arXiv:2301.06224v1 [math.GM] 16 Jan 2023

在科学中寻找有意义且可行的应用数学问题

Yue Wang

摘要

在这篇简短的综述中，我将总结自己在应用数学三个领域的研究经历：数学生物学、应用概率和应用离散数学。具体来说，我将介绍每个项目是如何启动的，以及曾采用过哪些错误的方法。这些细节对于学习如何开展研究十分重要，但无法从研究论文中读出来。我希望应用数学领域的学生和青年研究人员能从这份总结中有所收获。

1 引言

自2012年起，我一直从事应用数学研究。尽管我的研究或许并不十分引人注目，但我的研究经历应该能对应用数学领域的学生和初级研究人员有所帮助，无论是正面的（该做什么）还是反面的（应避免什么）。这篇综述的重点是讲述每篇论文背后的故事：我是如何发现这个问题的，又采用了哪些正确或错误的方法？这些经验教训无法仅靠阅读论文获得。如果我年轻得多，我会很乐意阅读这类综述。这正是我撰写这篇总结的初衷。在这篇综述中，我将总结自己在数学生物学（第2节）、应用概率（第3节）和应用离散数学（第4节）方面的研究。

为方便读者，我决定将最后的讨论移到这里。以下是你可能从这篇总结中得到的一些启示，你可以直接跳转到相应的小节：证明了一些结果后，可以考虑更一般或更现实的情形，尝试加以强化。如果某项改进是现实世界所必需的，那么即使改进幅度很小，也值得去做（第2.1小节）。作为一名应用数学研究者，有时不可避免地要在不熟悉的领域开展工作。在这种情况下，找到相应领域的合作者非常重要。因此，结交更多不同背景的朋友总是有益的（第2.2小节）。解决一个问题后，试着从相反的方向思考，以便利用已有结果提出另一个问题（第2.3小节）。着手解决难题之前，查找相关论文并与不同专家交流，以避免重复造轮子（第3.1小节）。撰写论

文时，要考虑潜在的读者（也包括审稿人），尤其是当论文涉及多个领域时（第3.1小节）。阅读提出新奇方法的论文时，要在反常情形下检验这些方法，并尝试构造反例（第3.2、3.3小节）。开发出新方法后，应始终尽力将其应用于一些真实数据，或至少应用于模拟数据。否则，审稿人总会对此提出质疑（第2.2小节）。此外，要评估或比较算法，最好的方法是实现并测试它们，因为你的理论分析可能会忽略某些因素。学习了本领域之外的知识后，尝试运用你的数学思维模式对其进行分析（第4.1小节）。提高你的编程技能。你需要实现自己开发的方法，而这有时并非易事，尤其是在需要设计快速算法时（第4.2小节）。否则，你就需要一位合作者来完成这项工作。希望你能找到有意义且切实可行的问题，并避免重蹈我犯过的错误。

2 数学生物学

2.1 种群动力学

在某些类型的癌症中，存在多种可以相互转化的表型。研究发现，无论从哪一种表型开始，经过一段时间后，不同表型的最终比例总会收敛到相同的常数。这种现象称为表型平衡。我读到一篇论文，作者使用马尔可夫链模型来解释这一现象 [7]。这个模型在数学和生物学方面都存在一些缺陷。我建立了一个线性常微分方程模型，并证明表型平衡现象可解释为收敛到唯一吸引不动点 [40]。这项工作也促使其他研究开始探讨癌症中的多种表型 [15, 3]。后来，我建立了一个在生物学上更合理的分枝过程模型。我尝试证明该分枝过程模型中的大数定律，以解释表型平衡现象。我在一篇概率论文中找到了一些结果 [8], ，但这些结果要求满足一个特定条件：所有表型构成一个单一的沟通类，而现实中未必如此。基于这篇论文中的结果，我采用了一些复杂的论证来去除这一条件，并证明了我所需要的大数定律 [10]。

我与一个培养白血病细胞的湿实验室合作。实验中有许多细胞群体，起始细胞数分别为 10/4/1，每天都会测量各群体的细胞面积。由于一些技术问题，小群体的细胞面积数据不太准确。因此，分析只能针对相对较大的群体进行。我处理了数据并绘制了不同的图表。最后，我找到了最能体现关键特征的图表。在这张图中，我对所有生长曲线进行了平移，使它们都从细胞数达到 2500 的时间点开始。在这张图中，所有起始细胞数为 10 的群体都快速生长，而一些起始细胞数为 1 的群体则生长得慢得多。我构建了许多玩具模型，并找到了最可能的一个模型，其中包含多个生长速率不同的表型。尽管它们来自同一细胞系，但可能存在多种

