

SINDy-GSN:面向社交群体行为的网络动力学稀疏识别模型

王钰涵¹ 马涪元² 马世旋³ 王 英⁴

1 吉林大学软件学院 长春 130012

2 吉林大学人工智能学院 长春 130012

3 吉林大学计算机科学与技术学院 长春 130012

4 符号计算与知识工程教育部重点实验室(吉林大学) 长春 130012

(yuhanw23@mails.jlu.edu.cn)

摘 要 社交网络中群体行为的演化过程通常呈现出非线性、多主体耦合与结构异质性等特征,传统建模方法在揭示其潜在演化规律方面存在局限性。为深入刻画社交网络中群体行为的动态演化机制,提出一种结构耦合函数库驱动的改进型动力学识别模型——SINDy-GSN(Sparse Identification of Network Dynamics for Group Behavior in Social Graphs)。该方法以特征驱动的离散仿真为基础,融合用户节点行为状态、邻接传播结构与主题信息构建三元状态向量,生成适用于社交网络环境的高维非线性函数库。所构建函数库集成了一阶邻居影响、结构归一化扩散项及主题传播耦合机制,全面捕捉个体行为与网络拓扑之间的复杂动态耦合关系。基于真实社交平台特征数据构建仿真网络,并通过离散演化模拟群体立场传播过程,从而实现对群体行为演化方程的稀疏建模与机制识别。实验结果表明,SINDy-GSN 在保持建模可解释性与结构稀疏性的同时,能够有效识别社交网络中的群体传播机制与演化规律,为社交系统中复杂行为的建模与预测提供了通用化的理论工具与方法框架,展现出良好的适应性与拓展潜力。

关键词: 社交网络;函数库构建;SINDy;网络动力学;稀疏回归

中图分类号 TP393

SINDy-GSN: Sparse Identification of Network Dynamics for Group Behavior in Social Graphs

WANG Yuhan¹, MA Fuyuan², MA Shixuan³ and WANG Ying⁴

1 College of Software, Jilin University, Changchun 130012, China

2 College of Artificial Intelligence, Jilin University, Changchun 130012, China

3 College of Computer Science and Technology, Jilin University, Changchun 130012, China

4 Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education, Changchun 130012, China

Abstract The evolution of group behavior in social networks is often marked by nonlinearity, multi-agent coupling, and structural heterogeneity, posing challenges for traditional modeling methods in uncovering the underlying dynamics. To better capture these dynamics, this paper proposes an improved method for dynamic identification based on a structure-coupled function library—SINDy-GSN. The method leverages feature-driven discrete simulation, integrating user behavior states, adjacency structures, and topic information to construct a tripartite state vector, generating a high-dimensional nonlinear function library suited for social networks. The library incorporates first-order neighbor influence, normalized diffusion, and topic propagation coupling, effectively capturing the dynamic interplay between individual behavior and network structure. Using real-world social platform data, a simulation network is created, and discrete evolution models the propagation of group stances, enabling sparse modeling and identification of group behavior dynamics. Results show that SINDy-GSN maintains interpretability and sparsity while accurately identifying group propagation mechanisms, offering a versatile framework for modeling and predicting complex social behaviors, with strong adaptability and scalability.

Keywords Social networks, Function library construction, SINDy, Network dynamics, Sparse regression

1 引言

动力学方程的识别是揭示自然、工程及社会系统潜在演化机制的核心手段。近年来,随着数据采集能力和计算资源

的持续提升,数据驱动的动力学建模逐渐成为物理、生物及复杂系统等多个领域的研究热点^[1-3]。作为复杂系统的重要代表,社交网络呈现出显著的结构异质性与动态时变性,其节点行为(如发帖、评论)受到拓扑结构与个体交互机制的耦合驱

基金项目:复杂动态系统智能理论与方法研究(2021ZD0112500);国家自然科学基金(62272191,62372211)

This work was supported by the Foundation of the Major Project of Science and Technology Innovation 2030—New Generation of Artificial Intelligence(2021ZD0112500) and National Natural Science Foundation of China(62272191,62372211).

通信作者:王英(wangying2010@jlu.edu.cn)

动^[4]。然而,传统动力学建模方法在迁移至社交网络中复杂事件传播的建模任务时存在显著局限。为刻画社交传播过程中的复杂演化规律,有必要发展可适配图结构特征、具备建模效率与解释性的建模方法。

尽管如此,将动力学方程识别方法有效应用于社交传播动力学建模仍面临以下三方面的挑战。1)时间尺度不一致性:社交网络中的传播行为常以离散事件(如转发、评论)为驱动,具有非平稳性与跳跃性,难以直接用传统微分方程建模。2)建模语义的结构差异性:社交传播系统的演化高度依赖图结构信息,传统函数库难以适配图结构驱动的高维异构演化规律。3)缺乏真实机制的验证标准:由于社交传播的底层机制难以直接观测,模型识别结果只能通过预测性能间接验证,限制了模型的可信性与可推广性。为应对上述挑战,本文围绕社交传播中离散性强、结构异质、动力学机制不可观测等问题,构建了一套以离散建模为核心的动力学识别框架,结合图结构信息建模与稀疏回归方法,提升了对复杂社交传播过程的结构刻画与机制挖掘能力。本文的具体贡献如下:

1)提出了一种以离散建模为核心的社交传播建模框架:以意见动力学为理论基础,重点通过离散仿真构建传播过程中的跳跃式行为演化,并在此基础上引入连续建模手段辅助识别动力学结构,有效提升了模型对离散性强、不可微传播过程的刻画与解析能力。

2)设计了结构感知的基函数构造机制:融合邻接矩阵、度矩阵等图结构信息,构建适用于社交网络的多尺度函数库,有效应对社交系统中传播机制由结构驱动所带来的建模语义差异,提升了函数库的表达能力和适配性。

3)提高了仿真数据验证方法的可解释性与稳健性:在加入噪声扰动的观测数据下,通过 MSE 评估所识别方程的准确性,验证了该方法在不同群体事件演化场景中的可解释性,并显著提高了计算效率。

通过本研究的创新方法,为社交网络群体事件演化的建模和分析提供了新的技术途径,也为更复杂系统的研究奠定了方法基础。

2 相关工作

在动力学方程识别领域,早期研究主要集中于物理系统的建模与解析^[29-30],代表性方法包括物理启发的神经网络(Physics-Informed Neural Networks, PINNs)^[30], PdeNet^[31], PDE-Read^[32]以及 DEADONet^[33],均致力于从观测数据中恢复潜在的物理偏微分方程。这类方法在物理建模任务中展现出显著优势,推动了数据驱动物理建模的发展。然而,为了拓展动力学识别方法在更广泛复杂系统中的适用性,引入了遗传编程(Genetic Programming)与符号回归(Symbolic Regression)等通用建模范式^[10,34-35],以摆脱对先验物理结构的依赖^[36-37],从而提升方法在社会、生命、经济等非物理领域中识别系统演化规律的能力。Bongard 等^[25]以及 Schmidt 等^[26]利用符号回归和遗传编程从观测数据中自动提取微分方程^[28]。这些方法通过进化算法优化数学表达式的结构,能够揭示复杂的非线性动力学关系。然而,在高维复杂系统的处理中,计算开销巨大,且模型的可扩展性较差^[34],限制了其在社交网络等大规模系统中的应用。

近年来,稀疏识别(Sparse Identification)方法因其高效建

模和可解释性优势得到了广泛关注^[5]。其中,SINDy(Sparse Identification of Nonlinear Dynamics)方法^[5]通过构建包含多项式、三角函数等算子的候选函数库,将微分方程识别任务转换为稀疏回归问题,并引入系统先验和物理约束来提高建模的准确性与泛化能力。该方法的核心在于稀疏回归策略的选择:LASSO(Least Absolute Shrinkage and Selection Operator)^[11]尽管 L1 正则化有助于构建稀疏、可解释的模型结构,但在高维特征间存在较强线性依赖时,稳定性较弱。为应对这一问题,STLSQ(Sequential Thresholded Least Squares)方法^[25]通过迭代阈值筛选与最小二乘回归相结合,提升了数值稳定性与收敛效率。后续研究进一步优化了回归策略,如引入凸松弛方法^[12]、贝叶斯正则化^[13]、ADMM^[9]等,增强了模型的鲁棒性与表达能力。

