

Preliminary Study on Transmission Model and Prediction of New Coronavirus

Fuyi Yang

Anshan Normal University, Anshan Liaoning
Email: yangfuyi@sina.com

Received: Feb. 9th, 2020; accepted: Feb. 20th, 2020; published: Feb. 27th, 2020

Abstract

In this paper, novel coronavirus (COVID-19) transmission model is studied and predicted. Complex network theory and six degrees association model, mathematical statistics and regression analysis are used for comprehensive prediction analysis. The application of the basic mathematical theory of analysis in the research and calculation of virus propagation model is given. The forecast analysis gives the short and medium term future data and change trend curve, for analysis and reference of relevant personnel.

Keywords

COVID-19, Prediction, SM Small World Network, Directed Graph, Complex Network

新型冠状病毒传播模型与预测的初步研究

杨福义

鞍山师范学院, 辽宁 鞍山
Email: yangfuyi@sina.com

收稿日期: 2020年2月9日; 录用日期: 2020年2月20日; 发布日期: 2020年2月27日

摘要

本文对新型冠状病毒疾病(COVID-19)的传播模型进行研究和预测,采用复杂网络理论和六度关联模型以及数理统计和回归分析进行综合预测分析,给出分析的数学基础理论在病毒传播模型研究计算中的应用。预测分析给出了中短期的未来数据和变化趋势曲线。供有关人员分析参考。

关键词

2019年冠状病毒病, 预测, SM小世界网络, 有向图, 复杂网络

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

当前, 新型冠状病毒(COVID-19)在中国肆虐并已经传播到国外。在中国国家领导的高度重视下, 全民总动员。投入抗击新型冠状病毒传染的战斗中, 作为科技工作者有必要尽快开展病毒传播特性研究并用于抗击新型冠状病毒的战斗中。本文所使用的数学模型和参数供领导、全国科技工作者和有关人员参考, 期望对进一步防控新型冠状病毒的工作有参考作用。

新型冠状病毒(COVID-19) (以下简称新冠病毒)的传播模型采用复杂网络理论的六度关联模型进行预测分析。是小世界网络模型在流行病传播与预测研究中的现实应用。

中国科学技术大学汪秉宏等人指出: 小世界网络中的流行病传播问题。传染病在人群中的流行、病毒在计算机网络上的传播和谣言在人类社会中的扩散等、都可视为流行物(传染病、谣言、舆论、计算机病毒等)在网络上的传播问题。

美国学者穆尔(C. Moore)和纽曼最早对小世界网络上的传播行为进行了较系统的研究。他们研究了NW小世界网络上的SIR模型(易感-感染-恢复模型)的传播行为。他们把小世界网络上的疾病传播问题等价于逾渗问题, 如座愈渗问题-删除一定比例的网络中的节点、健愈渗问题-删除一定比例的网络中的连边、或座健混合愈渗问题-删除一定比例的网络中的节点与连边, 以此来观察删除网络中一定比例的节点与连边对阻碍疾病在网络中传播的效果。穆尔和纽曼给出了愈渗的临界值与网络长程边(捷径)密度的函数关系, 发现随着长程边(捷径)的出现, 流行病在网络上的传播阈值(即愈渗的临界值)迅速下降, 即使少量的长程边(捷径)就可以显著地增加流行病在网络中的传播能力[1]。

阿根廷学者库珀曼(M. Kuperman)和艾布拉姆森(G. Abramson)研究了WS小世界网络上的SIRS模型(易感-感染-恢复-再易感模型)的传播行为。发现当重连概率 p 很小的时候, 疾病可以在网络中长期存在, 但患病比率很小且波动不大, 可以近似地看作收敛到一个不动点; 而当重连概率 p 很大的时候, 患病人数会出现周期性的波动。即随着重连概率 p 的逐渐增加, 对应于一定的人群结构, 网络中被传染的人数从不规则的、小幅度的增加发展到自发的、大范围的振荡状态, 其中在 $p=0.1$ 附近传染人数明显增加, 显示出小世界效应。

流行病在小世界网络上的传播行为与小世界网络的规模、结构、传播系数、网络各结点动态地传播行为、传播结点状态变化(结点消亡, 成为孤点或叶结点)等参数有关, 这些网络静态和动态参数对流行病在小世界网络上的传播规律和趋势及其控制有很大影响。掌握网络拓扑结构形态与动态参数的数值, 可以得到流行病在不同情况下的传播情况, 显示流行病的发展过程, 预测流行病的传播趋势, 分析流行的原因和关键因素, 寻找对其进行控制和预测的最佳策略。

笔者在充分研读相关材料的基础上, 根据对几种复杂网络的数据研究结果和概率统计预测的理论知识。对新冠病毒的传播模型进行模拟仿真, 求取数学模型, 验证, 依据数学模型开展短期与长期预测工作。

