

终极笔记
Research Note

修格致

2020 年 4 月 6 日-2020 年 5 月 9 日

目录

1 序言	1
1.1 关于	2
1.1.1 空间流行病学	2
2 空间流行病学	3
2.1 时空传播规律	4
2.1.1 孤立人口中流行病传播的幂律分布	4
2.2 大空间尺度的地理学问题	6
2.2.1 麻疹行波的强迫林火模型	6
2.2.2 from individuals to epidemics	6
3 网络弹性	7
4 演化网络	9
4.1 Essentials	9
4.1.1 防范意识对随机网络上的疾病传播的影响	9
4.1.2 偏好与地理对流行病传播的影响	9
4.1.3 Epidemic thresholds in dynamic contact networks	9
4.2 随机图上疾病和信息的传播	11
4.2.1 适应性网络上的流行病动态	11
4.3 本章 Ideas	16
4.3.1 戴口罩的演化博弈	16
5 人类移动性	17
5.1 经典模型	18
5.1.1 重力模型	18

5.1.2	辐射模型/介入机会 (IO) 模型	18
6	分工的形成	21
7	物理模型	23
7.1	通过几何重正化对真实网络进行多尺度展开	23
7.1.1	真实网络中几何标度存在的证据	24
8	内卷化	27
8.1	介数中心性	28
8.2	进化动力学	29
8.3	Growth models	29
8.3.1	Protocol: spatial-constrained attachment	29
8.3.2	解析	29
8.3.3	跟真实结果不符合? 增加一个参数	31

Chapter 1

序言

序章：本书是我从博二下半年开始的文献笔记和 idea 拾遗。会十分不连贯，不过也不乏精彩的部分。这已经是我第二年的下半年了。而我仍然觉得距离第一年并没有过去多久，一切还是开始时的样子。希望回头看看时，已经做成了自己想做的工作！

20 年 5 月 6 号更新。写这个东西的时候还能想起好多年前背的《兰亭集序》。每览昔人兴感之由，若合一契，未尝不临文嗟悼，不能喻之于怀。固知一死生为虚诞，齐彭殤为妄作。后之视今，亦犹今之视昔，悲夫！故列叙时人，录其所述，虽世殊事异，所以兴怀，其致一也。后之览者，亦将有感于斯文。

写文献笔记和 Idea 整理，自然是列叙时人，录其所述。我把这个笔记开源放在 GitHub 上面，若是他日文章发得好，有人看得见，或许会损失一些本可以自己做出的东西。但生活就是这样，错过一些东西之后，要接受生活中没他一样。时常想或许日后我当导师，第一年就不会让学生接触任何跟论文写作的东西。就像不能买第一条你挑中的裙子一样，也不要去看你的第一个 Idea 能写出的文章。真的很想写了再写，是我觉得的好状态。

1.1 关于

更新于 2020 年 4 月 6 日。

现在我的研究兴趣集中在城市生态系统、城市韧性、以及相关的数学物理模型上。本文的主要内容也将是涉及这些内容。这个过程大概将持续到 2021 年的 7 月，作为我心目中博士阶段的三块基石中的第一块。

1.1.1 空间流行病学

2020 年是个地狱开局。大卫·斯特恩和科比布莱恩特去世了。紧接着 COVID 走入并改变了我们每个人的生活。这个时候我们没有人可以置身事外。我也想做一些有意义的相关研究。我不希望我这些研究仅仅是涉及本次疫情的特定情况。更重要的是我们如何在每次流行病发生的时候应对得好一点，再好一点。

流行病传播的时空特征

工具：

- 小波分析

Chapter 2

空间流行病学

代表作家：Grenfell、Durrett ([近期演讲主页](#), [Stochastic Spatial Models: A Hyper-Tutorial](#))

You won't stop it until you fully understand it. It won't just *magically* disappear.

2.1 时空传播规律

重要文献: Travelling waves and spatial hierarchies in measles epidemics, Bryan Grenfell et. al., Nature 2001

流行病传播的实际问题与理论模型的区别 实际问题: 要注意流行病的具体传播规律。不光是参数的设定, 更是大类模型的选择。要关注非线性、时空异质性、非马氏性等重要因素。

2.1.1 孤立人口中流行病传播的幂律分布

Power laws governing epidemics in isolated populations, C. J. Rhodes and R. M. Anderson

在生物时间序列中非线性和混沌模式的识别和分析的背景下, 发达国家**大城市**社区中麻疹病毒感染的时空变化一直是许多讨论的焦点。相反, 由于感染记录的频繁消失且高度不规则, 孤立的小岛屿人群用传统分析并不能提供有用的见解。本文使用流行病大小和持续时间分布的方式来证明这种系统动力学的**规律性** (regularities) 是明显的。具体而言, 这些生物学系统的特征是定义明确的幂律分布, 其方式类似于物理学中其他非线性的, 空间中展开的动态系统。本文进一步表明, 所观察到的幂律指数已通过基于网格的简单模型很好地描述, 该模型反映了各个宿主之间的社会交互。

人口学上来讲, 小岛上的现象通常可以不考虑与外界的交互, 从而是一个良好的社区化语境。法罗群岛 (知乎链接: [法罗群岛的绞肉机](#), 讲得是这个地方的大家去捕鲸的故事) 是丹麦的海外自治领地。地理位置介乎挪威海和北大西洋中间, 处于挪威到冰岛之间距离一半的位置。