随着 SINDy 框架的不断发展,其应用范围得到了显著扩展,已成功应用于多个领域,如光学系统^[16,24]、传染病传播建模^[8,20,22]、生物医学^[6]、机械系统识别^[18-19,21]等。例如,Liang 等^[8]利用 SINDy 模型模拟 COVID-19 的全球传播动力学,并通过 maximin 聚合提出稳健的全球动态模型,克服了国家层面报告数据偏差的问题;Mangan 等^[6]提出数据驱动的隐式 SINDy 方法,通过稀疏优化选择非线性相互作用,成功推断了 3 种典型的生物网络模型,包括迈克利斯-门腾酶动力学、细菌的调控网络和酵母糖酵解代谢网络;Kaiser 等^[7]提出了 SINDY-MPC 框架,通过有限的噪声数据增强模型预测控制(MPC)性能,并展示了其在在线训练、系统快速变化响应和抗噪声能力方面的优势。

特别是当 SINDy 与深度学习结合后^[14],进一步增强了对强非线性行为的建模能力,同时保留了物理可解释性,展现了卓越的适应性。例如,Pan 等^[14]提出了神经隐式流(NIF)框架,通过改进的多层感知器(MLP)实现大规模、参数化、时空数据的低秩表示,解决了复杂时空动力学的可解释表示与压缩问题,并提高了稀疏重建的泛化性能;Servia 等^[15]提出了 ADoK-S 和 ADoK-W 框架,通过遗传编程和稳健的模型选择标准,自动生成催化动力学模型,成功从有限的噪声数据中提取基础动力学速率模型,展示了其在化学反应工程中的应用潜力;Li 等^[18]利用 SINDy 从对流问题的仿真数据中识别简化模型,揭示了一个捕食-被捕食型动力学系统,该系统近似描述了压力分布、湍流流动和分区流的能量状态,并通过分叉分析验证了该系统与仿真数据中的分叉结构一致。然而,原始的 SINDy 框架默认候选函数之间相互独立,未显式引入图结构信息,因此在社交网络等结构化系统的建模中存在一定局限性。

本文在 SINDy 框架基础上,结合社交网络的结构特性,提出了一种改进方法——SINDy-GSN(Sparse Identification of Network Dynamics for Group Behavior in Social Graphs)模型,以更精准地识别复杂传播动力学规律。

3 社交网络下群体演化的仿真建模

3.1 意见动力学

意见动力学(Opinion Dynamics)^[22-23]研究个体在社交网络中的互动如何形成集体意见并推动其演化。意见(Opinion)指个体对某议题的看法,而动力学(Dynamics)则描述这些看法如何在社交互动中变化与传播。此类模型揭示了个体如何相互影响,形成集体行为或群体意见,并帮助预测市场趋

势和品牌影响力。

意见动力学模型分为数据驱动模型与模型驱动模型。数据驱动模型通过分析数据提取演化规律,但由于真实数据难以获取,研究者常依赖自然语言处理技术。与此相对,模型驱动方法通过简化规则与多智能体仿真推演集体意见演化,尤其在数据不足时展现出较强的灵活性和预测能力,适用于本研究的仿真路径。

进一步地,意见模型可分为离散与连续模型。离散模型适用于有限选项的情境,而连续模型则描述意见在数轴上的渐进变化,适合描述持续变化的社会现象。演化研究过程中,社交网络通常表示为图 $G=(V,E)$,节点 $v_i, v_j \in V$ 为用户,边 $e_{i,j} \in E$ 为交互关系,权重反映信息流强度。节点状态的演化受邻居状态影响,反映了个体状态沿图连接边传递的基本特性。

本研究采用离散与连续仿真路径,分别挖掘社交网络中意见演化的离散与连续方程,通过离散模拟用户特征变化与连续微分方程建模意见演化,为网络演化规律提供细致分析框架。

3.2 离散演化

基于意见动力学的演化思想,本文将用户的意见状态表示为 $X \in R^{T \times N \times d}$,其中 T 表示演化的时间步数, N 为网络中节点的数量, d 为每个节点的特征维度。每个节点的意见状态在每个时间步的更新反映了该节点对主题信息的响应以及其在社交网络中的互动。

为模拟节点在每个时间步中的动态响应,模型引入了节点级别的输入事件特征 X_{in} 和输出事件特征 X_{out} ,分别表示节点接收到的输入信息和生成的输出信息。具体而言, $X_{in}(t)$ 表示用户在时间步 t 从其邻居接收到的主题信息,反映了信息的传播和扩散。此时,节点尚未对接收到的信息进行传播,因此 $X_{in}(t)$ 为节点的输入状态。随后,节点会根据当前的意见和激活的主题生成输出信息 $X_{out}(t)$,并传播到邻居节点,从而影响其他节点的意见。信息的接收与传播通过归一化邻接矩阵控制 $\hat{\mathbf{A}} = \mathbf{D}^{-1} \mathbf{A}$,其中 $\mathbf{A}$ 为邻接矩阵, $\mathbf{D}$ 为度矩阵。节点的输入特征由邻居节点的输出 X_{out} 经加权聚合得到:

$$X_{in}(t) = \hat{\mathbf{A}}^T X_{out}(t) \quad (1)$$

其中, $\hat{\mathbf{A}}^T$ 表示归一化后的邻接矩阵。随后,节点的输出特征 $X_{out}(t)$ 是由其对接收到的输入信息 $X_{in}(t-1)$ 进行整合并生成,公式如下:

$$X_{out}(t) = X_{in}(t-1) + \alpha_{out} (\mathbf{B}(t) \odot X(t-1)) \quad (2)$$

其中, α_{out} 是全局事件影响因子, $\mathbf{B}(t) \in \{0,1\}^N$ 是时间步 t 被激活的节点门控, $\odot$ 表示逐元素乘法操作, $X(t-1)$ 为节点的历史立场向量。此方法结合节点当前接收到的信息与自身的历史立场生成输出信息,并传播给网络中的其他节点。节点状态更新则由历史立场与输出信息共同决定,在每一轮的输入输出循环中,节点的意见不断演化,表现为信息的生成、扩散以及集体意见的演进。具体来说,节点状态的更新公式为:

$$\begin{aligned} X(t) &= X(t-1) + \alpha_{in} X_{out}(t) \\ &= X(t-1) + \alpha_{in} X_{in}(t-1) + \alpha_{in} \alpha_{out} \cdot (\mathbf{B}(t) \odot X(t-1)) \end{aligned} \quad (3)$$

其中, α_{in} 为节点的传播接收能力的统一参数。通过这样的输入输出循环,节点的立场不断根据外部信息和自身状态进行

更新,模拟了信息在网络中的传播和集体意见的形成过程。

该建模过程通过离散时间步递推展开,实现事件特征的扩散与节点状态的动态演化。节点激活参与传播取决于邻居节点状态及传播边结构。节点一旦激活,开始接收、整合并传播主题信息。所有节点共享相同的传播与影响参数,传播原则为“只要相连即传播”,有效降低了参数学习的复杂度,为后续的动力学建模与稀疏识别提供统一的传播框架。最终,该模型为基于图结构的离散仿真数据提供支持,为识别网络群体行为的演化机制奠定了基础。