小世界网络的定义由米尔格兰姆提出，六度关联模型就是一个反映自然规律的一个模型，只要世界上的事物发生联系，就构成小世界的关联现象，感染者和被感染者都可以抽象为集合中的点，感染过程的传递就是病毒由一个结点向另一个结点的传播关系连线。从而抽象为点与线相关联的传染与被传染关系，就构成了图论中有向图的研究对象。病毒的传播，个体的感染和被感染，已经由最初的个体的传播关系转变构成了群体传播行为。其中所爆发的涌现现象模型更值得充分认真的重视，鉴于此项研究关系到当前防控的预测和决策。特介绍有关的初步研究结果。

2. 小世界网络模型的研究

2.1. 小世界网络概述

1967年，美国社会心理学家 Stanley Milgram (斯坦利·米尔格兰姆, 1933~1984) 提出了小世界理论。这是一个数学领域的猜想，名为 Six Degrees of Separation，中文翻译包括以下几种：六度分割理论或小世界理论等。这个理论可以通俗地阐述为：“任何两个陌生人之间所间隔的人不会超过六个，也就是说，最多通过六个人，就可以让任何两个陌生人认识[2]”。这就是六度分割理论，也叫小世界理论。

Stanley Milgram 的实验结果由于设计和操作上的缺陷，受到了一些心理学家的质疑，但是其他的实验结果也表明世界的确不大。

1998年，Watts 和 Strogatz 采用环形网随机化重连的方法，构建了新的网络，随着边数的增加网络具有较短的平均路径又具有较高的聚类系数。这个网络模型被称为 WS 小世界模型。

WS 小世界模型构造算法中随机化重连过程有可能破坏网络的连通性，1999年，瓦茨又与英国物理学家提出了另一个稍有不同的模型，称为纽曼-瓦茨模型(NW 模型)。

而后随着大量学者对复杂网络模型的研究，又提出无标度网络层次网络、确定型网络和自相似网络等一系列网络模型。

中国学者汪小帆提出可变聚类系数无标度网络舆论传播模型。

为了纪念美国心理学家斯坦利·米尔格兰姆的卓越贡献。笔者在本论文中围绕六度分割理论核心：度(层)分布，对 10 余个复杂网络数据集进行度分布和距离计算等分析。受英国剑桥大学布鲁巴斯关于随机图论证定理的启发，用六度关联模型分析了有限图单个结点对网络的所有结点的连通属性。在此基础上提出小世界网络模型的定义。

小世界网络定义：给定论域的随机图中，各结点相互可达，6 层关联的结点包括几乎所有节点有向网络，称为米尔格兰姆小世界网络，简称 SM 小世界网络。复杂网络中满足小世界网络特性的子网，称为复杂网络的核心网。

米尔格兰姆小世界网络最重要的特性就是网络最大的直径几乎都不超过 7。网络的平均度总是在 4 附近。

杨炳儒指出：“图论通过点和线的构型来构成模拟各类系统的数学模型，并根据图的性质进行分析，提供研究各种系统的科学的、巧妙的方法。任何一个包含了某种二元关系的系统都可以用图论的方法分析，而且它往往具有形象直观的特点。”图论的研究具有十分广阔的客观原型，并在许多领域中有着极其广泛的应用([3], p. 184)。

病毒的传播可以用个体接受感染，而后在感染别的个体。是典型的二元关系。可以用图论研究小世界理论与数学模型，为了和图论中结点的出度、入度的概念相区别。对小世界网络的度看作层，将六度关联与度分布中的“度”一律加括号中的“层”标明，以避免科技术语概念实质的混淆。

病毒传播的数学模型，目前研究最为彻底应用最为广泛的传染病模型是 SI 模型、SIS 模型、SIR 模型 SIRS 模型。这些模型的特点如下([4], p. 100)。

SI 模型：系统个体分为两类：易感个体 S 和感染个体 I ；

SIS 模型：在上面情况中出现被感染治愈又受到感染的微分方程分析；

SIR 模型：引入免疫个体：含个体 R (不受感染获得免疫力)；

SIRS 模型：在移动状态获得免疫力的个体 R 以概率 r 失去免疫力的情况；

SEIR 模型：易感个体 S 在被感染后，以一定概率 α 变为潜伏状态 E ，再以一定概率 β 变为感染状态 I 。

笔者认为：这五个模型的重要区别在于把复杂网络图论中的结点个体，进行了属性分类。这种分类也造成了大范围传染病获得数据的统计难度。尤其后 4 个模型，具有重要的理论指导意义，但缺少宏观预测与定量分析的详实统计数据而增加工程实施的难度。而 **SI 模型**，实质就是逻辑斯蒂模型，稍加变换，就得出与本文一样的形式。

笔者认为复杂网络是各种网络链接生成的网络，每个网络都具有以小世界网络为核心。这个核心网络所具有的集合基数彼此相关的特性，构成了基于模型的宏观统计分析预测的基础。根据这一集合的基数概率统计分析，预测模型理论中以及数学知识，对新型冠状病毒的统计资料进行短期与较长期的预测与分析成为可能。其预测超前，预测精度可以事后分析检验并随时间序列反映的系统状态调节参数，提高预测水平。

