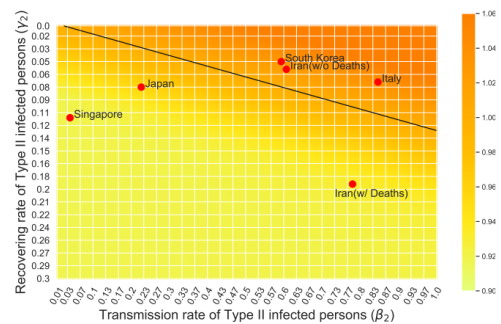


图 7. 爆发相对于 β_2 和 γ_2 的相变图。黑色曲线为渗透阈值。橙色区域表示疾病将爆发，而黄色区域表示疾病已得到控制。

利政府宣布封锁, 禁止人们聚集。值得一提的是, 图中伊朗有两个标记。高于渗透阈值的是未将死亡人数加到康复人数中的死亡人数。低于渗透阈值的另一个是将死亡人数加到康复人数中。由于某种未知的原因, 伊朗的死亡率高于其他国家。高死亡率似乎阻止了疫情在伊朗的爆发。

为了表明 $w_1(t)$ 可以随着时间的推移而发生显著变化, 以反映政府防疫政策的变化, 我们在图 8 中绘制了 2020 年 2 月 23 日至 2020 年 3 月 18 日在韩国估计的 $w_1(t)$ 。使用本节相同的实验设置, 我们将 β_1, γ_1 和 γ_2 设为中国之前研究的实测值, β_2 设为韩国的实测值。我们将 5 天作为时间间隔, 并利用 (25) 中的递归公式来估计 $w_1(t)$ 随时间的值。

在每个时间区间内, 通过最小化式 (25) 下残差平方和得到 w_1 [29]。如图 8 所示, 从 2020 年 2 月 23 日开始的前三个时间间隔中, w_1 明显增加。如果我们将这一结果与 Our World in Data[30]提供的数据集进行比较, 我们可以发现, 韩国从 2020 年 2 月 24 日开始对公众进行大量 COVID-19 检测, 并且每天都在增加检测次数。因此, 我们可以推断, 像大规模检测这样的防疫政策可以有效地增加 w_1 的概率, 并大大降低未检测到的人的比例。从上面的讨论, 我们知道 $w_1(t)$ 可以反映对新的防疫政策的影响。此外, 大规模检测对减少无法检测到的人的比例有重大影响。

4.7 社交距离的影响

在本小节中, 我们展示了社交距离方法的数值结果, 该方法通过移除边缘数量大于或等于 k_0 的节点来取消大规模聚集。如定理 5 所示, 有效

复制数减少 $\frac{\sum_{k=0}^{k-2} kq_k}{\sum_{k=0}^{\infty} kq_k}$, 其中 q_k 为 p_k 的超额度分布。

在本实验中, 我们使用从 Facebook 收集的数据集 [31]。该数据集表示艺术家类别的经过验证的 Facebook 页面 (带有蓝色复选标记) 网络。蓝色复选标记意味着 Facebook 已经确认该账户是其所代表的公众人物、名人或全球品牌的真实存在。网

络中的每个节点代表一个页面, 两个节点之间的边是相互点赞的。该数据集由 50,515 个节点和 819,306 条边组成。其他一些属性如下: 平均度 32.4, 最大度 1469, 直径 11, 聚类系数 0.053。在图 9 中, 我们显示了该数据集的度分布和超额度分布的对数-对数图。度分布似乎是指数为 1.69 (图中的斜率) 的 (截断的) 帕累托分布。

在图 10 中, 我们绘制了还原比 $\frac{\sum_{k=0}^{k-2} kq_k}{\sum_{k=0}^{\infty} kq_k}$ 作为

k_0 的函数。比值在 0 和 1 之间, 在 k_0 中单调递增。还原比表示禁止多大的集合还原对应的 $R_e(t)$ 的比值。使用表 1 最后一列中的 $R_e(t)$ 值 (2020 年 3 月 31 日), 我们还显示, 在意大利、美国 and 韩国, 防止爆发的最小 k_0 s。

