利用神经符号回归发现网络动力学

作者： Zihan Yu [1,2], Jingtao Ding [1,2] & Yong Li [1]

[1] 清华大学电子工程系，信息科学技术北京国家研究中心，中国北京。
[2] 这些作者贡献相同：Zihan Yu, Jingtao Ding。
电子邮件： dingjt15@tsinghua.org.cn; liyong07@tsinghua.edu.cn
论文信息：

- **期刊：** Nature Computational Science
- **接收日期：** 2025年1月31日
- **接受日期：** 2025年9月19日
- **在线发表日期：** 2025年10月23日
- **DOI:** https://doi.org/10.1038/s43588-025-00893-8

摘要

原文翻译
网络动力学是分析高维复杂系统属性和理解其行为的基础。尽管许多领域积累了大量的观测数据，但只有少数具有清晰底层原理的领域存在数学模型。在这里，我们展示了一种神经符号回归方法可以通过自动从数据中推导公式来弥合这一差距。我们的方法将高维网络上的搜索简化为等效的一维系统，并使用预训练的神经网络来指导精确的公式发现。应用于十个基准系统，它恢复了底层动力学的正确形式和参数。在两个经验性的自然系统中，它修正了现有的基因调控和微生物群落模型，分别将预测误差降低了59.98%和55.94%。在跨越不同尺度的人类移动网络的流行病传播中，它发现了在不同尺度上表现出相同节点相关性幂律分布的动力学，并揭示了国家层面的干预效果差异。这些结果表明，机器驱动的网络动力学发现可以增进对复杂系统的理解，并推动复杂性科学的发展。

深度解读
这篇摘要为我们描绘了一幅激动人心的科学图景：我们正处在一个数据爆炸但理论匮乏的时代。想象一下，我们拥有关于社交网络、生态系统、大脑神经元乃至全球流行病传播的海量数据，就像拥有了一部记录宇宙万物运行的巨著，但这本书是用我们不认识的语言写成的。我们能看到"发生了什么"，却不知道"为什么会发生"。这里的"网络动力学"就是我们试图寻找的"语法规则"——即描述这些复杂关联系统如何随时间演化的数学方程式。传统上，发现这些规则依赖于少数天才科学家的灵感和直觉，这是一个缓慢且充满偶然性的过程。

这篇论文的核心贡献，在于提出了一种名为"神经符号回归"的自动化方法，它就像一个人工智能版的"罗塞塔石碑"，能够自动将海量数据"翻译"成简洁、深刻的数学公式。它的第一个"独门秘籍"是将极其复杂的高维网络问题（想象一下成千上万个节点相互作用）巧妙地"降维打击"，变成一个等效的一维问题来处理，这极大地简化了问题的难度。第二个"杀手锏"是利用一个经过大规模数据"预训练"的神经网络（你可以把它想象成一个博览群书、经验丰富的AI导师）来指导公式的搜索过程，避免了在无限的可能性中盲目摸索。

摘要中提到的三个关键成果，就像是这场科学探索的"三幕剧"，层层递进地展示了该方法的强大威力。第一幕"恢复"：在10个我们已经知道正确答案的"模拟宇宙"中，这个方法百分之百准确地找到了已知的物理或生物学定律，这证明了它的可靠性。第二幕"修正"：在真实的基因调控和微生物群落研究中，它不仅找到了模型，还指出了现有权威模型中的错误，并将预测误差降低了超过一半（59.98%和55.94%）。这个数字不仅仅是性能的提升，它几乎是在宣告：我们过去对这些生命系统的理解存在根本性的缺陷。第三幕"发现"：在面对尚无现成理论的COVID-19传播这一复杂社会问题时，它从数据中发现了全新的动力学规律，揭示了不同国家干预政策效果的深层数学原理，并找到了一个跨越所有空间尺度都存在的普适性"幂律分布"。这表明，该方法不仅能验证和修正已知，更能探索未知，成为推动科学前沿的强大引擎。最终，这篇文章的结论是，我们可能正站在一个由机器驱动的科学发现新时代的门槛上。

引言

宇宙的互联之网与理解的瓶颈

原文翻译

在复杂系统中，许多组件之间的局部相互作用会产生无法仅由单个组件解释的涌现性全局行为。理解这些现象需要网络动力学，即使用数学公式通过节点间的相互作用来描述网络节点的状态变化。网络动力学为分析复杂系统属性（如可控性、韧性、稳定性以及对扰动的响应）提供了理论基础，涵盖了从神经、生态到社交网络等多种系统。传统上，网络动力学的发展依赖于专家的直觉和经验，这使得数学公式的建立仅限于少数被充分研究的领域，如生物学和生态学，而大多数缺乏第一性原理知识的系统仍然没有合适的模型。幸运的是，深度学习和神经符号人工智能的发展推动了一种数据驱动的范式，旨在基于网络结构和节点活动的观测数据，发现组件如何自我调节以及如何与其它组件相互作用。

深度解读

这段引言为我们揭示了现代科学面临的一个核心挑战：复杂性。想象一下一群椋鸟在空中形成的壮观鸟群（即"椋鸟群舞"），或是城市高峰期的交通堵塞。没有任何一只鸟在指挥整个鸟群的飞行模式，也没有任何一辆车在策划整个交通的拥堵。这些宏大而有序（或混乱）的全局现象，是从每一个个体遵循简单局部规则（例如，鸟与邻近几只鸟保持距离和方向一致）中"涌现"出来的。这就是"涌现性全局行为"。科学家的终极目标，就是找到描述这些局部规则的数学语言，即"网络动力学"。

拥有了这些动力学方程，就如同掌握了系统的"源代码"。我们可以借此分析一个电网在遭受攻击后能否快速恢复（韧性），预测一个生态系统能否维持物种多样性（稳定性），甚至设计策略来精准地控制癌细胞的扩散（可控性）。然而，文章指出了一个关键的"人类瓶颈"：发现这些"源代码"的过程，在过去几乎完全依赖于科学家的"直觉和经验"。这就像是要求牛顿或爱因斯坦级别的人物不断出现，才能在新的领域取得突破。因此，我们的理论知识库更新得非常慢，只局限在生物、生态等少数被几代科学家反复研究过的"经典"领域。对于更多新兴的、缺乏所谓"第一性原理"（即最基本、不证自明的物理定律）的系统，比如金融市场、社交网络舆论传播等，我们基本上处于"有数据，无理论"的尴尬境地。

而现在，转机出现了。深度学习和人工智能的兴起，为我们提供了一条全新的路径。我们不再仅仅依赖人类的智慧火花，而是可以构建一个"数据驱动"的范式。这个新范式的目标是，让机器直接从海量观测数据（比如，社交网络中每个用户的点赞、转发记录和他们的好友关系图）中，自动学习和发现那些隐藏在背后的、控制系统演化的数学规则。这标志着科学研究方法的一次根本性转变：从"人脑思考、数据验证"的传统模式，迈向"机器发现、人脑理解"的全新纪元。

寻找公式：在无限可能中航行

原文翻译

寻找最拟合给定数据的公式的过程被称为符号回归。它通过搜索数学算子（例如，+、× 和 sin）的组合来推导函数关系，以最佳地拟合数据，并已广泛用于发现非网络化的单体或少体系统的控制函数。然而，在复杂的网络化系统中，搜索空间随网络维度n呈超指数增长。随着n的增加，变量数量和公式长度分别按nd和nl的规模增长，其中d是每个节点的特征数，l是相互作用定律的长度。这产生了一个包含所有可能变量排列的搜索空间$O((nd)^{nl})$，远远超出了现有通常为n等于1或一个非常小的常数的非网络化系统设计的方法。除了符号回归，非线性动力学的稀疏辨识（SINDy）也被应用于网络动力学的辨识。尽管SINDy计算效率高，但它依赖于预定义的函数库，这限制了其在没有先验知识的系统中发现前所未知动力学的能力。据我们所知，目前还没有方法可以在复杂网络系统中实现符号回归。

深度解读

这里，文章深入到了问题的技术核心，并为我们呈现了故事中的"大反派"——组合爆炸。首先，我们需要理解什么是"符号回归"（Symbolic Regression, SR）。你可以把它想象成教计算机扮演一位像开普勒或牛顿那样的早期物理学家。我们给计算机一堆数据（比如行星的运行轨迹），然后让它尝试用最简单的数学符号（加、减、乘、除、sin、log等）组合出一个公式，来完美地解释这些数据。最终目标不是一个复杂的黑箱模型，而是一个简洁、优雅、人类可以理解的数学表达式，比如$F=ma$。

对于只有一个或几个物体（比如一个摆钟或太阳系中的几个行星）的简单系统，符号回归是有效的。但当问题变成一个由成百上千个节点（比如一个社交网络中的用户）组成的复杂网络时，情况就急转直下。文章提到的"超指数增长"的搜索空间，其恐怖程度可以用一个比喻来形容：想象一下用乐高积木拼一个特定的模型。如果只有10块积木，你很快就能试完所有组合。如果有100块积木，任务就变得非常困难。而对于一个仅有100个节点的"微型"网络，其可能的动力学公式（也就是"乐高模型"）的数量，会轻易地超过已知宇宙中所有原子的总和！这就是公式$O((nd)^{nl})$所描述的"组合爆炸"的威力，其中n是节点数，d是每个节点的状态数，l是描述相互作用的公式长度。这个天文数字般的搜索空间，是任何传统符号回归方法都无法逾越的"叹息之墙"。

文章还提到了一个"竞争者"——SINDy（非线性动力学的稀疏辨识）。SINDy试图通过一个聪明的捷径来解决这个问题。它的思路是：我们先准备一个包含各种可能函数（如x, x^2, x^3, $\sin(x)$, $\cos(x)$等）的"候选函数库"，然后假设真正的动力学方程只是这个库中少数几项的简单线性组合。这样问题就从"创造公式"变成了"挑选函数"。这大大提高了效率。但它的致命弱点在于，如果真正的物理定律恰好不在你预先准备的那个"函数库"里，SINDy就永远也找不到正确答案。打个比方，SINDy像一个只能照着一本固定食谱做菜的厨师，虽然速度快，但无法创造出任何新菜式。而本文提出的ND²方法，其雄心在于成为一个能从基本食材（数学算子）开始，创造全新菜谱（动力学公式）的"创意大厨"。最后，作者明确指出，在他们之前，还没有人能成功地将这种"创意大厨"式的方法应用于复杂网络，这凸显了他们工作的开创性。

结果

ND² 框架：发现网络动力学的全新工具箱

原文翻译

许多复杂系统，从基因调控网络到流行病传播，都可以被描述为动态过程，其中N个节点的状态，x=[x_1,…,x^N]^T，根据编码在网络结构A_{ij}中的相互作用演化，遵循动力学f=[f_1,…,f^N]^T。一个典型的例子是成对模型：

$$\frac{d}{dt}x_i = f_i(x_i) = W(x_i) + \sum_{j=1}^{N} A_{ij} Q(x_i, x_j) \quad (1)$$

其中，等式右侧的两项分别描述了节点i的自主动力学和与其邻居的相互作用动力学。非线性函数$W(x_i)$和$Q(x_i,x_j)$编码了系统的控制定律，而连接矩阵A_{ij}∈{0,1}（对于加权网络则为A_{ij}∈ℝ）捕捉了节点间的（加权）相互作用。为了从节点活动$x(t)$和网络结构A_{ij}中发现f，我们引入了ND²，一种神经符号回归方法，它通过网络动力学算子减少搜索空间，并通过一个预训练的NDformer实现高效、稳健的发现。

我们设计了三种网络动力学算子：源（ϕ_s）、目标（ϕ_t）和聚合（ρ），其灵感来源于图网络中的消息传递。这些算子不是独立处理每个节点状态x_i，而是作用于整个状态向量x∈$\mathbb{R}^N$。如图1b所示，ϕ_s和ϕ_t通过选择每条边的源节点和目标节点，将节点级变量映射到边级变量y∈$\mathbb{R}^E$（E表示边的数量），而ρ则通过对输入边求和，将边级变量聚合回节点（形式化定义见方法部分）。图1d展示了一个三节点示例，其中节点活动$x_i(t)$被向量化为$x(t)$（左图）。算子ϕ_s和ϕ_t提取每条边的源节点和目标节点变量，然后使用正弦和减法函数进行逐元素转换（中图）。算子ρ将边状态聚合回节点，对于没有输入边的节点2和3，其值为零（右图）。如图1c所示，这些算子使得网络动力学的表达与网络维度n无关，从而将高维网络的搜索空间简化为等效的一维系统。

