

# 冠状病毒研究不均衡性分析

杜明浩 熊凯路 席浴辰 刘昱含

**摘要** 本文以论文数作为研究程度的度量,综合运用时间序列分析、网络分析、动力学模拟等方法,方差、变异系数等统计量,分析揭示了冠状病毒研究在时间、病毒门类、国家上的不均衡性,建立了流行性病毒来袭时和控制后相关研究的增长和衰退模型。

**关键字** 冠状病毒 不均衡 时间 门类 国家 动力学

## 1 引言

2020 年,一场席卷全球的新型冠状病毒(covid-19)来袭,截止 2020 年 7 月 18 日,已经有上百个国家超过 1400 万人患病,死亡人数高达 60 余万([https://www.worldometers.info/coronavirus/?utm\\_campaign=homeAdvegas12](https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas12)),这可以说是近一个世纪以来全球范围内影响最大的事件。这已经不是冠状病毒第一次在世界范围内产生如此重大的影响,2003 年 SARS,2012 年 MERS 等冠状病毒都曾给人类带来巨大灾难。

随着病毒的飞快传播,感染人数的增加,越来越多的人开始关注冠状病毒,也有越来越多冠状病毒的有关研究出现(Haghani and Bliemer, 2020),甚至有的大学专门成立了相关的公共卫生学院。冠状病毒的研究是否存在不均衡的现象,在疫情时和疫情结束后,流行性病毒的研究趋势如何等问题尚不清楚。对上述问题的解答不仅有助于在疫情到来时更好地指导疫情的攻关工作,也有利于在疫情结束后更好地进行疫情的防控工作。

本文研究了冠状病毒在时间、病毒门类、研究国家上存在的不均衡现象,简要分析了造成不均衡可能的原因,揭示了冠状病毒等流行性病毒在病毒来袭和病毒控制后其研究趋势变化的统一规律,为这类流行性病毒研究和防控工作的指导提供了参考依据。

## 2 数据和方法

本文所有的使用论文数数据均来自 PubMed (<https://pubmed.ncbi.nlm.nih.gov/>),PubMed 上所收录的数据主要为医学数据,这也就保证了我们在进行分析时,主要是分析病毒的具体研究,并不涵盖一些如经济学、社会学等的研究。所搜集的方法均是以某一病毒为关键字进行检索,凡在论文的标题、作者、摘要、关键字中出现的,均为关键字,不区分大小写,这样可能会产生重叠,比如一篇论文的关键字中可能含有两种病毒,则该论文将出现在两种病毒的研究中,我们认为其所产生的误差可以忽略。

我们的任务是分析冠状病毒研究的不均衡性,故应首先寻找出可以定义“研究”的物理量来,我们选择使用论文的数量来衡量研究的程度,所有论文地位平等(不分期刊级别、引用次数),只看论文数量。即发表论文数越多,证明对冠状病毒的研究程度越高。

关于文献计量学的研究，还有一个很重要的因素，时间。且我们想要分析冠状病毒在时间上的不均衡，构建其来袭、控制后时的研究趋势模型，时间也是必不可少的参数，因此就必然要进行时间序列分析。在建模时，也可能要进行论文数增长、衰退的动力学分析。而且，既然我们要分析冠状病毒门类、国家上的研究不均衡性，病毒门类和论文作者的国家也是十分重要的数据。

定义了“研究”之后，现在最大的问题是怎样证明出研究的不均衡性甚至不均衡的程度。故而不均衡，我们也需要用指标来衡量。分位数、方差、变异系数，以及一些新定义的指数（见下文），都是可以用来衡量数据离散和不均衡性的统计量，我们可以借助他们来定量的刻画出冠状病毒研究不均衡的程度。

我们的整体思路为，先收集不同年份的冠状病毒论文数量，同时需要论文所研究的病毒门类，论文作者的国家等数据。进而用一系列统计量来衡量出冠状病毒研究的不均衡，使用增长、衰退的动力学模型来构建病毒研究来袭、控制后的研究趋势。

### 3 结果

#### 3.1 冠状病毒时间不均衡研究

我们分别搜集了以“coronavirus”，“coronavirus”和“sars”，“coronavirus”和“mers”为关键字的论文数，以时间为横轴，论文数位纵轴，绘制了论文数随时间的变化曲线（图 1），进行了简单的时间序列分析。

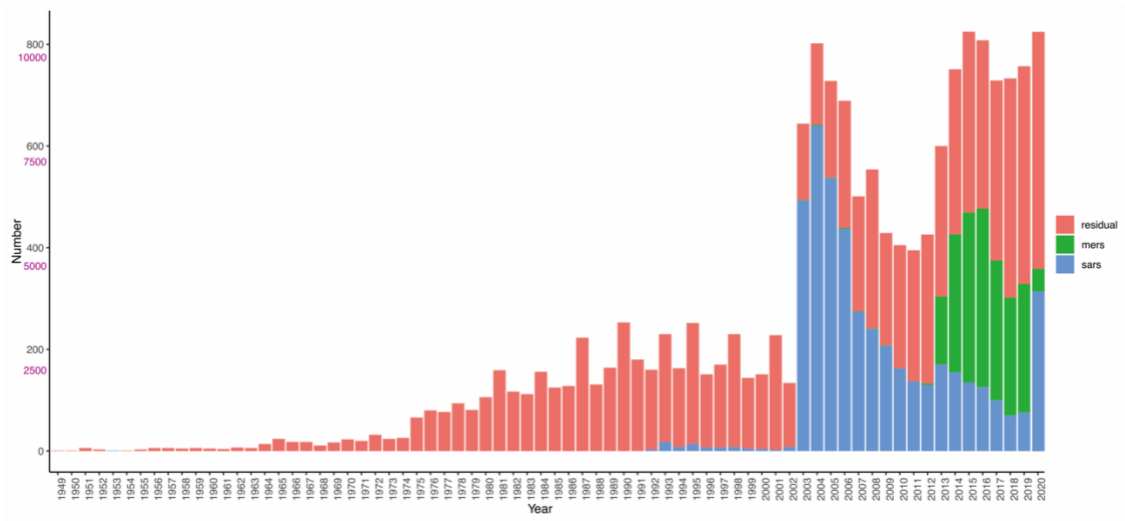


图 1 冠状病毒研究论文数随时间变化的直方图，其中横轴为年份，纵轴为论文数，蓝色的表示关键字中含“sars”的论文，绿色的为关键字表示中含“mers”的论文，红色的为除去这两者的其余论文，紫色的刻度表示 2020 年的论文数