表型 [16]。生长速率的差异可以简单解释：从 10 个细胞开始时，很可能至少有一个快速生长的细胞，它会占据主导地位，因此整体生长速率较高；从 1 个细胞开始时，它可能属于生长缓慢的类型，因此整体生长速率较低 [19, 34]。对数学研究人员来说，刚开始与实验生物学家交流和合作可能不太顺畅，因为双方的思维方式不同。我还参与了其他有关种群动力学的项目 [1, 23, 6, 36]。

2.2 基因调控

基因转录为信使RNA，随后信使RNA被翻译成蛋白质。一些蛋白质会影响其他基因的表达（相互调控），或影响其自身对应基因的表达（自我调控）。由于通过生化方法直接确定基因调控关系极其困难，因此人们提出了许多基因调控推断方法 [41, 2]。我阅读了一些关于基因调控推断的论文，发现其中的设定多种多样：不同方法采用不同模型，而这些模型基于不同的假设。此外，不同方法可能需要不同的数据类型。面对这种复杂情况，我逐渐形成了一套分类体系。数据类型可以按照四个维度分类：测量基因表达还是表型；在单细胞水平还是整体水平测量；在一个时间点还是多个时间点测量；测量时是否进行干预。对于多个时间点的单细胞数据，我们还需要区分能否测量联合概率分布，还是只能测量边缘概率分布。总共有20种数据类型。对于每种数据类型，我总结了已知方法、开发了新方法，或证明该数据类型无法用于基因调控推断。我实现了这些新方法，并发现其中一种方法对微小扰动过于敏感，因此不适用于真实数据。我原本只打算开发一种新的推断方法，最终却写成了一篇很长的论文 [30]。为了完成这篇长文，我与一位实验生物学领域的朋友合作，由他撰写生物学部分。我希望这篇论文能对不同背景的研究人员都有帮助。因此，我设法将不同章节模块化，并添加了一张图，标明各章节之间的依赖关系。这样，数学家和生物学家就可以跳过不同章节，直接阅读自己感兴趣的内容。

在研究相互调控的推断时，我还想到了一种推断自我调控是否存在的方法。在基因表达的马尔可夫链模型中，如果不存在自我调控，那么表达水平的方差应大于均值[24]。因此，如果均值大于方差，就可能存在自我调控。由于这个模型有无穷多个状态，我没能证明自己想要证明的结论。于是，我向一位纯数学领域的朋友求助。我们合作完成了所有证明。作为一篇应用概率论文，结果不够深入；作为一篇数学生物学论文，推断方法又没有实际应用。审稿人对方法学论文常会提出一个自然的问题：它有什么应用？因此，我不得不实现这些方法，并将其应用于一些实验数据。

2.3 发育生物学

考虑组织移植实验：将一块供体组织移植到宿主组织中。在宿主发育过程中，我们观察移植组织的命运。结果可能是正常/异常、类似宿主/类似供体。我得到了一些汇总的实验结果，目标是开发一些方法来推断未知的实验结果。由于数据量太少，我尝试了几种复杂的模型，但都没有成功。最后，受伊辛模型中能量函数的启发，我采用了一种直接的方法：我们可以估计实验之间的相似度。设计一个简单的惩罚函数。对于每一种对未知结果的猜测，计算总惩罚值：结果不同（无论是已知结果还是猜测结果）的一对相似实验将受到更高的惩罚。惩罚值最小的猜测就是概率最高的猜测 [31]。推断出未知结果后，我从相反的方向思考这个问题：如果我们想用这种方法推断实验结果，并且可以选择进行哪些实验，那么如何设计才能使成本最低？这个问题最终转化为 n 维格上的一个着色问题。在一位纯数学领域朋友的帮助下，我解决了这个实验设计问题。

我曾接手一个关于发育生物学中位置信息的未完成项目。在胚胎发育过程中，不同位置的细胞似乎知道自己身处何处，并发育成不同（但有序）的组织。为了解释这一现象，生物学家创造了“位置信息”一词 [35]。不同生物学家对这一概念的理解各不相同。我仔细分析后，提出了界定何为位置信息的三条判据。我还证明了没有任何一条判据是多余的，因为去掉其中任意一条，都会导致某些非位置信息也被纳入其中。通过这种逻辑方式，位置信息得到了明确界定。最后，我完成了一篇数学风格的论文，却没有用到任何数学[25]。我还参与了另一个发育生物学项目 [26]。