尽管空间被缩小，但对于现有的符号回归算法来说仍然过于庞大。因此，我们提出了一种NDformer引导的符号回归算法，它结合了神经引导和符号搜索（图1e）。NDformer整合了图神经网络（GNN）和Transformer来编码网络结构和节点活动数据，估计用于公式构建的符号上的概率分布（参见方法部分的"NDformer的设计"）。蒙特卡洛树搜索（MCTS）在这些概率的引导下选择符号来构建候选公式，然后由一个奖励计算器根据其准确性和简洁性对这些公式进行评估。由此产生的奖励进一步指导MCTS生成更好的候选公式，以更短的公式长度实现更高的准确性（参见方法部分的"NDformer引导的符号搜索"）。与经典的MCTS方法相比，ND²利用NDformer捕捉数据中的潜在模式，并引导MCTS优先考虑有前途的符号组合，从而加速搜索过程。与无法捕捉节点相互作用的现有基于表格的方法不同，ND²通过其架构和预训练方案专为网络动力学设计，从而能够高效地指导网络动力学的发现（详细讨论见补充材料第4.5节）。如图1f所示，NDformer在一个包含一百万个样本的大规模通用数据集上进行预训练，该数据集由随机生成的网络动力学公式、合成的网络结构和相应的节点活动数据组成。为确保多样性，我们生成了具有不同大小和度的Erdős-Rényi、Watts-Strogatz、Barabási-Albert和完全图网络。节点活动从高斯混合模型中抽取，以嵌入真实系统中常见的低维吸引子。在预训练期间，NDformer学习在给定网络结构和节点活动作为上下文输入的情况下，从前面的符号预测下一个公式符号（参见方法部分的"预训练NDformer"）。预训练后的NDformer随后赋能符号搜索算法，以在由网络动力学算子表达的简化搜索空间中高效地发现网络动力学公式。

深度解读

这一部分是论文的技术心脏，它详细介绍了ND²这个"自动科学发现机器"是如何被设计和构建出来的。我们可以把它拆解成三个革命性的核心部件来理解。

部件一：维度"收缩射线"（网络动力学算子）

这是本文最核心的理论创新，它解决了前面提到的"组合爆炸"问题。想象一下，你要为100位客人写一份制作蛋糕的食谱。一种笨办法是为每一位客人的那一小块蛋糕都单独写一份制作说明，这会产生一本厚得无法阅读的书。而聪明的办法是只写一条普适的规则："对于任意一块蛋

糕，它的风味取决于它周围相邻蛋糕的风味。" 这篇论文发明的三个算子——ϕ_s（源）、ϕ_t（目标）和ρ（聚合）——就是实现这种"聪明办法"的数学语言。

- **ϕ_s (source) 和 ϕ_t (target):** 这两个算子负责将信息从"节点"传递到"边"上。在图1d的例子中，对于从节点1指向节点2的边，ϕ_s会提取出节点1的状态，而ϕ_t会提取出节点2的状态。这就像是在描述一种相互作用时，明确指出"谁"对"谁"产生了影响。
- **ρ (aggregate):** 这个算子负责将所有作用于某个节点的信息汇总起来。它将"边"上的信息重新聚合回"节点"。比如，节点1同时收到了来自节点2和节点3的影响，ρ算子就会将这两个影响（比如，通过求和）整合起来，计算出节点1受到的总影响。

这套算子的天才之处在于，无论网络有多大（100个节点还是100万个节点），描述动力学规则的公式长度是固定的。如图1c所示，一个原本随节点数n急剧变长的复杂公式，被压缩成了一个不随n变化的、简短的表达式。这相当于用一把"收缩射线"，将一个看似无限庞大的问题，缩小到了一个可控的、固定大小的问题。

部件二：AI"领航员"（NDformer引导的搜索）

即使问题被缩小了，可能的公式组合仍然是一个巨大的迷宫。如果盲目搜索，依然会耗费海量时间。这时，就需要一个聪明的"领航员"来指引方向，这就是NDformer的作用。整个搜索过程是蒙特卡洛树搜索（MCTS）和NDformer的精妙协作（见图1e）。

- **蒙特卡洛树搜索 (MCTS):** 想象你在玩一个有无数种走法的棋类游戏。MCTS的策略不是去计算每一种走法之后的所有可能性，而是快速地、随机地"试玩"成千上万盘棋局直到终局，然后统计从某个初始走法出发，最终获胜的概率有多大。通过这种方式，它能很快地识别出哪些是"好棋"，哪些是"臭棋"，从而集中算力去探索那些最有希望的路径。
- **NDformer:** 如果说MCTS是一个勤奋的棋手，那么NDformer就是一位经验丰富的"棋圣"，为棋手提供"棋感"和"直觉"。它是一个强大的深度学习模型，其内部又包含两个关键部分：
  - **图神经网络 (GNN):** 专门用来理解网络的"地图"，即节点之间是如何连接的。它能捕捉到网络的拓扑结构信息。
  - **Transformer:** 这是处理序列数据的王者，最初因在自然语言处理（如GPT模型）中的卓越表现而闻名。在这里，它被用来理解节点状态随"时间"变化的序列，就像理解一句话中单词的顺序一样。
- **协同工作:** NDformer接收网络结构和节点活动数据后，会预测出在当前不完整的公式后面，接上哪个数学符号（比如"+"或"sin"）的可能性最大。MCTS则根据NDformer给出的这个"建议"（概率分布），优先去探索那些高可能性的公式组合，极大地提高了搜索效率。

部件三："模拟宇宙"训练场（预训练过程）

NDformer的"棋感"从何而来？答案是"预训练"（见图1f）。在正式解决任何实际问题之前，研究人员让NDformer在一个包含100万个不同"模拟宇宙"的巨大数据集中进行学习。每个"模拟宇宙"都包含一个随机生成的网络结构和一套随机生成的动力学公式（物理定律）。通过学习预测这100万套完全不同的"物理定律"，NDformer掌握的不是任何一条具体的定律，而是创造定律的普适"语法"和"模式"。它学会了什么样的公式结构在数学上是"合理"的，什么样的组合是"有意义"的。这种大规模的预训练，使得NDformer在面对一个全新的、未知系统时，能够凭借其丰富的"经验"做出精准的判断，为MCTS的搜索提供高质量的指引。

为了让这个复杂的架构更清晰，我们可以用一个表格来总结：

组件	类型	在ND²中的角色	类比
网络动力学算子	数学理论	将无限维的搜索空间降至固定的、 可管理的维度。	一把"维度收缩射线"， 将一个巨大的问题变得小巧可控。
MCTS	搜索算法	通过模拟运行来探索所有可能的公式空间。	一位棋手，通过快速下数千盘棋来探索最佳走法。
NDformer	神经网络	指导MCTS， 预测哪个数学符号最有希望构成正确公式。	一位国际象棋大师，为棋手提供"直觉"和"棋感"。
GNN (NDformer内部)	神经网络层	编码网络的连接结构（图）。	大师大脑中负责理解城市地图的部分。
Transformer (NDformer内部)	神经网络层	编码节点活动的时间序列数据。	大师大脑中负责理解城市交通流量随时间变化的部分。

总之，ND²框架通过"算子降维"、"AI领航"和"模拟训练"这三步，构建了一个前所未有的强大工具，旨在自动化地从数据中发现控制复杂世界的根本规律。

图1 | ND² 框架

图1a

给定网络结构和节点活动，ND²方法旨在发现目标公式。此处的示例是在一个经验网络（北欧电网网络，包含236个节点和320条无向边）上的Kuramoto动力学。

图1b-d

网络动力学算子的演示，包括所提出的三个算子的定义（b），这些算子减少Kuramoto动力学搜索空间的示例（c），以及一个三节点网络示例，用以说明信息从节点到边，最终聚合回节点的路径（d）。

图1e-f

NDformer引导的符号搜索算法和NDformer的架构（e），以及NDformer的预训练过程（f）。

图表描述

- **图1a (ND²的目标):** 左侧是网络结构图，展示了节点和边。中间是节点活动的时间序列图，横轴是时间，纵轴是节点的相位，不同颜色的曲线代表不同节点。右侧是ND²的目标：从左侧和中间的数据中，自动发现描述该系统演化的数学公式，例如Kuramoto动力学公式：$\frac{d}{dt}x_i = \omega_i + \sum_{j=1}^{n} A_{ij} sin(x_j - x_i)$。
- **图1b (网络动力学算子):** 定义了三个核心算子：
  - $\phi_s(x) \in \mathbb{R}^E$: 选择源节点（节点 → 边）。
  - $\phi_t(x) \in \mathbb{R}^E$: 选择目标节点（节点 → 边）。
  - $\rho(x) \in \mathbb{R}^N$: 聚合（边 → 节点）。
- **图1c (搜索空间缩减):** 这是一个关键的对比图，展示了算子的威力。
  - **目标公式:** Kuramoto动力学中的相互作用项 $\sum_{j=1}^{n} A_{ij} sin(x_j - x_i)$。
  - **缩减前:** 如果直接对高维网络进行搜索，公式中涉及的变量有$3n+3$个，公式长度为$6n+1$，搜索空间复杂度为$O((3n+3)^{6n+1})$，这会随着节点数n超指数增长。
  - **缩减后:** 使用算子，可以将公式表达为 $\rho(sin(\phi_s(x) - \phi_t(x)))$。此时，涉及的变量（符号集）大小为8，公式长度为9，搜索空间复杂度为$O(8^9)$，完全与节点数n无关。
- **图1d (三节点示例):** 直观地展示了算子的工作流程。
  - **左侧:** 一个三节点的有向图，节点1有两条出边，分别指向节点2和3。每个节点有其状态x_i。
  - **中间:** 算子ϕ_s和ϕ_t作用于节点状态向量x。对于从1到2的边，$\phi_s(x)$取出x_1，$\phi_t(x)$取出x_2。然后对这些边上的值进行数学运算，如$sin(\phi_s(x) - \phi_t(x))$，得到每条边上的一个新状态。
  - **右侧:** 算子ρ将边上的状态聚合回目标节点。由于只有节点1是目标节点（有入边），所以它得到了聚合后的值。节点2和3没有入边，所以聚合值为0。
- **图1e (NDformer引导的符号搜索流程图):**
  - 输入是"系统观测"（结构和活动数据）。
  - 数据首先进入NDformer（包含GNN和Transformer编码器）。
  - NDformer输出一个"策略"（Policy），即下一个最可能的符号。
  - 这个策略指导"蒙特卡洛树搜索"（MCTS）模块。
  - MCTS生成一个"不完整的公式"，一方面反馈给NDformer继续预测，另一方面送入"奖励计算器"。
  - 奖励计算器根据公式的拟合优度和简洁度给出一个"奖励"（Reward），这个奖励也用来指导MCTS。
  - 经过多次迭代，最终在"帕累托前沿"上选出最优的"已发现公式"。
- **图1f (NDformer预训练流程图):**
  - 输入是随机生成的"结构"和"活动"数据，以及一个"不完整的公式"。
  - 结构和活动数据通过GNN和Transformer编码器进行嵌入。
  - 不完整的公式通过嵌入器（Embedder）处理。
  - 这些信息被送入一个Transformer解码器，其任务是预测公式中的"下一个符号"。
  - 通过对比预测的符号和真实的符号，计算损失并训练NDformer。

在模拟系统中高效恢复基准网络动力学

原文翻译

为了验证ND²从数据中发现精确控制定律的能力，我们模拟了十个代表性的模型系统，涵盖了众所周知的网络动力学，包括Kuramoto、耦合

Rössler振荡器、同质耦合Rössler、FitzHugh-Nagumo、Wilson-Cowan、基因调控、Michaelis-Menten、Lotka-Volterra、互惠种群和易感-感染-易感模型（详见补充材料第4.2节）。
图2a展示了在Erdős-Rényi网络上用耦合Rössler振荡器动力学模拟的节点活动，其中每个节点有三个状态变量（x_i, y_i, z_i），以随机条件和固有频率ω_i进行初始化，并在邻居影响下演化。ND²准确地恢复了所有三个状态的控制公式，捕捉了函数形式和参数，并再现了导数和长期轨迹（图2b-d）。除了这个例子，ND²还在另外九个系统上，无论是在合成网络还是经验网络上，都识别出了正确的动力学（补充材料第4.4节）。

我们将ND²与两种最先进的方法进行了比较：物理嵌入图网络（PIGN）和两阶段推断（实验设置见补充材料第4.3节）。如图2e所示，ND²在恢复公式和预测导数方面均优于这两种方法。基于非线性动力学稀疏辨识（SINDy）的两阶段方法，仅限于在预定义的库中拟合动力学。一个较小的库可能缺乏表达能力，而一个较大的库则需要足够的数据以避免欠定问题，当先验知识不足时，通常无法捕捉到正确的公式。相比之下，ND²使用符号回归来发现任意形式的公式，从而能够准确恢复网络动力学。我们在补充材料第4.4节中提供了对实验结果的进一步讨论以及更多的基线方法。

我们进一步在不利条件下使用相同设置测试了ND²，发现它可以在观测噪声信噪比低至-10到约-5 dB、动力学噪声信噪比为25到约35 dB，以及高达45-50%的缺失或虚假边的情况下恢复控制定律（补充材料第6.2节）。值得注意的是，在不依赖先验知识的情况下，ND²的性能与拥有强大先验知识的两阶段方法相当。ND²在更具挑战性的场景中也取得了成功。当网络结构未知时，仅使用节点活动作为输入，它通过将边的存在视为可优化参数，恢复了Kuramoto动力学和底层的236个节点的网络（参见方法部分的"发现超越成对模型的动力学"）。它还能处理边权重未知、不同节点类别遵循不同动力学的异质网络，以及在NDformer预训练期间未见过的社区结构网络，展示了强大的泛化能力（补充材料第6.5节）。