我们将冠状病毒的研究分为五个阶段：

- （1） 1919-1974：这一阶段的特征为关于冠状病毒的研究较少，总体上呈现增长的趋势，但是增长不多。
- （2） 1975-2002：1975 年较之前的研究有了一个突然的上升，且在之后，逐年缓慢上

升。

- (3) 2003-2012: 2003 年再次有一个突然的上升, 冠状病毒的研究在 2004 年达到一个高峰, 然后减缓, 在 2006 年下降最快。
- (4) 2013-2019: 2013 年有一个突然的上升, 随后论文数基本不变。
- (5) 2020: 2020 年论文数再次突然增长, 在不到半年的时间, 2020 年发表的论文数已经快与之前发表的论文数总和持平。

### 3.2 冠状病毒门类不均衡研究

我们分别搜集了以 PubMed 上以“HCoV-229E”, “HCoV-NL63”, “HCoV-HKU1”, “SARS-CoV”, “MERS-CoV”, “COVID-19”为关键字的论文, 绘制论文数随时间的曲线图 (图 2)。

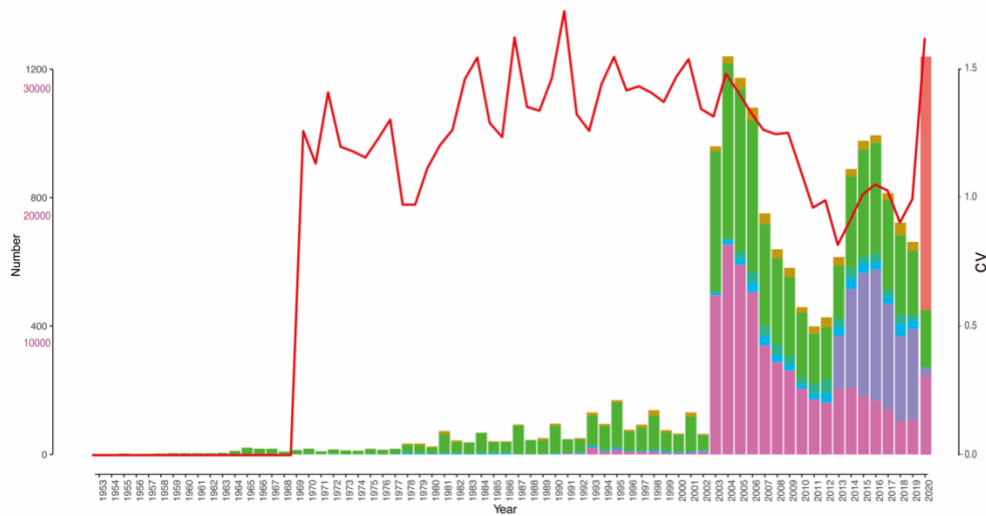


图 2 不同门类冠状病毒论文数随时间变化的直方图和变异函数曲线, 其中横轴为年份, 左纵轴为论文数, 紫色的刻度表示 2020 年的论文数, 右纵轴为变异系数值

我们可以很显然的看到冠状病毒门类每年的研究都是不均衡的, 表现在对于“HCoV-HKU1”, “SARS-CoV”, “MERS-CoV”的研究数远高于其他 (图 2)。我们使用变异系数来衡量不均衡的年变化。变异系数的好处是可以忽略每年论文数总量的不同, 而只考虑冠状病毒不同门类研究的相对差异, 其公式如下:

$$CV = \frac{\sigma}{\mu}$$

变异系数 1969 年之前为 0, 其余的值基本都在 1-1.5 之间。1969-1991 年呈逐步上升趋势, 门类研究不均衡逐渐变大; 1992-2002 年变异系数基本不变; 2003-2013 年, 变异系数逐渐变小, 研究不均衡变小; 2013-2015 年, 变异系数变大, 2016-2019 年又变小, 2020 年变异系数较之前升高很多。

### 3.3 冠状病毒国家研究不均衡分析

我们分别搜集了 PubMed 上以“coronavirus”, “coronavirus”和“sars”, “coronavirus”和“mers”为关键字的论文, 构建了冠状病毒研究国家合作网 (图 3), 绘制了国家-论文数直方图 (图 4), 计算了不同国家论文数的分位数、平均数、标准差、变异系数等统计量, 并使用