3.3 连续演化

离散传播模型刻画了节点状态在时间步的跳跃式演化,难以全面反映用户立场在持续交互中的动态变化。为更真实地模拟用户认知演化机制,本文进一步引入连续时间建模视角,构建基于多维意见动力学系统的传播模型,描述用户与接收信息之间的连续交互与立场更新。在该连续仿真框架中,用户在主题事件激活后,生成的信息以一定概率对邻居产生激活效应,触发后续传播。社交网络被形式化为包含用户节点与内容节点的双层有向图:1)每个内容节点唯一归属于某用户,表达该用户在特定时间的立场输出;2)在非整数时刻,仅存在用户节点与其内容节点之间的双向边,表示用户在生成内容时的内部认知加工;3)在整数时刻,内容节点将信息通过外部传播边扩散至邻居用户,模拟信息在网络中的传播。信息生成与立场更新过程基于多维意见动力学系统建模,假设每个用户的内部立场为一个 d 维向量 $\mathbf{x}_i \in R^d$,接收到的外部事件信息为另一向量 $\mathbf{x}_j \in R^d$,两者的动态交互可由如下耦合微分方程描述:

$$\begin{aligned} \frac{d\mathbf{x}_i}{dt} &= h I(\mathbf{x}_j - \mathbf{x}_i) \\ \frac{d\mathbf{x}_j}{dt} &= h I(\mathbf{x}_i - \mathbf{x}_j) \end{aligned} \quad (4)$$

其中, h 为意见偏移系数,表示立场调整的敏感程度; $I(\cdot)$ 为意见整合函数,描述立场之间的交互影响机制。该模型模拟了用户在吸收外部观点并生成自身立场时的动态调整过程,体现了立场更新的方向性与连续性。为了将微观交互机制扩展至社交网络的图结构层面,进一步刻画网络中节点群体在多轮传播与加工过程中的集体演化,构造如下连续时间动力系统:

$$\frac{d\mathbf{X}}{dt} = h(\mathbf{A}\mathbf{X} - \mathbf{D}\mathbf{X}) \quad (5)$$

其中, $\mathbf{X} = \{X_V, X_E\} \in R^{T \times (N+M) \times d}$ 表示网络中所有用户与事件节点在不同时间步上的状态集合, $X_V \in R^{T \times N \times d}$ 表示 N 个用户节点的内部立场向量演化过程, $X_E \in R^{T \times M \times d}$ 则表示由用户生成的 M 个内容节点所承载的主题信息; $\mathbf{A}$ 和 $\mathbf{D}$ 则为时间相关的邻接矩阵及其对应的度矩阵,用于刻画网络用户在邻接节点意见的影响下,逐步收敛至均衡态的动态演化过程。

为统一连续与离散仿真过程,并提取结构相关的演化方程结构,本文将上述微分方程以步长 δt 进行离散化,构造如下用户与事件的演化方程:

$$\begin{aligned} X(t) &= X(t-1) + h(\mathbf{A}(t)X(t-1) - \\ &\quad \mathbf{D}(t)X(t-1)) \cdot \delta t \end{aligned} \quad (6)$$

同时,网络中输入与输出的主题分别定义为:

$$\begin{aligned} X_{in}(t) &= \mathbf{A}(t)X(t-1) \\ X_{out}(t) &= \mathbf{B}(t)X(t-1) \end{aligned} \quad (7)$$

邻接矩阵 $\mathbf{A}(t)$ 和度矩阵 $\mathbf{D}(t)$ 的动态性通过激活门控 $\mathbf{B}(t) \in \{0, 1\}^{N+M}$ 建模, 即:

$$\mathbf{A}(t) = \mathbf{A}^T \mathbf{B}(t), \mathbf{D}(t) = \mathbf{B}(t) \quad (8)$$

其中, $\mathbf{B}(t)$ 表示在时间 t 被激活的节点集合, 指示当前参与传播的用户及其对应内容节点。构造中的 $\mathbf{A}^T \mathbf{B}(t)$ 通过从静态图结构中动态筛选出满足传播条件的子图, 确保传播关系仅由被激活的内容节点向邻接用户传播。该机制基于仿真原理——每个内容节点仅归属于一个用户, 激活状态仅依赖于该用户是否生成内容, 因此 $\mathbf{D}(t) = \mathbf{B}(t)$ 成立。通过此种构造, 模型有效捕捉时变性、稀疏性与可传播性。

尽管迭代过程形式上是离散的, 但从数学原理上, 它可视作为连续演化动力系统的数值近似。在实验中, 采用解析解生成连续演化数据, 步长 δt 越小, 所逼近的连续系统解越接近真实演化轨迹。该建模方式保持了连续模型的动态一致性, 并为后续结构识别与动力学建模提供了统一可解释的仿真数据支持。

4 网络群体演化下的稀疏回归

4.1 SINDy 模型

稀疏识别非线性动力学 (SINDy) 是一种基于时间序列数据自动挖掘系统演化方程的稀疏建模方法。该方法假设系统动力学可由一组稀疏的非线性函数组合驱动, 从观测数据中提取具备物理或行为含义的结构性模型, 广泛适用于复杂系统的动力学建模。具体而言, SINDy 将输入的时间序列构造为状态矩阵 $\mathbf{X}$ 及其状态变量的微分矩阵 $\dot{\mathbf{X}}$, 其形式如下:

$$\mathbf{X} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_m) & x_2(t_m) & \cdots & x_n(t_m) \end{bmatrix} \quad (9)$$

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \cdots & \dot{x}_n(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \cdots & \dot{x}_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_1(t_{m-1}) & \dot{x}_2(t_{m-1}) & \cdots & \dot{x}_n(t_{m-1}) \end{bmatrix} \quad (10)$$

其中, 设系统在 m 个离散时间步上进行观测, 包含 n 个状态变量 (或组件)。由此构建的状态矩阵 $\mathbf{X} \in \mathbb{R}^{m \times n}$ 描述了系统在各时间点的状态演化轨迹。状态导数矩阵 $\dot{\mathbf{X}} \in \mathbb{R}^{(m-1) \times n}$ 通常通过有限差分法近似计算, 即通过对状态矩阵在时间维度上逐行差分, 并除以相邻时间步的间隔 δt , 从而得到导数矩阵。具体而言, 对于第 i 个变量, 其在时间点 t_j 处的导数 $\dot{x}_i(t_j)$ 表示为:

$$\dot{x}_i(t_j) = \frac{x_i(t_{j+1}) - x_i(t_j)}{\Delta t}, \Delta t = t_{j+1} - t_j \quad (11)$$

这种导数近似方法可能在时间步 δt 较大或数据噪声较强时引入一定的误差。因此, SINDy 方法进一步假设状态变量 $x(t) \in \mathbb{R}^n$ 的演化受一组隐式非线性函数的主导:

$$\dot{x}(t) = f(x(t)) \quad (12)$$

其中, SINDy 为了能够自适应挖掘出非线性函数 $f(x)$, 提出构建包含多项式项、交叉项、三角函数等非线性项的候选函数库 $\Theta(X)$, 然后将系统建模任务转换为如下稀疏线性系统的求解:

$$\Theta(X) = \begin{bmatrix} | & | & | & | & & | & | & | \\ 1 & X & X^{p_2} & X^{p_3} & \cdots & \sin(X) & \cos(X) & \cdots \\ | & | & | & | & & | & | & | \end{bmatrix} \quad (13)$$

$$\Xi = [\xi_1 \quad \xi_2 \quad \cdots \quad \xi_n] \quad (14)$$

$$\dot{X} = \Theta(X) \Xi \quad (15)$$

其中, $X^{p_2}, X^{p_3}, \cdots$ 表示状态变量 X 的高阶多项式, 例如 X^{p_2} 表示二次非线性多项式; Ξ 为待估计的系数矩阵, 为了提升可解释性与模型简约性, 对 Ξ 施加稀疏性约束, 确保每个状态变量的演化仅依赖于少量函数项。这种“以稀疏性约束结构复杂性”的策略, 与群体行为演化中“少量关键交互驱动系统演变”的理论高度契合。以社交网络为例, 个体意见的变迁往往受到其邻居态度、多源刺激或突发事件的局部驱动, 而非整体网络状态的均匀影响; 同样, 信息的传播更常表现为“骨干链路”引导下的结构性扩散, 而非全网随机传播。