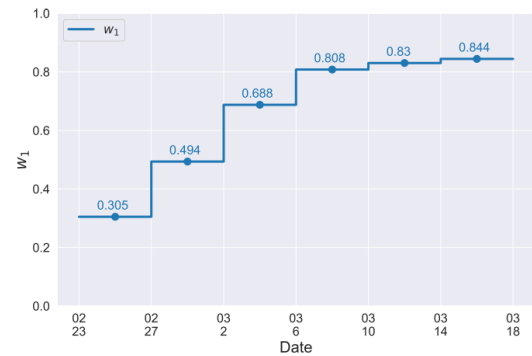


图 8. 估计的 w_1 是韩国每隔 5 天的函数。

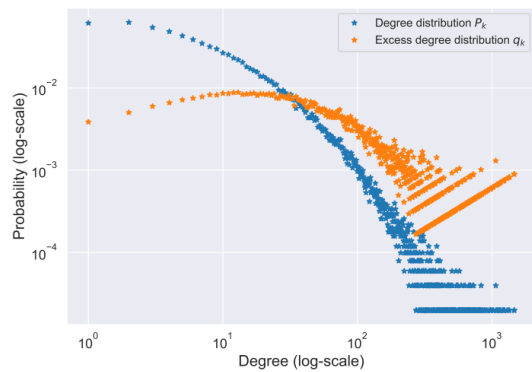


图 9. Facebook 数据集的度分布和超额度分布。

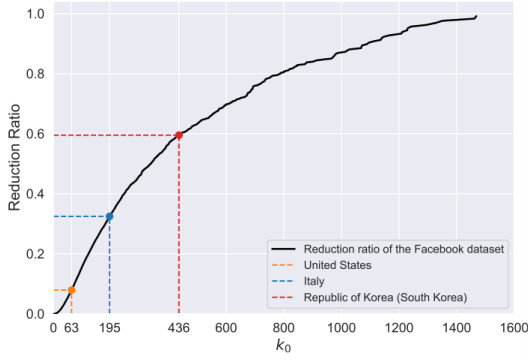


图 10. 还原比 $\frac{\sum_{k=0}^{k-2} kq_k}{\sum_{k=0}^{\infty} kq_k}$ 作为 $k_0 s$ 的函数。在意大利、美国和

韩国，预防疫情的最低限度分别是 63、195 和 435 人。

分别为 63、195 和 435；尾分布的影响率分别为 13.1%、2.2% 和 0.4%。特别是，如果取消大规模聚集是 2020 年 3 月 31 日 $R_e(t)$ 值为 12.59 的美国控制新冠肺炎的唯一措施，那么可以通过“删除” (Facebook 数据集中) 所有大于或等于 63 度的节点来防止爆发，删除可能会影响 Facebook 数据集中 13.1% 的节点。

6. 讨论与建议

截至 2020 年 3 月 2 日，自 2020 年 1 月中国发布全市封锁等防控政策以来，中国的新冠肺炎疫情似乎已逐步得到控制。虽然交通暂停、小社区管理、全市封锁等政策可以有效降低传播率 β_1 和 β_2 ；然而，这些相对极端的政策不仅限制了人身自由的权利，也影响了社会的正常运行。这些极端的政策迫使一些公司和工厂停产，影响了经济的所有部门。因此，在预防疾病和确保社会正常运行之间取得平衡至关重要，提出所谓的“最优”控制政策是很重要的。

为此，我们想根据我们的系统模型结果的观察，对控制 COVID-19 的传播提出一些讨论和建议。根据定理 1、推论 2、推论 3、推论 4 和定理 5 的结果，我们知道，如果有效繁殖数为 0，则不存在疾病爆发

$$R_e = h \cdot s \cdot \left(w_1 \frac{\beta_1}{\gamma_1} + w_2 \frac{\beta_2}{\gamma_2} \right) < 1, \quad (65)$$

其中， h 为推论 2 中有 $1-h$ 部分个体对该疾病免疫时的群体免疫降低因子， s 为推论 4 或定理 5 中的社会距离降低因子 w_1 (或 w_2) 是指被感染者能 (或不能) 够感染的概率被检测到， β_1 和 γ_1 (或 β_2 和 γ_2) 是感染者的传播率和恢复率。不能被检测到。对于 COVID-19, $\beta_1 \leq \beta_2$ 和 $\gamma_1 \geq \gamma_2$ 作为感染者，一旦发现，可以进行治疗 (以缩短恢复时间) 和隔离 (以降低传播率)。由于 $w_2 + w_1 = 1$ ，因此 w_1 比 w_2 大得多是可取的。

为了防止爆发，应该最小化 (65) 左边的值。这里我们将讨论几种方法。

提高治愈率 γ_1 : 提高治愈率最有效的方法是寻找抗病毒药物; 然而，这需要时间。因此，在这个阶段，我们应该把重点放在控制疾病传播的其他方法上

降低群体免疫降低因子 h : 最有效的方法是寻找疫苗来降低易感人群的比例。同样，这需要时间，我们应该把重点放在现阶段控制疾病传播的其他方法上。

降低传播率 β_1 [32]: 一旦发现感染者，应进行隔离，避免对社会造成额外感染，降低传播率 β_1 。对疑似与感染者有接触的人进行隔离也可以降低传播率 β_1 和 β_2 。