本预测的初始条件做如下限定：在论域内的个体只包含两类人群，被传染者和传染者以及随时间由被传染者转变为传染源。即网络的点做如下限定，网络中的任何点都具有接受病毒感染以及传出病毒的能力，即图论中结点的入度和出度都不为 0，不考虑自环和重边。结点具有输入和输出的能力。以有向边表示传播途径与层次。

传染病小世界的度(层)分布，反映了群体之间的关系，研究度(层)分布的特性曲线对实际应用具有重要意义。由于传染过程统计以时间划分因此在直角坐标系的 X 轴表示度(degrees)或传染层次， Y 轴表示被感染人群总数。

SM 小世界网络的一个重要应用基础是研究复杂网络中的涌现现象。涌现现象的数学模型对于快速的物理与化学过程尤其重要。其关键是在极其短的时间内，产生的大量动力学现象，在时间切割为微秒、纳秒的时间段，物质分子的动力学特性都与网络的特性与数学参数相关。传染速率最大的时间结点是度(层)分布的峰值。计算表明，此时的感染人数是总人数的 50%。

笔者在分析了多个复杂网络后，采用集合的集合。以子集合基数之间的关系为主线，分析了网络层关联现象，对集合的基数进行度(层)分布的相关分析。用于解决文献智能分类，概念群体之间的距离分析等实际问题。获得了一些进展。从而为本次预测分析提供验证的实际经验。

2.2. 六度关联模型预测新冠病毒传播的理论基础

2.2.1. 数学模型和方法

数学模型是“针对或参照某种事物系统的特征或数量相依关系，采用形式化数学语言，概括地或近似地表述出来的一种数学结构。”

“**数学模型方法**”(mathematical modeling method)，它不仅是处理数学理论问题的一种经典方法，而且也是处理科技领域中各种实际问题的一般数学方法。特别，现代电子计算机的广泛应用和科学技术的数学化趋势，使得数学模型方法已经非常广泛地应用于自然科学、工程技术科学和社会科学的一切领域中([5], p. 15)。

关系映射反演原则(简称为 **RMI**)原则是指一种分析处理问题的普遍方法或准则([5], p. 24)。

RMI 原则实际可以理解作为一种包罗万象的科学方法论原则。数学关系是在数学对象中可以确切定义的关系。**映射(或变换)**是在两类数学对象或两个数学集合元素之间建立了一种“对应关系”，就定义了一

个映射。特别，如果是一一对应关系。则称为**可逆映射**([5], p. 28)。

微分方程的求解就是把微分方程取拉普拉斯变换将微分方程变换为象函数的代数方程，求解代数方程，解后再把解出的象函数取拉普拉斯逆变换得到象原函数，即微分方程的解。

将被传染的各代群体的基数作为群体的特征参数，就可以分析网络涌现特征的数学表达式，计算爆发区的范围和爆发期的传染规模。这些都在 SM 小世界网络数学模型研究中得到体现。

2.2.2. 时间序列与预测

著名控制论专家维纳在《控制论》第三章“时间序列，信息和通讯”中关于时间序列预测问题时指出：时间序列“它的全部过去能够用可数个量的集来决定。例如，对于范围很广的一类函数 $f(t)$ ($-\infty < t < +\infty$)，当已知一组量

$$a_n = \int_{-\infty}^0 e^{t^n} f(t) dt \quad (n = 0, 1, 2, \dots)$$

时， f 就完全被决定。”特别，如果 a_0, \dots, a_n 全部给定， A 的分布就可以决定。这里，我们要用到到尼克杜(Nikodym)关于条件几率的著名定理。由这个定理可知，在很普遍的情况下，这分布当 $n \rightarrow \infty$ 时收敛到一极限，而这个极限分布将提供我们关于任一未来量分布的全部知识，同样，如果知道了过去，我们就可以决定任何一组未来量分布的全部知识，同样，如果知道了过去，我们就可以决定任何一组未来量的数值的同时分布，或任何一组决定于过去和未来两方面的量的数值的同时分布。因此，如果我们对任何统计参数或统计参数集的“最优值”都能做出适当的解释——多半是在平均值或中值或众值的意义上，我们就能由这个已知分布把他们计算出来，得到一个就其优良程度而言能够满足我们所要求的预测准则的预测([6], p. 70~71)。

维纳又指出：“我们能够计算因固定过去而提供我们的关于任一统计参数或统计参数集的信息量。由关于过去的知识，我们甚至能计算某一瞬间以后全部未来的总信息量。由于我们能由一般的过去推知未来，所以，即使上述的瞬间就是现在，我们关于现在的知识也会包含无限大的信息量([6], p. 71)。”

维纳还指出：“离散时间序列的理论在许多方面都比连续时间序列理论更简单([6], p. 91)。”

这一系列的预测理论与公式，为建立仿真模型和预测提供了有力的理论根据。

2.3. 新型冠状病毒传播模型的仿真

2.3.1. 仿真方法的应用

仿真是通过有所研究的经济系统估计的模型计算值来确定系统的行为和性质。如果模型充分地说明了各种经济关系，那么就可以用计算机进行数值研究，根据不同的假设可以确定系统的数值状态。在仿真预算中，内生变量的数值是利用所有的政策变量及其他外生变量、模型参数的估计值、随机扰动项及模型参数的估计值等数据由模型方程计算出来的[7]。