尽管 SINDy 最初主要应用于连续时间系统, 其建模框架亦可自然扩展至离散时间系统, 尤其适用于社交行为演化中广泛存在的离散时间建模需求。在该离散建模框架下, 系统状态的更新由上一时刻状态的函数组合预测当前状态, 其形式可表示为:

$$X(t) = \Theta(X(t-1)) \Xi \quad (16)$$

其中, $\Theta(X(t-1))$ 为上一时刻状态构建的特征函数库, Ξ 为稀疏结构系数矩阵。相较于连续模型中导数估计可能引入的数值误差, 该离散建模形式在处理跳跃演化、时序传播等过程时具备更好的数值稳定性与建模适应性。在社交网络中, 个体状态的更新往往受到主题信息接收、意见调整及信息传播等瞬时交互事件的驱动, 表现为离散的事件序列。因此, 采用离散建模方式更加契合这一数据生成机制, 同时避免了对导数计算的依赖及其可能带来的误差积累问题。

然而, 尽管 SINDy 模型通过稀疏化策略有效提取系统演化中的关键驱动因素, 并能够通过诸如正弦、余弦等数学项进行预测, 但是它并未考虑社交网络中固有的图结构信息。这一局限使得模型所建构的方程无法准确反映基于社交网络的实际传播机制。具体而言, 虽然 SINDy 可以通过其函数库中的数学项捕捉系统的动态行为, 但这些项并未模拟出社交网络传播中的核心, 即节点间的相互作用和传播路径。因此, SINDy 在社交网络中的应用面临一定挑战。

为了弥补这一不足, 本文提出在 SINDy 的离散建模框架和稀疏化策略基础上, 进一步引入社交网络的图结构信息。通过构建更加契合社交网络特征的函数库, 希望能够更精准地捕捉节点间的相互作用与传播机制, 并提升模型在社交网络中的适用性。此举不仅为传播机制的识别提供了更为细致的视角, 也为演化方程的提取提供了更加精确和可解释的建模路径。

4.2 针对群体演化的 SINDy-GSN

在复杂系统动力学建模领域, 揭示系统内在的演化规律是理解其行为机制、实现精确预测的核心。社交网络中群体行为的演化本质上是一个复杂的动态过程, 体现为节点状态在网络拓扑结构及信息交互作用下的非线性、耦合性和时变性。传统的基于先验假设的建模方法, 尽管能够捕捉一定的系统特征, 但往往难以充分揭示高维复杂系统中的潜在非线性耦合结构, 尤其是在社交网络这样的动态环境中。因此, 迫

切需要依赖于数据驱动的机制识别方法,实现对群体行为演化的自动化、精准建模。

为此,SINDy-GSN(见图1)创新性地将 SINDy 引入社交

网络,建模群体行为演化动力学,在离散识别方法的基础上,引入图算子作为图结构信息捕捉社交网络的潜在演化机制,从而提高模型的可解释性与预测能力。

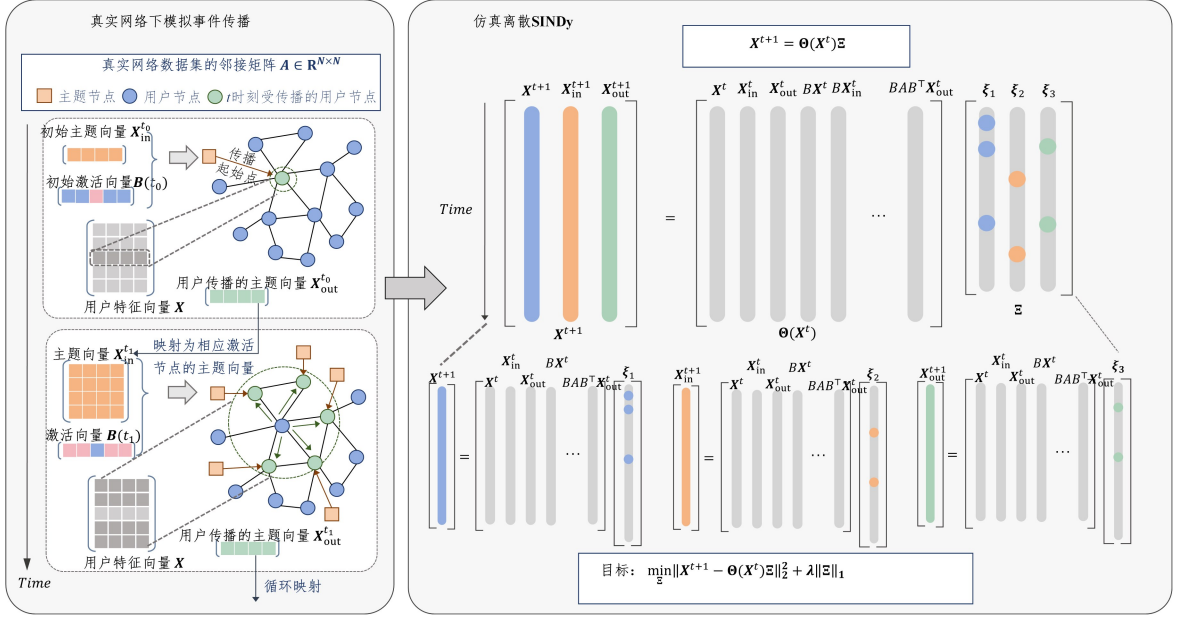


图1 SINDy-GSN 模型架构

Fig.1 Overall structure of SINDy-GSN model

4.2.1 函数库构建

现有的 SINDy 方法主要应用于低维物理系统,其函数库通常以多项式、三角函数等标准基函数为主。然而,在社交网络中,节点状态的演化往往受到复杂的网络拓扑结构和节点间互动关系的强烈影响。与传统的连续时间系统不同,社交网络中的节点状态变化通常表现为离散交互事件驱动的动态演化,且这一过程受限于网络的结构和节点间的信息传播路径。因此,尽管现有的 SINDy 方法在一定程度上能够进行预测,但其识别出的方程往往难以充分反映社交网络中群体行为演化的复杂拓扑结构和时变激活状态,导致模型的可解释性和适用性存在一定局限性。

为了解决该问题,本文提出融合社交网络拓扑特征的图结构感知函数库构建策略,增强模型对网络动力学的表达能力。具体而言,通过引入图算子集合 $O = \{I, \mathbf{A}, \mathbf{A}^T, \mathbf{D}, (\mathbf{D}^{-1}\mathbf{A})^T\}$,构建了一个能够有效融合网络结构和节点间传播机制的函数库。在拓扑结构建模方面,所设计的图算子集合具有明确的网络语义,在建模社交网络中的动态传播过程时,扮演着不同的角色,能够捕捉到网络中节点间复杂的相互关系。其中,身份算子 I 保留了节点的自身状态,确保模型能够维持节点自身的特征表达;邻接矩阵 $\mathbf{A}$ 与其转置 $\mathbf{A}^T$ 刻画邻居节点的相互影响与社交网络中的同质性传播效应,揭示了节点之间的传播依赖关系,使得模型能够捕捉节点间信息的直接传递与共享,进而映射出网络结构在信息传播中的关键作用;度矩阵 $\mathbf{D}$ 量化了节点的连接强度,其对角元素表征节点在网络中的传播影响力,特别是在高度异质的网络中,节点的度数往往直接决定其传播效能,此时度矩阵的作用尤为关键。因此,度矩阵有助于捕捉节点的局部和全局传播影响,增强了模型对社交网络中节点作用力的表达能力。另外,归一化邻接转置 $(\mathbf{D}^{-1}\mathbf{A})^T$ 则通过度数归一化机制,有效缓解高度数节点在信息传播中的主导效应,这一点尤其在社交网络中

至关重要,因为在许多实际网络中,某些高连接度节点往往会对传播过程产生过于显著的影响,可能导致模型对传播动态的过度关注。归一化后的算子使得模型能够更加均衡地考虑所有节点的传播效应,从而更准确地捕捉异质网络中信息的流动和动态变化。这种多尺度图算子的协同作用,能够同时捕捉到局部连接特征与全局拓扑属性,从而克服了传统方法在复杂传播机制建模中的局限性。具体地,本文构建了包含这些图算子的函数库,从而为每个节点的状态更新提供了基于网络结构的非线性变换。函数库的表达式为:

$$\Theta(X_{\text{all}}) = \{G(X_{\text{all}}), G(B \odot X_{\text{all}}) | G \in O\} \quad (17)$$