e 系数和相关系数来度量国家 GDP 和患病数对国家冠状病毒研究程度的影响（图 5，表 1）。

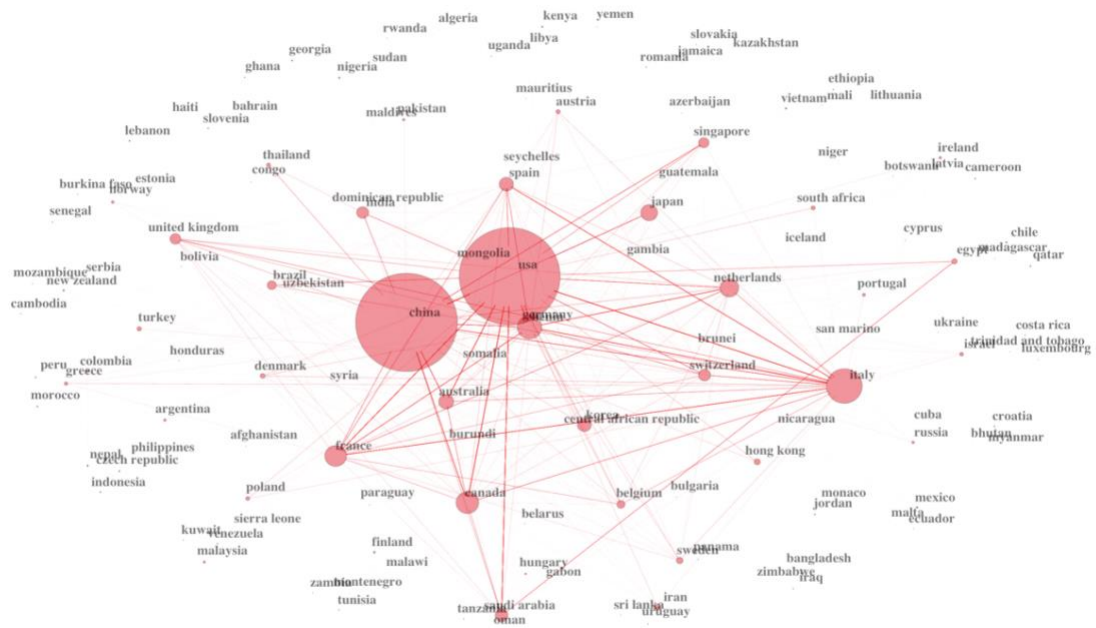


图 3 冠状病毒国家论文合作网络

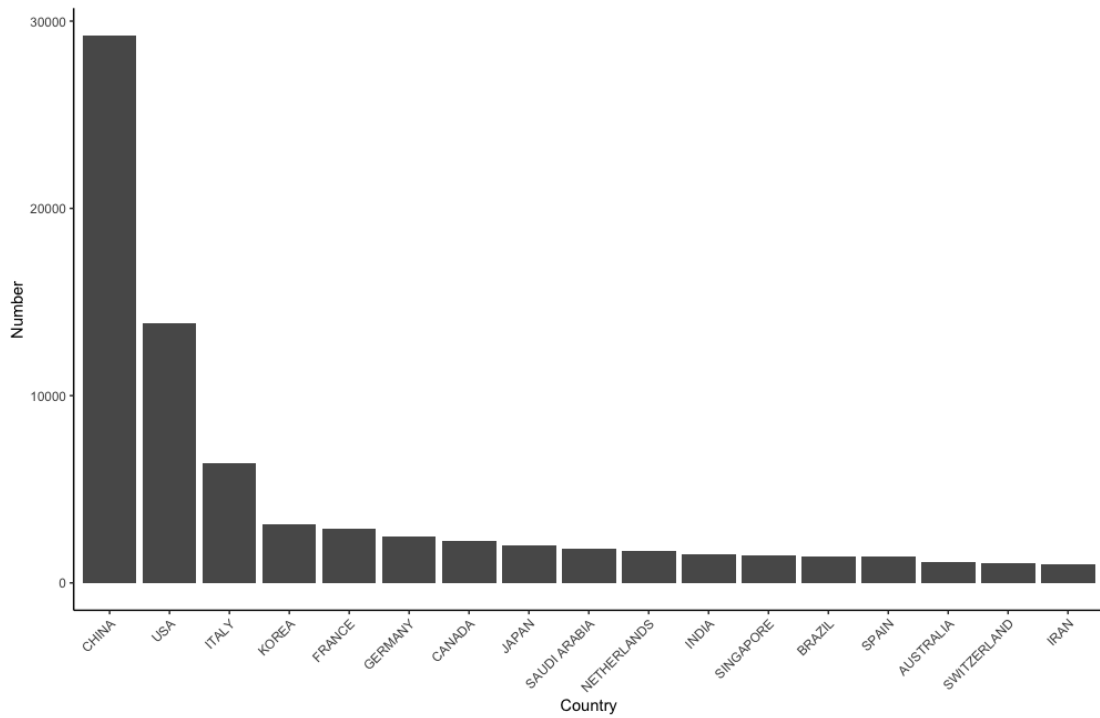


图 4 研究论文数大于 1000 的国家的冠状病毒论文数直方图

患病数可能引起不同国家研究不均衡，一个国家患病数越多，其对冠状病毒的研究理应越多，我们定义  $e_d$  的计算公式如下：

$$e_d = \begin{cases} -\frac{d-p}{\sqrt{d}} & \text{当 } d > p \text{ 时} \\ 0 & \text{当 } d = p \text{ 时} \\ \frac{p-d}{\sqrt{p}} & \text{当 } d < p \text{ 时} \end{cases}$$

$$e_g = \begin{cases} -\frac{g-p}{\sqrt{d}} & \text{当 } g > p \text{ 时} \\ 0 & \text{当 } g = p \text{ 时} \\ \frac{p-g}{\sqrt{p}} & \text{当 } g < p \text{ 时} \end{cases}$$

$e_d$  的值越接近 0，说明某一国家患病数和论文数排名差别不大， $e_d$  的值为正，且越大，说明该国家论文数高于患病数，证明这个国家冠状病毒的研究程度高。 $e_d$  的值为负，且越小，说明该国家论文数小于患病数，明这个国家对冠状病毒的研究程度低。

同理，GDP 的  $e_g$  和患病数一样，只不过将患病数  $d$  改为 GDP 的值  $g$ ，这里不再赘述。

国家合作网络的建立方式如下：

节点：国家

连边：两个国家的研究人员曾发表过同一篇论文，则这两个国家之间形成连边。

表 1 各病毒模型患病数/GDP 与论文的相关关系

模型		斯皮尔曼相关系数	肯德尔相关系数
Sars	患病数与论文数	0.4705	0.3663
	GDP 与论文数	0.7751	0.5993
患病数大于 3 的 Sars	患病数与论文数	0.4057	0.2732
	GDP 与论文数	0.7185	0.5550
Mers	患病数与论文数	0.0545	0.0341
	GDP 与论文数	0.7279	0.5286
患病数大于 3 的 Mers	患病数与论文数	0.3667	0.2222
	GDP 与论文数	0.6000	0.3889
Covid-19	患病数与论文数	0.6616	0.4811
	GDP 与论文数	0.8354	0.6584
论文数大于 20 的 Covid-19	患病数与论文数	0.5587	0.3896
	GDP 与论文数	0.7714	0.5859