仿真运算有多种形式，预测仿真是根据历史数据，计算现在时点以后的数值，用于预测未来。

新冠病毒传播的预测，就是根据已有数据预测未来近期和远期的数据。而根据每日发布的各省市数据，实时加入新数据，在与时俱进的状态下，微调系统预测模型的参数，逐渐滤除扰动项而逼近系统变化的真实状态，从而提高预测的实用价值。在笔者的大量预测图表中，分析，比较精度误差和真实数据的状态。因此。误差数据也是财富，具有洞察复杂网络系统中病毒传播现象的规律趋势的参考意义。最求过高的精度带来宏观预测的难度和计算数据缺失困惑，而实时更新统计数据的原始资料才是科学预测前提的意义所在。这里，时间序列系统数据的增加提高了系统的测度，给予系统更多的信息熵，因此，如同天气预报，在动态跟踪下，产生依据客观规律的预测报告。

由于仿真是模拟客观真实物状态发展的规律。很多现象，例如：病毒在生物种群的传播，有的是人

类所不能控制的，因此，必须慎重考虑人为数据变量与增加项。基本预测算法可以多样，但数据事实和规律不变。

笔者按以上原则，进行了几种传播速率的仿真计算，以用于仿真比较。

2.3.2. 新冠病毒传播的仿真计算

新型冠状病毒网络传播的模拟采用以下假设与计算进行。

1) 区域易感人群总数为 K ，设定为 1 万人每个人接触传播的人数为 3~5 人。采用随机关系的任意 3~5 个成员为传染对象则平均传染系数 $R = 4$ ，即一人随机传染 4 人。

可使用计算机模拟生成各个点的传染数据(3 人，4 人，或 5 人)的传染过程，从而形成有向网络。类似于朋友圈，每一层是新增感染人数，则构成 $1 \sim n$ 个不相交的集合。各顺序集合的基数构成 $Y(X_1, X_2, X_3, \dots, X_n)$ 的分布列，对这一分布列按皮尔曲线模型进行非线性回归分析，就获得参数 a ， b 。

2) 模拟计算第 1 代传播数据集，第 2 代传播数据集，……，直到第 n 代，不再有增量为止。

当 n 达到一定正整数时系统全部人员被感染。这里传播感染的代，也就是 SM 小世界的网络直径。

3) 统计计算数据集集合的基数。

4) 以集合基数为 Y ，以集合层数为 X ，进行数学模型的计算，根据以前的研究，采用皮尔曲线模型非线性回归计算获得属性参数 a 与 b 。仿真使用的 SM 小世界模型用皮尔曲线描述。

皮尔曲线是 1938 年比利时数学家哈尔斯特(P. F. Verhulst)首先提出的一种特殊曲线。后来，近代生物学家皮尔(R. Pearl)和 L. J. Reed 两人把此曲线应用于研究人口生长规律。所以这种特殊的曲线也称之为皮尔增长曲线，简称皮尔曲线。其时间序列倒数的一阶差分的环比为一个常数，可以用皮尔曲线来描述。皮尔曲线又称为 logistic(中文译为逻辑斯蒂克)曲线，其数学表达式又称为 Logisitc 数学模型。其曲线如图 1 所示。

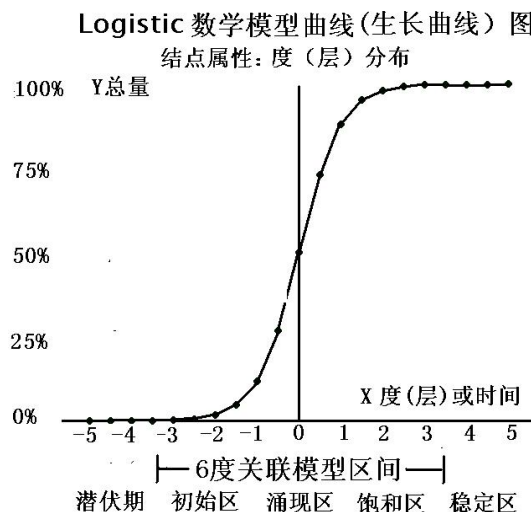


Figure 1. Logistic mathematical model curve

图 1. Logistic 数学模型曲线

皮尔曲线广泛用于各种自然科学与社会科学的研究中，笔者发现：皮尔曲线与六度关联模型具有高度的类比性，皮尔曲线用于描述小世界网络模型是一种很好的数学方法。当经济变量的发展变化表现为初期增长速度缓慢，随后增长速度逐渐加快，达到一定程度后又逐渐减慢，最后达到饱和状态的趋势，可以用皮尔曲线来描述。因此，皮尔曲线的预测法是根据预测对象具有皮尔曲线变动趋势的历史数据，拟合成一条皮尔曲线，通过建立皮尔曲线模型进行预测的方法。皮尔曲线在水平轴为互连网络有向图的