其中, $X_{\text{all}} = \{X(t-1), X_{\text{in}}(t-1), X_{\text{out}}(t-1)\}$ 包含社交网络里所有的历史观测信息; $G(\cdot)$ 表示对节点状态的非线性变换; $\odot$ 表示逐元素乘积; B 是节点激活门控变量,表示节点的状态调制,该机制在空间维度上支持特征选择性地激活 $B \odot X_{\text{all}}$,增强模型对稀疏特征的感知能力,在时间维度上可动态调节信息传播强度,模拟社交网络中信息传播的衰减效应。通过该函数库,不仅能够有效整合节点原始特征,还能够融合网络传播路径和动态激活信息,为建模社交网络中的复杂行为演化提供强有力的工具。

本文提出的结构感知函数库不仅考虑了节点间的直接交互,还能够充分挖掘网络结构对节点间信息传播的深层次影响,从而为社交网络中群体行为演化提供了更为精确和可解释的建模框架。通过引入图算子,模型能够在节点状态演化过程中准确地融入网络结构和传播机制,从而为社交网络中的潜在动力学规律提供深刻洞察。这一方法不仅突破了传统 SINDy 方法的局限,也为社交网络建模开辟了一条新的研究路径,为群体行为的预测与干预提供了更加准确和可解释的动力学建模框架。

4.2.2 稀疏建模

为了从构建的函数库中筛选出与系统演化密切相关的特

征,采用了 STLSQ(Sequential Thresholded Least Squares)稀疏回归方法。该方法基于最小二乘回归,通过设定阈值对回归系数进行迭代稀疏化,有效剔除冗余和噪声,从而获得简洁且具有良好解释性的模型。与传统 LASSO 不同,STLSQ 通过多轮迭代,在每次迭代中仅针对当前较大系数子集进行最小二乘求解,不断更新系数,最终实现稀疏化。其优化目标可形式化为:

$$L = \|X^{t+1} - \Theta(X^t)\Xi\|_2^2 + \lambda \|\Xi\|_1 \tag{18}$$

其中, X^{t+1} 为目标变量, Θ 为包含节点状态及图结构特征的函数库, Ξ 为回归系数, λ 为控制稀疏性的正则化参数。STLSQ 的迭代策略既保证了模型的预测精度,也有效控制了复杂度,实现了稀疏性与拟合性能的平衡。

此外,为评估模型的实用性,对回归过程的计算时间进行了量化分析。结果表明,STLSQ 在保持较高精度的同时,显著优于传统 LASSO 方法的计算效率,尤其在高维特征库和大规模数据集上表现出更好的适应性与稳定性。该方法不仅提高了特征选择的有效性,也提升了模型在复杂动态网络中的泛化能力和可靠性。

5 实验与分析

在两个不同的网络上模拟了两个动态模型,并测量了改进的 SINDy 的准确性和效率。该算法的准确性是使用发现的微分方程与输入数据相比的均方误差(MSE)来测量的,如式(18)所示。重复每个实验 3 次,并记录发现的系数的平均值,MSE 和算法的运行时间。

5.1 数据集

本文的仿真实验基于两个真实世界社交网络数据集,以保障模拟过程在结构合理性与行为异质性方面的现实关联性。这些数据集的节点拓扑与时间结构为仿真过程提供了真实基础,便于模型对群体演化机制的识别与验证。

1)Enron 邮件通信网络数据集:该数据集记录了一家中型企业员工之间的电子邮件通信行为,包括 184 个节点和 3125 条边。数据以三元组(发送者 ID,接收者 ID,发送时间)的形式呈现,具备明确的时间戳信息,能够用于还原群体之间的信息传播过程。

2)Facebook 社交网络数据集:该数据集收集自 Facebook 用户的社交关系,包含 4039 个节点和 88234 条无向边。尽管缺乏时间信息,本文通过行为规则生成和状态演化机制对其进行仿真,构建群体行为演化过程。

5.2 仿真数据采集

在本实验中,分别按照时间步长[0.1,0.2,0.5,1.0]进行采样,为了清晰展示仿真过程中的状态建模设置,表 1 列出了不同数据集在仿真中采用的节点表征维度以及仿真相关系数。

表 1 仿真数据的基本信息

Table 1 Basic information of the simulation data			
仿真数据集	特征维度	相关系数	
		离散仿真	连续仿真
Facebook Enron	16	$\alpha_{in}=0.20$	$h=0.05$
		$\alpha_{out}=0.15$	

为了评估模型在噪声环境下的稳定性,本文在仿真过程中引入了基于正态分布的噪声,旨在模拟离散与连续仿真数据中的随机扰动。噪声的强度由比例因子 k 控制,其中噪声的标准差与原始数据的绝对值成正比。具体而言,噪声通过以下公式生成:

$$\epsilon(t,x) \sim N(0,k \cdot |u(t,x)|) \tag{19}$$

带噪声的数据 $\tilde{u}(t,x)$ 由原始数据 $u(t,x)$ 与噪声 $\epsilon(t,x)$ 相加得到,即 $\tilde{u}(t,x)=u(t,x)+\epsilon(t,x)$ 通过调整比例因子 k ,可以控制噪声强度,从而模拟不同噪声水平对模型表现的影响。这一方法为分析模型在不同噪声环境下的鲁棒性和稳定性提供了有效手段,能够全面评估噪声对模型预测性能的干扰程度。

5.3 实验环境

为确保实验结果的稳定性与可重复性,本文对每组实验均独立运行 10 次,并采用其平均性能指标。所有实验均在配备 Intel Core i5-8265U 处理器(1.60 GHz)与 NVIDIA RTX 3090 GPU 的计算平台上完成。

5.3.1 主要实验结果

为系统评估所提方法在复杂社交传播建模任务中的有效性,本文设计了 4 组仿真实验,基于 Facebook 与 Enron 两个真实社交网络的拓扑结构,分别开展连续传播与离散传播过程的数值仿真。通过对仿真过程中的系统状态进行采样,构建了相应的动力学数据集,为后续动力学方程识别与模型验证提供了数据支持。实验中,采用经典 SINDy 与本文提出的结构感知建模方法 SINDy-GSN 对 3 类方程对象(系统状态 X 、中间变量 X_{in} 、输出结果 X_{out})进行识别,使用均方误差(Mean Squared Error,MSE)作为性能评估指标,以衡量识别方程对真实数据的拟合精度。实验结果 MSE 如表 2 所列。

表 2 各实验场景下方均误差(MSE)评估比较

Table 2 Comparison of mean squared error(MSE) evaluation across experimental scenarios					
方程对象	模型	均方误差(MSE)			
		Facebook(离散)	Facebook(连续)	Enron(离散)	Enron(连续)
X	SINDy	1.56×10^{-9}	5.17×10^{-7}	2.84×10^{-7}	4.67×10^{-6}
	SINDy-GSN	1.77×10^{-16}	1.09×10^{-17}	1.62×10^{-15}	6.01×10^{-17}
X_{in}	SINDy	6.13×10^{-6}	4.37×10^{-3}	1.42×10^{-6}	2.42×10^{-2}
	SINDy-GSN	4.19×10^{-18}	9.09×10^{-19}	3.94×10^{-20}	2.22×10^{-30}
X_{out}	SINDy	4.13×10^{-8}	1.58×10^{-18}	4.37×10^{-3}	9.54×10^{-19}
	SINDy-GSN	2.84×10^{-7}	1.06×10^{-15}	4.67×10^{-6}	6.00×10^{-17}

从结果中可以看出,SINDy-GSN 相较于传统 SINDy 方法,在拟合误差表现与建模能力方面展现出显著优势,具体体现在以下几个方面。1)拟合误差显著优越且传统方法表现受限:SINDy-GSN 在所有仿真场景及方程对象中均取得数量级

