

2025 年 9 月 22 日至 9 月 28 日周报

何瑞杰
中山大学, 大湾区大学

1. 项目进展

1.1. 使用神经网络学习生命游戏的演化动力学

本周受台风影响和其他因素，我的精神状态欠佳，正在努力调整中。本周计划中缺失的部分将尽量在下周补齐。

1.1.1. 对模型和训练的改动

1.1.1.1. 增大卷积核大小

增加到 5 或者 7，并期待训练好的模型第一层权重在九宫格以外的接近于零。

1.1.1.2. 缩小并行模型的参数量

1.1.1.3. 权重稀疏化

在优化器中添加 `weight_decay` 项，并注意一下能不能设置成 L1

1.1.2. 不变性

1.1.2.1. 对卷积权重的硬约束

1.1.2.2. 权重旋转不变性损失

1.1.2.3. 群等变 CNN

1.1.3. 生命游戏规则的变种

本周取 Golly 文档中若干邻域大小为 3 的九种其他规则进行实验，每种规则的具体信息列表如下

规则名称	邻域大小	邻域类型	特性描述
B36/S23	3	Moore	和 Conway 的原版生命游戏相似，但有自我复制结构
B3678/S34678	3	Moore	活细胞群中的死细胞的行为与死细胞群中的活细胞的行为相同
B35678/S5678	3	Moore	有不可预测行为的菱形斑点
B2/S	3	Moore	活细胞每代都会死亡，但该系统常常爆发
B234/S	3	Moore	单个的 2×2 会演化为一个波斯地毯
B345/S5	3	Moore	周期极长的振荡器可以自然地出现
B13/S012V	3	Von Neumann	
B2/S013V	3	Von Neumann	

1.1.4. 数据生成

经过进一步研究发现，Golly 虽然支持大量规则，但无法作为包导入 Python 中使用，只限于其程序之内。一番搜索后我找到了 `pyseagull`，并对所需的关键部分进行了检查。其运行模式十分简单，下面是一段官网给出的模拟代码：

```
import seagull as sg
from seagull.lifeforms import Pulsar

# Initialize board
board = sg.Board(size=(19,60))

# Add three Pulsar lifeforms in various locations
board.add(Pulsar(), loc=(1,1))
board.add(Pulsar(), loc=(1,22))
board.add(Pulsar(), loc=(1,42))

# Simulate board
sim = sg.Simulator(board)
sim.run(sg.rules.conway_classic, iters=1000)
```

相比于先前代码中的逻辑，它要简单得多。即使运行过程中被包装成一个函数，我们依然可以通过 `sg.Simulator` 中的 `get_history()` 方法得到这一次模拟的所有历史数据的 `ndarray`，其形状为 `[iters+1, w, h]`，其中 `iters` 为迭代轮数，`w` 和 `h` 为网格尺寸大小。

另外，`pyseagull` 还支持自定义的简单规则。生命游戏的简单规则可以写为 `B[...]/S[...]` 的字符串格式。最经典的生命游戏的规则为 `B3/S23`，意为死细胞邻居存活数为 3 时，下一时刻复活；活细胞邻居存活数为 2 或 3 时下一时刻继续存活，其他情况下一时刻细胞死亡。自定义函数签名如下

```
seagull.rules.life_rule(X: ndarray, rulestring: str)
```

该函数在 `pyseagull` 的源码实现中通过正则表达式提取 `rulestring` 中的规则信息。由于其灵活性，在需要时我们可以将其拓展为其他更加复杂的规则，例如改变邻域的形状、将邻域的贡献从各向同性改为各向异性。

我正使用该库重写数据生成部分的代码，正在调试中。

1.1.5. 对模型权重的解释

1.1.5.1. 训练完毕的神经网络作为生命游戏模拟器

1.1.5.2. 通过 CNN 权重的直接解释方案

参考资料

1. <https://pyseagull.readthedocs.io/>
- 2.