### 3.4 流行病毒增长、衰退动力学模型

我们搜集了以 PubMed 上以“Severe Acute Respiratory Syndrome”，“Middle East respiratory syndrome coronavirus”，“Influenza A virus subtype H1N1”，“Ebola virus disease”以及“COVID-19”为关键字的论文，分别以月和年为横坐标，绘制了论文数随时间的变化曲线（图 6）。月份截止到疫情基本控制的月（SARS：2003 年 5 月；MERS：2015 年 8 月；H1N1：2010 年 8 月；Ebola：2015 年 1 月；COVID-19：2020 年 6 月），年则以上述疫情得到控制的月开始，往后计 12 个月，如 SARS 的第一个年为 2003 年 5 月-2004 年 4 月。

论文数在最初的几个月（如 SARS 的第一个月，MERS 的第一个月，Ebola 的前三个月）数量很多，随后减少，进而开始呈现一个先快速增长，后缓慢增长，最后保持平稳的趋势。据此特点，我们使用逻辑斯蒂增长模型来描述疫情到来时论文数随时间的增长模型。

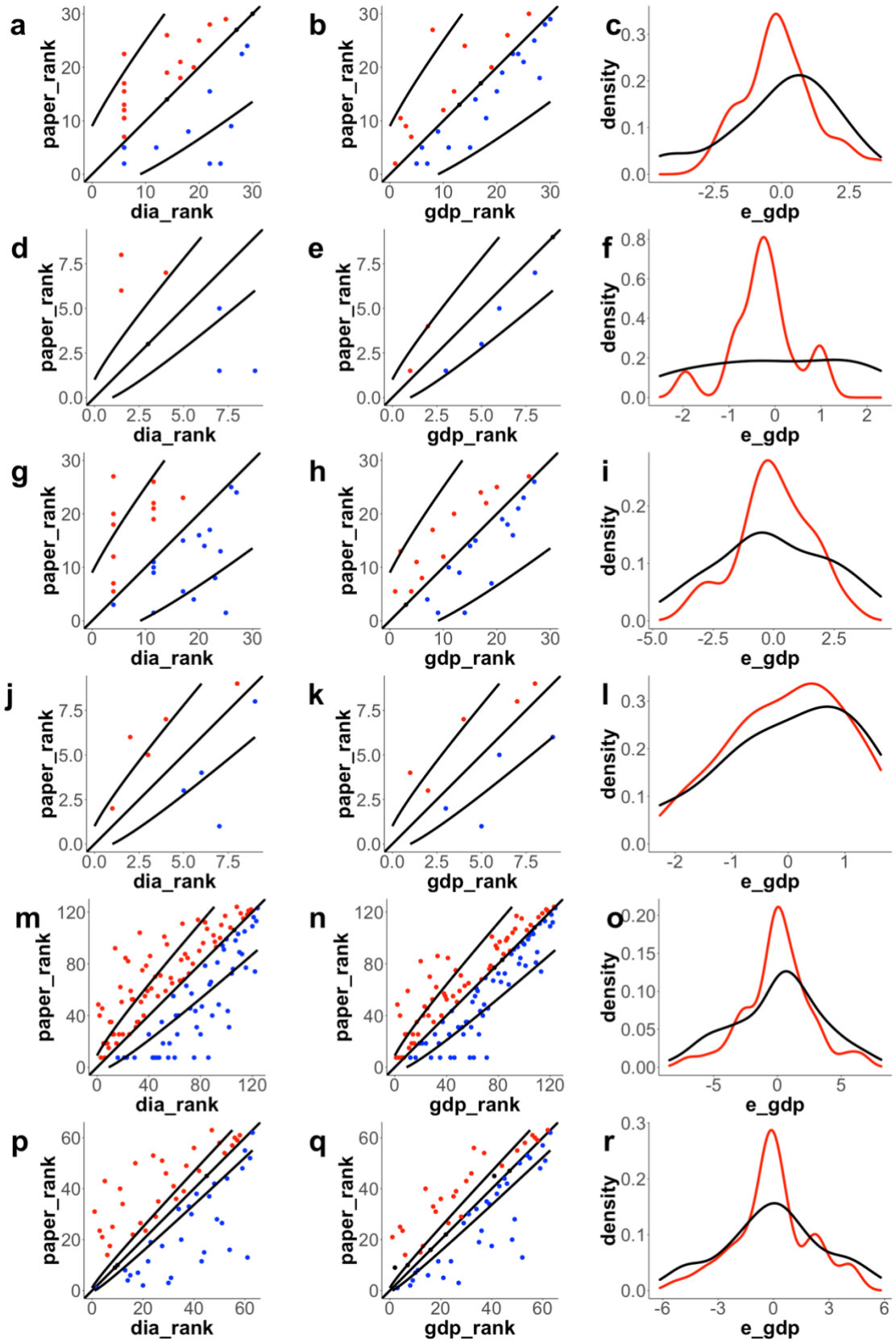


图5 患病数/GDP 和论文数排名的散点图和  $e_d$ ,  $e_g$  概率分布曲线, 第一列散点图中黑色的曲线表示  $e=3$ , 第二列中黑色的曲线表示  $e=1$ 。(a) Sars 的患病数和论文数排名散点图 (b) Sars 的 GDP 和论文数排名散点图 (c) Sars 的  $e_d$ ,  $e_g$  概率密度分布曲线 (d) 患病数大于 3 的 Sars 的患病数和论文数排名散点图 (e) 患病数大于 3 的 Sars 的 GDP 和论文数排名散点图 (f) 患