法罗群岛的人口总数为 25,000。外部人口与之交互会带来流行病麻疹。我们认为输入病例是非常精确的, 因为法罗群岛面积小而交互局部化, 也因为每个病例在这里都会引起很大重视。在 58 年中, 有 43 个互不相同的 epidemic event (**定义**为连续的 $t = \tau_{\text{end}} - \tau_{\text{start}}$ 个月都出现有限个病例记录, 这段时间的前后月份都不出现麻疹病例)。每个 event 的规模 (size) 定义为病例数的求和: $s = \sum_{\tau_{\text{start}}}^{\tau_{\text{end}}} C(\tau)$. 地震学里面有个 **Gutenberg-Richer 定律**, 说的是地震频率与烈度之间关系的幂律分布, 长得就跟法罗群岛流行病的发病情况差不多, 都是 $\log N(> s) = a - b \log s$ 。流行病的频数和持续时间的幂律关系在本文找到了:

$$N(s) \propto s^{-1-b}, \quad b \simeq 0.28 \quad (2.1)$$

$$N(t) \propto t^{-1-c}, \quad c \simeq 0.8 \quad (2.2)$$

这两个关系对我们估计短期内流行病规模和持续时间的概率分布是十分有用的。小而短的麻疹疫情比大而长的麻疹疫情要更频繁出现。这种幂律关系的好处是它对子样本数据同样成立, 可以用前一半数据来预测后一半数据。作者利用博恩霍尔姆岛和雷克雅未克的精确麻疹病例来估计来幂律指数 b 和 c 。连同法罗群岛, 这三个地方的幂律指数是高度吻合的。

这种幂律现象为什么会出现还不能很好的被理解。有一些作者没有提到的空间模型做得还不错。于是作者使用了一个基于格点的模型, 该模型之前被用为林火传播的模型, 用在这个空间 $S-I$ 的场景也合适 (相

关文献: [phys. rev. lett. 自组织临界的林火模型](#)、[phys. rev. e 林火模型中的相变](#)、[phys. rev. e 林火模型的标度律和模拟结果](#) 这几个是同一个团队的作品)。模型叙述如下: periodic 的离散的 $L \times L$ 格点图上, 每个点有三种可能的状态: 被感染、易感、空点, 随着离散的时间更新。更新规则是:

1. 如果易感者 S 的最近邻居有患者, 就可能被感染;
2. 患者 I 失去活性, 所处位置清空;
3. 易感者以概率为 μ 移动到空格子上;
4. 新患者 I 不时以概率 γ 被感染。

模型有效反映了代表罕见外来病例作用的迁移项。模拟用 $L = 250$, $\mu = 2.6 \times 10^{-5}$, $\nu = \mu/300$, 得到了 $b \simeq 0.29$, $c \simeq 1.5$ 的结果。模型对十个月以上的长期流行病的数量是低估的。网络模型的占用情况可以模拟真实社区。平均人口密度是 25,000, 这意味着每 4 个位置有一个人。均衡状态下, 平均寿命如果是 70 岁, 我们期望每天有 ~ 1 个新生儿降生。真实的 ν/μ 应该是 1/400, 而不是模拟使用的 1/300. 新生的易感人群和患病的人的迁移由均值为 1 和 1/300 泊松过程刻画, 这样一个 step 相当于一天的步长。模拟结果受长期误差影响非常大, 但是对 5 个月以内的流行病的分布有着很好的复现。

作者还使用了随机 SEIR 模型进行对比。此时人口被假设为均匀同质混合, 外来人口比例很低。同样计算了时间和规模的分布。这个模型高估了大流行病的频率, 而且与真实分布吻合不好。SEIR 模型很可能不适合小人口的不频繁的流行病。

我们的结果表明, 在孤立的海岛麻疹数据集中存在流行病的规模和持续时间的标度律。这将这些流行病的动力学与其他空间扩展的非线性动力学系统归为同一类, 在该非线性动力学系统也观察到标度律。实际上, 这促进了一种预测方法, 可以计算给定大小和持续时间的流行病的发生频率。

一个简单的空间模型所产生的指数几乎与我们的数据分析所见的相同。幂律现象也可能与高度接种疫苗的社区和发展中国家偏远农村人口中麻疹不经常爆发的研究有关。这里讨论的方法是完全通用的, 可以应用于少数人群中任何其他时间序列的传染病暴发。

[Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza](#)

量化流行病的长程传播是疾病动力学和疾病控制的主要因素。本文使用美国超过 30 年 (1972-2002) 流感周期数据来分析两次爆发之间变化规律。流感高发季度的传播能力更快更强。

2.2 大空间尺度的地理学问题

麻疹传播的波状结构和空间层次时空行波是捕食者-猎物和宿主-寄生虫动力学的显着表现。然而，很少有系统的文献足以检测重复的波并解释它们与人口结构和人口统计学的时空变化的相互作用。在这里，我们用详尽的时空数据集证明了英格兰和威尔士的流行性流行行波。我们使用小波相位分析，可以实现动态非平稳性-解释这些以及其他许多生态时间序列中的时空模式很复杂。在疫苗接种之前的时期，明显的分层感染浪潮从大城市向小城镇转移。麻疹疫苗的引入限制，但并没有消除这种等级传染病。一个机理上的随机模型为从大的“核心”城市到较小的“卫星”城镇通过传染性“火花”传播的海浪提供了动力学解释。因此，宿主种群结构的空间层次是这些感染浪潮的先决条件。