距离和关系时使用的整数。六度关联模型使用的可就是 7 个分布在横轴的整数点划分形成 6 个段 3 个区间。几乎所有的集合结点，都在这六度(层)区间内(99.74%)。所以，六度关联是一个自然科学与社会科学的普遍规律。

皮尔生长曲线的一般模型为：

$$y = \frac{k}{1 + e^{f(x)}}$$

式中， K 为常数， $f(x) = a + bx + cx^2 + \dots$ 在应用中略去平方项以后数据，则 $f(x) = a + bx$ 。这就是 SM 小世界网络数学模型，即有限集合群与群体关系变化的规律。模型参数解释如下：

- K : 论域中元素总数；
- x : 关系层次(度)(或时间)，
- y : 被感染总数相当于有向网可达点；
- a : 水平偏移量，数据初始属性；
- b : 曲线飞升速率与感染速率相关。

这个模型前面已经定义为 SM 小世界网络模型。所以曲线称为小世界模型曲线。SM 小世界模型函数的曲线也是皮尔曲线的水平右移后的区间应用，使用，定义 X 变量从 1 开始，表示 1 度(层)传染数，按平均传染系数 R 传播到第 2 层、第 3 层。

因此，SM 小世界网络模型。本质上是皮尔模型在一定条件下的特例。

SM 小世界网络模型就是根据皮尔曲线进行网络的度(层)及其分布计算的。

这也就是把皮尔模型曲线和六度关联模型曲线以及高斯分布函数曲线等一起类比分析的初步研究。用高斯正态分布函数曲线的 7 个离散整数数据列，以六度相关模型类比的而生成的图形曲线是网络六度关联属性。见图 2——复杂网络结点六度关系图。

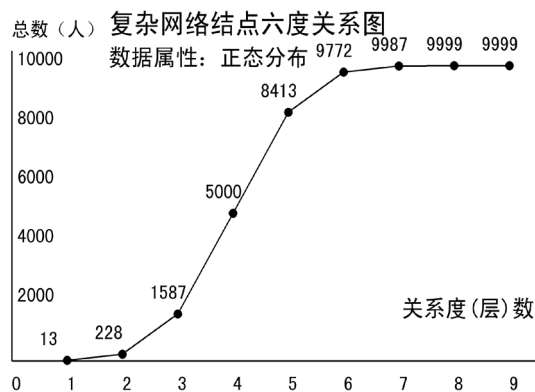


Figure 2. Six degrees diagram of normal distribution Small World of complex network

图 2. 复杂网络正态分布小世界六度关系图

SM 小世界网络模拟生成方法，以结点出度为 3~5 的概率随机联系其他随机点所构成。

算法如下：从 1 到 1000，每次加 1，按概率选用出自己(感染输出点)以外的 9999 点中的点作为感染输入点，选择以 3 至 5 的传染系数进行，这时的边成为有向边予以存储备算，被感染点的选择随机，传染个数的选择随机。生成 SM 小世界网络模拟数据集。图 3 和图 4 两图是模拟传染模型一个人传染 3~5 人的总量曲线图和增量度(层)分布图。用以模拟真实网络随机传染病毒流动的过程。从图中可以明显看出传染病爆发涌现段的数据以及变化过程。图 3 是传染系数为 2.5 时病毒传播模拟感染病例总人数趋势图，模型函数参数为：

系数 $K = 10000$;
 系数 $a = 9.7117265$;
 系数 $b = -1.6775001$;
 回归系数 $r = -0.9953854$;
 方差: 14583.395054708。

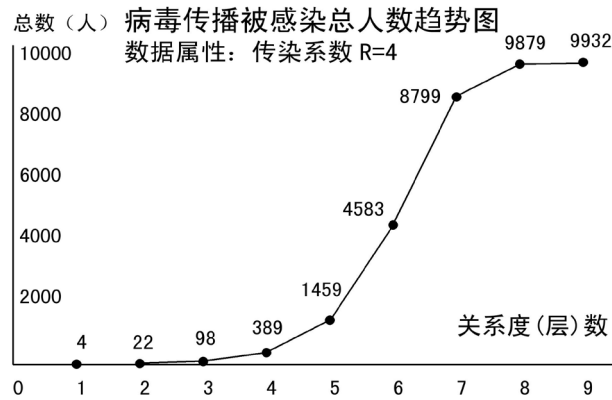


Figure 3. Trend of the total number of simulated cases of virus transmission
图 3. 病毒传播模拟感染病例总人数趋势图

病毒传播模型的增量模型如图 4 所示，实质是总量函数的微分曲线。

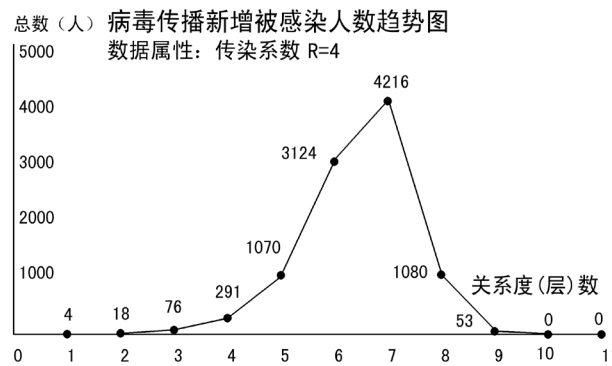


Figure 4. Trend chart of the number of new infected people in virus transmission
图 4. 病毒传播新增被感染人数趋势图