病数大于 3 的 Sars 的  $e_d$ ,  $e_g$  概率密度分布曲线 (g) Mers 的患病数和论文数排名散点图 (h) Mers 的 GDP 和论文数排名散点图 (i) Mers 的  $e_d$ ,  $e_g$  概率密度分布曲线 (j) 患病数大于 3 的 Mers 的患病数和论文数排名散点图 (k) 患病数大于 3 的 Mers 的 GDP 和论文数排名散点图 (l) 患病数大于 3 的 Mers 的  $e_d$ ,  $e_g$  概率密度分布曲线 (m) Covid-19 的患病数和论文数排名散点图 (n) Covid-19 的 GDP 和论文数排名散点图 (o) Covid-19 的  $e_d$ ,  $e_g$  概率密度分布曲线 (p) 论文数大于 20 的 Covid-19 的患病数和论文数排名散点图 (q) 论文数大于 20 的 Covid-19 的 GDP 和论文数排名散点图 (r) 论文数大于 20 的 Covid-19 的  $e_d$ ,  $e_g$  概率密度分布曲线

因为疫情刚发生的前几个月论文数有一个小峰值，为减小模型拟合误差，在进行参数拟合时，剔除这几个月数据，我们也暂不对这一阶段进行建模。论文数增长模型如下：

$$N = \frac{N_{max}}{1 + e^{mt-n}}$$

其中， $N_{max}$ 表示模型预测的论文数的最大值；

$n$ 没有过多的物理意义；

$m$ 表示模型预测的论文数增长速率，值越大，表示论文数增长越快。

模型拟合的参数如下表所示（表 2）：

表 2 逻辑斯蒂增长模型拟合参数，估计值，标准差

病毒	模型参数	估计值	标准差	t 值
Sars	$N_{max}$	495.26	224.317	2.21
	$n$	5.99	0.348	-17.20
	$m$	1.47	0.282	5.22
Mers	$N_{max}$	32.95	2.028	18.71
	$n$	4.38	2.521	-1.85
	$m$	0.73	0.412	1.86
H1N1	$N_{max}$	281.17	14.692	25.68
	$n$	1.61	0.647	-3.05
	$m$	0.67	0.235	3.31
Ebola	$N_{max}$	318.55	22.054	26.12
	$n$	9.66	2.004	-6.99
	$m$	1.05	0.232	6.74
Covid-19	$N_{max}$	11655.31	119.06	17.31
	$n$	5.14	0.166	-8.54
	$m$	1.82	0.064	7.58

在疫情得到控制之后，论文数呈先快速减少，后缓慢减少的趋势。前人的研究表明，任何一种研究的衰减都服从由口头记忆和文化记忆控制的双指数衰减模型(Candia et al., 2019; Higham et al., 2017; Wang et al., 2013)，但是由于流行性病毒的特殊情况，其在病毒得到控制开始衰减后，因而没有口头记忆，只有文化记忆，应服从一个单指数衰减，我们尝试构建