旅行波来自于激活-抑制动力学研究，并被很多的 host-natural enemy 系统所预测。但是除了入侵动力学，经验观测被研究的还是比较少。

2.2.1 麻疹行波的强迫林火模型

2.2.2 from individuals to epidemics

Chapter 3

网络弹性

Chapter 4

演化网络

4.1 Essentials

4.1.1 防范意识对随机网络上的疾病传播的影响

防范意识对随机网络上的疾病传播的影响。文章中说流行病期间产生的防范意识不会对 basic reproduction number 产生影响。从其他人群中产生的防范意识会降低 R_0 。防范意识可以显著减小流行病的规模，切断与病人的连边比降低感染率的效果要明显得多。流行病规模不会随着防范意识和感染率单调减小。局部和全局的防范意识是否对流行病规模产生影响取决于网络的类型和感染率。

4.1.2 偏好与地理对流行病传播的影响

偏好与地理对流行病传播的影响：随机网络上的 SIS 模型，每个时刻，空间网络增加一个有 m 条边的结点。它与度为 k_j ，距离为 d_{ij} 的结点 j 连接的概率正比于 k_j^A/d_{ij}^B ， A 和 B 都是正常数。如果 $A = 0$ ，就是有临界值的普通疾病传播，疾病最终会消失； $B = 0$ 时，(a) $A = 1$ 时没有临界行为，其他时候仍然有临界行为。如果两个因素同时存在，临界值之上的 A 和临界值之下的 B 控制的网络对于传染是有鲁棒性的。

4.1.3 Epidemic thresholds in dynamic contact networks

Epidemic thresholds in dynamic contact networks再生率 R_0 是流行病学中的基本量，它决定了易感群体中传染病的一阶增长。在大多数流行病的**理论模型**中，都有一个特定的值 R_0 ，即流行病阈值，高于该值可能会大流行，但低于此值就不会发生大流行。随着流行病模型的复杂性增加，计算流行病阈值的难度也增加。本文通过简单的动态随机网络类推导了 SIR 流行病的 R_0 和流行阈值。与大多数流行病学模型一样， R_0 取决于两个基本流行病参数，即传播率和恢复率。我们发现， R_0 还取决于社会参数，即描述并发

联系人数量异质性的**度分布**以及给出联系人启动和终止速率的混合参数。我们表明，社会融合从根本上改变了流行病学格局，因此，用静态网络来近似动态网络可能是不足够的。

几个简写：neighbour exchanges (NEs)

任意度分布的静态随机网络已经有很多研究在做了 (Callaway et al. 2001; Newman et al. 2001, 2002; Chung & Lu 2002; Catanzaro et al. 2005)。我们假设一个度分布为 $\{p_k\}$ 的网络。对于大型的网络，给定一条 ego 到 x 边， $x = alter$ 的概率正比于 $alter$ 的度。用 $\{ego, alter\}$ 来记录一条边; $(ego, alter)$ 来记录 ego 到 alt 的联系。

4.2 随机图上疾病和信息的传播

Some features of the spread of epidemics and information on a random graph.

三种随机图 小世界图是 Watts 和 Strogatz 引入的。在一维圆环 $\mathbf{Z} \bmod n$ 上, 连接所有距离小于等于 m 的点。然后以概率 p , 将边的一端连到纯随机的另一个结点上。**BA 模型**依次加入结点, 连接 m 条边到已有结点上, 概率正比于该结点的度。该模型的度分布收敛于

$$p_k = \frac{2m(m+1)}{k(k+1)(k+2)} \quad \text{for } k \geq m \quad (4.1)$$

如果不是正比于度, 而是一个平移项 $k+a$ ($a > -1$), 那么度分布将会是 $p_k \sim Ck^{-3+a}$. 如果 $a \geq 0$, 这种方式等价于: 新边以概率为 $a/(a+1)$ 随机选一个结点, 以概率为 $1/(a+1)$ 使用偏好依附机制。**NSW 模型**, 给定度分布 $\{p_k, k \geq 0\}$, 独立采样出真实的度序列 $\{d_k\}_{k=1, \dots, n}$, 假设采样出的度的和为偶数, 然后结点依次随机连 d_i 条边, 构建一个 n 个结点的随机图 G_n .

两种流行病 **SIR** 类似于分枝过程, 以速率为 λ 感染, 以速率为 1 恢复, 恢复之后不易感。该模型与渗流模型有很好的对应。如果 x, y 在图里是邻居, 就以概率为 $\lambda/(\lambda+1)$, 也就是二者之一患病期间感染另一人的概率, 连边。流行病的规模就是含有初始被感染结点的连通图的大小。**SIS** 模型要更复杂一些。感染和恢复速率分别是 λ 和 1. 这又称为接触过程。

小世界网络的简单版本 $m=1$ 在 Watts 和 Strogatz 的小世界十年前就被 Bollobas 和 Chung 提出了。这种小世界的半径是 $\sim \log_2 n$. 称为 BC 小世界。我们希望每个个体有且只有一个长程联系, 这样就可以定一个相关的大世界图 \mathcal{B}_m .