更优的拟合误差,其均方误差(MSE)远低于传统 SINDy 方法,显示出更高的拟合精度。相比之下,传统 SINDy 模型的误差水平普遍维持在较高区间,表明其在高维结构驱动的传播系统中存在显著的建模局限。2)输入与输出间的中间状态

建模优势明显:该性能差距在传播过程的中间状态变量建模中尤为突出,反映出 SINDy-GSN 更加擅长捕捉由图结构主导的传播动态及其复杂演化行为。

上述结果验证了所提模型在拟合精度与鲁棒性方面的显著优势。然而,SINDy 模型构建于通用函数库(如多项式、三角函数)之上,缺乏对图结构的表达能力,候选函数项之间的独立性导致其难以适用于结构高度耦合的社交系统。这使得其所识别的动力学方程虽然在数值上可能误差较小,却在机制表达上缺乏结构语义,难以揭示节点交互与拓扑依赖等核心传播特征。相比之下,SINDy-GSN 明确将邻接矩阵、度矩

阵等图算子纳入函数库,形成结构感知的多尺度表示空间,使识别结果在数值精度与结构解释性方面均具有显著优势。为进一步比较两种方法在结构建模上的差异,选取 Facebook 网络中的离散与连续两类典型场景进行分析。一方面,Facebook 网络结构清晰、节点活跃度高,具备良好的代表性;另一方面,离散与连续模拟在时间建模粒度上具有天然差异,能从多维度反映模型识别能力与稳定性。接下来将展示两种方法在不同噪声强度下所识别出的具体方程形式(见表 3、表 4),进一步揭示图算子引入对社交传播建模机制的结构性影响。

表 3 Facebook 离散仿真识别结果可视化对比

Table 3 Comparative visualization of identification results from Facebook discrete simulation

噪声强度 $k/\%$	模型	识别方程			预期 方程
		$\alpha_{in}=0.20$	$\alpha_{out}=0.15$	$\alpha_{in}\times\alpha_{out}=0.03$	
0	SINDy	$X(t)=-1.728X_{in}(t-1)+0.316X_{out}(t-1)-1.748X(t-1)X_{in}(t-1)+0.546X_{in}(t-1)^2$			(3)
	SINDy-GSN	$X(t)=X(t-1)+0.20X_{in}(t-1)+0.03\cdot B\cdot X(t-1)$			
0.1	SINDy	$X(t)=0.536X_{in}(t-1)-1.537X(t-1)X_{in}(t-1)-0.637X_{in}(t-1)^2+0.504\sin(X_{in}(t-1))$			
	SINDy-GSN	$X(t)=X(t-1)+0.67X_{in}(t-1)+0.03\cdot B\cdot X(t-1)+0.05(AD^{-1})^T\cdot B\cdot X_{in}(t-1)$			
0.5	SINDy	$X(t)=2.590X_{in}(t-1)+2.589X(t-1)X_{in}(t-1)+3.102X_{in}(t-1)^2+1.172\sin(2X_{in}(t-1))$			
	SINDy-GSN	$X(t)=X(t-1)+0.29X_{in}(t-1)+0.03\cdot B\cdot X(t-1)+0.63(AD^{-1})^T\cdot B\cdot X_{in}(t-1)+0.20(AD^{-1})^T\cdot B\cdot X_{out}(t-1)$			
1	SINDy	$X(t)=12.826X_{in}(t-1)+12.83X(t-1)X_{in}(t-1)-3.036\sin(X_{in}(t-1))+12.611\sin(X_{in}(t-1))$			
	SINDy-GSN	$X(t)=X(t-1)+0.58X_{in}(t-1)+0.03\cdot B\cdot X(t-1)+0.08(AD^{-1})^T\cdot B\cdot X_{out}(t-1)+0.03(AD^{-1})^T\cdot X_{in}(t-1)$			
2.5	SINDy	$X(t)=9.562X_{in}(t-1)+9.602X(t-1)X_{in}(t-1)-13.364X_{in}(t-1)^2-1.955\sin(X_{in}(t-1))+18.976\cos(X_{in}(t-1))-18.398\sin(2X_{in}(t-1))$			
	SINDy-GSN	$X(t)=X(t-1)+0.31X_{in}(t-1)+0.34\cdot B\cdot X_{in}(t-1)+0.06(AD^{-1})^T\cdot B\cdot X(t-1)+1.11(AD^{-1})^T\cdot B\cdot X_{out}(t-1)$			

表 4 Facebook 连续仿真识别结果可视化对比

Table 4 Comparative visualization of identification results from Facebook continuous simulation

噪声强度 $k/\%$	模型	识别方程		预期 方程
		$h=0.05$		
0	SINDy	$X(t)=0.500X_{in}(t-1)+0.500X(t-1)X_{in}(t-1)$		(6)
	SINDy-GSN	$X(t)=X(t-1)+0.048A^T\cdot B\cdot X(t-1)-0.048B\cdot X(t-1)$		
0.1	SINDy	$X(t)=0.500X_{in}(t-1)+0.500X(t-1)\cdot X_{in}(t-1)+0.025\sin(2X_{in}(t-1))$		
	SINDy-GSN	$X(t)=X(t-1)+0.048A^T\cdot B\cdot X(t-1)-0.048B\cdot X(t-1)-0.038\cdot B\cdot X_{in}(t-1)+0.030(AD^{-1})^T\cdot B\cdot X_{in}(t-1)$		
0.5	SINDy	$X(t)=0.500X_{in}(t-1)+0.500X(t-1)X_{in}(t-1)$		
	SINDy-GSN	$X(t)=X(t-1)+0.048A^T\cdot B\cdot X(t-1)-0.048\cdot B\cdot X(t-1)+0.061A^T\cdot B\cdot X_{in}(t-1)-0.093(AD^{-1})^T\cdot B\cdot X_{in}(t-1)$		
1	SINDy	$X(t)=0.653X_{in}(t-1)+0.653X(t-1)\cdot X_{in}(t-1)+1.458X_{in}(t-1)^2+1.415\sin(X_{in}(t-1))-0.906\cos(X_{in}(t-1))-0.147\sin(2X_{in}(t-1))$		
	SINDy-GSN	$X(t)=X(t-1)+0.048A^T\cdot B\cdot X(t-1)-0.048\cdot B\cdot X(t-1)+0.024A^T\cdot B\cdot X_{in}(t-1)-0.025(AD^{-1})^T\cdot B\cdot X_{in}(t-1)$		
2.5	SINDy	$X(t)=0.500X_{in}(t-1)+0.500X(t-1)X_{in}(t-1)-0.003\sin(X_{in}(t-1))+0.002\sin(2X_{in}(t-1))$		
	SINDy-GSN	$X(t)=X(t-1)+0.048A^T\cdot B\cdot X(t-1)-0.048\cdot B\cdot X(t-1)+0.027B\cdot X_{in}(t-1)-0.065(AD^{-1})^T\cdot B\cdot X_{in}(t-1)+0.136A^T\cdot B\cdot X_{out}(t-1)-0.170(AD^{-1})^T\cdot B\cdot X_{out}(t-1)$		

从表 3 与表 4 所展示的识别方程可视化结果中可以观察到以下几点。

1)两种方法均对噪声具有一定敏感性:在不同噪声水平下进行建模时,无论是 SINDy 还是 SINDy-GSN,均可能引入一定程度的冗余项,表明这两种方法对噪声具有一定的敏感性。

2)结构保真度存在显著差异:尽管都受到噪声干扰,SINDy-GSN 所识别的方程在结构上依然高度接近真实的传播动力学模型,能够较为准确地还原图结构驱动下的传播机制;相比之下,传统 SINDy 模型由于缺乏对图结构的建模能力,所得方程结构杂乱,难以揭示潜在的拓扑关系与节点交互规律。

3)SINDy-GSN 中结构感知函数库提升了解释性与可识别性:该对比结果进一步表明,引入结构感知函数库有助于在噪声环境下保持模型的结构解释性,并增强传播机制的可识别性,具有显著的建模优势。