增加检测概率 w_1 (从而降低 w_2): 大规模测试肯定会增加 w_1 。事实上，韩国在“得来速”测试方面做得非常出色。如果由于医疗资源的限制而无法进行大规模检测，那么测量体温也是一种有效的替代方法，因为 87.9% 的新冠肺炎确诊病例有发烧症状。此外，还可以对确诊病例的旅行史、职业、接触者、聚集性 (TOCC) 等进行追踪，缩小可能的传染源范围。这些来源可能包含无法检测到的感染者，因此检测这些可能来源的密切接触者可以通过检测无法检测到的感染者来增加 w_1 。

降低传播率 β_2 : 健康教育知识的宣传可以大幅降低传播率 β_2 。例如，在公共场所和封闭场所戴口罩，洗手，避免接触口、眼、鼻，这不仅是保护自己不被无法检测到的感染者感染的好方法，也是避免感染他人的好方法。

降低社会距离缩减系数 s : 根据推论 4 和定理 5, 有两种切实可行的方法可以降低社会距离缩减系数 s : (i) 允许每个人在正常接触人数的一小部分范围内保持人际接触, (ii) 取消大规模聚会。

7. 结论及未来工作

在本文中, 我们对 COVID-19 进行了数学和数值分析。我们的时变 SIR 模型不仅比传统的静态 SIR 模型具有更强的自适应性, 而且比直接估计方法具有更强的鲁棒性。我们的数值结果表明, 对于从中华人民共和国国家卫生健康委员会(NHC)收集的数据集, 感染人数 $X(t)$ 和康复人数 $R(t)$ 的一天预测误差在(几乎)3%以内[12]。此外, 我们能够跟踪传播率 $\beta(t)$ 和恢复率 $\gamma(t)$ 相对于时间 t 的特征, 并准确预测未来中国 COVID-19 疫情的趋势。

为了解决 COVID-19 中无法检测到的感染的影响, 我们扩展了 SIR 模型, 考虑了两种类型的感染者: 可检测的感染者和不可检测的感染者。这种模型中是否存在爆发的特征是与基本繁殖数 R_0 密切相关的 2×2 矩阵的谱半径。除了我们对中国的数值分析, 我们进一步将我们的研究扩展到其他国家, 包括日本、新加坡、韩国、意大利和伊朗。

为了了解社会距离措施的影响, 包括减少人际接触和取消大规模聚会, 我们分析了配置模型中疾病传播的 IC 模型。通过将 IC 模型中的传播概率与 SIR 模型中的传播率和恢复率联系起来, 我们发现这些社交距离方法可以导致有效繁殖数 R_e 的降低。

最后, 根据实验结果, 从模型的角度对防疫工作提出了一些讨论和建议。一个可能的未来工作是将我们的确定性 SIR 模型扩展到随机模型, 以便将超传播事件(在低患病率下显示高随机性的事件[33])纳入分析。这样可以进一步提高预测结果的精度。

参考文献

- [1] H.Nishiura et al., "Estimation of the asymptomatic ratio of novel coronavirus infections (covid-19)," *Int. J. Infectious Dis.*, vol. 94, p. 154, 2020.
- [2] "Coronavirus disease (covid-19) outbreak," Jan.

2020.[Online].Available:<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>

[3] I. Nesteruk, "Statistics based predictions of coronavirus 2019-ncov spreading in mainland china," *MedRxiv*, 2020.

[4] Y. Chen, J. Cheng, Y. Jiang, and K. Liu, "A time delay dynamical model for outbreak of 2019-ncov and the parameter identification," 2020, arXiv:2002.00418.

[5] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of Covid-19 in china by dynamical modeling," 2020, arXiv:2002.06563.

[6] T. Zhou et al., "Preliminary prediction of the basic reproduction number of the wuhan novel coronavirus 2019-ncov," *J. Evidence-Based Med.*, 2020.

[7] B. F. Maier and D. Brockmann, "Effective containment explains subexponential growth in confirmed cases of recent covid-19 outbreak in mainland china," 2020, arXiv:2002.07572.

[8] Q. Lin et al., "A conceptual model for the outbreak of coronavirus disease 2019 (Covid-19) in Wuhan, China with individual reaction and governmental action," *Int. J. Infectious Dis.*, vol. 93, pp. 211–216, Apr. 2020.