自由积是个不错的形式。 $\mathbf{Z} * \{0, 1\}$. 其中的元素的形式是 $z_0 1 z_1 1 \dots 1 z_k$. 也就是移动 z_0 连到 1, 然后到 z_1 再到 1 以此类推。这个文章太难了我以后再看吧。不过真的好精彩。

演化网络上的疾病传播

4.2.1 适定性网络上的流行病动态

Epidemic Dynamics on an Adaptive Network

易感者能够通过重新连接其网络连接来避免与感染者接触。引起、分类度相关性, 振荡, 磁滞和一阶跃迁。

考虑例如在社交网络上传染病的传播。人类倾向于通过**避免与感染者接触**来应对这种流行病。局部连接的这种重新布线可对疾病的动力学产生强烈影响, 进而影响重新布线过程。因此, 出现了时变网络拓扑与节点动态之间的复杂交互。自适应网络上的 SIS 模型可以证明, 用于网络连接的简单直观的重连规则对网络具有深远的影响, 并且能够生成特定的网络属性, 例如广度分布, 分类度相关性以及两个松散连接的

子隔间的形成。**动力学上的结果**是出现了新的流行阈值（对应于一阶跃迁），多个稳定平衡的共存（导致磁滞）以及振荡机制的出现，所有这些在静态 SIS 网络中都是不存在的。

考虑结点数为 N ，边数为 K 的网络上的 SIS 过程。每个时间节点上，每个 S-I 边以概率 p 变成 I-I；每个 I 以概率 r 恢复健康。**重连**体现在：S 可以通过重连来保护自己。SI 以概率 w 被打断，连接到一个随机的健康 S 结点。

流行条件 考虑临界感染概率 p^* 。在不重连的网络上，基本再生数 $R_0 := p\langle k \rangle / r$ ，其中 $\langle k \rangle = 2K/N$ ，它表示由其他易受感染的网络上的单个感染节点引起的继发感染数。所以，给定网络结构，需要一个二级感染的话， $p^* = r / \langle k \rangle$ 。如果考虑网络重连的话，一个 I 结点平均每次会失去 w 比例的邻居。邻居数将是 $k(t) = \langle k \rangle \exp(-wt)$ 。由于 I 的周期是 $1/r$ ，阈值将是

$$p^* = \frac{w}{\langle k \rangle [1 - \exp(-w/r)]} \quad (4.2)$$

由于 $w \gg r$ 时， $p^* = w / \langle k \rangle$ ，所以高重连率会使得流行病阈值提高很多，进而减小流行病的程度。

分几种情况：

重连与状态独立 度分布是泊松分布，给定结点二阶邻居的平均度 $\langle k_{nn} \rangle$ 与度 k 是独立的。

上面的重连规则 先让 SIS 机制关掉，即 $r = p = 0$ 。设 I 的比例是 i ，S 的比例是 $s = 1 - i$ ，都是常数。但是 SI 边随时间不断减少，逐渐变成不在一起的两个组团。假设我们从一个随机图开始，各种类型边的密度分别是 $l_{SS} = s^2 \langle k \rangle / 2$ ， $l_{II} = i^2 \langle k \rangle / 2$ ， $l_{SI} = \langle k \rangle / 2 - l_{II} - l_{SS} = si \langle k \rangle$ 。有适定性重连规则之后，稳定状态下所有的 SI 边都会转为 SS 边，密度是 $l_{SS} = (1 - i^2) \langle k \rangle / 2$ 。因此，S 与 I 在不同的度分布 ρ_k 假定下，S 的平均度 $\langle k_S \rangle = (1 + i) \langle k \rangle$ ，I 的平均度 $\langle k_I \rangle = i \langle k \rangle$ 。虽然各自还是泊松的，但是 S 组团的联通度更高。由于 $\langle k_{nn} \rangle$ 与两个集团中分别的度 k 都是独立的，组团内部的度关联也消失了。但是如果两个组团考虑在一起，一个净度关联 $r_{\text{corr}} > 0$ 可以出现，因为 $\langle k_{nn} \rangle$ 对于 S 组团更大。

最后，同时考虑适定性重连与流行病动力。虽然重连没有快到可以完全分离 S 与 I，但是还是能形成两个比较松散连接的组团，比如说 $l_{SI} \approx 0.01 \langle k \rangle$ 。组团间联系在被切断，有的 I 结点也在康复，有的 S 结点也在被感染。这导致结点度随时间的巨大变化。只要一个结点是 S，它的度几乎就以线性增长，

$$\dot{k} = w l_{SI}. \quad (4.3)$$

反之，感染的结点的度以指数递减，

$$\dot{k} \sim -wk. \quad (4.4)$$

所以存在一个复杂的均衡，组团间和组团内的边，S 与 I 的密度都是一个常数。这个均衡中，连续的重连会维护一组 S 与 I 的度分布，以及一个正的度关联（高高相连，低低相连）。

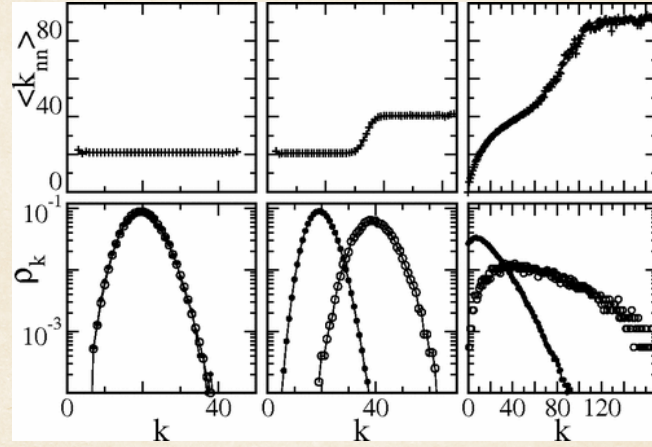


图 4.1: 适应性网络的结构。上边一行是平均最近邻度 $\langle k_{nn} \rangle$, 下面一行是度分布。从左到右分别是无差别重连/没有 SIS 关系的/适应性网络。参数设置是 $w = 0.3$, $r = 0.002$, $p = 0.008$, $N = 10^5$, $K = 10^6$.