3 应用概率

3.1 随机过程中的热力学

我的博士课题组有研究随机过程热力学的传统[13, 38, 17, 18, 37, 4, 5]。我研究了一个问题：比较马尔可夫链及其覆盖空间的热力学性质。这个问题与拓扑学和群论有关，而我对这两者都不太熟悉。在尝试了许多错误的思路之后，我找到了应该证明的正确定理。这个定理看起来是对的，但我却卡了很久，并尝试向不同的数学家请教。最后我意识到，自己被这个问题吓住了，而一个非常宽松的不等式就足够了。

后来，我计划将这一结果推广到更一般的覆盖空间。我遇到了瓶颈，向许多数学家请教，还修了一门相关课程。最后，我偶然得知，之前的一篇论文 [11] 已经涵盖了我想做的所有内容，甚至还包括一些我已经

证明过的结果。因此，我的推广计划彻底失败了。那篇论文的作者给我研究的对象起了一个截然不同的名称，所以我在查找参考文献时没有找到它。此外，这个小领域的研究者大多在欧洲，因此我请教的数学家都不知道这篇论文。幸运的是，他们的方法与我的不同，所以我的论文最终得以发表 [28]。这篇论文的发表过程也很艰难。它涉及统计物理、概率和代数拓扑。审稿人可能不熟悉其中某个领域，从而难以理解论文内容。凭借我在概率方面的技巧，我还证明了有限马尔可夫链中的定理 [27]。

3.2 因果推断

有一次，我的导师（他并不研究因果推断）听了一场关于因果效应量化的精彩报告，之后把一些相关论文发给了我。传统的因果量在两个可能原因非常相似时会遇到问题。有一些新方法利用这种细微差异来提升性能 [9, 39]。从定义来看，当两个可能原因完全相同时，这些因果量便没有定义。我尝试使用延拓法，也就是在联合分布上加入小扰动，使这些因果量有定义，然后再将扰动逐渐减小至零。这却引出了许多麻烦：在这些因果量没有定义的分布附近，它们会剧烈波动。因此，当两个可能原因完全相同（或者更一般地说，马尔可夫边界不唯一）时，这些因果量从根本上说是定义不良的。我尝试设计新的因果量来避开这个问题，但没有成功。随后我意识到，问题本身或许就是定义不良的。我为因果量提出了三个判据，并证明当分布存在多个马尔可夫边界时，这些判据彼此不相容[29]。由于存在多个马尔可夫边界的分布会带来问题，我还设计了快速算法，以便从有限数据中检测出这类分布。在实现这些算法之前，我做了一些理论分析，并认为其中一个算法最好。实现之后，我却发现另一个算法才是最好的，因为我在理论分析中低估了某些因素。由于我的导师和我本人都没有接受过因果推断方面的训练，我花了很长时间才找到一位合作者。在这类数学性很强的领域，专家的指导对于把控研究方向和采用正确的写作风格都至关重要。

3.3 强化学习

有一次，一位研究运筹学的朋友给我发来几篇关于在线学习的论文。在这些论文中，玩家反复选择一个动作并获得随机奖励（同时也会部分了解系统参数）。玩家的目标是采用明智的策略，使累积奖励最大化。我试着理解一篇颇为精妙的论文，却发现用于评估策略表现的指标存在问题。该指标只考虑最坏情况的表现，因此会导致一些奇怪的行为。具体而言，

历史数据无法帮助找到最优策略。这个领域对具有概率背景的人来说相当友好，我只花了两个月就解决了这个问题 [32]。尽管如此，这类负面结果很难发表。后来，我还研究了历史数据受到污染的情形 [33]。这篇论文的问题在于模型过于简单（因此可以显式计算出最优策略），而更复杂的模型只能通过数值方法研究，无法提供足够的洞见。

4 应用离散数学

4.1 继承法

我在继承法方面写过一篇论文 [20], ，尽管我从未修过法律课程，甚至没有学过社会科学。这个项目的起因出乎意料。有一次，我在社交媒体上读到一个关于遗产继承的故事：Alice 的父亲先于她的祖父去世。因此，她父亲的一部分遗产归她的祖父所有；随后（在她的祖父去世后），这部分遗产又归她的叔叔和姑姑所有。如果她的父亲在她的祖父母之后去世，那么她的叔叔和姑姑就不会继承她父亲的遗产。原因在于，在中国，民法典规定父母、子女和配偶可以平等分割遗产。