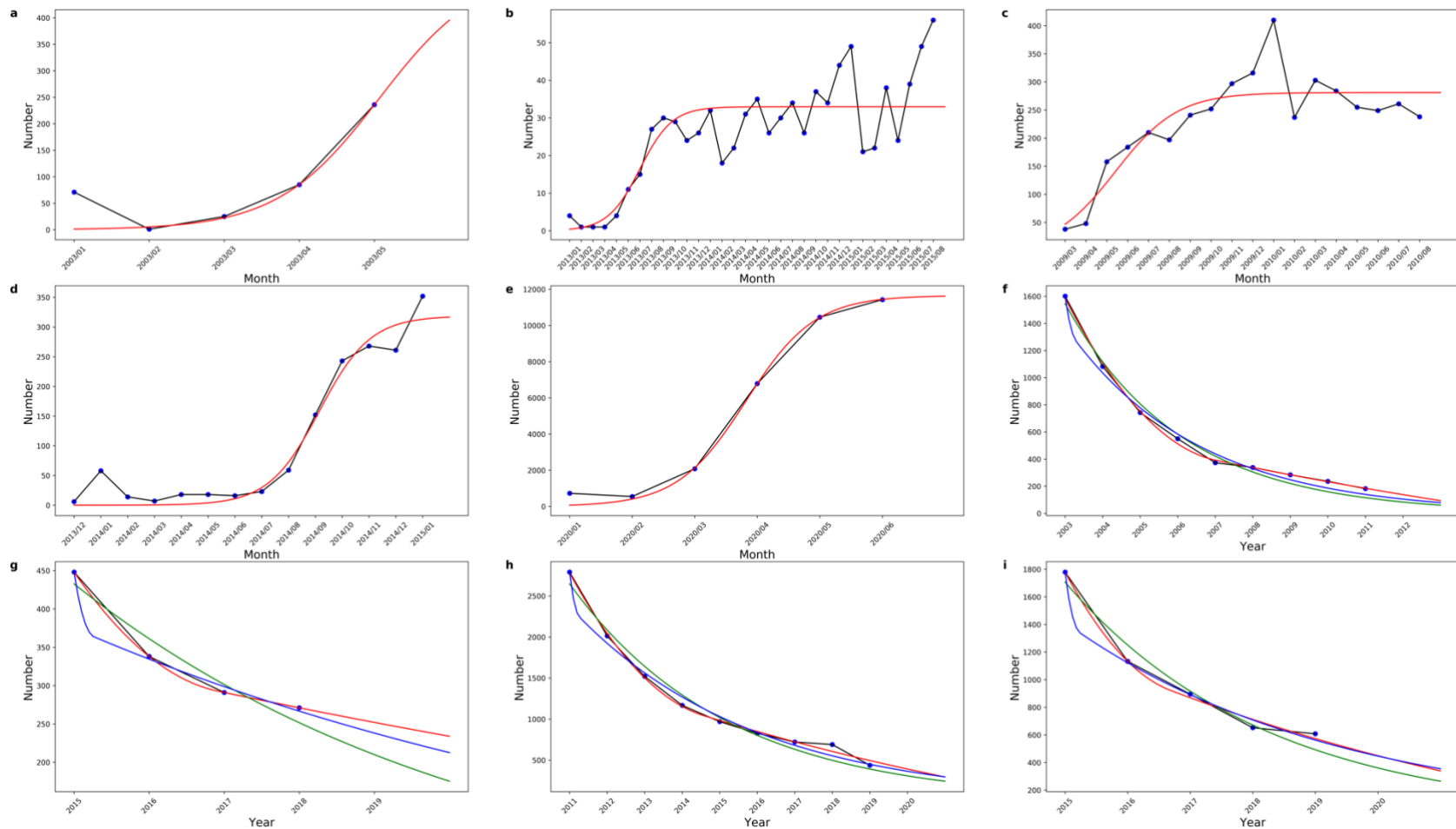


图 6 流行性病毒增长和衰退的模型拟合。蓝色的点为实际的论文数，红色、绿色、蓝色的曲线分别为为指数-线性、单指数、双指数模型拟合曲线。a-e，逻辑斯蒂增长模型，横轴为月份，纵轴为论文数 (a) Sars 增长阶段论文数 (b) Mers 增长阶段论文数 (c) H1N1 增长阶段论文数 (d) Ebola 增长阶段论文数 (e) Covid-19 增长阶段论文数 (f) Sars 衰减阶段论文数 (g) Mers 衰减阶段论文数 (h) H1N1 衰减阶段论文数 (i) Ebola 衰减阶段论文数



单指数衰减模型。模型如下：

$$n = N_0 e^{-mt}$$

其中， $N_0$ 表示开始衰减第一年的论文数，可以在一定程度上该病毒的研究程度，值越大，则该病毒的研究程度越大，但并不绝对；

$m$ 表示论文衰减速度，值越大，表示论文数衰减越快。

模型拟合的参数如下表所示（表 3）：

表 3 单指数衰减模型拟合参数，估计值，标准差

病毒	模型参数	估计值	标准差	t 值
Sars	$N_0$	1545.5	54.00	2.21
	$m$	0.325	0.01956	-17.20
Mers	$N_0$	432.7	22.37	18.71
	$m$	0.181	0.0344	-1.85
H1N1	$N_0$	2647.5	100.35	25.68
	$m$	0.239	0.0162	-3.05
Ebola	$N_0$	1706.4	94.81	26.12
	$m$	0.311	0.0380	-6.99

但是却发现单指数模型在前期衰减较慢，而在后期衰减过快，拟合效果并不好。这可能是由于流行性病毒前期研究太多，故而在病毒得到控制后，其研究迅速衰减，进而再缓慢衰减。据此特点，我们尝试使用两阶段衰减来构建论文的衰减模型：

阶段一：由文化记忆控制的指数衰减和另外一种衰减机制共同控制论文的衰减，另一种衰减机制的衰减速度随时间逐渐减少，最终到 0，该机制消失，简记其为后疫情衰减。

阶段二：仅有文化记忆控制的单指数衰减。

现在的问题是另一种衰减机制到底是否是什么衰减尚且未知，我们分别尝试使用线性衰减和指数衰减来进行拟合。模型如下：

指数-线性衰减模型：

后疫情衰减的衰减速度呈线性减少，最终为 0，记其衰减速度为  $q_t$ ，其开始的衰减速度为  $q_0$ ， $t$  时刻的衰减速度为

$$q_t = \begin{cases} q_0 - \frac{q_0}{a}t & \text{当 } t \leq a \text{ 时} \\ 0 & \text{当 } t > a \text{ 时} \end{cases}$$

论文数随衰减速度线性减少，类似一个匀减速直线运动，有：

$$N = \begin{cases} N_0 e^{-pt} - q_0 t + \frac{q_0}{2a} t^2 & \text{当 } t \leq a \text{ 时} \\ N_0 e^{-pt} - \frac{q_0 a}{2} & \text{当 } t > a \text{ 时} \end{cases}$$

模型拟合的参数如下表所示（表 4）：

表 4 指数-线性衰减模型拟合参数，估计值，标准差

病毒	模型参数	估计值	标准差
Sars	$N_0$	1586.4	22.13
	$p$	0.420	0.0080
	$q_0$	469.9	25.57
	$a$	4.039	0.2553
Mers	$N_0$	448.0	Inf
	$p$	0.051	Inf
	$q_0$	118.9	Inf
	$a$	1.914	Inf
H1N1	$N_0$	2778.5	47.97
	$p$	0.064	0.0122
	$q_0$	654.2	63.79
	$a$	3.564	0.4224
Ebola	$N_0$	1778.5	74.37
	$p$	0.119	0.0650
	$q_0$	636.5	228.85
	$a$	1.678	0.8278

双指数衰减模型：

受控于后疫情衰减的衰减速度仍然和上面一样呈线性减少，有：

$$q_t = \begin{cases} q_0 - \frac{q_0}{a}t & \text{当 } t \leq a \text{ 时} \\ 0 & \text{当 } t > a \text{ 时} \end{cases}$$

但是这种衰减受控于指数衰减，衰减量除了和衰减速度有关，还和当前论文数有关。

文化记忆的衰减速度为常数  $p$ ， $N$  为论文数，故有

$$N_{t+1} = \begin{cases} N_t(1-p) - N_t q_t = N_t(1-p-q_0) + \frac{q_0}{a}N_t t & \text{当 } t \leq a \text{ 时} \\ N_t(1-p) & \text{当 } t > a \text{ 时} \end{cases}$$

可得

$$N = \begin{cases} N_0 * e^{\frac{q_0}{2a}t^2 - (p+q_0)t} & \text{当 } t \leq a \text{ 时} \\ N_0 * e^{-pt} & \text{当 } t > a \text{ 时} \end{cases}$$