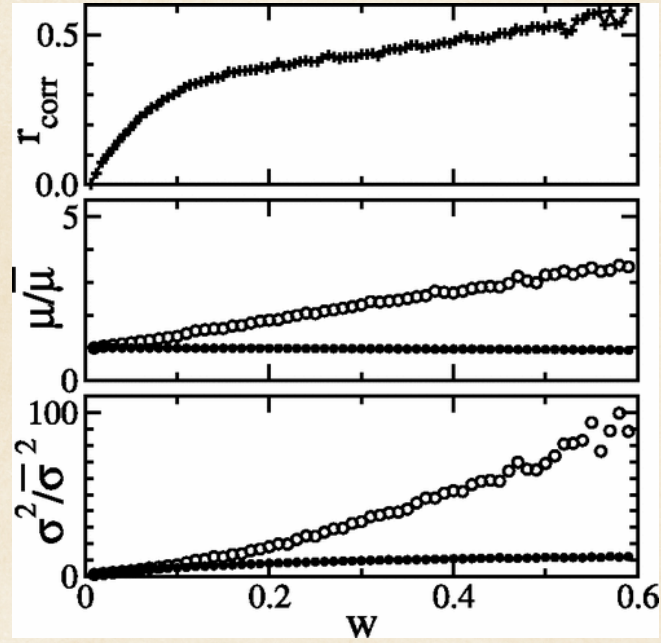


图 4.2: 上图是 r_{corr} 作为 w 的函数, 下图是度分布的均值和方差作为 w 的函数。空心是 S, 实心是 I. 都相对于没重连的情况标准化了。参数是 $N = 10^5$, $K = 10^6$, $r = 0.002$, $p = 0.008$.

进一步的性质 图 4.2 进一步量化了自适应重新布线对新兴网络结构的影响。随着 w 的增加, 相关度迅速增加。此外, 易感者的平均程度增加而感染程度则略有下降。甚至更明显是在易感的度分布的方差的增加。比如 $w = 0.6$ 时的方差 σ^2 大了 100 倍。这表明形成了牢固连接的中心和临时孤立的节点, 这些节点由重连而迅速重新连接。

适应性重连促进了受感染个体的隔离, 这可以大大提高流行阈值。然而, 这样做时, 连引入了人口中连接的混合, 并且还导致形成高度连接的易感簇, 其特征在于度分布的方差较大, 因此具有较低的流行阈值。因此, 重连的局部作用趋于抑制流行, 而产生的拓扑作用则促进流行。所以重连有两个相反的作用。

为了研究这个问题, 我们考虑一个低维问题中的 l_{SS} 和 l_{II} 。利用 moment closure approximation 方法。对于 $a, b, c \in [S, I]$ 做如下近似: $l_{abc} = l_{ab}l_{bc}/b$ 。这样的系统上面就有如下的 ODE:

$$\frac{d}{dt}i = pl_{SI} - ri \quad (4.5)$$

$$\frac{d}{dt}l_{II} = pl_{SI} \left(\frac{l_{SI}}{s} + 1 \right) - 2rl_{II} \quad (4.6)$$

$$\frac{d}{dt}l_{SS} = (r + w)l_{SI} - \frac{2pl_{SI}l_{SS}}{s} \quad (4.7)$$

它可以与数值模拟的结果, 也就是图 4.3。在不重新布线的情况下, 只有一个连续的动力过渡发生在流行阈值 p^* 上。随着重连, 该阈值与等式 4.2 完全吻合。但同时, 也会出现一个**较低的鞍点分叉**。该点之上, 已有的流行病会继续存在。与没有重连相比, 两个流行病机制之间出现了不连续一阶相变 (导数就不连续了)。其之间有双稳定状态, 健康和流行病可能都是稳定的。所以, 会形成一个**滞后曲线**。

数值模拟表明, 滞后回线和一阶跃迁的存在是自适应模型的通用特征, 可以在所有有限的重连速率下观察到。虽然增加重连速率几乎不会减少局部流行病的规模, 但滞后的阈值的性质会在更高的重连速率下发生变化。首先, 亚临界 Hopf 分支产生了不稳定的极限循环, 从而取代了鞍点分支。在更高的重连速率下, Hopf 分支变得超临界。由于现在出现的极限周期是稳定的, 因此 Hopf 分叉标志着第三阈值, 在该阈值处会发生连续向振荡动力学的过渡。但是, 这些振荡只能在遇到持久性阈值之前在相对较小的参数区域 (请参见图 4.4) 中观察到, 该阈值现在对应于循环的折叠分叉。