为了评估NDformer的加速能力，我们实现了一个没有NDformer的ND²版本，发现该搜索很少能在合理时间内恢复目标公式（图2f）。NDformer将符号回归加速了三个数量级（图2g），这得益于其预训练过程，在该过程中，它从一百万个不同动力学系统的网络结构和节点活动$x(t)$中学习预测公式符号（参见方法部分的"预训练NDformer"）。在预训练期间，节点状态是独立采样的，$x_i(t_n) \sim_{i.i.d.} p_\theta(x)$（其中$\sim$表示"服从分布"），避免了迭代$x_i(t_n) = x_i(t_{n-1}) + f_i(x(t_{n-1}); A)\Delta t$（方法）带来的数值不稳定性。为了捕捉动力学系统中典型的低维流形，如不动点、极限环和其他吸引子，$p_\theta(x)$被建模为高斯混合模型。与均匀分布相比，这种设计在恢复具有一维极限环状态空间的FitzHugh-Nagumo动力学时，实现了58倍的加速（补充材料图10）。

深度解读
在上一节中，我们了解了ND²这个强大工具的内部构造。现在，这一节将通过一系列严格的"能力测试"来证明它的价值。在科学研究中，一个新方法在声称能发现未知事物之前，必须首先证明它能准确地找到我们已知的事物。这就像在派一名侦探去破悬案之前，先让他解决几个我们已经知道答案的模拟案件，以检验他的基本功是否扎实。

第一关：模拟宇宙的"期末考试"

研究人员构建了十个不同的"模拟宇宙"，每个宇宙都遵循一套已知的、经典的物理或生物学定律（如描述振荡的Kuramoto模型、描述神经元放电的FitzHugh-Nagumo模型、描述种群竞争的Lotka-Volterra模型等）。他们让这些系统运行，生成数据，然后把这些数据喂给ND²，看它能否"逆向工程"出最初设定的那些定律。

- **耦合Rössler振荡器案例（图2a-d）：** 这是一个具体的成功案例展示。耦合Rössler系统描述了一组相互连接的振荡器，行为复杂，有点像混乱的钟摆。ND²不仅成功地找回了描述每个振荡器状态（x, y, z三个维度）演化的那三个核心数学公式，而且连公式中的具体参数（比如0.5, 0.165, 5.5这些数字）都分毫不差地复原了。图2b展示了ND²恢复的公式所预测的"瞬时变化率"（导数）与真实值的对比，三条完美的$R^2 = 1.000$的对角线意味着预测与真实完全一致。更令人信服的是图2c，它展示了用ND²找到的公式来模拟系统未来的长期行为，生成的轨迹与真实的轨迹几乎一模一样。这证明ND²不仅"知其然"（拟合了数据点），更"知其所以然"（抓住了系统演化的本质）。
- **横向大比拼（图2e）：** 这张图是ND²的"战绩榜"。它将ND²与其它两种顶尖方法（PIGN和Two-phase）进行了正面交锋。结果一目了然：在所有十个测试中，ND²全部成功（每个柱子上方的括号里显示了成功次数/总次数，如(3/3)），并且准确率达到了完美的1.0。而竞争对手们则在多个系统中"翻车"，成功率和准确率都远低于ND²。这清晰地表明，ND²不是众多工具中的一个，而是在这个特定任务上具有压倒性优势的"冠军"方法。

第二关：压力测试与极限挑战

科学不仅要在理想条件下工作，更要能在现实世界的"脏数据"中生存。研究人员对ND²进行了"压力测试"，向数据中添加了大量的噪声、随机删除或添加网络连接。结果显示，即使在数据质量很差的情况下，ND²依然能稳健地找出正确的规律。更具挑战性的是，当研究人员假装"不知道网络结构"（这在很多真实生物学问题中是常态）时，ND²仅凭节点的活动数据，就同时推断出了网络连接和其背后的动力学规律，展现了惊人的"破案"能力。

第三关：验证核心引擎的威力

ND²的强大，很大程度上归功于其内部的AI"领航员"——NDformer。为了证明这一点，研究人员做了一个"控制变量实验"：他们移除了NDformer，让剩下的MCTS算法"盲目"搜索。

- **速度与激情的对比（图2f, 2g）：** 结果是戏剧性的。图2f中的蓝色方块（带NDformer）几乎都快速且成功地到达了准确率1.0的顶端。而橙色圆圈（不带NDformer）则在巨大的搜索空间中苦苦挣扎，要么耗费极长时间，要么在规定时间内彻底失败。图2g的柱状图更直观地显示，在三个代表性任务上，NDformer带来了三个数量级（约1000倍）的速度提升。这无可辩驳地证明了，NDformer的智能引导是ND²成功的关键，它将一个几乎不可能完成的搜索任务，变成了一个高效、可行的过程。
- **训练数据的"点睛之笔"：** 这里揭示了一个非常精妙的细节。在预训练NDformer时，研究人员没有使用在状态空间中"均匀分布"的随机数据，而是使用了"高斯混合模型"生成的数据。这背后是对真实物理世界的深刻洞察。真实世界中的系统，比如钟摆，并不会随机出现在任何可能的位置和速度上，它们倾向于被"吸引"到某些稳定的状态（如静止不动）或稳定的运动模式（如极限环振荡）。高斯混合模型生成的数据能够模拟这种"聚集"在吸引子周围的特性。用这种更"真实"的数据训练出来的NDformer，自然也更擅长理解和发现真实动力学系统的规律。这个看似微小的技术选择，实际上是连接人工智能与物理学思想的桥梁，也是ND²性能卓越的秘诀之一。

总而言之，这一系列的模拟实验，从多个维度、多个层次上，系统性地验证了ND²方法的准确性、鲁棒性、高效性及其核心组件的不可或替代性，为其在后续章节中挑战真实世界的未知问题奠定了坚实的信誉基础。

图2 | 合成实验

图2a-d
在合成的Erdős-Rényi网络上模拟的耦合Rössler动力学的恢复结果（a），包括恢复公式预测的时间导数（b）和生成的长期活动（c）以及恢复的公式（d）。

图2e
我们提出的ND²方法、移除了NDformer的修改版ND²、两阶段推断和PIGN在五类动力学上的实验结果，包括Kuramoto (KUR)、耦合Rössler (CR) 振荡器、同质耦合Rössler (HCR)、FitzHugh-Nagumo (FHN)、Wilson-Cowan (WC)、基因调控 (GR)、Michaelis-Menten (MM)、Lotka-Volterra (LV)、互惠种群 (MP) 和易感-感染-易感 (SIS)。柱子上方的数字表示成功恢复的公式数量（即找到正确或数学上等价的公式形式）与总数的比值，而柱子的高度代表结果的平均R²。我们的方法成功发现了所有十个模型，R²值为1，表明其恢复网络动力学公式的能力。

图2f
使用带和不带NDformer的ND²在十个模拟系统上恢复动力学的合成实验。图中描绘了搜索时间和发现公式的R²，标记的形状表示是否使用NDformer，大小表示公式长度。ND²成功恢复了所有十个动力学。灰色线条显示了Michaelis-Menten (MM) 动力学的搜索过程，其中NDformer引导的搜索耗时减少了99.91%。

图2g
使用带和不带NDformer恢复三个模型的搜索时间，其中NDformer的引导将搜索速度加快了三个数量级。

图表描述

- **图2a (模拟数据):** 左图展示了耦合Rössler系统在一段时间内的节点活动轨迹，不同颜色的曲线代表不同节点的某个状态变量（如x, y, z）。右图是承载这些动力学的合成网络结构（Erdős-Rényi图）。
- **图2b (导数预测):** 三个散点图，分别对应状态变量x, y, z的导数（$dx/dt, dy/dt, dz/dt$）。横轴是真实的导数值，纵轴是ND²恢复的公式计算出的导数值。所有的点都完美地落在对角线上，决定系数R^2均为1.000，表示预测与真实完全一致。
- **图2c (长期轨迹):** 展示了使用ND²恢复的公式从初始状态开始，模拟生成的长期动态轨迹（彩色曲线），并与真实的轨迹（灰色曲线）进行对比。两者高度重合，说明恢复的公式准确地捕捉了系统的长期演化行为。
- **图2d (公式恢复结果):**
  - **目标公式:** 列出了耦合Rössler系统的三个真实微分方程。
  - **搜索结果:** 展示了ND²使用其内部算子语言找到的公式形式。
  - **重写为节点形式:** 将搜索结果翻译回我们熟悉的、针对单个节点的标准数学形式。
  - **获得的公式:** 最终得到的、与目标公式在形式和参数上都完全一致的结果。
- **图2e (性能对比条形图):**
  - 横轴是五大类共十种动力学模型。
  - 纵轴是平均R^2值，衡量预测准确度。
  - 四种不同颜色的条形代表四种方法：ND²（蓝色）、ND² without NDformer（橙色）、Two-phase（绿色）、PIGN（红色）。

    - ND²的蓝色条柱在所有模型上都达到了1.0的高度，且成功率均为100%（如(3/3)）。其他方法的条柱则高低不一，且成功率远非100%。
- **图2f (搜索效率对比散点图):**
    - 横轴是搜索时间（秒，对数坐标），纵轴是发现公式的R^2值。
    - 蓝色方块代表使用ND²（带NDformer）的结果，橙色圆圈代表不带NDformer的结果。
    - 标记的大小代表公式的长度。
    - 可以清晰地看到，蓝色方块普遍集中在左上角区域（耗时短、精度高），而橙色圆圈则散布在右侧，甚至很多任务失败（未显示）。灰色箭头线生动地展示了对于MM模型，使用NDformer后搜索时间的大幅缩短。
- **图2g (加速效果对比条形图):**
    - 横轴是三个不同的动力学模型。
    - 纵轴是搜索时间（秒，对数坐标）。
    - 蓝色条柱（ND²）远低于橙色条柱（ND² without NDformer），直观地显示了NDformer带来的巨大速度提升（超过1000倍）。

修正经验系统中的现有网络动力学以揭示科学见解

基因调控网络：从独立影响到集体决策

原文翻译

适当的细胞功能依赖于及时、特定情境的基因表达。基因通过其表达来调节细胞过程，而其表达又被这些过程所塑造，形成一个复杂的自调节系统。理解基因如何相互促进或抑制对方的表达是分子生物学的核心挑战。与具有直接相互作用的典型网络不同，基因通过环境介导的mRNA表达间接影响其它基因（图3a）。传统上，基因动力学使用希尔方程建模：

$$\frac{d}{dt}x_i(t) = s_i - \gamma_i x_i(t) + \sum_j A_{ij} S(x_j(t))$$

其中s_i和γ_i是基础合成和降解率，A_{ij}表示成对相互作用强度，$S(x) = x^2/(x^2 + \theta_j^2)$是一个有界函数（$|S(x)| \leq 1$），参数为$\theta_j$。这个数学模型最初源于双组分系统，并通过加性假设扩展到多基因网络，因此无法准确捕捉基因调控的真实动力学。尽管观测数据日益增多，但更准确的模型仍然缺乏，这阻碍了依赖于动态模型的关键问题研究，如网络推断和稳定性分析。
在这里，我们应用ND²于经验基因表达数据以解决这一局限。我们使用酵母细胞分裂周期数据，基因表达水平在119分钟内每7分钟测量一次（约两个细胞周期；详见补充材料第5.1.1节）。基因按生物学功能分组，每组内活动同步，我们将其视为不同的组件（图3b）。由于底层网络结构不可观测，我们假设一个全连接网络，并让我们的方法拟合连接权重A_{ij}。ND²发现了一个修正后的公式：

$$\frac{d}{dt}x_i(t) = s_i - \gamma_i x_i(t) + \bar{S}(\sum_{j=1}^{N} A_{ij} x_j(t)) \quad (2)$$

其中s_i, γ_i, β和A_{ij}是参数，$\bar{S}(x) = \beta(1 + exp(-x))^{-1}$是逻辑斯蒂函数（图3c；详见补充材料第5.1.2节）。与$S(x)$类似，$\bar{S}(x)$也是有界的，但它作用于邻居的总和，而不是单个邻居（图3d）。因此，$\frac{\partial \dot{x}i}{\partial x_j} = \beta A i j \bar{S}(\sum_k A_{ik}x_k)(1 - \bar{S}(\sum_k A_{ik}x_k))$依赖于i和j之外的节点，反映了一种高阶效应，即一个基因对另一个基因的影响可以被其他基因调节。这与在许多具有环境介导相互作用的自然系统和其他高级基因调控模型中广泛观察到的高阶相互作用相符（补充材料第5.1.4节）。如图3e, f所示，修正后的公式准确预测了基因表达水平，并再现了真实数据中观察到的振荡动力学。它将均方根误差（RMSE；参见方法部分的"从经验数据中发现动力学"）降低了59.98%，对称平均绝对百分比误差（sMAPE）降低了50.82%，并将贝叶斯信息准则（BIC）从908降至310，表明其不仅提高了准确性，而且具有更好的函数形式，而不仅仅是增加了模型复杂性。它还优于其他高级基因调控模型和GNN（补充材料第5.1.4节），并展示了在不同基因数量的基因网络间的泛化能力（补充材料第5.1.5节）。

深度解读

在证明了自己能完美解决"已知问题"后，ND²现在开始挑战"未知领域"，它的第一个目标就是修正我们教科书中关于生命如何运作的核心模型。这一节将ND²从一个强大的验证工具，提升为了一个能够产生全新科学见解的发现引擎。