1.2. 微型抗癌机器人在血液中的动力学

1.2.1. 项目目的

微型抗癌机器人是通过癌症细胞散发出的化学吸引物 (chemoattractant) 趋化性驱动 (chemotaxis-driven) 运动，与癌细胞进行配体-受体结合后定向释放药物，达到治疗的目的。本项目研究理想状况下的微型抗癌机器人集群在血液中的动力学。

1.2.2. 建模

目前项目对血液中的化学吸引物、游离的微型机器人和与癌细胞结合的微型机器人分布进行建模。设 t 时刻，位于血液中 x 位置的化学吸引物浓度为 $c(x, t)$ ，化学吸引物正常的消耗或讲解速率为 k ，则有

$$\frac{\partial c}{\partial t} = D_c \nabla^2 c - kc + S_{\Omega_t}(x) \quad (1)$$

其中

- D_c 为化学吸引物在血液中的扩散系数
- k 为化学吸引物正常的消耗或讲解速率
- Ω_t 为癌细胞所在区域， $S_{\Omega_t}(x)$ 为癌细胞区域中 x 位置向血液中释放化学吸引物的速度

类似地，设 $\rho(x, t)$ 为游离机器人血液中的分布密度， $b(x, t)$ 为非游离的机器人的分布密度，有

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= D_\rho \nabla^2 \rho - \nabla \cdot (\chi \rho \nabla c) - k_b \rho \delta_{\Omega_t} + k_u b \\ \frac{\partial b}{\partial t} &= k_b \rho \delta_{\Omega_t} - k_u b \end{aligned} \quad (2)$$

2. 文献阅读

2.1. The Yeast Cell-cycle Network is Robustly Designed

- Fangting Li, Tao Long, Ying Lu, Qi Ouyang, and Chao Tang
- <https://pnas.org/doi/full/10.1073/pnas.0305937101>

本文来自杨武岳老师的推荐阅读，通过对酵母菌细胞周期调控网络构建简化 Bool 网络，然后研究该网络的演化特征、吸引子、主要演化路径和对外界扰动的稳定性。

2.1.1. 调控酵母菌细胞周期的核心因素

酵母菌细胞周期或增殖可以大致划分为四个时期：

- G₁**: 细胞生长，当细胞大小足够、外部营养充足时开始进入分裂过程
- S**: DNA、染色体复制
- G₂**: S 期和 M 期之间的间隔
- M**: 染色体分离、细胞分裂

即使酵母菌和细胞周期相关的基因数量众多，但其中的核心角色十分有限，它们可以分为四类

- 与激酶 (kinase) 结合的细胞周期蛋白 (cyclins) : Cln1,-2,-3、Clb1,-2,-5,-6
- 上述复合体的抑制 (inhibitors)、降解 (degraders) 和竞争蛋白 (competitors) : Sic1、Cdh1、Cdc20、Cdc14
- 转录因子 (transcription factor) : SBF、MBF、Mcm1/SFF、Swi5
- 检查点 (checkpoints): 细胞大小、DNA 的复制和损伤、纺锤体组装 (spindle assembly)

下面是一些激活和抑制的例子：

- 细胞生长到一定大小，且环境营养充足时，激活 **细胞周期蛋白(cyclin) Cln3** 和 **激酶(kinase) Cdc28** 的复合体 Cln3/Cdc28
- Cln3/Cdc28 通过磷酸化激活 **转录因子(transcription factor) SBF 和 MBF**
- SBF 和 MBF 分别激活 **细胞周期蛋白 Cln1,-2** 和 **Clb5,-6**
- Sic1 蛋白可与 Cln3/Cdc28 复合体结合，抑制后者的功能
- Clb1, -2 可以使 Swi5 磷酸化，抑制后者进入细胞核
- Cdh1 可以降解 Clb1, -2