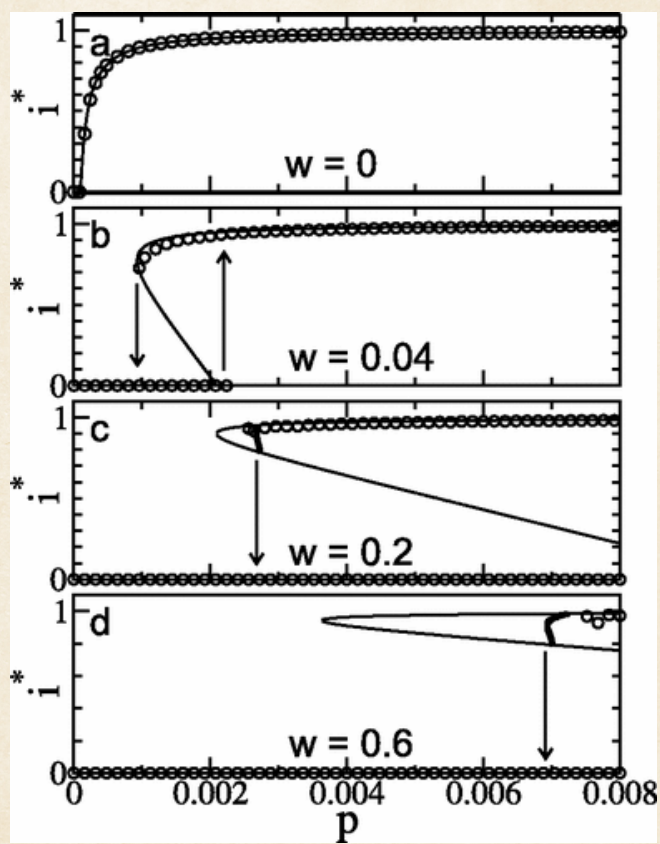


图 4.3: I 的密度作为传染概率 p^* 的函数, 对于不同的重连概率 w .

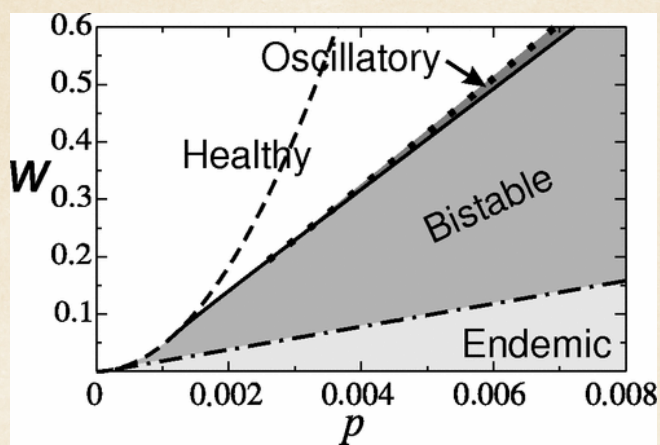


图 4.4: 两个参数的相图。

4.3 本章 Ideas

4.3.1 戴口罩的演化博弈

演化机制

Chapter 5

人类移动性

移动性其实分为很多种，人类移动性、商品的移动性、信息的移动性都被研究了很多。Physics Reports 有一篇综述，[Human mobility: Models and applications](#)

5.1 经典模型

王铮老师有一本书大家可以看一下，叫《理论经济地理学》里面有大量的地理学模型。虽然不算新，也没有什么实证分析，不过想法很值得借鉴。

5.1.1 重力模型

重力模型可能是地理学中最有名的模型了。形式如下：

$$T_{ij} = \frac{m_i^\alpha n_j^\beta}{f(r_{ij})} \quad (5.1)$$

重力模型对火车运货量Zipf, G. K. The P 1P 2/D hypothesis: On the intercity movement of persons. Am. Sociol. Rev. 11, 677–686 (1946).、地铁乘客Goh, S., Lee, K., Park, J. S. and Choi, M. Y. Modification of the gravity model and application to the metropolitan Seoul subway system. Phys. Rev. E 86, 026102 (2012).、韩国高速公路Jung, W. S., Wang, F. and Stanley, H. E. Gravity model in the Korean highway. EPL 81, 48005 (2008).、航空网络Grosche, T., Rothlauf, F. and Heinzl, A. Gravity models for airline passenger volume estimation. J. Air Transp. Manag. 13, 175–183 (2007).、通勤Viboud, C. et al. Synchrony, waves, and spatial hierarchies in the spread of influenza. Science 312, 447–451 (2006).和人口迁移Tobler, W. Migration: Ravenstein, thornthwaite, and beyond. Urban Geogr. 16, 327–343 (1995).等问题的拟合很不错。

通过目的地选择来推导重力模型是一个学术套路，我觉得还能玩二十年。目前目的地选择的理论有这样一些根源：确定性效用理论Niedercorn, J. H. and Bechdolt, B. V. Jr. An economic derivation of the "gravity law" of spatial interaction. J. Regional Sci. 9, 273–282 (1969).、随机效用理论Domencich, T. A. and Mcfadden, D. Urban travel demand: A behavioral analysis. (North-Holland, Amsterdam, 1975).、博弈论Yan, X. Y. and Zhou, T. Destination choice game: A spatial interaction theory on human mobility. Sci. Rep. 9, 1–9 (2019).等。文章最近也还有，不过都是集中在 SR 之类的期刊上。

一个规律如果基本是客观存在且不很精确的话，不同的发现方法总是能发出文章的。毕竟人类对于规律的探索就像对孤独的回避。

5.1.2 辐射模型/介入机会 (IO) 模型

辐射模型记载在 Simini 和 Marta 等人一篇名声不大好的 Nature 论文里，A universal model for mobility and migration patterns. Nature 484, 96–100 (2012).