案例研究一：酵母基因调控的"悄然革命"

- **背景知识：** 细胞是我们身体的基本单位，其内部运作像一个高度精密的工厂。基因就是这个工厂的"生产指令"，它通过"表达"（制造蛋白质）来控制细胞的各种活动（如图3a, 3b）。基因之间会相互影响，有的促进（激活）别的基因表达，有的则抑制。理解这个复杂的调控网络是现代生物学的圣杯之一。

- **传统模型的缺陷：** 长期以来，科学家使用一种叫做"希尔方程"的数学模型来描述这个过程。这个模型的**核心假设**是"线性叠加"或"独立影响"。如图3d左侧所示，它认为一个基因（比如基因i）受到的总影响，等于所有其他基因（基因j）对它独立影响的简单加总（$\sum A_{ij}S(x_j)$）。打个比方，这就像认为一个房间里的总音量，等于每个单独说话的人的音量之和。这在很多情况下是合理的，但可能过于简化了。
- **ND²的颠覆性发现：** ND²在分析了真实的酵母细胞周期数据后，提出了一个看似微小但意义深远的修改。在它发现的新公式（方程2）中，求和符号（$\sum$）被移到了非线性函数$\bar{S}$的**内部**，变成了$\bar{S}(\sum A_{ij}x_j)$（见图3d右侧）。
- **全新的科学洞见——"高阶相互作用"：** 这个数学上的位置调换，彻底改变了我们对基因调控的理解。它意味着，基因j对基因i的影响，**不再是独立发生的**，而是取决于**所有其他基因k的当前状态**。这不再是简单的音量叠加，而更像一场复杂的"圆桌讨论"。你在会议上对某个人发言的反应，不仅取决于那个人的观点，还取决于整个会议的氛围、之前其他人的发言以及整体议题的走向。一个基因对另一个基因的"激活"或"抑制"作用，会受到"旁观者"基因的调节。这就是所谓的"高阶相互作用"，它揭示了细胞内部决策过程的集体性和情境依赖性，远比我们之前想象的要复杂和智能。
- **证据的力量：** 这种新理论不是空想。图3e展示了ND²发现的模型（彩色曲线）与真实数据（灰色圆点）的惊人拟合度。图3f中的数据更具说服力：与传统模型相比，新模型的预测误差（RMSE和sMAPE）降低了超过50%，而BIC（一个衡量模型好坏的指标，越低越好，因为它同时惩罚了不准确和不简洁）从908骤降到310。这强有力地证明了ND²发现的模型不仅更准确，而且在数学形式上更优越、更接近真相，而不是通过堆砌参数来强行拟合。这不仅仅是一次模型更新，而是一次对生命系统运作原理的认知升级。

图3 | 基因调控动力学

图3a
在酵母细胞的分裂周期中，基因之间通过环境介导的调控可以被描述为一个全连接网络。

图3b
基因表达水平（通过表达的mRNA量的对数来衡量）作为节点活动。

图3c, d
现有公式与修正后公式（c）在聚合与非线性算子作用次序上的比较（d）。

图3e, f
生成的节点活动与真实值拟合得很好（e），使RMSE降低了59.68%（应为59.98%），sMAPE降低了50.82%。f中的箱线图显示了中位数（中心线）、25-75百分位数（箱体）、最小值/最大值（须线）和平均值（黄色菱形），并覆盖了个体数据点（n=7）。

图表描述

- **图3a (系统示意图):** 展示了酵母细胞内，基因通过表达mRNA来间接相互影响，形成一个复杂的调控网络。
- **图3b (数据图):** 纵轴是基因表达水平（对数值），横轴是时间（分钟）。七条不同颜色的曲线代表七个功能基因组的平均表达水平随时间的变化，呈现出周期性振荡。
- **图3c (公式对比):**
  - **现有公式:** 列出了传统的希尔方程模型。
  - **ND²发现的公式:** 列出了ND²找到的新模型，其核心区别在于非线性函数$\bar{S}$作用于所有邻居影响的总和。
- **图3d (作用机制对比图):**
  - **左侧 (受约束的成对影响):** 对应现有公式。每个邻居x_j先经过非线性函数S处理，然后这些处理后的结果再加权求和。
  - **右侧 (受约束的整体影响):** 对应ND²发现的公式。所有邻居x_j先进行加权求和，得到一个总的输入信号ξ，然后这个总信号再经过一个非线性函数$\bar{S}$处理。
- **图3e (拟合结果对比图):** 七个小图分别对应七个基因组。每个图中，灰色圆点是真实的实验数据，彩色曲线是ND²发现的模型生成的预测轨迹。曲线与数据点高度吻合。
- **图3f (误差对比统计图):**
  - **左侧 (RMSE):** 均方根误差的箱线图对比。ND²（右侧）的误差中位数和分布范围远低于现有模型（左侧），显示误差降低了59.98%。
  - **右侧 (sMAPE):** 对称平均绝对百分比误差的箱线图对比。ND²同样表现出显著优势，误差降低了50.82%。
  - **BIC值:** 在图下方标注了两个模型的贝叶斯信息准则值，ND²的BIC值（310）远低于现有模型的（908），说明模型在准确性和简洁性上取得了更好的平衡。

微生物群落：揭示生态系统的稳定性法则

原文翻译

微生物群落是另一个重要的自然系统，其中物种的种群增长通常由Lotka-Volterra方程建模：

$$\frac{d}{dt}x_i(t) = (r_i - s_i x_i(t))x_i(t) + \sum_{j=1}^{n} A_{ij} Q(x_j(t), x_i(t))$$

其中r_i和s_i是生长和自我调节率，A_{ij}是物种间的相互作用，$Q(x, y) = xy$是相互作用项。我们分析了n=6个细菌物种在10天内生长的经验数据（图4a,b），使用一个具有可优化边权重的全连接图（补充材料第5.2.1节）。应用ND²，我们得到了一个修正后的公式：

$$\frac{d}{dt}x_i(t) = (r_i - s_i x_i(t))x_i(t) + \sum_{j=1}^{n} A_{ij} \bar{Q}(x_j(t), x_i(t)) \quad (3)$$

其中r_i, s_i和A_{ij}是参数，$\bar{Q}(x, y) = (1 + x^{-y})^{-1}$是相互作用项（图4c；详见补充材料第5.2.2节），它会随着x饱和（图4d），这与Lotka-Volterra的现有扩展一致（补充材料第5.2.4节）。然而，$\bar{Q}(x_j, x_i)$随着x_i的增加而减少，揭示了在敏感性上的一个关键差异。我们考虑$T_{ij} \triangleq A_{ij} Q(x_j, x_i)$，它衡量物种j如何影响i的生长。如图4e所示，在Lotka-Volterra模型中，$\frac{\partial}{\partial x_i}|T_{ij}| \geq 0$，意味着物种随着其种群增长而变得更加敏感。然而，在我们修正的动力学中，$\frac{\partial}{\partial x_i}|T_{ij}| \leq 0$，表明更大的种群受其他物种的影响更小（详见补充材料第5.2.3节）。这对Lotka-Volterra模型背后的假设提出了挑战。在图4f,g中，我们比较了模型准确性，发现我们的公式将RMSE降低了55.94%，sMAPE降低了56.72%，BIC从-41降至-201。它还优于其他高级生态模型和GNN（补充材料第5.2.4节），并在更大群落（12、24和48个物种）和不同营养浓度下的测试证实了其泛化性（补充材料第5.2.5节）。与基因网络的结果一起，这些发现表明ND²可以修正现有的网络动力学公式，不仅实现更高的准确性，而且提供超越当前理解的科学见解。

深度解读

继修正了微观的基因调控模型后，ND²将目光投向了介观的生态系统，再次挑战了一个百年经典理论——Lotka-Volterra方程，即著名的"捕食者-被捕食者"模型。

- **背景知识：** 微生物群落（比如我们肠道里的菌群）是一个复杂的生态系统，不同物种之间存在竞争、合作或捕食关系。Lotka-Volterra方程是描述这种动态关系的基础模型，它被广泛应用于生态学研究中（如图4a, 4b）。
- **传统模型的隐含假设：** 经典的Lotka-Volterra模型中，一个物种j对另一个物种i的影响（T_{ij}）与物种i自身的种群数量x_i成正比。这意味着，一个物种的种群规模越大，它对外界的干扰就**越敏感**。如图4e中的"Existing"线所示，随着x_i增加，受到的影响强度$|T_{ij}|$也随之增加或保持不变。这在直觉上似乎有些反常。
- **ND²的深刻洞察：** ND²通过分析真实的细菌群落生长数据，发现了一个全新的相互作用项$\bar{Q}$（见方程3和图4c, 4d）。这个新项揭示了一个与传统模型**完全相反**的规律。
- **全新的科学法则——"种群规模反相关的敏感性"：** 在ND²发现的新动力学中，一个物种的种群规模越大，它受到其他物种的影响就**越小**（见图4e中的"ND²"线，随着x_i增加，$|T_{ij}|$显著下降）。这个发现非常符合直觉和现实：一个拥有百万个体的庞大、健康的种群，其稳定性远高于一个只有几十个个体的脆弱、濒危种群。前者对外部环境的微小扰动具有更强的"抵抗力"或"缓冲能力"。ND²首次从数据中提炼出了这个关于生态系统稳定性的基本原则，即**"强者恒强，大而不倒"**的内在机制，这是经典Lotka-Volterra模型所忽略的关键因素。
- **压倒性的证据：** 与基因调控的案例类似，这一新发现同样有坚实的数据支持。图4f展示了新模型（右）与旧模型（左）在拟合真实数据上的巨大差异。图4g的统计数据显示，新模型的预测误差（RMSE和sMAPE）再次降低了超过55%，并且BIC值从-41大幅改善至-201。这表明ND²不仅是发现了一个更准确的模型，更是揭示了一个更深刻、更符合自然规律的生态学原理。

这两个案例共同证明，ND²不仅仅是一个"数据拟合"工具，它是一个真正的"科学洞察"引擎。它能够深入到复杂系统的内部，挖掘出被传统模型和人类直觉所忽视的核心机制，从而修正甚至颠覆我们对世界的既有认知。这体现了由机器驱动的科学发现的巨大潜力。

图4 | 微生物群落动力学

图4a, b

六种细菌物种共同生长10天，形成一个具有全连接网络结构的动态系统（a）。每天测量每种物种的生物量（光密度OD）作为节点活动（b）。

图4c-e

现有公式与修正后公式（c）的比较。修正后的结果引入了一个不同的相互作用项（d；等值面颜色表示Q值），其中物种j对i的影响强度随着i的种群增长而减小（e）。e中的中心线和须线显示了10天内30条边的平均值和最小值/最大值（n=300）。

图4f, g
由修正公式生成的节点活动比现有公式更好地拟合了经验数据（f），使RMSE降低了55.94%，sMAPE降低了56.72%（g）。g中的箱线图显示了中位数（中心线）、25-75百分位数（箱体）、最小值/最大值（须线）和平均值（黄色菱形），并覆盖了个体数据点（n=6）。

图表描述

- **图4a (实验示意图):** 展示了在培养皿中多种细菌共同生长的实验场景。
- **图4b (数据图):** 纵轴是生物量（OD值），横轴是时间（天）。六条不同颜色的曲线代表六种细菌的种群数量随时间的变化。
- **图4c (公式对比):**
  - **现有公式:** 列出了经典的Lotka-Volterra模型。
  - **ND²发现的公式:** 列出了ND²找到的新模型，其核心区别在于相互作用项从$Q(a,b)=ab$变成了$\bar{Q}(a,b)=(1+a^{-b})^{-1}$。
- **图4d (相互作用项可视化):** 一个三维曲面图，展示了新相互作用项$\bar{Q}(x_j,x_i)$的值如何随x_j和x_i的变化而变化。颜色代表$\bar{Q}$值的大小。
- **图4e (敏感性对比图):**
  - 横轴是物种i的种群数量x_i。
  - 纵轴是物种j对i的影响强度$|T_{ij}|$的偏导数，即敏感性。
  - 橙色线（Existing）代表现有模型，其值大于等于0，表示敏感性随种群增大而增强或不变。
  - 蓝色线（ND²）代表新发现的模型，其值小于等于0，表示敏感性随种群增大而减弱。
- **图4f (拟合结果对比图):**
  - **左侧两图:** 真实数据（灰色）与现有模型预测（彩色）的对比，可以看出拟合效果不佳。
  - **右侧两图:** 真实数据（灰色）与ND²发现的模型预测（彩色）的对比，拟合效果非常好。
- **图4g (误差对比统计图):**
  - **左侧 (RMSE):** 均方根误差的箱线图对比。ND²（右侧）的误差远低于现有模型（左侧），降低了55.94%。
  - **右侧 (sMAPE):** 对称平均绝对百分比误差的箱线图对比。ND²同样表现出显著优势，降低了56.72%。
  - **BIC值:** 在图下方标注了两个模型的BIC值，ND²的BIC值（-201）远低于现有模型的（-41）。