本文为了简化模型，仅仅考虑细胞大小的一个检查点，并在一些节点上添加自降解回路。具体如 Figure 1 所示。

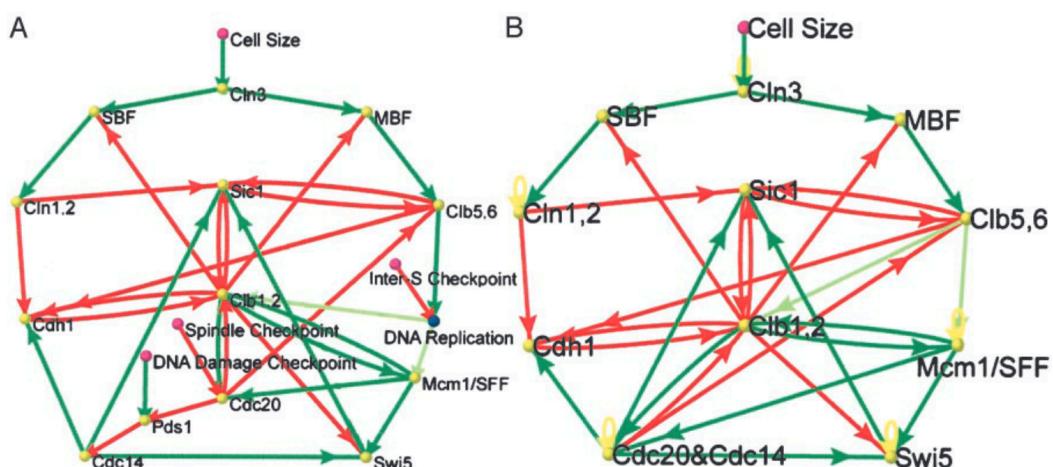


Fig. 1. (A) The cell-cycle network of the budding yeast. (B) Simplified cell-cycle network with only one checkpoint “cell size.”

Figure 1: (A) 发芽酵母的细胞周期调节网络 (B) 仅有细胞大小这一个检查点的简化网络

2.1.2. 网络动力学建模和演化实验

2.1.2.1. 动力学模型

本文的设定中，上述网络的每个节点仅拥有两种状态。具体地，设第 i 个节点在 t 时刻的状态为 $S_i(t)$ ，那么有 $S_i(t) \in \{0, 1\}$ ，并且有下面的动力学

$$S_i(t+1) = \begin{cases} 1 & \sum_{j \in \mathcal{N}(i)} a_{i,j} S_j(t) > 0 \\ 0 & \sum_{j \in \mathcal{N}(i)} a_{i,j} S_j(t) < 0 \\ S_i(t) & \sum_{j \in \mathcal{N}(i)} a_{i,j} S_j(t) = 0 \end{cases} \quad (3)$$

其中 $a_{i,j}$ 表示图中从节点 i 指向节点 j 的箭头之权重。在本文设定中，对于红色（抑制）箭头，对应的 $a_{i,j}$ 均被设置为 a_r ，绿色箭头均被设置为 a_g ，且有 $a_g = -a_r = 1$ 。对于带有自降解回路的节点，其自降解规则如下。假如某个带自降解回路的节点 i 的外部输入（也就是前面的和式）在 $t+1$ 时刻到 $t+t_d$ 时刻都为零，则该节点的状态在 $t+t_d$ 时刻变为零，即 $S_i(t+t_d) = 0$ 。

2.1.2.2. 吸引子

作者让网络由所有 2^{11} 种初始状态开始演化，发现所有的起始状态都将演化到七个吸引子中的一个，其中最多初始状态演化到达的那个吸引子对应的是占细胞周期时间最长的 G1 期。各吸引子的状态如下图所示。

Table 1: 每个吸引子容纳的起始状态大小及吸引子对应的各节点状态

Table 1. The fixed points of the cell-cycle network

Basin size	Cln3	MBF	SBF	Cln1,2	Cdh1	Swi5	Cdc20	Clb5,6	Sic1	Clb1,2	Mcm1
1,764	0	0	0	0	1	0	0	0	1	0	0
151	0	0	1	1	0	0	0	0	0	0	0
109	0	1	0	0	1	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0	0
7	0	1	0	0	0	0	0	0	1	0	0
7	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0

Each fixed point is represented in a row. The first column is the size of the basin of attraction for the fixed point; the other 11 columns show the protein states of the fixed point. The protein states of the biggest fixed point correspond to that of the G₁ stationary state.

2.1.2.3. 生物演化路径

下一步作者通过启动检查点信号（细胞大小）手动启动“细胞”的生命周期，发现网络的演化方向和真实世界中酵母菌的演化过程相同，即 $G_1 \rightarrow S \rightarrow G_2 \rightarrow M \rightarrow G_1$ ，如 Table 2 所示。除此之外，其他不在 G_1 期对应状态开始演化的情形，也会随着演化过程回归到这条路径，然后沿着这条路径演化至回到 G_1 期，如 Figure 2 所示。

Table 2: 每个吸引子容纳的起始状态大小及吸引子对应的各节点状态

Table 2. Temporal evolution of protein states for the simplified cell-cycle network of Fig. 1B