文章中的配图甚是有意思，图一小人的表情是有出处的。法国大作家雨果写毕名著《巴黎圣母院》，与出版商有了这番史上最短通信：

“? —雨果”

“! —出版商”

辐射模型刻画了：

辐射模型的表述如下：

$$\langle T_{ij} \rangle = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \quad (5.2)$$

其中， T_{ij} 表示 i 到 j 的流量， m_i, n_j 代表两个位置的人口， r_{ij} 为两地距离， s_{ij} 代表以 i 为中心， r_{ij} 为半径的圆内的人口总数。

严小勇同志又搞了一篇 Scientific Reports, [Liu, E., Yan, X. A universal opportunity model for human mobility. Sci Rep 10, 4657 \(2020\).](#) 介入机会 (intervening opportunity) ¹模型说的是：个体选择目的地与两个因素相关，终点的机会与起点终点之间介入的机会 (?)。此类模型可以给出特定时空尺度上的准确预测，但是不同尺度上都合适的 IO 模型在上面这篇文章是第一次给出的。本文考虑的是人类行为的两个倾向：探索倾向和谨慎倾向。模型的形式是

$$Q_{ij} = \int_0^\infty \Pr_{m_i + \alpha \cdot s_{ij}}(z) \Pr_{\beta \cdot s_{ij}}(< z) \Pr_{m_j}(> z) dz, \quad (5.3)$$

¹Stouffer, S. A. Intervening opportunities: A theory relating mobility and distance. Am. Sociol. Rev. 5, 845–867 (1940).

Chapter 6

分工的形成

分工是城市理论中的重要一环。近期的研究工作有：

- The Spatial Division of Talent in City Regions: Location Dynamics of Business Services in Copenhagen

Chapter 7

物理模型

7.1 通过几何重正化对真实网络进行多尺度展开

- 多尺度与小世界的矛盾
 - 欧式长度和对称性的缺陷
- 复杂网络的几何重正化群
 - 将真实网络插入到一个度量空间，会体现出 geometric scaling 特征

复杂网络中，多尺度也是共存的，但是它们被一些其他的事限制住了，并不能直接讨论自相似性和标度无关性。原因是我们没有一种有效的手段来对网络的 length scale 进行变换。

- 以前的手段：
 - 拓扑/粗粒化/random walks
 - /box-covering/: 证明了真实网络有着有限的分形维数，有自相似性
 - 拓扑 scaling 性质只体现在度分布、平均度、最大度上面
 - 尽管有很好的度量，最短路径的集合作为研究 length-based scaling factor 还是很不好的数据集。（原因是 small-world 的存在）
- In this work, we introduce a geometric renormalization group for complex networks (RGN). The method is based on a geometric embedding of the networks to construct renormalized versions of their structure by coase-graining neighbouring nodes into supernodes and defining a new map which

progressively selects longer range connections by identifying relevant interactions at each scale. The RGN technique is inspired by the block spin renormalization group devised by L. P. Kadanoff [18].

7.1.1 真实网络中几何标度存在的证据

研究对象：复杂网络到 hidden 度量空间的映射： $\mathcal{M}(T, G)$

- 定义一个几何重正化算子 \mathbb{F}_\setminus ，得到一个新的拓扑 T' 和一个新的几何图 G' ，由此定义一个新的重正化映射 \mathcal{M}' : $\mathcal{M}(T, G) \xrightarrow{\mathbb{F}_\setminus} \mathcal{M}'(T', G')$
- The transformation zooms out by changing the **minimum length scale** from that of the original network to a larger value.
- 这个过程可以迭代 $O(\ln N)$ 次。

例子：

- 最简单的度量空间：一维圆周： $\{\theta_i : i = 1, 2, 3, \dots, N\}$
 - 重正化步骤：
 1. 定义 block：圆周上挨着的 r 个点。
 2. 粗粒化为超级结点（不管是否连接）每个超级结点都控制一个角区域。所以它们的序关系得以保留。
 - * 原连接：
 - 超级结点内
 - 超级结点间：建立边
- 用到的例子：
 - Internet
 - Airports
 - 新陈代谢
 - scripts.....

S_1 模型：将结点放在一个圆周上，以一定概率分布连接每两个点。两个点越远链接概率越低（similarity），度乘积越大连接概率越高（popularity）。

应用：The RGN enables us to unfold scale-free complex networks in a self-similar multi-layer shell which unveils the coexisting scales and their interplay. Beyond

- Mini-me network replicas.
 - networked communication systems
 - 可以保持微观结构的同时，不破坏介观结构
 - Mini-me replicas can also be used to perform finite size scaling of critical phenomena taking place on real networks
 - Typically, the renormalized average degree of real networks increases in the flow, since they belong to the small-world phase (see inset in Fig. 3B), meaning that the network layer at the selected scale is more densely connected.

Chapter 8

内卷化

内卷化是我最近看到的一个高频词。偶然间想到这个概念与我们平时讲城市体系的层次结构似乎有着很好的对应关系，就决定在笔记里开一章做一些思考。

8.1 介数中心性

指标定义我们从一些已有的文献开始看。比如介数中心度的这篇 [\[1\]](#)

8.2 进化动力学

8.3 Growth models

生长模型是处理这类问题的一个重要出发点。下面是之前的一个文献笔记。

8.3.1 Protocol: spatial-constrained attachment

1. 给定一个有界 d 维欧氏空间 $S = \mathbf{x}_1, \dots, \mathbf{x}_d \leq \frac{L}{2}$. $t = 0$ 时在原点插入一个结点.
2. 结点生成: 每个 t 时刻, 以均匀分布在 S 中放置一个新结点 P_t . 记它的坐标为 x_t . 如果存在一个结点 $P_q, q \in \{1, 2, \dots, t-1\}$, 使得 $x_p - x_q < r$, 则结点 P_t 存活。否则 P_t 死亡。
3. 连边: 将新加入结点与其 r - 临域的所有结点相连。
4. 重复这个过程, 直到存活的结点达到 N 个。

这个几何网络会加速增长。因为新加入的结点可生存的区域测度（为所有存活结点的 r - 临域的开覆盖）会越来越大。

8.3.2 解析

- 所需要的极限: **热力学极限**: $L, t \rightarrow \infty$.