发现具有可解释性的流行病传播动力学

原文翻译
世界不同地区蔓延的流行病对社会产生了深远影响，扰乱了经济，阻碍了教育，并夺走了无数生命。理解其通过人类移动的传播至关重要，但由于系统复杂性和未知的底层原理，精确的动力学公式仍然难以捉摸。流行病动力学在不同尺度上有所不同：城市级别的传播由日常互动驱动，而全球传播则依赖于政策、社会经济和环境因素。非药物干预措施的差异进一步使传播复杂化，使得传统的元种群模型不足以应对。在这里，我们应用ND²来发现全球多个空间尺度的传播动力学。
如图5a所示，我们研究了2019冠状病毒病（COVID-19）在四个空间尺度（城市、州、国家和全球）的七个区域的传播，其中163周的每周确诊病例$x_i(t)$作为节点活动，交通流量M_{ij}定义了网络结构（补充材料第5.3.1节）。ND²发现了每个区域的传播动力学，尽管存在区域差异，但它们共享一个通用结构，如表5b所列，即一个自演化项$W(x_i)$，一个相互作用项$Q(x_j,x_i)$，以及一个作用于总相互作用$S_i=\sum_{j=1}^{n}A_{ij}Q(x_j,x_i)$的函数$\Phi$：

$$x_i(t+1)=W(x_i(t))+\Phi(\sum_{j=1}^{n}A_{ij}Q(x_j(t),x_i(t))) \quad (4)$$

其中$A_{ij}=\mathbb{I}_{(M_{ij}>0)}$（详见补充材料第5.3.2节）。为了验证发现的动力学，图5c比较了真实和预测的周病例数，显示所有区域的$R^2>0.85$。这种准确性也推广到未见过的时期，通过沿时间维度的训练-测试划分得到证实（补充材料第5.3.7节）。

这些公式的物理意义可以逐项解释。例如，在纽约市区域，发现的公式$x_i(t+1)=x_i(t)\times(a_0+a_1x_i(t)-a_2\bar{x}_i(t)+a_3\frac{\sum_{j=1}^{n}A_{ij}x_j(t)}{D_i^{in}\bar{x}_i(t)+\epsilon})$将$x_i(t+1)$估计为$x_i(t)$乘以一个四项校正因子。前三项使用过去7天和14天的感染情况调整基础校正率a_0，而最后一项描述了输入性感染，并受到分母中入度D_i^{in}和历史14天病例$\bar{x}_i(t)$的调节，这可能反映了在交通繁忙和近期爆发严重的地区采取的封锁或学校关闭等干预措施。完整的解释在补充材料第5.3.2节中提供，显示了与现有元种群模型的一致性，并提供了额外的见解（补充材料第5.3.3节）。发现的动力学还使得能够分析成对节点相关性，这是复杂网络系统中的常用工具。如图5d所示，对节点j的扰动dx_j会传播到节点i，得到$G_{ij}=|\frac{dx_i/x_i}{dx_j/x_j}|$。在所有七个区域中，G都遵循相同的幂律$P(G)\propto G^{-0.80}$（图5e和补充材料第5.3.4节），尽管公式形式有所不同。这一发现揭示了系统对扰动响应在世界不同国家和空间尺度上的共同模式，阐明了流行病传播的传播特性，并支持了所发现动力学的有效性和普适性。

为了展示ND²在帮助理解系统属性方面的价值，我们比较了在美国和中国发现的动力学。两国的$W(x_i)$都是$x_i(t)$乘以一个校正因子，该因子在中国随着$x_i(t)$的增长表现出自我抑制，但在美国则保持不变（图5f）。相互作用动力学$Q(x_j,x_i)$也不同。在美国，它依赖于邻近地区的感

染情况$\langle x_i \rangle := \sum_j M_{ij} x_j$，而在中国，它与$x_j$无关，表明相互作用效应可以忽略不计。这在图5g中得到证实，其中$\langle x_i \rangle$与$\Delta x_i = x_i(t+1) - x_i(t)$之间的皮尔逊相关系数的平方（详见补充材料第5.3.5节）在美国很高（高达0.8），但在中国则低得多（小于0.1）。这些差异可能反映了中国积极的干预措施，包括封锁、接触追踪和旨在限制高感染区传播并最小化跨区域感染的针对性措施。最后，我们使用Gao等人的方法分析稳态特性，该方法推导了在给定平均跨区域交通$\beta_{eff} = \sum_i w_i M_i^{in}$下平均感染$x_{eff}(t) = \sum_i w_i x_i(t)$的动力学，其中$w_i = M_i^{out} / \sum_j M_j^{out}$，$M_i^{in} = \sum_j M_{ij}$，$M_j^{out} = \sum_j M_{ji}$是总流入和总流出（补充材料第5.3.6节）。如图5h所示，中国的相图有一个临界点β_{eff}^c：当$\beta_{eff} < \beta_{eff}^c$时，$x_{eff}$稳定在0，对应于受控的流行病状态。当$\beta_{eff} > \beta_{eff}^c$时，$x_{eff}$发散，表明如果跨省交通超过此阈值，将出现严重爆发。然而，在美国，稳定状态随β_{eff}线性变化并保持有限，因此对交通的干预仅线性影响感染，影响小于中国。这些发现与这两个国家非药物干预措施的不同效果相一致，并展示了ND²揭示系统属性的能力。

深度解读

这是ND²的"毕业大戏"，也是其能力的终极展示。在前两幕中，它分别证明了自己能"恢复已知"和"修正已知"。现在，它将挑战一个完全未知的、极其复杂的真实世界问题：COVID-19大流行的传播动力学。这里没有标准答案，ND²将从零开始，直接从数据中发现科学规律，并提供具有现实意义的深刻洞见。

- **宏大的实验舞台（图5a）：** 研究人员选取了横跨全球的七个不同地区，涵盖了从城市（纽约、芝加哥）到州（纽约州、伊利诺伊州），再到国家（美国、中国）乃至全球的四个空间尺度。数据包括了长达163周的每周确诊病例数（节点活动）和人与人之间的流动数据（网络连接）。这个设置的目的是检验ND²能否在不同文化、政策和地理尺度下，发现普适或特定的传播规律。
- **发现可解释的"病毒传播手册"（图5b和下表）：** ND²为每个地区都找到了一套独特的动力学方程。这些方程并非无法理解的"黑箱"，而是可以逐项解读的"说明书"。例如，纽约市的公式显示，下一周的病例数是本周病例数乘以一个"增长因子"，这个因子由四部分构成：一个基础增长率、一个由本地现有病例驱动的加速项、一个由历史病例（过去14天）驱动的减速项（可能反映了免疫或行为改变），以及一个由邻近地区输入的病例项。这个输入项还被本地的"交通繁忙度"和"历史疫情严重程度"所调节，这清晰地反映了封锁等干预措施的效果。这些发现的公式不仅准确（图5c显示所有地区的预测R^2都超过0.85），而且具有深刻的物理和社會学意义。

表2：发现的流行病传播动力学

尺度	地区	自演化项 W(xi)	相互作用函数 Φ(Si)	相互作用项 Q(xj,xi)	R2
城市	NYC	$x_i(a_0 + a_1 x_i - a_2 \bar{x}_i)$	$\frac{a_3 x_j}{D_i^{in} \bar{x}_i + \epsilon} S_i$	x_j	0.8740
	CHI	$x_i(b_0 + b_1 x_i - b_2 \bar{x}_i)$	$b_3 S_i$	x_j	0.8917
州	NY	$c_0 x_i - c_1 \bar{x}_i$	$c_2 x_i \sigma(S_i)$	$c_3 M_j^{in} - c_4 D_j^{in} \bar{x}_j$	0.8508
	IL	$d_0 x_i - d_1 \bar{x}_i$	$d_2 x_i \sigma(d_3 S_i^{-2})$	$\bar{x}_j$	0.9234
国家	USA	$x_i(\frac{e_0 x_i + e_1}{x_i + e_2} - \frac{e_3 x_i}{x_i + e_4})$	S_i	$M_{ij} x_j$	0.8595
	CN	$x_i(g_0 - g_1 \bar{x}_i - g_2 x_i^2)$	$g_3 x_i^3 S_i$	M_{ij}	0.9187
全球	Global	$-h_0 + h_1 x_i - h_2 \bar{x}_i$	$\frac{h_3}{h_4 + S_i} \sigma(h_5 S_i)$	M_{ij}	0.9363
注：$\bar{x}_i(t) = x_i(t) + x_i(t-1)$，$\sigma(x) = 1/(1+exp(-x))$是sigmoid函数，其他字母为拟合参数。					

- **惊人的普适性发现（图5d, 5e）：** 这是本节最令人震撼的发现。尽管每个地区的具体公式千差万别，但它们共同遵循一个隐藏的、更深层次的普适定律。当研究人员在模型中模拟一次微小的"扰动"（比如在一个城市增加一个病例），然后观察这个扰动在网络中传播的"涟漪效应"大小时，他们发现，无论在哪个国家、哪个尺度，这些"涟漪"的大小分布都精确地遵循同一个数学形式——幂律分布$P(G) \propto G^{-0.80}$。幂律分布是复杂系统（如地震、股市崩盘、物种灭绝）中"临界现象"的标志。这一发现强烈暗示，全球人类流动网络天然地处

在一个极其脆弱的"临界状态"，就像一堆精心堆砌的沙堆，一粒沙子的落下都可能引发一场规模不可预测的连锁崩塌。这为我们理解为何大流行病如此难以预测和控制，提供了一个根本性的物理解释。

- **政策制定的"数学显微镜"——中美对比（图5f, 5g, 5h）：** ND²的价值不仅在于发现抽象理论，更在于为现实决策提供依据。通过对比为美国和中国发现的动力学公式，可以清晰地看到两国抗疫策略差异的数学根源：
  i. **内生增长模式不同（图5f）：** 在中国的公式中，存在一个强大的"自我抑制"项（$-g_2x_i^2$），意味着病例数越多，增长反而越慢。这反映了随着疫情升级，内部管控（如社区封锁、大规模检测）会急剧加强。而在美国的公式中，这种抑制效应不明显，增长更趋于常数。
  ii. **外部输入影响不同（图5g）：** 在美国，本地病例的增长与邻近地区的输入病例高度相关（相关性高达0.8），说明跨区域传播是主要驱动力。而在中国，这种相关性几乎为零（小于0.1），这雄辩地证明了其严格的旅行限制和"熔断"政策在切断传播链上的巨大成功。
  iii. **系统稳定性的根本差异（图5h）：** 这是最深刻的洞察。分析显示，中国的传播系统存在一个"临界点"或"引爆点"（tipping point）。只要跨省流动强度控制在这个临界点以下，疫情就能被彻底扑灭（稳定在0）。一旦超过，就会失控爆发。而美国的系统则没有这样的临界点，控制人口流动只能线性地、有限地减少病例数。这一发现从数学上解释了为什么中国的"清零"策略在理论上是可行的，而同样的策略在美国则难以奏效。

总而言之，ND²在流行病这个案例中，完成了一次从数据到知识、再到智慧的完美飞跃。它不仅发现了可解释的动力学模型，揭示了隐藏的普适性科学定律，还为评估和理解不同国家的公共卫生政策提供了前所未有的定量化、系统化的视角。

图5 | 流行病传播动力学

图5a
数据集包括全球四个空间尺度（城市、州、国家和全球）的七个不同区域，曲线的颜色代表不同区域之间的交通流量。

图5b
在所有七个区域发现的流行病动力学，其中$x_i(t)$表示第i个区域在第t周的确诊病例，$\bar{x}i(t) = x_i(t) + x_i(t-1)$表示历史14天的确诊病例，$Aij$表示网络结构，$D_i^{in} = \sum_j A_{ij}$表示i的入度，$M_{ij}$表示从j到i的人员流动，$M_i^{in} = \sum_j M_{ij}$表示i的总流入量，$\sigma(x) = 1/(1 + exp(-x))$是sigmoid函数，$\epsilon$是一个小的非零常数，$a_i \dots h_i$是参数。

图5c
发现的动力学与真实结果拟合得很好。

图5d, e
扰动实验（d）显示所有发现的动力学都具有相同的幂律分布特性$P(G) \propto G^{-0.80}$（e）。

图5f-h
在自演化动力学（f）、感染动力学（g）和稳态分析（h）方面，对美国和中国发现的COVID-19动力学进行比较。g中的中心线和须线标记了平均值和最小值/最大值（n=15）。

图表描述

- **图5a (研究区域示意图):** 四张地图分别展示了城市、州、国家和全球尺度的人类流动网络。节点是地理区域，连线的颜色深浅代表流动强度。
- **图5b (发现的公式列表):** 一个表格，详细列出了为七个不同区域发现的动力学公式的三个组成部分：自演化项$W(x_i)$、相互作用函数$\Phi(S_i)$和相互作用项$Q(x_j, x_i)$，以及每个模型的R^2值。
- **图5c (拟合结果散点图):** 七个对数坐标下的散点图，每个图对应一个区域。横轴是真实的周确诊病例数，纵轴是模型预测的病例数。点密集地分布在对角线附近，且R^2值均较高，表明拟合效果良好。
- **图5d (扰动传播示意图):** 一个网络图，形象地展示了在一个节点j上施加一个扰动dx_j，会如何通过网络传播，导致另一个节点i产生响应dx_i。
- **图5e (幂律分布图):** 对数-对数坐标图。横轴是扰动响应的大小G，纵轴是其出现的概率P(G)。来自七个不同区域的数据点（用不同颜色和形状标记）惊人地汇聚在一条直线上，表明它们都遵循相同的幂律分布，其斜率约为-0.80。
- **图5f (自演化动力学对比):** 横轴是周确诊病例数$x_i(t)$，纵轴是自演化部分的增长率$W(x_i)/x_i$。蓝线（USA）基本是一条水平线，表示增长率恒定。红线（China）是一条向下倾斜的曲线，表示增长率随病例数增加而下降（自我抑制）。
- **图5g (相互作用动力学对比):** 纵轴是$\Phi(S_i)$和Δx_i之间的R^2值，衡量外部输入对病例增长的贡献度。蓝色的美国数据点显示出很高的相关性，而红色的中国数据点则接近于零。