4)从稳定性角度来看,本文基于多次实验下识别系数的决定系数 $R^2R\wedge 2R^2$ 计算其波动性,并通过变异系数(CV)加以量化。具体地,决定系数定义为:

$$R^2=1-\frac{(y-\hat{y})^2}{(y-\bar{y})^2}$$

(20)

其中, y 为真实系数, $\hat{y}$ 为预测系数, $\bar{y}$ 为真实系数的均值。多次实验得到的 R^2 均值和标准差分别记作 μ_{R^2} 与 σ_{R^2} ,据此计

算变异系数 CV:

CV=\frac{\mu_{R^2}}{\sigma_{R^2}}\times 100\% (21)

为了将波动性转换为稳定性指标,进一步使用 1-CV 作为模型在不同噪声条件下的稳定性度量。CV 越小,即 1-CV 越接近 1,表示模型在噪声环境下的识别结果越稳定。由于传统 SINDy 的函数项与社交网络传播机制存在显著差异,难以在相同语义下对其结果进行统计,因此本文仅报告 SINDy-GSN 的稳定性。如表 5 所列,在 Facebook 数据集的离散与连续场景中,即便噪声强度达到 2.5%,模型的稳定性仍保持在约 90%,充分验证了 SINDy-GSN 在噪声干扰下的鲁棒性与可靠性。

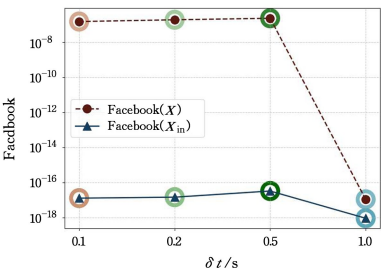
表 5 Facebook 数据集方程系数稳定性分析

Table 5 Stability analysis of equation coefficients on the Facebook dataset

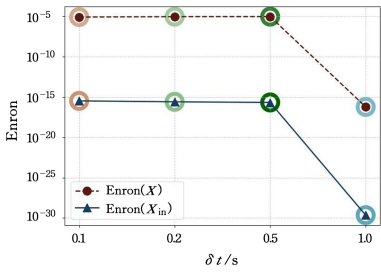
噪声强度 k/%	稳定性 1-CV/%	
	Facebook(离散)	Facebook(连续)
0	97.30	98.44
0.1	86.19	97.89
0.5	83.23	95.89
1	73.18	94.01
2.5	67.57	92.20

5.3.2 采样时间间隔对比实验

为了评估时间间隔对模型性能的影响,在连续仿真中对不同采样时间间隔[0.1,0.2,0.5,1.0]进行了实验,使用了 Facebook 和 Enron 数据集。主要评估指标为均方误差结果如图 2 所示。



(a)Facebook vs. $\delta t/s$



(b)Enron vs. $\delta t/s$

图 2 采样时间间隔对比结果

Fig. 2 Comparison results of sampling time intervals

实验结果表明,通常情况下采样间隔越短,Facebook 数据集的 MSE 越小,表现越好。然而,在较长的采样时间间隔(1s)下,也观察到某些参数(如 X 和 X_{in})的 MSE 表现非常接近最短时间间隔的表现。这是因为本文引入了瞬时变化仿真机制,较长的时间间隔能够捕捉到瞬时变化,导致效果与短间隔相似。Enron 数据集上也显示了类似的趋势,但运行时间随采样间隔减小而显著增加,表明较小的时间间隔会影响模

型的计算效率。

尽管短时间间隔通常能提升精度,但在包含瞬时变化的仿真中,较长时间间隔有时能提供类似的效果。这为模型的优化提供了更多灵活性。

5.3.3 稀疏建模方法对比实验

为比较不同稀疏建模方法的性能,本文对 LASSO 与 STLSQ 在 3 类特征表示(X,X_{in}和 X_{out})上进行了实验对比,如表 6、表 7 所列。从结果来看,STLSQ 在所有场景下均表现出更优的建模精度,其 MSE 显著低于 LASSO,误差水平多在 10⁻¹⁶~10⁻³⁰ 范围内,远优于 LASSO 的 10⁻⁷~10⁻⁸。同时,STLSQ 在大部分实验中的运行时间更短,尤其在 X_{out}上差异更为明显,说明其在准确性和效率之间实现了更优平衡,适用于快速高精度建模场景。

表 6 离散仿真场景下不同算法的均方误差(MSE) 对比

Table 6 Comparison of mean squared error(MSE) for different algorithms under discrete simulation

变量	稀疏建模 算法	均方误差(MSE)	
		Enron	Facebook
X	LASSO	6.674×10 ⁻⁸	3.237×10 ⁻⁸
	STLSQ	6.001×10 ⁻¹⁷	1.323×10 ⁻¹⁶
X _{in}	LASSO	2.443×10 ⁻⁷	5.211×10 ⁻⁸
	STLSQ	2.216×10 ⁻³⁰	5.590×10 ⁻²²
X _{out}	LASSO	6.760×10 ⁻⁸	6.269×10 ⁻⁷
	STLSQ	2.228×10 ⁻³⁰	4.526×10 ⁻²⁰

表 7 连续仿真场景下不同算法的均方误差(MSE)对比

Table 7 Comparison of mean squared error(MSE) for different algorithms under continuous simulation

变量	稀疏建模 算法	均方误差(MSE)	
		Enron	Facebook
X	LASSO	2.277×10 ⁻⁷	5.300×10 ⁻¹⁰
	STLSQ	1.495×10 ⁻²⁰	9.094×10 ⁻¹⁹
X _{out}	LASSO	8.624×10 ⁻⁸	3.126×10 ⁻¹⁰
	STLSQ	4.055×10 ⁻²⁰	9.544×10 ⁻¹⁹

结束语 本文针对社交网络事件传播的建模挑战,提出了一种基于图结构算子的稀疏动力学方法 SINDy-GSN。该方法在传统 SINDy 框架上引入邻接矩阵、度矩阵等多尺度图算子,构建结构感知的先验函数库,有效捕捉节点状态受网络拓扑与传播机制共同驱动的复杂动态演化。通过离散-连续耦合机制,模型能够统一刻画非平稳跳跃传播与连续动力学过程,增强了在复杂传播场景下的表达能力与适应性。实验显示,该方法在噪声扰动下仍具备良好的稳定性和预测准确性。

尽管如此,当前函数库依赖于静态先验设计,难以灵活适应传播机制的动态演变。值得借鉴的是,物理领域已有遗传编程等自动方程识别技术^[38-39],能从观测数据中自适应挖掘关键函数项。未来的研究可从以下几个方向展开。

1)自动构建与优化函数库:引入自动方程识别与符号回归等方法,实现从社交网络观测数据中自动生成并筛选与传播机制相关的函数项,避免人工设计的局限性,近期文献^[40-41]在自适应子图对比学习和图卷积状态选择机制方面的研究为自动构建函数库提供了新的方法论启发。

2)适应时变与多层网络结构:针对现实社交平台中拓扑结构的动态演变与多层交互特征,将时变图算子与跨层信息融合机制及弱监督自训练策略^[42]纳入函数库构建过程,使模型能够刻画多尺度、多关系下的传播动力学过程。

3)面向实际平台的部署与验证:探索在真实社交平台数据上的高效部署方案,包括数据采样、噪声抑制、在线更新与推理加速等技术路径,以确保模型在大规模网络环境下依然保持较高的识别精度与计算效率。