Time	Cln3	MBF	SBF	Cln1,2	Cdh1	Swi5	Cdc20 and Cdc14		Clb5,6	Sic1	Clb1,2	Mcm1/SFF	Phase
							Clb5,6	Sic1					
1	1	0	0	0	1	0	0	0	1	0	0	0	START
2	0	1	1	0	1	0	0	0	1	0	0	0	G ₁
3	0	1	1	1	1	0	0	0	1	0	0	0	G ₁
4	0	1	1	1	0	0	0	0	0	0	0	0	G ₁
5	0	1	1	1	0	0	0	0	1	0	0	0	S
6	0	1	1	1	0	0	0	0	1	0	1	1	G ₂
7	0	0	0	1	0	0	1	1	0	1	1	1	M
8	0	0	0	0	0	1	1	0	0	1	1	1	M
9	0	0	0	0	0	1	1	0	1	1	1	1	M
10	0	0	0	0	0	1	1	0	1	0	1	1	M
11	0	0	0	0	1	1	1	0	0	1	0	0	M
12	0	0	0	0	1	1	0	0	0	1	0	0	G ₁
13	0	0	0	0	1	0	0	0	0	1	0	0	Stationary G ₁

The right column indicates the cell-cycle phases. Note that the number of time steps in each phase do not reflect its actual duration.

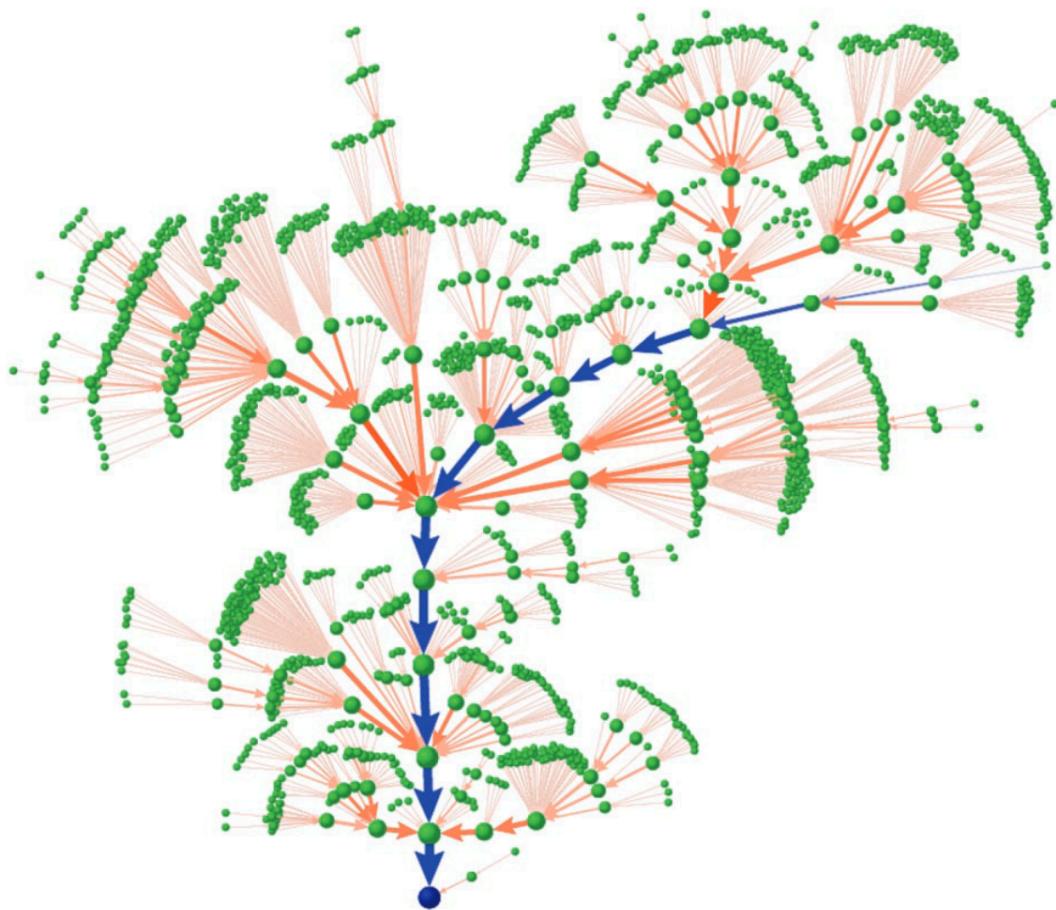


Figure 2: 最终流向吸引子 G_1 (蓝色圆点) 的 1,764 个状态 (绿色圆点) 的状态转移图。橙色箭头表示从一个状态演化至另一个状态。箭头和圆点的粗细或大小表示经过它们的演化路径数目。蓝色箭头指示现实中酵母菌的系统演化路径。