- **图5h (稳态分析相图):** 横轴是平均跨区域交通强度β_{eff}，纵轴是系统最终达到的稳态平均感染水平x_{eff}。蓝线（USA）显示稳态感染水平随交通强度线性增加。红线（China）则显示存在一个临界点β_{eff}^{c}，低于此点疫情会消亡（$x_{eff}=0$），高于此点则会失控爆发。

讨论

原文翻译
符号回归，即发现揭示数据底层模式的公式，是自动化科学发现的一个关键方向。虽然它已广泛应用于低维动力学和非动力学系统，但一直被认为对高维网络无效。在这里，通过设计一种神经符号回归方法，我们表明高维网络上的符号搜索可以简化为等效一维系统上的搜索。我们的工作代表了一项理论突破，将符号回归的范围扩展到复杂网络系统，从而显著增强了其适用性和潜力。

除了符号回归，Gao和Yan提出了一种基于SINDy的网络动力学辨识方法，该方法将系统行为表示为来自预定义库的函数的稀疏线性组合。他们的方法也已应用于随机复杂系统。虽然在识别主导系统动力学的主要项方面效率很高，但其表达能力受到库的限制，而库的大小本身又受限于可用数据，这限制了其在未知系统中的适用性。相比之下，我们的方法更具表达力和自动化，能够在没有先验知识的情况下发现任何形式的公式，从而能够在复杂的、未见过的系统中发现网络动力学。

ND²将传统的人类驱动的发现推进到机器驱动的网络动力学发现，这对于交通、灾害管理和经济学等复杂系统尤其有价值，因为在这些系统中，经验复杂性和有限的知识限制了传统的人类驱动方法。发现的动力学可以为这些系统提供见解，包括推断节点相关性、揭示韧性属性和相变，正如我们在对COVID-19传播动力学的分析中所展示的那样。ND²还可以通过在广泛的复杂系统中发现展现更丰富行为的多样化网络动力学公式，来推动复杂性科学的研究。

尽管其性能强大，但将ND²扩展到具有高阶相互作用、非加性聚合和未知网络结构的场景仍需进一步研究。高阶相互作用，在脑皮层动力学、物种相互作用和社会习俗中均有观察，可以通过将网络动力学算子扩展到超边来建模（参见方法部分的"发现超越成对模型的动力学"）。然而，调整NDformer以适应高阶动力学并非易事。首先，高阶公式的多样性使得随机公式不足以用于预训练，需要使用大型语言模型生成更真实的公式。其次，需要将NDformer与超图表示学习相结合，以捕捉高阶相互作用的潜在特征。最后，鉴于获取超图结构的困难，ND²可以与超图推断方法协同，共同发现超图结构和高阶动力学。聚合算子也可以扩展到最大值或乘积，以捕捉非加性行为，并对NDformer进行相应调整（方法）。当网络结构不可观测时，ND²可以通过优化边的存在来恢复数百个节点的网络动力学和结构（方法），但这种方法的可扩展性较差。线性回归需要的时间步数多于节点数，而简单地增加时间步数会减慢大型网络的计算速度。未来的改进可能涉及稀疏回归或整合网络推断方法。

我们的方法选择在准确性和复杂性之间达到最佳平衡的公式（帕累托前沿）。然而，基于上下文的标准，如与现有理论的一致性或物理直觉，可能更可取，特别是在数据不足或有噪声的情况下。在这种情况下，简单的函数可以达到与基准公式（如果存在）相当的准确性，使得它们不可辨识（详细讨论见补充材料第2.4节）。在这个阶段，我们引入人类智能来从发现的帕累托前沿中进行评估和选择。将大型语言模型整合到符号回归中提供了一种有前途的方法，它可以结合领域知识来构建或评估符合人类偏好的公式。

深度解读
这部分是对整项研究的总结、反思与展望，它将研究成果置于更广阔的科学背景下，探讨其意义、局限和未来方向。

- **核心突破的重申：** 作者首先强调，他们完成了一件"前无古人"的工作——成功地将符号回归这一强大的科学发现工具，从只能处理简单、低维问题的"玩具"，改造成了能够驾驭高维、复杂网络的"利器"。其关键的理论创新，即通过网络动力学算子实现"降维打击"，打破了长期以来阻碍该领域发展的"维度诅咒"。这不仅是一次技术上的进步，更是一次应用范围上的巨大扩张。
- **与竞争者的对比与优势：** 文章再次将ND²与基于SINDy的方法进行比较，并清晰地阐明了二者的哲学差异。SINDy是"保守派"，它在一个已知的、有限的知识框架（函数库）内寻找最佳解释，效率高但无法带来颠覆性创新。而ND²是"探索家"，它不依赖任何先验的知识框架，有能力发现全新的、前所未见的数学形式。在科学探索的征途中，尤其是在那些我们几乎一无所知的领域（如经济危机、意识的产生），ND²这种更具表达力和创造力的方法，无疑拥有更大的潜力。
- **从"人类驱动"到"机器驱动"的科学革命：** 这项工作被定位为科学发现范式转变的一个里程碑。对于像交通系统、灾害管理、宏观经济这类极其复杂、变量众多、因果链条模糊的系统，传统依赖人类科学家构建模型的方法已经力不从心。ND²提供了一种全新的可能性：让机器直接从海量数据中挖掘出动力学规律，然后由人类科学家来解读、验证和应用这些规律。正如在COVID-19案例中看到的那样，机器发现的公式不仅能预测，更能提供深刻的洞见，比如揭示系统的临界点、评估政策的有效性。这预示着一个人类科学家与AI发现引擎协同工作的新时代。
- **坦诚的自我审视——未来的挑战：** 一项诚实的科学工作不仅要展示成就，也要承认局限。作者指出了ND²未来需要攻克的几个难关：

i. **超越"成对"关系（高阶相互作用）：** 现实世界中，许多相互作用不是简单的"A影响B"，而是"A、B、C共同影响D"（例如，三个朋友一起决定去看电影）。这需要将模型从处理"边"（连接两个节点）升级到处理"超边"（连接多个节点），这对算法和模型架构都提出了更高的要求。

ii. **超越"求和"聚合（非加性行为）：** 当前模型假设邻居的影响是简单相加的。但在某些系统中，最终效果可能取决于"最强"的那个影响（最大值聚合），或是所有影响的"协同"效应（乘积聚合）。扩展聚合算子的形式是未来的一个重要方向。

iii. **处理未知网络结构的可扩展性：** 虽然ND²能在几百个节点的规模上同时推断网络结构和动力学，但当网络规模达到数百万甚至数十亿时（如整个互联网），现有方法将难以为继。这需要与更高效的网络推断算法相结合。

iv. **"正确"与"有用"的权衡：** AI找到的可能是在数学上最精确拟合数据的公式，但这个公式可能极其复杂，或者不符合物理直觉。在数据有噪声或不充分时，可能会有多个同样"准确"的公式。此时，就需要"人类智慧"的介入，从AI提供的"帕累托前沿"（一系列在准确度和简洁度上达到最佳平衡的候选公式）中，挑选出最符合领域知识、最具有解释力的那一个。

- **未来的融合之路——AI与人类的深度协作：** 最后，文章展望了将大型语言模型（LLM）等更先进的AI技术融入符号回归的未来。这可能意味着未来的AI发现系统不仅能生成数学公式，还能用自然语言解释其物理意义，甚至结合已有的科学文献来评估其合理性，从而实现更高层次的人机协作，加速科学发现的进程。

方法

原文翻译

在本节中，我们详细描述我们的ND²方法，包括网络动力学算子的形式化定义（第1部分）、由NDformer引导的符号搜索方法（第2部分）、NDformer的模型架构（第3部分）和预训练过程（第4部分），以及从经验数据中发现公式的方法（第5部分）。

网络动力学算子

为了减少随系统变量数量呈超指数增长的搜索空间，我们将所有节点的标量状态或边的标量权重视为一个整体的"向量化"变量。因此，向量化变量被分为两类：节点级变量$v \in \mathbb{R}^N$（例如，节点的状态和入度）和边级变量$e \in \mathbb{R}^E$（例如，边的权重和相邻节点的状态差异），其中N和E分别表示网络中节点和边的数量。所提出的三种网络动力学算子，其灵感来源于图网络中的节点更新块和边更新块，可以按以下方式在节点级和边级之间映射变量：

$$\rho : \mathbb{R}^E \to \mathbb{R}^N, v = \rho(e) \Leftrightarrow v_i = \sum_{k:j\to i} A_{ij} e_k \quad (\forall i = 1 \dots N) \quad (5)$$

$$\phi_s : \mathbb{R}^N \to \mathbb{R}^E, e = \phi_s(v) \Leftrightarrow e_k = v_j \quad (k = 1 \dots E) \quad (6)$$

$$\phi_t : \mathbb{R}^N \to \mathbb{R}^E, e = \phi_t(v) \Leftrightarrow e_k = v_i \quad (k = 1 \dots E) \quad (7)$$

其中第k条边从第j个节点连接到第i个节点，A代表网络的邻接矩阵。直观地说，ρ算子通过聚合每个节点的输入边来产生一个节点级变量，而ϕ_s / ϕ_t算子通过拾取每条边的源/目标节点来产生一个边级变量（图1b）。为了将向量化变量和提出的网络动力学算子与其他数学算子集成，我们进一步定义，相同类别的向量化变量可以直接以逐元素的方式相互操作（例如，$v_1 + v_2, e_1 \times e_2, sin(v)$）并产生一个相同类别的新向量化变量。相反，不同类别的变量在被映射到同一类别之前不能相互操作（例如，$v + \rho(e)$和$e \times \phi_s(v)$）。图1d展示了一个三节点网络示例，演示了如何组合所提出的网络动力学算子来表示Kuramoto动力学中的相互作用项，即$\sum_{j=1}^{N} A_{ij} sin(x_j - x_i)$。左图显示了每个节点的状态$x_i$，它们可以堆叠成一个向量化的节点级变量$x = [x_1, x_2, x_3]$。中图演示了算子$\phi_s$和$\phi_t$如何通过选择每条边的源节点和目标节点，将节点级变量x映射到边级。得到的两个边级变量$\phi_s(x)$和$\phi_t(x)$，然后通过减法和正弦函数进行逐元素转换，以计算跨边的相位差的正弦值，即$sin(x_2 - x_1)$和$sin(x_3 - x_1)$。右图说明了这些边上的正弦相位差如何被聚合到目标节点，从而产生一个捕捉邻居相互作用的节点级变量。对于节点1，聚合产生其输入边的正弦相位差之和，这正对应于该节点上Kuramoto动力学的相互作用项。对于节点2和3，聚合结果为零，因为它们没有输入边，这也与它们位置上的Kuramoto相互作用项一致。

NDformer引导的符号搜索

在这项工作中，我们提出了一种神经-符号算法来搜索目标公式，其中符号部分，MCTS，在神经部分，一个NDformer的引导下搜索公式。尽管先前的工作已证明MCTS可以有效地在低维系统上执行符号搜索，但在高维网络系统上的符号搜索——尽管我们努力通过引入网络动力学

算子来减少搜索空间——对于纯MCTS算法来说仍然过于庞大。为了解决这个问题，我们引入了一种NDformer引导的MCTS算法，它包括一个用于搜索的符号部分和一个用于引导搜索的神经部分。神经部分，NDformer，学习捕捉系统底层动力学的潜在特征，并估计用于构建公式的每个符号的概率分布。符号部分，MCTS，根据NDformer预测的概率选择符号来构建候选公式。对于每个候选公式，一个奖励计算器使用Broyden-Fletcher-Goldfarb-Shanno算法将未定系数（如果有）拟合到数据，并返回一个评估准确性和简洁性的奖励。更好地拟合数据且长度更短的公式会获得更高的奖励，引导MCTS构建更好的候选公式。

MCTS维护一个搜索树，其内部节点和终端节点分别代表不完整的公式（例如，$sin(\cdot), x+(\cdot)$）和完整的公式（例如，$sin(x), x+y$）。完整的公式可以通过其长度和对观测数据的拟合优度来评估，使用Sun等人提出的奖励函数：

$$r(\mathrm{MSE}, c) = \frac{1}{(1+\sqrt{\mathrm{MSE}/\sigma_{out}^2})(1+\eta c)} \quad (8)$$

其中$\eta \leq 1$是一个超参数，c是公式的长度，σ_{out}^2是真实输出的方差，MSE是公式输出与真实值之间的均方误差。因此，一个更准确地拟合数据且形式更简洁的公式将获得更高的奖励。对于不完整的公式，尽管无法评估其拟合优度，但NDformer可以估计不同符号填入不完整部分以拟合数据的可能性，称为策略$\Pi \in \mathbb{R}^S$（S是补充材料第2.1节中列出的符号集的大小）。策略可以在无法计算奖励的内部节点指示有希望的搜索方向，从而引导和加速搜索过程。