综上,有必要发展能够适配图结构特征、在计算效率与结构解释性之间取得平衡的建模方法,使结构驱动的稀疏动力学框架能够在更广泛的复杂系统中实现有效应用。

参 考 文 献

- [1] GAO J, BARZEL B, BARABÁSI A L. Universal resilience patterns in complex networks[J]. *Nature*, 2016, 530(7590): 307-312.
- [2] HARUSH U, BARZEL B. Dynamic patterns of information flow in complex networks[J]. *Nature Communications*, 2017, 8(1): 2181.
- [3] HENS C, HARUSH U, HABER S, et al. Spatiotemporal signal propagation in complex networks[J]. *Nature Physics*, 2019, 15(4): 403-412.
- [4] BARZEL B, BARABÁSI A L. Universality in network dynamics[J]. *Nature Physics*, 2013, 9(10): 673-681.
- [5] BRUNTON S L, PROCTOR J L, KUTZ J N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems[J]. *Proceedings of the National Academy of Sciences*, 2016, 113(15): 3932-3937.
- [6] MANGAN N M, BRUNTON S L, PROCTOR J L, et al. Inferring biological networks by sparse identification of nonlinear dynamics[J]. *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, 2017, 2(1): 52-63.
- [7] KAISER E, KUTZ J N, BRUNTON S L. Sparse identification of nonlinear dynamics for modelpredictive control in the low-data limit[J]. *Proceedings of the Royal Society A*, 2018, 474(2219): 20180335.
- [8] LIANG J, ZHANG X, WANG K, et al. Discovering dynamic models of COVID-19 transmission[J]. *Transboundary and Emerging Diseases*, 2022, 69(4): e64-e70.
- [9] QUADE M, ABEL M, NATHAN KUTZ J, et al. Sparse identification of nonlinear dynamics for rapid model recovery[J]. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2018, 28(6): 063116.
- [10] JIANG Y, SUN J. Evo-SINDy: Universal Discovery of Partial Differential Equations Using Cooperative Evolutionary Computation[C]// *Proceedings of the Genetic and Evolutionary Computation Conference*. 2025: 791-799.
- [11] HASTIE T, TIBSHIRANI R, WAINWRIGHT M. Statistical learning with sparsity[J]. *Monographs on Statistics and Applied Probability*, 2015, 143(143): 8.
- [12] ZHENG P, ASKHAM T, BRUNTON S L, et al. A unified framework for sparse relaxed regularized regression: SR3[J]. *IEEE Access*, 2018, 7: 1404-1423.
- [13] FASEL U, KUTZ J N, BRUNTON B W, et al. Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control[J]. *Proceedings of the Royal Society A*, 2022, 478(2260): 20210904.
- [14] PAN S, BRUNTON S L, KUTZ J N. Neural Implicit Flow: a mesh-agnostic dimensionality reduction paradigm of spatio-temporal data[J]. *Journal of Machine Learning Research*, 2023, 24(41): 1-60.
- [15] SERVIA M Á C, SANDOVAL I O, HELLGARDT K, et al. The Automated Discovery of Kinetic Rate Models--Methodological Frameworks[J]. *arXiv*: 2301. 11356, 2023.
- [16] SOROKINA M, SYGLETOS S, TURITSYN S. Sparse identification for nonlinear optical communication systems: SINO method[J]. *Optics Express*, 2016, 24(26): 30433-30443.
- [17] DAM M, BRØNS M, JUUL RASMUSSEN J, et al. Sparse identification of a predator-prey system from simulation data of a convection model[J]. *Physics of Plasmas*, 2017, 24(2): 022310.
- [18] LI R, SUN C, DONG M, et al. The controllability analysis of brain networks during rhythmic propagation[J]. *IEEE Transactions on Network Science and Engineering*, 2024, 11(4): 3812-3823.
- [19] NAKARMI U, RAHNAMAY-NAEINI M, KHAMFROUSH H. Critical component analysis in cascading failures for power grids using community structures in interaction graphs[J]. *IEEE Transactions on Network Science and Engineering*, 2019, 7(3): 1079-1093.
- [20] KOLEY S. Critically reckoning spectrophotometric detection of asymptomatic cyanotoxins and faecal contamination in periurban agrarian ecosystems via convolutional neural networks[J]. *Trends in Sciences*, 2024, 21(12): 8528-8528.
- [21] ZHANG Q, YU K, GUOZ, et al. Graph neural network-driven traffic forecasting for the connected internet of vehicles[J]. *IEEE Transactions on Network Science and Engineering*, 2021, 9(5): 3015-3027.
- [22] DONG Y, ZHAN M, KOU G, et al. A survey on the fusion process in opinion dynamics[J]. *Information Fusion*, 2018, 43: 57-65.
- [23] DELABAYS R, DE PASQUALE G, DÖRFLER F, et al. Hypergraph reconstruction from dynamics[J]. *Nature Communications*, 2025, 16(1): 2691.
- [24] ZHANG L, SCHAEFFER H. On the convergence of the SINDy algorithm[J]. *Multiscale Modeling & Simulation*, 2019, 17(3): 948-972.
- [25] BONGARD J, LIPSON H. Automated reverse engineering of nonlinear dynamical systems[J]. *Proceedings of the National Academy of Sciences*, 2007, 104(24): 9943-9948.
- [26] SCHMIDT M, LIPSON H. Distilling free-form natural laws from experimental data[J]. *Science*, 2009, 324(5923): 81-85.
- [27] KOZA J R. Genetic programming; On the programming of computers by means of natural selection[M]. MIT Press, 1992.
- [28] GEAR C W, HYMAN J M, KEVREKIDID P G, et al. Equation-free, coarse-grained multiscale computation; Enabling microscopic simulators to perform system-level analysis[J]. *Communications in Mathematical Sciences*, 2003, 1(4): 715-762.
- [29] RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Physics-informed neural networks; A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. *Journal of Computational Physics*, 2019, 378: 686-707.
- [30] LONG Z, LU Y, DONG B. PDE-Net 2. 0: Learning PDEs from data with a numeric-symbolic hybrid deep network[J]. *Journal of Computational Physics*, 2019, 399: 108925.
- [31] STEPHANY R, EARLS C. PDE-READ; Human-readable par-

tial differential equation discovery using deep learning[J]. Neural Networks, 2022, 154: 360-382.

[32] LU L, JIN P, PANG G, et al. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators[J]. Nature machine intelligence, 2021, 3(3): 218-229.

[33] CAMPS-VALLS G, GERHARDUS A, NINAD U, et al. Discovering causal relations and equations from data[J]. Physics Reports, 2023, 1044: 1-68.

[34] XU H, CHANG H, ZHANG D. DLGA-PDE: Discovery of PDEs with incomplete candidate library via combination of deep learning and genetic algorithm[J]. Journal of Computational Physics, 2020, 418: 109584.

[35] CHEN Y, LUO Y, LIU Q, et al. Symbolic genetic algorithm for discovering open-form partial differential equations(SGA-PDE) [J]. Physical Review Research, 2022, 4(2): 023174.

[36] MASLYAEV M, HVATOV A, KALYUZHNYAYA A V. Partial differential equations discovery with EPDE framework: Application for real and synthetic data[J]. Journal of Computational Science, 2021, 53: 101345.

[37] ATKINSON S, SUBBER W, WANG L, et al. Data-driven discovery of free-form governing differential equations[J]. arXiv: 1910. 05117, 2019.

[38] IVANCHIK E, HVATOV A. Knowledge-aware differential equation discovery with automated background knowledge extraction[J]. Information Sciences, 2025, 712: 122131.

[39] MASLYAEV M, HVATOV A. Equation discovery framework EPDE: Towards a better equation discovery [J]. arXiv: 2501. 14768, 2025.

[40] WANG Y, LIU Y, LIU N, et al. AdaGCL+: An Adaptive Sub-graph Contrastive Learning Towards Tackling Topological Bias [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(9): 8073-8087.

[41] HE X, WANG Y, FAN W, et al. Mamba-based graph convolutional networks: Tackling over-smoothing with selective state space[J]. arXiv: 2501. 15461, 2025.

[42] JUAN X, PENG M, WANG X. Dynamic self-training with less uncertainty for graph imbalance learning [J]. Expert Systems with Applications, 2025, 271: 126643.



WANG Yuhang, born in 2001, postgraduate. Her main research interests include machine learning and deep learning.



WANG Ying, born in 1981, Ph.D, professor, Ph.D supervisor, is a member of CCF(No. 18369S). Her main research interests include machine learning, social networks, data mining and search engines.