2.1.3. 和随机网络之比较

2.1.3.1. 吸引子和演化路径重合程度

作者也将酵母菌的调控网络与相同节点数量的随机网络进行了比较，后者的吸引子关于其容纳的初始状态多少呈现幂律分布。作者还定义了一个变量 w_n 其中 n 指示某个系统初始状态。该变量指示不同初始状态的演化路径之间的重叠程度，如 Figure 3 所示。具体而言，先定义 $T_{n,i-1,i}$ 为系统所有初始状态演化路径中，经过当前初始状态在 $t-1$ 时刻演化至 t 时刻经过的有

向边 $A_{j,k}$ (这里的有向边是指如 Figure 2 那样的系统状态迁移图的有向边，不是系统本身的图结构) 的演化路径条数， L_n 定义为该初始状态下演化路径到达吸引子经过的状态数目 (Figure 2 中到达吸引子之前的跳数)。则 w_n 定义为

$$w_n = \frac{1}{L_n} \sum_{i=1}^{L_n} T_{i-1,i} \quad (4)$$

通过图中可以看到，随机网络产生像酵母菌调控网络那样的一个巨大吸引子的概率很低，且后者的 w_n 值总体更高，这说明大部分的演化路径重叠度高、殊途同归。

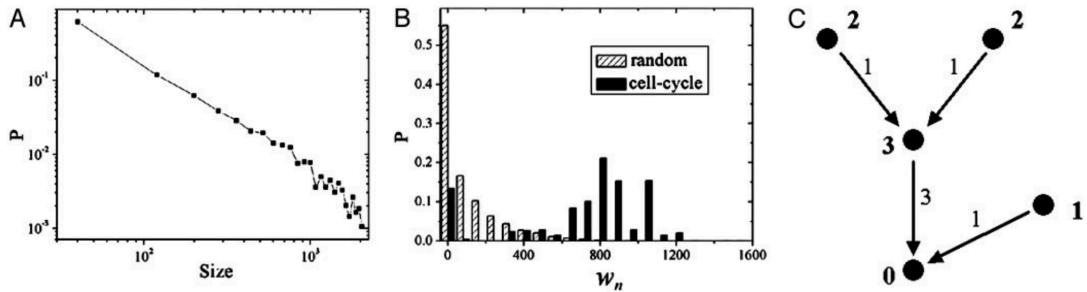


Figure 3: 酵母菌调控网络和相同节点数的随机网络之比较。(A) 随机网络吸引子大小的分布 (B) 酵母菌网络和随机网络的 w_n 值分布 (C) w_n 值的计算规则。

2.1.3.2. 网络稳定性

作者接下来对酵母菌网络和用于与之对比的随机网络做了扰动实验，观察扰动后系统演化中的最大吸引子容量的大小变化 $\Delta B/B$ 。其中 B 为扰动前的最大吸引子容量， ΔB 为扰动后最大吸引子容量的变化量。本文作者采用的扰动方式 (对系统本身的网络，而不是状态演化图) 具体为：删除一条边、增加一条本不存在的边 (可能是绿色或红色)、将一条边变成另一种颜色。作者对这三种扰动造成的 $\Delta B/B$ 大小以及三种方法的平均效应进行统计，结果如 Figure 4 所示。