MCTS算法通过一个四步循环迭代地向搜索树中添加节点（在补充材料第2.2节中说明）：
(1) **选择 (Selection):** 在此步骤中，MCTS根据记录的策略和奖励选择最有希望的节点添加到搜索树中。我们遵循Kamienny等人的方法，使用预测的树的上置信界（PUCT）来评估有希望的节点：

$$u(s,a) := q(s,a) + c_{PUCT}\frac{\sqrt{1+\sum_{a'} n(s,a')}}{1+n(s,a)}\Pi_s(a) \quad (9)$$

其中$q(s,a)$和$n(s,a)$分别代表节点s的动作a的记录最大奖励和访问次数（这里的"动作"指"将符号a填入s的不完整部分"），$\Pi_s(a)$是策略Π_s的第a个值，c_{PUCT}是一个平衡探索-利用权衡的超参数。PUCT的第一项$q(s,a)$反映了搜索历史，而第二项，与$\Pi_s(a)$成正比，反映了NDformer的引导。这种引导使得MCTS能够做出有效的选择，而无需为每个动作a探索q(s, a)，从而加速搜索过程。此外，为了在扩展步骤中充分利用NDformer的并行处理能力，我们采用束搜索技术同时选择K个节点，而不是像其他工作那样只选择一个节点。具体来说，我们维护两个优先队列，$\mathcal{P} = \{(s_0, 0)\}$和$Q = \emptyset$，分别用于存储具有前K个"路径平均"PUCT值$v_s$的节点s和要添加到搜索树中的节点。这里，$s_0$是根节点（例如，一个空公式），其PUCT值为0。我们通过以下方式更新它们：

$$Q \leftarrow \mathrm{Enqueue}((s', (1-\alpha)v_s + \alpha u(s,a)) | \forall (s, v_s) \in \mathcal{P}, \forall a, \mathrm{s.t.}\ s' \notin \text{Search tree}) \quad (10)$$

$$\mathcal{P} \leftarrow \text{Top-K of}((s', (1-\alpha)v_s + \alpha u(s,a)) | \forall (s, v_s) \in \mathcal{P}, \forall a, \mathrm{s.t.}\ s' \in \text{Search tree}) \quad (11)$$

直到$|Q| = K$，其中$\alpha = d_s^{-1}$是节点s在搜索树中深度的倒数，动作a作用于节点s产生子节点s'，v_s是从s_0到s的路径上的平均PUCT值。Q中选定的K个节点被添加到搜索树中，并用于后续步骤。
(2) **扩展 (Expansion):** 对于添加的K个节点，MCTS将其q(s,:)和n(s,:)初始化为零，并将它们输入到NDformer以获取其策略Π_s。上一步中采用的束搜索技术使得NDformer能够一次处理K > 1个节点，从而充分利用其并行处理能力，当K=10时，这将处理时间减少并使整体搜索速度提高了3-11倍（补充材料第2.3节）。
(3) **模拟 (Simulation):** 为了评估添加的节点，MCTS估计每个添加节点s对应的子树中终端节点的最大奖励R_s。对于每个添加的节点s，它遵循NDformer的策略在s的子树内采样M个终端节点，然后计算R_s为这些终端节点的最大奖励。值得注意的是，我们使用最大值而不是通常使用的平均值，因为符号回归旨在找到最优公式，而不是一系列平均性能最佳的公式。
(4) **反向传播 (Backpropagation):** 利用添加节点的估计R_s，MCTS沿从选定节点到根节点的路径更新q和n，如下所示：

$$q(s,a) = \max\{q(s,a), R_s\} \quad (12)$$

$$n(s,a) = n(s,a) + 1 \quad (13)$$

其中a是路径上节点s的动作。

MCTS重复这个四步循环，直到满足终止条件，例如找到一个足够准确的公式、达到时间限制或用户手动终止。终止后，MCTS检查其搜索树中的所有终端节点以识别帕累托前沿。具体来说，它选择不同长度的最优公式，然后丢弃那些长度更长但准确性更低的公式，从而获得不同长度约束下的最佳拟合公式。

NDformer的设计

我们设计NDformer来从观测数据中捕捉潜在模式，并引导MCTS的搜索方向。NDformer将观测到的网络结构和节点活动以及不完整的公式作为输入，并生成一个在符号集上的概率分布，称为策略，以评估每个符号填入输入公式的不完整部分以拟合观测数据的可能性。它由三个主要部分组成：（1）一个网络编码器，（2）一个公式编码器，和（3）一个策略解码器（在补充材料第3.1节中说明）。

网络编码器将网络结构$A \in \mathbb{R}^{N\times N}$和节点活动$X \in \mathbb{R}^{T\times N}$嵌入到上下文嵌入$H \in \mathbb{R}^{N_{max}\times d_f}$中，其中N是节点数，T是时间步数，$d_f$是模型维度，$N_{max}$是一个超参数。现有的专注于非网络系统$y = f(x)$的符号回归方法通常将每个时间步的x-y对$(x(t), y(t))$嵌入到一个固定维度的向量$v(t) \in \mathbb{R}^{d_v}$中，这代表了目标公式$y(t) = f(x(t))$的一个样本。然而，在网络化系统中，尽管我们可以用同样的方式将x-y对$(x_i(t), y_i(t))$嵌入到一个固定维度的向量$v_i(t) \in \mathbb{R}^{d_v}$中，但它并不包含足够的信息来作为目标公式$y_i(t) = f(x; A, i)$的样本，因为这个目标公式不仅依赖于$x_i(t)$，还依赖于其邻居和网络结构A，这些都没有包含在$v_i(t)$中。为了解决这个问题，我们引入了一个GNN，其中节点通过消息传递机制与邻居共享状态：

$$v_i \leftarrow \frac{1}{D_i^{in}} \sum_{k:j\to i} M_{edge}(v_j, v_i, e_k) \quad \forall i \in 1 \dots N \quad (14)$$

$$e_k \leftarrow M_{node}(v_j, v_i, e_k) \quad \forall k \in 1 \dots E \quad (15)$$

其中第k条边连接第j个节点到第i个节点，D_i^{in}是节点i的入度，M_{edge}和M_{node}是带有一个隐藏层的多层感知器（MLP），v_i是第i个节点的嵌入（我们省略了时间索引t），$e_k \in \mathbb{R}^{d_f}$是边的嵌入（由边级变量如边权重的嵌入初始化）。经过R轮更新后，获得的v_i和e_k包含足够的信息，包括邻居状态信息，可以作为描述目标公式$y_i = f(x; A, i)$的样本。随后，更新后的节点嵌入$v_i(t)$由一个Transformer编码器编码，以产生上下文嵌入H。为了避免Transformer中自注意力机制的$O(N^2)$复杂性，我们从$v_i(t)$中采样不超过N_{max}个样本点。此外，我们在此处省略了位置编码，因为样本点的顺序不重要。

长度为l的输入公式由公式编码器处理，生成一个公式嵌入$F \in \mathbb{R}^{(l+4)\times d_f}$。具体来说，一个公式可以被看作一个公式树，其内部节点是算子，叶节点是操作数（即变量、未定参数或常数）。通过对树进行前序遍历，我们得到一个符号序列，称为公式的前缀表示法。我们添加一对标记（和）来表示序列的开始和结束，一个标记（[N]或[E]）来表示公式的类型（节点级或边级），以及一个<QUERY_POLICY>标记来指定输出序列中获取策略的位置。尽管前缀表示法已经包含了原始公式的所有信息，我们引入了两种辅助嵌入来强调公式的树结构，如补充材料图4所示，包括（1）公式树中每个节点的父节点索引和（2）每个节点的子树类型（节点级或边级）。这些辅助嵌入与位置编码一起，被添加到前缀表示法的嵌入中，以获得最终的公式嵌入。

以上下文嵌入和公式嵌入为输入，策略解码器使用一个Transformer解码器生成一个输出序列$E \in \mathbb{R}^{(l+4)\times d_f}$。在<QUERY_POLICY>标记指定位置的嵌入随后被送入一个MLP，以生成策略$\pi \in \mathbb{R}^S$，这是一个在符号集上的概率分布，评估每个符号被用作下一个符号来完成输入公式并拟合输入数据的可能性。

预训练NDformer

为了获得在各种系统中引导搜索的能力，NDformer在一个大规模通用数据集上以自监督的方式进行预训练（补充材料第3.2节）。该数据集包含100万个随机生成的网络动力学公式f、网络结构A和节点活动X(t)。为了生成f，我们从一个初始的不完整公式（例如，一个空公式）开始，并用随机选择的符号迭代地填充它，直到它变得完整。此外，我们确保公式的每个部分（即，多于一个节点的子树）至少包含一个变量，从而防止生成仅由常数组成而没有任何变量的琐碎公式（详见补充材料第3.3.1节）。为了生成网络结构A，我们首先从Erdős-Rényi、Watts-Strogatz、Barabási-Albert和完全图中随机选择一种网络类型，然后生成一个相应的随机网络结构，节点数$N \sim \mathcal{U}\{10 \dots 100\}$（详见补充材料第3.3.2节）。为了生成节点活动$X(t) = [x_1(t), \dots, x_N(t)]$，我们没有使用从随机初始值$X(t_0)$开始的展开方法$x_i(t_{n+1}) = x_i(t_n) + f(X(t_n); A, i)\Delta t$，因为这可能导致节点状态在展开过程中发散，而是从一个D维分布中独立同分布地采样每个节点i在每个时间步t_n的状态：

$$x_i(t_n) \sim_{i.i.d.} p_\theta(x)$$

参数θ随机。然后我们计算每个节点i在每个时间步的公式输出$f(X(t); A, i)$来替代$\dot{x}i(t)$。考虑到真实动力学系统通常在其状态空间中具有吸引子（例如，不动点或极限环），观测到的节点状态通常被限制在一个低维流形上，而不是均匀分布在整个状态空间中。因此，我们使用一

个带有离散潜变量和L维连续潜变量（L<D）的高斯混合模型来生成具有多个质心和低维流形形状的分布$p\theta(x)$（详见补充材料第3.3.3节）。通过这种方式，我们可以在D维相空间内的低维（即L维）流形上采样节点状态，这更能反映真实动力学系统。该方法显著增强了NDformer在动力学系统上的引导能力。在发现具有一维极限环状态空间的FitzHugh-Nagumo动力学时，与使用均匀分布$p_\theta(x) = \mathcal{U}[-10, 10]^D$相比，使用高斯混合模型训练的NDformer将搜索速度提高了约60倍（补充材料图10）。

为了预训练NDformer，我们通过从公式的前缀表示法中顺序移除最后一个符号，将每个生成的公式分解为输入-标签对。然而，我们发现，由于一个长度为L的公式可以分解为L个输入-标签对，较长的公式会比较短的公式产生更多的样本，导致NDformer对它们过拟合。为了缓解这个问题，我们从每个公式的分解结果中采样不超过N_{exp}个输入-标签对，这防止了过拟合，并将NDformer的性能提高了10.87%（补充材料图7）。分解后的输入、网络结构和节点活动被送入NDformer以生成策略，然后计算其与标签的交叉熵损失用于反向传播。然而，由于NDformer内的GNN模块一次只能处理一个网络结构，批量大小被限制为1，从而限制了每次反向传播步骤的数据量，导致训练不稳定。为了解决这个问题，我们使用一个缓冲区来缓存上下文嵌入和公式嵌入。一旦缓冲区存储了超过N_{buf}个标记，保存的嵌入就用于训练模型一步。这种设计使得训练更加稳定，并将NDformer的性能提高了50.63%（补充材料图8）。

从经验数据中发现动力学

当从经验数据中发现动力学公式时，我们对经验数据的节点活动进行前向差分，以估计它们的$\dot{x}_i \approx \frac{x_i(t+\Delta t)-x_i(t)}{\Delta t}$，其中$\Delta t$是时间步的长度。然后，使用估计的$\dot{x}_i$作为因变量，我们采用ND²来发现公式$f =$，使得$\dot{x}_i \approx f(X; A, i)$，其中$X = [x_1, x_2, \ldots, x_N]$。每个$f_d$是一个由网络动力学符号集构建的符号序列，该符号集包括提出的网络动力学算子、常用数学算子、未定参数、数学常数和向量化变量（补充材料第2.1节）。

在经验系统中，目标公式中可能存在未知的边权重和异质参数。为了使ND²能够自动发现这些潜在参数，我们在符号集中引入了两种额外的"向量化"参数类型，除了标量参数$C \in \mathbb{R}$之外，包括（1）用于拟合边权重的边级参数$C_e \in \mathbb{R}^E$和（2）用于拟合异质参数的节点级参数$C_v \in \mathbb{R}^N$。例如，考虑Kuramoto模型：

$$\dot{x}_i = \omega_i + \sum_{j=1}^{N} K_{ij} sin(x_j - x_i) \quad (16)$$

其中自然频率ω_i和耦合强度K_{ij}是未知的。仅仅基于观测到的节点活动$x(t)$，我们的方法可以发现：

$$\dot{x} = C_v + \rho(C_e \times sin(\phi_s(x) - \phi_t(x))) \quad (17)$$

然后参数C_v和C_e可以使用奖励计算器的非线性优化器拟合$x_i(t)$数据，以恢复未知异质参数（ω_i）和边权重（K_{ij}）的值。我们的方法不要求每个节点具有相同的参数或每条边具有相同的权重。唯一的同构要求是每个节点的参数及其边的权重以相同的方式影响节点活动（即，遵循相同的公式）。此外，当只有一部分节点满足此同构要求时，我们也可以通过在NDformer和奖励计算器中添加掩码，使它们专注于这些节点，从而找到它们的动力学公式（补充材料第6.7节）。