在本文的新型冠状病毒的统计预测分析中，就使用了与以上的网络的数学模型。

网络模型用于新型冠状病毒的重要参数：

- K : 最终感染人群；
- A : 曲线传布的初始区偏移；
- B : 曲线斜率是与传播速率有关的参数，绝对值越大传染能力越强；
- X : 传播层次与传播时间；
- Y : 各时间段被感染的人总数。

由于最终感染人数需要根据各个时间节点的数据移动建模分析，根据已有数据和小世界网络特性曲线数据进行拟合计算，由于目前被感染人数已经达到 6000 例，因此根据小世界网络特性曲线规律，可以

预测总人数将超过 1 万 2 千人。笔者计算出乐观、一般和较悲观的 3 种情况的预测结果。可供随着新数据的出现而随时递推计算获取更新数据后的最新预测结果。

对于不同感染系数 R_0 在传播过程各代的传播模型预测数据见表 1 所示。

易感人数占总人口的百分比，各地区有所不同，本表只对易感人群中病毒传播的状态，按论域总数为 1 万人测算。可以看出，不加控制的感染结局是全体被感染。这里暂不研究治愈率和病死率的问题。

Table 1. Total number of new coronavirus infections in SM small world model (unit: person)

表 1. SM 小世界模型新冠病毒万人感染总数测算表(单位: 人)

感染系数	第一代	第二代	第三代	第四代	第五代	第六代	第七代	第八代
$R_0 = 2$	2	6	14	30	61	123	246	486
$R_0 = 3$	3	12	397	120	358	1042	1042	2776
$R_0 = 5$	5	30	154	752	3222	8455	9996	9996
$R_0 = 7$	7	56	392	2519	8623	/	/	/
$R_0 = 9$	9	90	788	5267	9980	/	/	/
$R_0 = 11$	11	132	1365	7984	/	/	/	/

注：“/”符号表示所有人(9999 人)被感染，开始病毒带入者人数 1。

3. 新型冠状病毒统计与预测分析

3.1. 新型冠状病毒的传播模型

2020 年 1 月 22 日中国疾控中心答记者问时，有记者问，“有一家英国的大学研究表明，根据上周的流行病学研究，预计有 2000 人在中国受到感染，有没有预测过，现在有多少还没有得到治疗的感染人群？”中国疾控中心回应：这位科学家的计算结果是一个数学模型，他算出的数字区间的最大数。欢迎大家提出各种模型。模型是不是和事实相符合，随着大家对这个病毒的认识，会去验证这个模型。目前掌握的情况，并不是这位科学家计算的模型情况。(资料来源：<http://www.chinanews.com/gn/2020/01-22/9066968.shtml>)

那么，从目前实际发展情况可以看出：那家英国大学的研究模型，不够准确，还是远远低估传染病的传播速率。

2020 年 1 月 24 日凌晨 2 点，世界卫生组织发布声明，中国已经报告了“武汉的第四代病例、武汉以外的第二代病例”。“第四代病例”指的是，病毒之间的四次人际传播，比如 A 传给 B，B 传给 C，C 传给 D，D 传给 E。世卫组织表示，人际传播正在发生，初步传播速率估计值为 1.4~2.5 之间，即一个宿主可以链接式传播 1.4~2.5 人。(资料来源：<https://tech.ifeng.com/c/7tZah66gW5w>)

基于建立多个模拟传播网络的参数基础上：

1 人传播 2 人则 $b = -0.8$;

1 人传播 3 人则 $b = -1.5$;

1 人传播 3~5 人则 $b = -1.67$ 。

武汉的第二代病例和武汉以外的第二代病例中的代，在复杂网络和数学图论中，就是传播过程中的层(代，次，度等)，在易感人群中的个体就是图论中的结点，传染与被传染关系就是离散数学中的集合论中的关系。也是图论里有向图中的有向边。

预测数据来源于国内公开发布资料，对每日发布的数据以 0 时~24 时国家卫健委发布为准。前期缺少的极个别数据，以数学方法内插和外推决定。

根据已有数据和笔者建立的小世界网络模型公式，可以进行统计分析：

SM 小世界网络模型的数学模型使用的 Y 是有限集合中的总体，X 是病毒传播的层次(或称为六度中的度)。在宏观大数据的条件下，很难获得每个个体的传播路径与层次，只能采用平均数。即使如此，也不便于进行时间序列数据的分析研究。采用时间序列变量为横轴 X，建立日期与传染层次的映像关系。方便了观察动态变化和趋势分析。因此，模型以度(层)为横坐标和以日期为横坐标的计算模型参数集有所区别，可以互相变换。