在这种情况下, 网络会渐近形成一个 d 维球。

- 半径 $R(t)$ (定义为 $\max\{P_0 - P_i, i = 1, 2, \dots, t\}$)

– 有 $R(t) \sim t$.

– 证明: 以 1 维正半轴为例, 新结点 $P(t+1)$ 落在 $[0, R(t)+r)$ 可以生存, 落在 $[R(t)-r, R(t)+r]$ 可以使得 $R(t+1) > R(t)$. 于是增加半径的量的期望为 (在 $R(t)$ 远大于 r 的前提下)

$$\int_0^L dx P(x_{t+1} = x) \Delta R_{x,t} \quad (8.1)$$

$$= \frac{2r}{L/2} \int_{R(t)-r}^{R(t)+r} dx \frac{1}{2r} (x + r - R(t)) \quad (8.2)$$

$$= \frac{2}{L} \{ (2r \cdot [r - R(t)]) + \frac{1}{2} [(R(t) + r)^2 - (R(t) - r)^2] \} \quad (8.3)$$

$$= \frac{2}{L} 2r^2 \quad (8.4)$$

$$= r^2 / L \quad (8.5)$$

$$\equiv C \quad (8.6)$$

所以网络的半径 $R(t)$ 匀速增长。所以 $R(t) \sim t$. q.e.d.

对于多维情形，在每个维度上的半径 R_1, \dots, R_d 的增长速率都相同，所以它们的线性组合的增长速率相同。即向任何一个方向的增长速率相同。证毕。

- 距中心距离为 ρ 时，密度 $\mu(\rho, \Theta, t) \sim \frac{R(t)-\rho}{L^d}$, $\mu(\rho, t) \sim \frac{(R(t)-\rho)\rho^{d-1}}{L^d}$. 到中心距离在 ρ 以内的结点个数 $\sim \rho^d$.
- 网络中结点的总数

$$N(t) = \int_0^{R(t)} \mu(\rho, t) d\rho \quad (8.7)$$

$$\sim R(t)^{d-1} \quad (8.8)$$

$$\sim t^{d-1} \quad (8.9)$$

- 总边数:
 - 边数 $v(\rho, \Theta, t) \sim u(\rho, \Theta, t)^2$ (局部每两个边都相连)
 - 总边数

$$E(t) = \int v(\rho, \Theta, t) d\sigma \quad (8.10)$$

$$\sim \int t^2 \cdot t^{d-1} dt \quad (8.11)$$

$$= t^{d+1}. \quad (8.12)$$

– 所以我们有 $E(t) \sim N(t)^{\frac{d+2}{d+1}}$.

- 标度率

1. 边数与结点数 $\gamma = \frac{d+2}{d+1}$.

– 证明：总边数：

边数 $v(\rho, \Theta, t) \sim u(\rho, \Theta, t)^2$ (局部每两个边都相连)

总边数

$$E(t) = \int v(\rho, \Theta, t) d\sigma \quad (8.13)$$

$$\sim \int t^2 \cdot t^{d-1} dt (\text{极坐标变换}) \quad (8.14)$$

$$= t^{d+1}. \quad (8.15)$$

所以我们有 $E(t) \sim N(t)^{\frac{d+2}{d+1}}$.

2. 体积与结点数 $\gamma = \frac{d}{d+1}$. (这个就不证明了, 跟上面差不多) 这是一个亚线性的现象 (Heap's law).

3. 加速增长效应 $t \sim N(t)^{\frac{1}{d+1}}$.

$$s.t. \frac{dN(t)}{dt} \sim N(t)^{\frac{d}{d+1}} \text{ 左边: } N(t) \sim t^{d+1} dN(t) \sim t^d dt \text{ and } t^d \sim N(t)^{\frac{d}{d+1}} \text{ so that } \frac{dN(t)}{dt} \sim N(t)^{\frac{d}{d+1}}.$$

8.3.3 跟真实结果不符合？增加一个参数

一个新加入的结点的生存概率是插入点的点密度的一个负指数倍: $P(\text{survive}) = \mu(\rho, \Theta, t)^{-\alpha}$.

与之前的推导类似, 在某位置 $d\sigma$ 附近的结点个数为 (从开始有结点开始, 对时间积分)

$$\mu(\rho, \Theta, t) d\sigma = \int_{\tau_\rho}^t \mu(\rho, \Theta, s)^{-\alpha} \frac{d\sigma}{L^d} ds$$

我们可以得到一个 PDE:
$$\begin{cases} \frac{\partial u}{\partial t} = \frac{1}{L^d} \mu^{-\alpha} \\ \mu(\rho, \Theta, \tau_\rho) = 0. (\text{初值条件}) \end{cases}$$

它的解是

$$\mu(\rho, \Theta, t) \sim (t - \tau_\rho)^{1/(1+\alpha)}$$

- 边: $R(t) \sim N(t)^{1 + \frac{1}{1+(1+\alpha)d}}$

- 体积 $V(t) \sim N(t)^{1+\frac{1+\alpha}{1+(1+\alpha)d}}$

$\alpha \rightarrow \infty$ 时, 上述两个幂律指数都会趋近于 1. 也就是随着 α 的增加, 超/亚线性性会降低至线性。

参考文献

- [1] Alec Kirkley, Hugo Barbosa, Marc Barthelemy, and Gourab Ghoshal. From the betweenness centrality in street networks to structural invariants in random planar graphs. *Nat Commun*, 9(2501), 2018.