为了评估发现的网络动力学公式f，我们使用RMSE和sMAPE指标，比较由发现的公式生成的节点活动与真实值。具体来说，节点活动是通过以下方式生成的：

$$\hat{x}_i(t + \Delta t) = \hat{x}_i(t) + f(\hat{X}(t); A, i)\Delta t \quad (18)$$

其中$\hat{X}(t) = [\hat{x}_1(t), \ldots, \hat{x}_N(t)]$表示在时间t生成的节点活动，$\Delta t$表示时间步的长度，$\hat{X}(t_0) = X(t_0)$表示真实的初始状态。在微生物群落动力学和基因调控动力学的实验中，由于原始数据中的Δt较大，我们使用$\frac{1}{10}$的Δt来生成连续轨迹（图3e和4f）。在流行病传播动力学的实验中，因为发现的公式f的输出具有物理意义（即，每日新增病例），我们额外使用R^2比较f的输出与真实值$\dot{x}_i$之间的相关性。此外，为了比较基因调控和微生物群落中现有的和我们修正的动力学公式，我们使用了BIC，它通过平衡拟合优度与模型复杂性来评估模型。RMSE、sMAPE、R^2和BIC的定义在补充材料第5.1.3节中提供。

除了涉及人类干预的专家选择步骤外，我们引入约束来引导搜索过程朝向可解释的结果。具体来说，在合成和经验实验中，我们将公式长度限制在最多30，并将参数数量限制在不超过5个。在经验实验中，我们额外要求未知权重C_e仅作为聚合算子内的乘法因子出现（即，$\rho(C_e \times \ldots)$），以便C_e可以被解释为边权重。在这些限制下，获得的公式表现出可解释性。因此，我们直接从帕累托前沿中选择最准确的公式（列于补充表7和9）。

发现超越成对模型的动力学

成对模型具有一般形式：

$$\dot{x}_i = W(x_i) + \sum_{j=1}^{N} A_{ij} Q(x_i, x_j)$$

通过改变W和Q的形式，它可以有效地描述广泛的系统。然而，它们在捕捉更复杂的动力学方面存在不足。例子包括具有非线性聚合的系统、异质节点动力学、隐藏的边权重、未知的网络结构、高阶相互作用和非加性聚合。我们的核心创新——网络动力学算子——不仅能够发现经典的成对模型，还能发现更复杂的、高阶的模型。

- **具有非线性聚合的系统。** 当相互作用项通过一个非线性函数影响节点动力学时，即$\dot{x} * i = W(x_i) + \Phi(\sum * j = 1^N A_{ij} Q(x_i, x_j))$，系统可以表现出多稳态或混沌行为。如图3c和5b所示，我们的方法能够发现此类动态模型。
- **异质网络上的系统。** 异质网络由多类节点组成，每类节点表现出不同的动力学。Hong和Strogatz研究了一个包含与邻居同步的整合者节点和反对它们的逆反者节点的Kuramoto系统。如补充材料第6.7节的实验所示，通过在NDformer和奖励计算器中应用掩码，使其一次专注于一个节点类别，我们的方法可以恢复每个节点类别的异质动力学。
- **具有隐藏边权重的系统。** 在许多现实世界案例中，边可能具有反映不同信息传输水平的隐藏权重。我们引入了一个解决方案，在符号集内包含可优化的向量化参数来拟合这些边权重。在补充材料第6.8节中，我们通过成功恢复Kuramoto系统中的Kuramoto动力学和隐藏的异质边权重，进一步证明了该方法的有效性。
- **具有未知网络结构的系统。** 当一个系统的网络结构未知时，仅从节点活动中发现系统动力学可能非常具有挑战性。为了解决这个问题，如补充材料第6.6节所详述，我们通过将A_{ij}视为全连接网络上的未知权重，扩展了隐藏边权重的解决方案，通过线性回归求解A_{ij}以提高准确性，并使用BIC来促进A的稀疏性。如补充材料第6.6节所示，我们的方法在合成实验中仅使用Kuramoto系统的节点活动，就成功地恢复了动力学公式和底层网络结构。
- **具有高阶相互作用的系统。** 现实世界的复杂系统通常包含高阶相互作用，其中节点通过连接多个节点的超边进行交互。如补充材料第1.3节所讨论，我们的核心创新，网络动力学算子，可以推广到描述高阶相互作用。具体来说，考虑到ϕ_s和ϕ_t通过选择每条边上的第一个（源）和第二个（目标）节点将节点级变量映射到边级，很自然地可以将它们推广到$\phi_1^{(n)}, \phi_2^{(n)}, \ldots, \phi_{n+1}^{(n)}$，通过选择每个n-超边上的第1、2、...、(n+1)个节点，将节点级变量映射到n-超边级。类似地，将边上信息聚合到节点的ρ可以推广到$\rho^{(n)}$，将n-超边上的信息聚合到节点。
- **具有非加性聚合的系统。** 某些系统可能具有非加性聚合机制，其中邻居影响通过非求和操作（如乘积或最大化）组合。如补充材料第1.1.4节所讨论，我们可以将聚合算子ρ扩展到使用最大值函数聚合的ρ_{max}，从而能够发现此类模型。

深度解读

这一部分是论文的"技术附录"，它为那些希望深入了解ND²内部工作原理的读者，提供了详细的蓝图和操作手册。虽然充满了数学公式和算法术语，但我们可以通过更直观的方式来理解其核心思想。

- **网络动力学算子（公式5-7）：** 这里给出了之前提到的"维度收缩射线"的精确数学定义。
  - **ϕ_s (源) 和 ϕ_t (目标):** 这两个算子的作用就像是"查水表"。对于网络中的每一条"管道"（边），它们会精确地记录下管道的"起点"（源节点）和"终点"（目标节点）的状态。
  - **ρ (聚合):** 这个算子的作用是"汇总信息"。对于每一个节点，它会检查所有指向它的"管道"（入边），并将这些管道中传递的信息（比如通过加权求和）整合起来，得到该节点受到的总影响。
  - **核心思想：** 这套算子将对"具体某个节点"的操作，转化为了对"所有节点"和"所有边"的普适性操作。这使得描述动力学的语言从依赖于网络大小的"绝对坐标系"，转变为不依赖网络大小的"相对坐标系"，从而实现了降维。
- **NDformer引导的符号搜索：** 这部分详细描述了"AI领航员"如何与"探索者"协同工作。
  - **奖励函数（公式8）：** 这是评价一个候选公式好坏的"评分标准"。它巧妙地平衡了两个目标：**准确性**（公式预测的结果与真实数据之间的误差MSE要小）和**简洁性**（公式的长度c要短）。这体现了奥卡姆剃刀原则——"如无必要，勿增实体"，好的科学定律总是简洁而优美的。
  - **PUCT公式（公式9）：** 这是MCTS算法在"选择"下一步探索方向时的"决策公式"。它也包含两个部分：
    a. **$q(s,a)$ (利用/Exploitation):** 代表从历史经验来看，走这一步（选择符号a）能得到的"已知最好成绩"。这鼓励算法沿着已知的、看起来不错的路径深入探索。
    b. **第二项 (探索/Exploration):** 这一项的大小与NDformer给出的"建议"($\Pi_s(a)$)成正比，与这一步被尝试过的次数($n(s,a)$)成反比。这意味着，即使某一步的历史成绩不佳，但如果NDformer强烈推荐它，或者它很少被尝试过（充满未知），算法也会倾向于去"冒险"探索一下。
  - **四步循环：** 选择、扩展、模拟、反向传播，是MCTS算法不断迭代、优化其搜索树的过程，就像一个棋手不断地进行"局面评估、思考新招、快速推演、复盘总结"的循环。
- **NDformer的设计与预训练：**

    - **架构：** NDformer的架构设计体现了对问题的深刻理解。它使用GNN来"看懂"空间上的网络结构，使用Transformer来"读懂"时间上的动态序列，将时空信息完美地融合在一起。
    - **预训练的智慧：** 预训练过程是NDformer获得"智能"的关键。通过在100万个随机生成的"模拟宇宙"中进行训练，它学会了构建物理定律的"通用语法"。这里再次强调了使用高斯混合模型生成"类真实"数据的重要性，这使得NDformer的"世界观"更接近现实，从而能更有效地指导对真实问题的搜索。
- **处理真实世界数据：**
    - **应对未知：** 当面对真实的、不完美的实验数据时，ND²展示了其强大的适应性。它不仅能发现公式，还能同时"反解"出系统中未知的参数，比如网络中每条连接的"强度"（边权重C_e）和每个节点自身的"特性"（异质参数C_v）。这就像一个侦探，不仅能推断出作案手法，还能同时刻画出每个嫌疑人的特征和他们之间的关系强度。
    - **可解释性约束：** 为了确保找到的公式不仅准确，而且有意义、可解释，研究人员加入了一些"软约束"，比如限制公式的最大长度和参数数量。这避免了算法找到一个极其复杂、难以理解但恰好能拟合数据的"怪物"公式，保证了最终结果的科学价值。
- **超越经典模型：** 最后，方法部分明确指出，ND²的框架具有极强的扩展性，其潜力远不止于恢复经典的成对相互作用模型。通过对算子进行推广，它可以被用来探索更前沿、更复杂的科学问题，如涉及多体相互作用的"高阶动力学"和具有非线性汇总机制的系统，为复杂性科学的未来发展打开了新的大门。

总之，"方法"部分是ND²的"设计图纸"和"操作说明书"，它向我们展示了这项工作的技术深度和严谨性，并揭示了其强大性能背后的精妙设计和深刻思考。

Works cited

1. s43588-025-00893-8.pdf
2. [2304.10336] Controllable Neural Symbolic Regression - arXiv, accessed on October 24, 2025, https://arxiv.org/abs/2304.10336
3. Network dynamics - Wikipedia, accessed on October 24, 2025, https://en.wikipedia.org/wiki/Network_dynamics
4. Symbolic Regression : r/learnmachinelearning - Reddit, accessed on October 24, 2025, https://www.reddit.com/r/learnmachinelearning/comments/ov0g7u/symbolic_regression/
5. Symbolic Regression with Genetic Algorithms, accessed on October 24, 2025, https://www.ml4science.com/static_files/lectures/06/symbolic-reg
6. Symbolic regression - Wikipedia, accessed on October 24, 2025, https://en.wikipedia.org/wiki/Symbolic_regression
7. What is Symbolic Regression and Why It's a Game Changer for Data Science - TuringBot, accessed on October 24, 2025, https://turingbotsoftware.com/blog/symbolic-regression-data-science/
8. Discovering governing equations from data by sparse identification of nonlinear dynamical systems | PNAS, accessed on October 24, 2025, https://www.pnas.org/doi/10.1073/pnas.1517384113
9. Monte Carlo tree search - Wikipedia, accessed on October 24, 2025, https://en.wikipedia.org/wiki/Monte_Carlo_tree_search
10. The Animated Monte-Carlo Tree Search (MCTS) | by Thomas Kurbiel - Medium, accessed on October 24, 2025, https://medium.com/data-science/the-animated-monte-carlo-tree-search-mcts-c05bb48b018c
11. Monte Carlo Tree Search - About, accessed on October 24, 2025, https://www.cs.swarthmore.edu/~mitchell/classes/cs63/f20/reading/mcts.html
12. Monte Carlo Tree Search: A Guide | Built In, accessed on October 24, 2025, https://builtin.com/machine-learning/monte-carlo-tree-search
13. ML | Monte Carlo Tree Search (MCTS) - GeeksforGeeks, accessed on October 24, 2025, https://www.geeksforgeeks.org/machine-learning/ml-monte-carlo-tree-search-mcts/
14. Monte Carlo Tree Search (MCTS) algorithm for dummies! | by michelangelo - Medium, accessed on October 24, 2025, https://medium.com/@_michelangelo_/monte-carlo-tree-search-mcts-algorithm-for-dummies-74b2bae53bfa
15. What is a GNN (graph neural network)? - IBM, accessed on October 24, 2025, https://www.ibm.com/think/topics/graph-neural-network
16. Graph neural network - Wikipedia, accessed on October 24, 2025, https://en.wikipedia.org/wiki/Graph_neural_network
17. Graph Neural Networks (GNNs) – Comprehensive Guide - Viso Suite, accessed on October 24, 2025, https://viso.ai/deep-learning/graph-neural-networks/
18. Graph Neural Network and Some of GNN Applications: Everything You Need to Know, accessed on October 24, 2025, https://neptune.ai/blog/graph-neural-network-and-some-of-gnn-applications
19. A Gentle Introduction to Graph Neural Networks - Distill.pub, accessed on October 24, 2025, https://distill.pub/2021/gnn-intro/