从这个故事中，我了解到，在中国的继承法中，不同的死亡顺序可能导致不同的继承结果。如果我们将死亡和继承视为作用于不同人财富的算子，那么这个算子就是非交换的。这引出了一个自然的问题：如果多位亲属在同一场灾难中死亡，而死亡顺序又无从得知，会怎样？

我向一位法学领域的朋友请教，得知中国继承法中考虑过这种情况。死亡顺序根据亲属关系确定：如果一位死者没有仍在世的可继承亲属，则推定此人先死亡；如果两位死者都有仍在世的可继承亲属，且属于同一辈分，则推定二人同时死亡，彼此不发生继承；否则，推定辈分较高者先死亡。例如，姐妹被推定为同时死亡，儿子则被推定为晚于父亲死亡。

这项关于死亡顺序的规定看起来不太严谨。首先，“辈分”一词可能没有明确定义。例如，诺贝尔奖得主加夫列尔·加西亚·马尔克斯娶了他叔叔的嫂子。在这种情况下，该如何比较加夫列尔、他的妻子和他叔叔的辈分？

即使不考虑“辈分”带来的麻烦，这项关于死亡顺序的规定仍然不太妥当。借助离散数学中的一些思路，我构造了一个反例：对于多位死者，已知其中一些人的死亡顺序，而对未知顺序作出的规定却与已知顺序相矛盾。

尽管 唉，这个反例既不自然，也不受那位朋友认可。 in

法学方面，我开始寻找一种能够确定死亡顺序、又不会导致矛盾的方法。经过几番尝试，我仍没找到好的解决方案，于是开始怀疑这样的方案是否存在。如果无法确定死亡顺序，还可以采取另一种思路：之所以需要确定死亡顺序，是因为继承法是非交换的。我尝试构造一套交换的继承法，但也失败了。

由于两种方法都行不通，我猜想死亡顺序问题无法解决。我为继承法提出了一些合理的判据。随后我证明，在这些判据下，不存在交换的继承法，而且无论如何指定死亡顺序，都会导致矛盾。这些证明只需要运用离散数学中的一些基本技巧。总之，任何继承法都是不完整的。提出合理的判据并证明它们彼此不相容，是推导不可能性结果的一种标准方法，类似于投票中的阿罗不可能定理 [14]。当然，实际过程恰恰相反：我先思考如何证明这类不可能性结果，再尝试提出必要的判据并剔除冗余判据。

交换的继承法与我提出的三项判据（性别平等、子女继承权和父母的条件继承权）不相容。由于无法指定死亡顺序，我转而尝试构造一种交换的继承法，使其只违反一项判据。我证明了，这种只违反一项判据的交换继承法只有四种。因此，该判据就是实现定义明确的继承法所必须付出的代价。

证明这个定理后，我开始撰写论文。为了补充引言部分，我查阅了不同国家的继承法。法国民法典规定，如果无法确定两个人的死亡顺序，他们就不能相互继承。这意味着，实际上并未指定死亡顺序（两者不可比较）。了解到这一点后，我意识到一个令人尴尬的事实：我只证明了指定死亡顺序会导致矛盾，但法国的这种做法并未引入新的顺序，因此不会产生矛盾。由此可见，它是解决死亡顺序问题的完美方案。于是，我修改了证明，将这种情况也纳入其中，最终得出的结论是：法国的做法是唯一有效的解决方案。此前关于仅有四种交换继承法只违反一项标准的结果，在实践中也因此变得毫无用处。由于这些结果仍有一定的理论价值，我便将它们移到了附录中。

这篇论文处境尴尬：法学研究者看不懂数学证明，而应用数学家对法律不感兴趣。因此，很难找到合适的期刊和审稿人（尽管最终还是找到了）。此外，也很难获得指导或找到合作者。