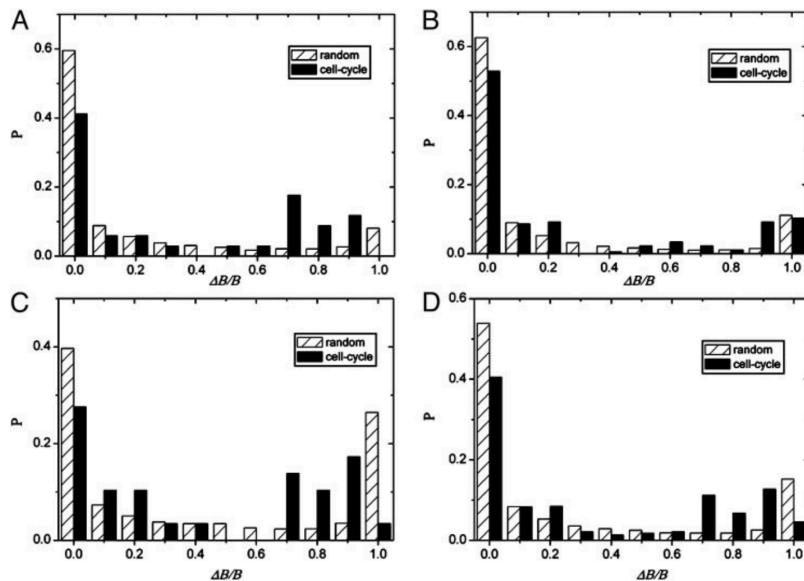


Figure 4: 酵母菌调控网络和相同节点数的随机网络面临扰动时 $\Delta B/B$ 的分布。(A) 随机删除 34 个连接 (B) 随机添加 174 条连接 (C) 随机改变 29 条连接的性质 (D) A-C 的平均

原论文在得到结果后声称相比于随机网络，酵母菌网络在扰动下 $\Delta B/B = 1$ 的频数显著下降，该数值对应着之前的吸引子在扰动后被完全抹去。原文作者根据上图结果还得出了“对大部分扰动

而言，最大吸引子的相对容量变化很小”，很遗憾我并未从图中得出：如(A)、(C)、(D)中的酵母菌网络在扰动后存在相当一部分 $\Delta B/B > 0.5$ 的情况，因此我认为这个推论有些差强人意。

作者对扰动后 w 值的变化分布也进行了统计，得到的结果与 $\Delta B/B$ 的分布相似。

2.1.4. 演化动力学的其它参数组合和其他检查点

作者尝试其他情况的 Equation 3 中的参数，发现系统动力学对参数 $a_{i,j}$ 的选取不敏感。对于其他检查点的情况，作者进行了单一开放一个其他检查点的实验，实验结果和上文类似。

2.1.5. 总结和讨论

本文发现酵母菌的调控网络具有一个超级吸引子，且生物意义上的标准系统演化过程对应着模拟实验中的吸引轨道。由于本文中对酵母菌系统进行了较大的简化，一个合理的猜测是考虑更多因素的系统依然会具有简化后系统的良好性质（即对应 G_1 期的超级吸引子、吸引轨道和对扰动的稳定性）。这样的性质可以使的细胞周期系统也许可以和其他系统更好的配合，并使酵母菌在进化中占据有利地位。

3. 问题解决记录

3.1. Typst 相关

3.1.1. grid() 中的多重对齐

多重对齐可以用 + 同时指定。例如想实现高度居中 (horizon) 和水平居中 (center)，则可以在对应函数的 align 参数中填入 horizon+center。下图中从左至右的对齐方式分别为 center、left+horizon 和 bottom+center。

a b c		
d e f	a b c	
g h i	d e f	
j k l	g h i	
m n o		

a b c
d e f
g h i
j k l

具体地，typst 文档中定义了下面的若干种对齐变量值：

- start: 对齐至文本方向起点
- end: 对齐至文本方向终点
- left: 对齐至左侧边界
- center: 对齐至水平方向的中央位置
- right: 对齐至右侧边界
- top: 对齐至上侧边界
- horizon: 对齐至水平中线
- bottom: 对齐至下侧边界

3.1.2. 三线表的绘制

```
#table(
  columns: 3,
  stroke: none,
  table.hline(),
  table.header([a], [b], [c]),
  table.hline(stroke: 0.5pt),
  [d], [e], [f],
  [g], [h], [i],
  table.hline(),
)
```

a b c
d e f
g h i

参考资料

1. <https://typst.app/docs/reference/model/table/#parameters-align>

3.1.3. 数学公式自动编号

实现简单的数字编号，只需添加下面的代码至文件开头

```
#set math.equation(numbering: "(1)")
```

3.2. Python 相关

3.3. 深度学习相关

4. 下周计划

论文阅读

1. 生成模型

- Denoising Diffusion Probabilistic Models 完成
- Sliced Score Matching: A Scalable Approach to Density and Score Estimation

2. 动力学

- 暂无

项目进度

1. 使用神经网络学习生命游戏的演化动力学

- 填补所有本周未完成的部分

2. 微型抗癌机器人在血液中的动力学

- 尝试学习并撰写 PDE 的数值模拟代码

理论学习

1. 随机过程课程

- 完成第一章（独立随机变量序列）和第二章（Poisson 过程）

2. 随机微分方程

- 完成第四章，并开启第五章

3. 机器学习理论

- Markov Chain Monte Carlo