3.2. 新冠病毒预测

以半月后最终感染总人数和钟南山的论述进行推断，进行了三种情况(乐观、一般和悲观)的逐日感染总数分析。结果见下表。为简化，以 2 日为间隔。

图 5 是 1 月 30 日感染总人数短期趋势预测图，图 6 是新增感染人数趋势预测图。

分析预测函数曲线图图形如下：

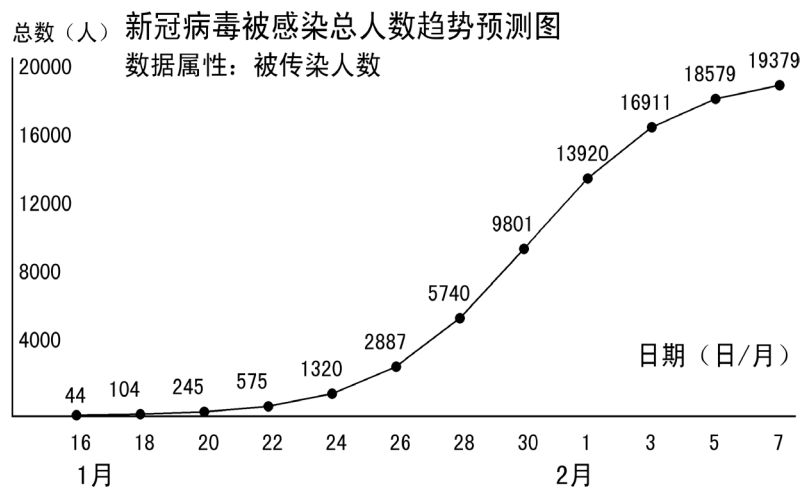


Figure 5. Novel coronavirus pneumonia infection total population trend forecast
图 5. 新冠肺炎病毒被感染总人数趋势预测图

函数曲线的数据如下： $a = 7$ ， $b = -0.87$ ， $k = 20000$ 。

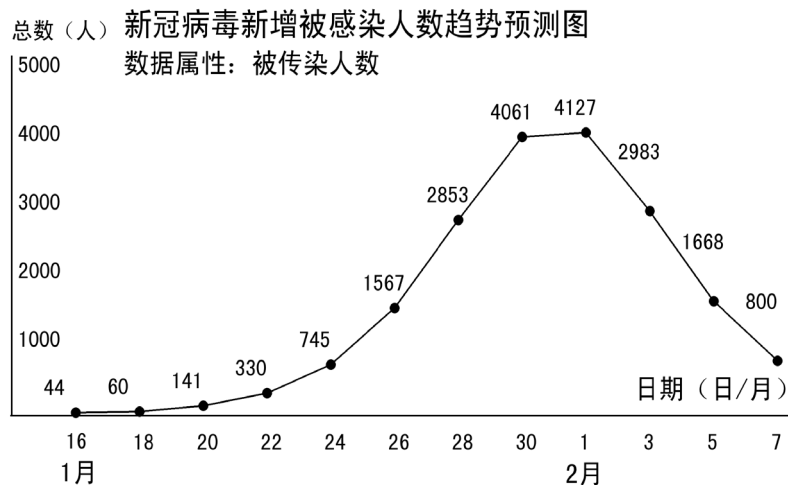


Figure 6. Novel coronavirus trend prediction chart for newly infected population
图 6. 新冠病毒新增被感染人数趋势预测图

3.3. 预测结果精度分析

预测得到的一组预测值同实际对应比较后,即可知预测结果的精确度。衡量预测值的精确程度的指标称为预测精度。预测精度有多种不同的定义,主要决定于如何“比较”测量值与实际值[8]。

本文采用预测值减去实际值的差与实际值比的百分数表示。短期预测方案见表 2,长期期预测方案见表 3。误差比较由后来时间序列数据补充后算出。预测时间:2020 年 1 月 29 日。所有实际值数据由中国国家卫健委权威发布。

Table 2. Accuracy analysis of short-term prediction of the total number of new cases of coronavirus (unit: person)

表 2. 新冠病毒累计确诊病例总人数近期预测精度分析表(单位:人)

时间	1 月 30 日	2 月 1 日	2 月 3 日
预测值	9801	13,928	16,911
实际值	9692	14,411	20,471
误差	+1.1246%	-0.9378%	-17.3905%

注:误差为+号表示高于实际,为-号低于实际值。

分析表 2 预测精度表明,近期未来预测时点越近,精确度越高,误差变化的速率说明。对新冠病毒的传播估计预测数过于乐观。远远低估了传染流行的严重性。

表 3 是 1 月 30 日对最终总传染人数近 4 万 5 千人的预测。从发展趋势,最终将超过 7 万人。

Table 3. Long term prediction accuracy analysis of total number of new cases of coronavirus (unit: person)

表 3. 新冠病毒确诊总人数远期预测与精度分析表(单位:人)

时间	2 月 4 日	2 月 5 日	2 月 6 日	2 月 7 日	2 月 8 日	2 月 9 日
预测值	25,054	28,838	32,269	35,218	37,640	37,556
实际值	24,363	28,060	31,211	34,598	37,251	40,235
误差	+2.84%	+2.77%	+3.40%	+1.04%	+1.79%	-1.69%