4.2 算法设计

虽然我没有接受过系统的计算机科学训练，但我运用离散数学技巧设计算法并证明其正确性。一个项目研究的是如何寻找转座子：给定来自不同个体的一些已注释基因序列，例如

$\{1,2,3,4,5,6\}$, $\{1,2,5,3,4,6\}$, $\{1,2,3,4,6,5\}$，应该如何找出位置发生变化的基因？在这个例子中，显然是基因 5 的位置发生了变化。不过，我们也可以认为基因 3、4、6 的位置发生了变化，以此解释差异。当然，我们可以尽量减少将所有序列转换为相同序列所需的操作次数。我提出了一个更简洁的解决方案：确定最长公共子序列，其余部分就是所有发生转位的基因。在一位计算机科学专业朋友的帮助下，我实现了一种快速算法，并用它来比较细菌的基因[12]。尽管确定最长公共子序列是计算机科学中的经典问题，我还研究了存在多个最长公共子序列的情况，并据此发表了一篇新论文 [22]。

另一项算法设计工作需要比较两棵标签可能重复的有根树，其中不同子顶点之间没有顺序。我阅读了许多相关论文，却没有找到可以直接应用于这种情况的方法。我首先设计了一种“最佳匹配”度量，用于计算所有可能的变换后树之间的最小距离。然而，我认为无法在多项式时间内计算这种度量，于是改用另一种计算速度更快的度量。后来，我发现最佳匹配度量可以在平方时间内计算。因此，我在论文中同时纳入了这两种度量，尽管最佳匹配度量明显更好 [21]。这些度量用于比较植物的早期发育。

参考文献

[1] Angelini, E., Wang, Y., Zhou, J. X., Qian, H., and Huang, S. 癌症治疗内在极限模型：治疗诱导的细胞死亡与治疗诱导的干性之二元性。《PLOS计算生物学》18, 7 (2022), e1010319.[2] Bocci, F., Zhou, P., and Nie, Q. spliceJAC：基于单细胞转录组数据识别转变基因和状态特异性基因调控。分子系统生物学 18, 11 (2022), e11176.[3] Chen, X., Wang, Y., Feng, T., Yi, M., Zhang, X., and Zhou, D. 可逆表型可塑性癌症动力学特征中的超调与表型平衡。《理论生物学杂志》390 (2016), 40–49.
[4] Cheng, Y.-C., and Qian, H. 随机扰动下非线性系统的随机极限环振荡。《统计物理学杂志》182, 3 (2021), 1–33.[5] Cheng, Y.-C., Qian, H., and Zhu, Y. 条件概率分布序列的渐近行为与正则系综。Annales Henri Poincar ´e 22, 5 (2021), 1561–1627。

[6] Dessalles, R., Pan, Y., Xia, M., Maestrini, D., D'Orsogna, M. R., and Chou, T. 异质性胸腺输出与稳态增殖如何塑造初始 T 细胞克隆计数。《免疫学前沿》12 (2021).[7] Gupta, P. B., Fillmore, C. M., Jiang, G., Shapira, S. D., Tao, K., Kuper-wasser, C., and Lander, E. S. 随机状态转变使癌细胞群体产生表型平衡。Cell 146, 4 (2011), 633–644.[8] Janson, S. 多类型分枝过程与广义波利亚 urn 的函数极限定理。《随机过程及其应用》110, 2 (2004), 177–245.[9] Janzing, D., Balduzzi, D., Grosse-Wentrup, M., and Schölkopf, B. 量化因果影响。《统计年鉴》41, 5 (2013), 2324–2358.[10] Jiang, D.-Q., Wang, Y., and Zhou, D. 多表型细胞群体动力学中的表型平衡作为概率收敛。PLOS ONE 12, 2 (2017), e0170916.[11] Kaimanovich, V. A., and Woess, W. 空间齐次马尔可夫链的边界与熵。《概率年刊》(2002), 323–363.[12] Kang, Y., Gu, C., Yuan, L., Wang, Y., Zhu, Y., Li, X., Luo, Q., Xiao, J., Jiang, D., Qian, M., et al. 原核生物基因组重排的灵活性与对称性揭示了与谱系相关、由核心基因定义的基因组组织框架。mBio 5, 6 (2014), e01867–14.[13] Ma, Y.-A., and Qian, H. 线性不可逆随机热力学的普适理想行为与宏观功关系。《新物理学杂志》17, 6 (2015), 065013.[14] Maskin, E., and Sen, A. 阿罗不可能定理。哥伦比亚大学出版社，2014。[15] Niu, Y., Wang, Y., and Zhou, D. 癌细胞的表型平衡：从平均水平稳定性到逐路径收敛。《理论生物学杂志》386 (2015), 7–17.[16] Qian, H., and Cheng, Y.-C. 单细胞计数及其异质性计算：从表型频率到定量生物标志物的均值。《定量生物学》8, 2 (2020), 172–176.[17] Qian, H., Cheng, Y.-C., and Thompson, L. F. 随机变化的三元表示以及熵及其涨落的起源。arXiv预印本 arXiv:1902.09536 (2019)。