[9] S. Zhao et al., "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak," *Int. J. Infectious Dis.*, vol. 92, pp. 214–217, Mar. 2020.

[10] T. Zeng, Y. Zhang, Z. Li, X. Liu, and B. Qiu, "Predictions of 2019-ncov transmission ending via comprehensive methods," 2020, arXiv:2002.04945.

[11] Z. Hu, Q. Ge, L. Jin, and M. Xiong, "Artificial intelligence forecasting Of covid-19 in china," 2020, arXiv:2002.07112.

[12] "Outbreak notification," Jan. 2020. [Online]. Available:http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml

- [13] R. M. Anderson and R. M. May, "Vaccination and herd immunity to infectious diseases," *Nature*, vol. 318, no. 6044, pp. 323–329, 1985.
- [14] S. Zhao et al., "Modelling the effective reproduction number of vector-borne diseases: The yellow fever outbreak in Luanda, Angola 2015–2016 as an example," *PeerJ*, vol. 8, p. e8601, 2020.
- [15] M. Newman, *Networks: An Introduction*. Oxford, U.K. Oxford Univ.Press, 2010.
- [16] CSSEGISandData and J. H. University, "COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hop-kins University," Feb. 2020. [Online]. Available: <https://github.com/CSSEGISandData/COVID-19>
- [17] B. F. Finkenstädt and B. T. Grenfell, "Time series modelling of child-hood diseases: A dynamical systems approach," *J. Roy. Statistical Soc.:Ser. C (Appl. Statist.)*, vol. 49, no. 2, pp. 187–205, 2000.
- [18] A. D. Becker and B. T. Grenfell, "tsir: An r package for time-series susceptible-infected-recovered models of epidemics," *PloS One*, vol. 12, no. 9, pp. 1–10, 09 2017. [Online]. Available: <https://doi.org/10.1371/journal.pone.0185528>
- [19] J. Dehning, J. Zierenberg, F. P. Spitzner, M. Wibral, J. P. Neto, M. Wilczek, and V. Priesemann, "Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions," *Science*, vol. 369, no. 6500, Jul. 2020, Art. no. eabb9789.
- [20] Y.-C. Chen, P.-E. Lu, C.-S. Chang, and T.-H. Liu, "A time-dependent sir model for covid-19 with undetectable infected persons," 2020, arXiv:2003.00122.
- [21] O. Wyman, "Covid-19 pandemic navigator core model: Overview of methodology and use cases," Apr. 2020. [Online]. Available: <https://hlc-hlc21901.s3-us-west-2.amazonaws.com/public/Whitepaperpercent20Pandemicpercent20Navigatorpercent20Corepercent20Model.pdf>
- [22] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. ACM, 2003, pp. 137–146.
- [23] A. Y. Ng, A. X. Zheng, and M. I. Jordan, "Stable algorithms for link analysis," in *Proc. 24th Ann. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2001, pp. 258–266.
- [24] B. S. Dayal and J. F. MacGregor, "Identification of finite impulse response models: methods and robustness issues," *Ind. Eng. Chem. Res.*, vol. 35, no. 11, pp. 4078–4090, 1996.
- [25] T. Ganyani et al., "Estimating the generation interval for Covid-19 based on symptom onset data," *medRxiv*, 2020.
- [26] H. E. Randolph and L. B. Barreiro, "Herd immunity: Understanding Covid-19," *Immunity*, vol. 52, no. 5, pp. 737–741, 2020.
- [27] C.-J. Chen, *Epidemiology: Principles and Methods*. Linking Publishing Company, 1999.
- [28] N. G. Becker, *Modeling to Inform Infectious Disease Control*. Boca Raton, FL, USA: CRC Press, 2015.
- [29] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [30] E. O.-O. Max Roser, Hannah Ritchie and J. Hasell, "Coronavirus pandemic (covid-19)," *Our World Data*, 2020, <https://ourworldindata.org/coronavirus>.
- [31] B. Rozemberczki, R. Davies, R. Sarkar, and C. Sutton, "Gemsec: Graph embedding with self clustering," in *Proc. 2019 IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining 2019*. 2019, pp. 65–

72.

[Online].

Available:<https://snap.stanford.edu/data/gemsec-Facebook.html>

[32] Q. Li et al., “Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia,” *New England J. Med.*, to be published, 2020, doi: 10.1056/NEJMoa2001316.

[33] B. M. Althouse et al., “Stochasticity and heterogeneity in the transmission dynamics of SARS-COV-2,” 2020, arXiv:2005.13689.