模型 2：衰减速度不变模型：

模型拟合的参数如下表所示（表 5）：

可以发现这两种衰减机制都在一定程度上克服了单指数衰减前期衰减速度过慢而后期过快的问题。但是在拟合过程中，双指数衰减的拟合结果很不稳定，可以拟合出多种  $q_0$  和  $p$  不同的参数，且后疫情衰减机制消失的很快。

表 5 双指数衰减模型拟合参数，估计值，标准差

病毒	模型参数	估计值	标准差
Sars	$N_0$	1601.0	54.15
	$q_0$	0.985	503
	$p$	0.303	154
	$a$	0.286	0.2289
Mers	$N_0$	448.0	Inf
	$q_0$	1.364	Inf
	$p$	0.262	Inf
	$a$	0.113	Inf
H1N1	$N_0$	2790.0	94.73
	$q_0$	1.249	Inf
	$p$	0.263	Inf
	$a$	0.208	0.0163
Ebola	$N_0$	1779.0	73.13
	$q_0$	1.918	242
	$p$	0.240	302
	$a$	0.230	0.0427

## 4 结论

### 4.1 冠状病毒时间不均衡分析

冠状病毒在 1975 年研究突然上升可能是因为 1975 年病毒命名委员会正式命名了冠状病毒科，从此引发了科学家们对于冠状病毒的关注。2003 年突然增长的原因是因为 2002 年年底非典的爆发，我们也可以从图中看出，2003 年超过 75% 的冠状病毒研究都集中于 SARS 的研究，而 SARS 很快得到控制，故对于 SARS 的研究减少，对于 SARS 研究的减少也导致了整个冠状病毒的研究呈现减少的趋势，且在病毒得到控制的 3 年后（2006 年），达到一个论文数的拐点。而在 2012 年 MERS 爆发后，冠状病毒的研究由于 MERS 再一次增多，且之后保持不变，这可能与 MERS 一直没有完全得到控制有关。2020 年新型冠状病毒的出现再次引起了冠状病毒研究的热潮，而由于病毒巨大的感染力，也造成相关的研究达到了惊人的数量。

我们可以发现，流行性病毒（SARS、MERS、COVID-19）的研究可以带动冠状病毒的整体研究，而在病毒得到控制后的 2-3 年，研究出现大幅度的减少。排除流行性病毒对冠状病毒研究的影响后，冠状病毒整体上是呈现一个缓慢增长的趋势。可以预测，在新型冠状病毒控制的 2-3 年内仍然会有较多的冠状病毒研究，但在之后的 1-2 年，将会迅速下降，进而缓慢下降。

## 4.2 冠状病毒门类不均衡分析

对于冠状病毒 HCoV-HKU1 的研究开始的早,研究数量也一直很多,其次研究较多的是“SARS-CoV”和“MERS-CoV”,这是因为之前两次大的疫情的影响,而对于“HCoV-229E”和“HCoV-NL63”的关注一直都很少。

1969 年之前变异系数为 0 是因为每年只有一种门类的研究;1970-1991 年其值变大,说明研究门类的不均衡开始出现,且有了“富者更富”的现象,研究差异越来越大;1991 年之后,其变异系数基本不变,说明各门类之间的研究保持平稳;2003 年,因为 SARS 的原因变异系数增大,在疫情之后,变异系数反而降低了,说明疫情结束后更多的人开始关注冠状病毒,关注其各个门类,这是一件好事。在 2013 年有一个上升,则是由于疫情导致了人们对于“MERS-CoV”更多的关注,而在疫情结束后,变异系数再次变小。这再次说明,疫情可以降低冠状病毒门类研究的不均衡。2020 年,由于新冠病毒,其变异系数上升是必然。可以预测,在新冠病毒过后,也会带动其他冠状病毒的研究,减少其不同门类研究的不均衡。

## 4.3 冠状病毒国家国家不均衡分析

由分位数、标准差和变异系数(表 6),我们可以看出不同国家对冠状病毒研究的差异性很大,最多的国家(中国)发表论文的人数近三万人,而一些国家则一篇论文都没有发表。从国家合作网络中我们也可以明显的看出,一些国家(中国、美国)的度值很大,而有一些国家则度值为零,度值是一个节点在网络中中心性的重要表现,这也可以反映出不同国家研究的不均衡。

我们认为造成这种不均衡的主要因素为患病数、GDP,当然也可能有其他的一些因素,比如人口、政治等,在这里我们仅分析患病数和 GDP 和论文数的相关关系(表 1)。

我们发现大多数国家患病数、GDP 的  $e$  值都在 3 以内,如果只计算患病数和论文数较多的国家,则多数国家  $e$  值都在 1 以内,相关性更好。患病数/GDP 和论文数的斯皮尔曼和肯德尔相关系数也表明患病数/GDP 和论文数存在一定的相关关系。从散点图和相关系数都可以看出, GDP 与论文数的相关性均大于患病数与论文数的相关性。这也说明了提升一个国家经济实力的重要性。

本来预期的是排除掉一些患病数、论文数较少的国家后,相关性会更强,结果反而显示排除后相关性变化不大甚至变化很小,这可能是因为患病数和论文数增加后,各个国家的研究都比较多,因此不均衡性减小。

其中 Mers 患病数和论文数的相关性最差,几乎为 0,而 GDP 与论文数的相关性则较好,这可能是因为 Mers 主要发病区在非洲,医疗研究条件相对落后,很多国家即使患病很多,想研究,也没有经济的支撑,只能由较发达的国家来援助,这可能也导致了上面问题二中 Mers 研究的论文数较少以及 Mers 疫情持续的时间较长。Covid-19 论文数与患病数和 GDP 的相关性都很高。