[18] Qian, H., Cheng, Y.-C., and Yang, Y.-J. 复杂动力学涌现能量学的运动学基础。EPL（《欧洲物理快报》）131, 5 (2020), 50002.[19] Wang, Y. 随机动力学与癌症单细胞生物学统计分析中的若干问题。博士论文，华盛顿大学，2018.[20] Wang, Y. 关于继承与死亡顺序的不可能性结果。PLOS ONE 17, 11 (2022), e0277430.[21] Wang, Y. 带标签有根无序树上的两种度量。分子生物学算法 17, 1 (2022), 1–17.[22] Wang, Y. 用于确定可转座基因的最长公共子序列算法及其应用。arXiv预印本 arXiv:2301.03827 (2023).[23] Wang, Y., Dessalles, R., and Chou, T. 建模生育控制政策对中国人口与年龄的影响：延迟生育和最低生育年龄限制的作用。《英国皇家学会开放科学》9, 6 (2022), 211619.[24] Wang, Y., and He, S. 基因表达自我调控的推断。arXiv预印本 arXiv:2201.03164 (2022).[25] Wang, Y., Kropp, J., and Morozova, N. 形态发生理论中位置信息／值的生物学概念。《国际发育生物学杂志》64, 10-11-12 (2020), 453–463.[26] Wang, Y., Minarsky, A., Penner, R., Soulé, C., and Morozova, N. 形态发生模型。《计算生物学杂志》27, 9 (2020), 1373–1383.[27] Wang, Y., Mistry, B. A., and Chou, T. SELEX离散随机模型：适配体捕获概率与方案优化。《化学物理学杂志》156, 24 (2022), 244103.[28] Wang, Y., and Qian, H. 克劳修斯和开尔文表述的第二定律及不可逆性的数学表示。《统计物理学杂志》179, 3 (2020), 808–837.[29] Wang, Y., and Wang, L. 退化系统中的因果推断：一项不可能性结果。载于国际人工智能与统计会议（2020），PMLR，第3383–3392页。[30] Wang, Y., and Wang, Z. 基因调控网络结构的推断。《理论生物学杂志》539 (2022), 111055.[31] Wang, Y., Zhang, B., Kropp, J., and Morozova, N. 组织移植实验的推断。《理论生物学杂志》520 (2021), 110645。

[32] Wang, Y., and Zheng, Z. 利用离线数据衡量在线定价中的策略表现。可见于SSRN 3729003 (2021).[33] Wang, Y., Zheng, Z., and Shen, Z.-J. M. 利用受污染的离线数据进行在线定价。可见于SSRN 4320324 (2023).[34] Wang, Y., Zhou, J. X., Pedrini, E., Rubin, I., Khalil, M., Qian, H., andHuang, S. HL60白血病细胞群体中的多种表型。arXiv预印本 arXiv:2301.03782 (2023).[35] Wolpert, L. 位置信息与图式形成。发育生物学当前主题 117 (2016), 597–608.[36] Xia, M., Li, X., and Chou, T. 年龄结构Lotka-Volterra模型与过度补偿的涌现。美国物理学会通报（2023）。[37] Yang, Y.-J., and Cheng, Y.-C. 连续马尔可夫过程的势与随机扰动。Journal of Physics A: Mathematical and Theoretical 54, 19 (2021), 195001.[38] Ye, F. X.-F., Wang, Y., and Qian, H. 随机动力学：马尔可夫链与随机变换。Discrete & Continuous Dynamical Systems-B 21, 7 (2016), 2337.[39] Zhao, J., Zhou, Y., Zhang, X., and Chen, L. 用于量化网络中直接关联的部分互信息。美国国家科学院院刊 113, 18 (2016), 5130–5135.[40] Zhou, D., Wang, Y., and Wu, B. 具有细胞可塑性的多表型癌症模型。《理论生物学杂志》357 (2014), 35–45.[41] Zhou, P., Wang, S., Li, T., and Nie, Q. 通过多尺度随机动力学从单细胞转录组数据中解析过渡细胞。Nature Communications 12, 1 (2021), 1–15。