注:误差为+号表示高于实际,为-号低于实际值。

使用六度(层)关联小世界数学模型分析的结果表明:被感染总人数与时间的关系函数的二阶导数为 0 的数据点,也就是新增传染趋势下降的关键点,也是一阶导数数值最大点。

通俗的解释:日新增感染人数最大时那天的确诊总人数,就是远期预测全部战胜病毒感染后确诊总人数的二分之一。以 2 月 9 日的数据为例。预测总感染人数将达到 7~8 万人。

根据 2 月 9 日最新数据分析,目前已经达到新增变量的峰值。未来半月总量增加趋缓,新增病例逐渐减少,是防控取得重要进展的重要标志。

防控速度与病毒传播速度赛跑。由群体传染特性,可以分析出目前的防控措施是有效的,在统一领导下,还有必要进一步强化切断传染途径。扩大孤点群体,即分社区强制隔离等。

4. 结语与讨论

在强大自然规律面前,人类只能认识规律,调整应对措施,新型冠状病毒传染规律与控制措施,将在人类与自然的相处中得到教训、规律和认识。

4.1. 小世界核心网络的度分布问题

综上所述,斯坦利·米尔格兰姆小世界六度关联网络模型是一个普遍的反映自然科学与社会科学的基本规律的模型,不同结点的度(层)分布完全可以模拟传染病传播规律。

笔者认为完全可以模拟传染病传播的随机现象,并用于有向图研究所构成传染层次的计量预测中。把小世界模型的分布规律与数理统计中的高斯分布,泊阿松分布,指数分布以及二项式分布等作维度空间的映像变换进行比较,得到曲线族群的距离。从而完成基于时间系列大型状态变量矩阵的进一步分析研究。

4.2. SM 小世界网络分布各个节点属性距离问题

小世界网络度各个节点度分布与高斯分布,泊阿松分布,二项式分布、指数分布有相关关联关系。不同的结点初始区和饱和区属性不同,而涌现区的数据则反映群体属性,也是集合各节点的共性。反映群体相关作用的动力学模型数据与所涉及曲线族空间的距离问题,小世界六度(层)关联的数据链预测问题。每一个结点方程数学模型的属性数据、空间关系与距离分析,还有待于进一步深入研究。高聚类系数结点对缩小网络直径和对网络属性的影响等的研究,将为指导和解决复杂网络模拟分析预测与工程控制(例如舆论传播模型,生物工程的控制模型)提供扎实的理论决策与思考。

4.3. 大型社会卫生事件现代控制工程的模型与综合预测决策问题

有必要引进大数据技术,将宝贵详实数据的金子挖掘出来。中华民族是一个伟大的民族,亿万人在国家领导决策下的控制工程,将为人类社会积累下宝贵的经验。有关涉及以立体多维矩阵的总量对各变量的偏微分研究。矩阵表述各个不同物理位置以时间序列状态方程组所列函数的详细解析问题,都有待于专家的进一步研究。从而为世界科学技术的发展做出中华民族的伟大贡献。

4.4. 大规模数据集问题

人类与病毒赛跑,迅速、详实较为准确全国各类基层统计大规模数据集的工作,是战胜病毒的预测决策需要。是一项涉及人们生命健康的第一大事,也是对全国新冠病毒传播状态的分析掌控的重要基础资料。应用现代控制理论和数学模型以及大数据技术分析新冠病毒的传播模型与现代大型社会控制工程策略,可以为国家最高决策层提供科学的决策依据。对系统构成大规模的数据集,可以集中国家最优秀的医务工作者、统计工作者和数学工作者等各行业专家群体智慧和国家计算机总体模型计算模拟仿真能力,为涉及全国各省市地区抗击新冠病毒提供更具体的参考意见。

对新型冠状病毒传播模型和与预测是一项时间性极强的工作。笔者的初步研究希望得到前贤、专家和读者的指正。有关论述的缺欠和瑕疵希望能得到批评,以有利于开展进一步的深入研究。

致 谢

本研究过程获得哈尔滨工业大学导师姜守旭的殷切指导与帮助,获得北京中医学院李晓莉导师收集整理医疗与统计资料支持。在此一并致谢。

参考文献

- [1] 汪秉宏,李平.小世界网络浅介[J].现代物理知识,2016(3):51-55.
- [2] 冯志伟,胡凤国.数理语言学[M].北京:商务印书馆,2012:332.
- [3] 杨炳儒.图论概要[M].天津:天津科学技术出版社,1990:184.
- [4] 孙玺菁,司守奎.复杂网络算法与应用[M].北京:国防工业出版社,2016:100.
- [5] 徐利治.数学方法论选讲[M].武汉:华中工学院出版社.1983:15,24,28.
- [6] 维纳.控制论[M].北京:科学出版社,1962:70-71+91.
- [7] 张寿.于庆文,编著.计量经济学[M].上海:上海交通大学出版社,1984:562.
- [8] 吴凤山,等,编.实用预测技术[M].北京:社会科学文献出版社,1986:250.