## 4.4 流行病毒增长、衰退动力学模型分析

在论文增长模型中,前几月的论文数增多,随之减少,可能是因为病情刚出现,引起大

家关注，将病毒的发现进行报道，然后开始进行研究，前期研究人不多，进展也比较缓慢随着关注的人越来越多，进行的研究也越来越多，呈现逻辑斯蒂模型前半部分快速增长的趋势。随着关注的人趋于饱和，研究的深入，论文增速开始逐渐减缓，最后论文数到达峰值，平稳增长。因为对病毒研究的人数是有限的，刚开始会有很多其他领域的人的介入，随后大家又回归到自己的本职研究中去。比较特殊的是 SARS 和 Ebola, SARS 在尚未达到峰值时，就已经停止增长，这是因为疫情控制的较快，疫情控制后，研究减少。Ebola 在前期的时间较久，是因为前期病例不多，人们并没有怎么关注，随着病例的增多，人们才正式开展大量研究。

表 6 各国家冠状病毒研究统计量（百分数表示分位数）

		Mers 全 部数据	Mers 患 病数大于 3 数据	Sars 全 部数据	Sars 患 病数大于 3 数据	Covid-19 全部数据	Covid-19 论文数大 于 20 数 据
0%	1.0	0.0	0.0	0.0	0.0	1.0	21.0
5%	1.0	0.6	1.6	0.0	0.0	1.0	25.4
10%	1.0	3.2	3.2	0.9	1.6	1.0	38.2
15%	3.0	4.9	6.4	1.0	4.1	3.0	41.9
20%	5.0	5.6	11.2	1.8	4.8	4.0	47.6
25%	6.0	12.0	16.0	4.3	17.5	5.0	55.5
30%	8.0	17.6	22.8	8.5	36.0	6.9	59.2
35%	12.6	20.1	29.6	11.5	57.0	8.1	73.4
40%	16.8	21.8	37.2	20.6	72.6	13.0	96.6
45%	22.4	30.0	45.6	27.2	81.6	17.4	105.9
50%	28.0	54.0	54.0	30.0	83.0	21.5	113.0
55%	46.4	80.9	65.6	55.7	133.4	38.7	132.7
60%	83.8	89.0	77.2	68.0	160.6	49.2	178.8
65%	107.4	93.9	85.0	80.9	169.3	59.9	229.3
70%	150.2	132.8	89.0	104.6	171.4	98.0	270.4
75%	192.0	234.0	93.0	155.0	204.5	114.3	575.0
80%	253.2	324.0	584.6	169.6	261.0	180.6	691.2
85%	482.4	511.1	1076.2	183.1	345.0	273.6	792.1
90%	1217.4	1386.8	1354.4	249.0	1218.6	692.4	1031.4
95%	2071.8	1841.0	1419.2	1146.8	2098.7	1034.3	1672.6
100%	29233.0	2232.0	1484.0	2812.0	2812.0	17896.0	17896.0
平均数	624.6	330.9	343.2	218.7	397.2	388.0	757.0
标准差	2830.6	632.6	603.1	588.5	803.5	1787.3	2460.9
变异系数	4.53	1.91	1.76	2.69	2.02	4.61	3.25

比较 5 个病毒逻辑斯蒂增长模型拟合的参数，我们发现 Covid-19 的论文数最高值远高于另外 4 个病毒，其论文数的增长速度也是 5 个病毒中最快的，Mers 的论文数最高值最低，说明研究最少，H1N1 的增速最慢，说明当时没有马上引起重视。可以预测，Covid-19 的论文数已接近最高值，接下来论文数变化不大，将会在每个月 11000 篇左右浮动，直到疫情得到控制。

指数-线性的衰减机制是这三种衰减机制里面最优的。单指数衰减并未考虑到流行性病毒的特殊情况，而双指数衰减模型的后疫情衰减机制衰减过快，且这种衰减机制与论文数有关的假设并不合理，应该是只与时间有关。

而由拟合出来的参数我们可以发现，H1N1 的研究量最多，且其文化记忆的衰减最快，其后疫情衰减机制结束的时间也较晚，因而持续研究时间最长。SARS 的文化记忆衰减最快，后疫情衰减机制结束的时间最晚。

## 5 疫情研究机制

综上，我们可以看到疫情的研究在时间、门类、国家上都存在一些不均衡现象，政府部门需要合理的对疫情研究进行调控，在不同时期对病毒的研究施以不同的资助，对不同门类的研究施以不同的资助，合理的控制病毒在时间、门类研究的不均衡性。同时，流行性病毒往往是全球范围内传播，各个国家应加强国际合作。病毒的增长、衰退动力学模型揭示了现阶段病毒来袭和控制后相关研究的统一机制。我们应该在病毒来袭时加大研究，尽快达到逻辑斯蒂增长的峰值，以期尽快的控制住病毒；而在病毒得到控制后，尽快的使后疫情衰减机制结束，减少后疫情衰减机制的论文数，以免造成研究的浪费。

## 参考文献

- Candia, C., Jara-Figueroa, C., Rodriguez-Sickert, C., Barabási, A.-L., Hidalgo, C.A., 2019. The universal decay of collective memory and attention. *Nat Hum Behav* 3, 82–91. <https://doi.org/10.1038/s41562-018-0474-5>
- Haghani, M., Bliemer, M.C.J., 2020. Covid-19 pandemic and the unprecedented mobilisation of scholarly efforts prompted by a health crisis: Scientometric comparisons across SARS, MERS and 2019-nCov literature. *arXiv:2006.00674 [physics, q-bio]*.
- Higham, K.W., Governale, M., Jaffe, A.B., Zülicke, U., 2017. Fame and obsolescence: Disentangling growth and aging dynamics of patent citations. *Phys. Rev. E* 95, 042309. <https://doi.org/10.1103/PhysRevE.95.042309>
- Wang, D., Song, C., Barabási, A.-L., 2013. Quantifying Long-Term Scientific Impact. *Science* 342, 127–132. <https://doi.org/10.1126/science.1237825>