

复现 Stepwise reconstruction of higher-order networks from dynamics

摘要

随着复杂网络研究的深入，高阶相互作用在网络结构和功能分析中的重要性日益凸显。然而，高阶网络重建面临维度诅咒问题，使得从动态数据中识别高阶交互极具挑战性。传统方法主要通过直接策略重建高阶网络，但其复杂性随着节点数量和交互阶数的增加而显著增长。本文结合逐步策略与优化技术，提出了一种从时间序列中重建高阶网络的新方法。该方法基于高阶交互的“低阶依赖”特性，显著减少了潜在高阶交互的搜索空间，提高了重建效率和鲁棒性。模拟实验验证了方法在多种网络和动力系统的有效性，展示了其在应对高阶网络稀疏性及维度诅咒方面的优势。

关键词：高阶相互作用；优化技术；逐步策略

1 引言

随着网络科学和复杂系统研究的深入，人们越来越关注网络中的高阶相互作用。传统的网络分析主要集中在节点对 (pairwise) 之间的交互，但现实世界中的许多系统，如社交网络、生态系统、脑网络等，其动态行为往往涉及三个或更多节点的高阶相互作用。这些高阶相互作用对于理解网络的结构和功能至关重要。然而，由于高阶网络重建面临的“维度诅咒”问题，即随着交互阶数的增加，潜在的交互数量呈指数级增长，使得从动态数据中重建高阶网络结构变得极其困难。

2 相关工作

网络重建的理论研究主要关注于如何从动态数据中推断网络的结构。这包括了基于适应性同步方法、Granger 因果关系技术 [3]、统计推断、压缩感知策略 [4] [5] 以及基于深度学习的优化方法 [9] 等多种技术。这些方法主要针对节点对之间的交互，为高阶网络重建提供了理论基础和技术支持。针对高阶网络重建，研究者们开发了基于优化的技术，这些技术通过将网络重建问题转化为求解方程组的问题，尝试找到网络中存在的高阶交互。在算法复杂性分析中，直接策略和逐步策略是两种主要的方法。直接策略尝试建模所有可能的交互，而逐步策略则通过降低模型复杂度来解决维度诅咒问题。[2]

2.1 传统高阶网络重建

传统高阶网络重建方法直接从时间序列中得到数据记为 \tilde{y}_i ，然后将动态方程写为紧凑形式，得到： $\tilde{y}_i = \bar{G}_i^T \bar{A}_i$ ，再将所有时间序列数据 $X \in \mathbb{R}^{m \times n}$ 代入方程 (3)，得到 $y_i = \bar{G}_i \bar{A}_i$ (文

中方程 (4))。在方程 (4) 中, y_i 是已知的时间序列数据, \bar{G}_i 是可以直接获得的动态模式, 而 \bar{A}_i 表示第 i 个节点与其他节点的交互, 是需要确定的参数 [6]。因此, 传统方法将网络重建问题转化为为每个节点单独求解 \bar{A}_i 的问题, 通过根据 \bar{G}_i 的行数和列数之间的关系, 使用一些经典方法来求解这个问题。然而, 传统的方法会产生一些问题, 即维度诅咒: 在直接策略中, 潜在交互 \bar{G}_i 的维度 (即 \bar{G}_i 的列数) 随着节点数量和最大交互阶数的增加而显著增加, 这对重建构成了巨大挑战。

2.2 从动力学中逐步重建高阶网络的方法

对于高阶网络, 其中高阶交互表现出低阶依赖性和更弱或更少的高阶连接, 从动力学中逐步重建高阶网络的方法开发了一种结合逐步策略和优化技术的重构方案 [8], 以从时间序列推断高阶网络。这种方法显著减少了高阶交互的潜在搜索空间。在各种网络和动力系统上的模拟实验表明了方法的有效性和鲁棒性。

3 本文方法

3.1 本文方法概述

该方法首先假设网络演变仅由成对交互驱动, 从而近似表示为:

$$y_i \approx G_i^{(2)} \hat{A}_i^{(2)}$$

其中, $G_i^{(2)}$ 表示第 i 个节点潜在的成对交互集合, $\hat{A}_i^{(2)}$ 代表近似的两体交互关系。由于高阶结构对重建的不可避免影响, 识别出的系数 $\hat{A}_i^{(2)}$ 包含一定误差, 其大小取决于高阶结构对网络动态的影响程度。鉴于高阶结构的“较弱或较少连接”特性, 高阶交互对网络动态的影响相对较小, 因此低阶结构的识别精度相对较高。通过设置阈值或使用节点的先验度, 可以确定哪些边更有可能存在。独立地为每个节点执行此步骤后, 可以获得近似的两体交互矩阵 $\hat{A}^{(2)}$ 。

基于高阶交互的“低阶依赖”特性, 可以将最初高维的三体交互空间 $\bar{G}_i^{(3)}$ 收缩为更小的空间 $G_i^{(3)}$, 其中 $s \leq \frac{n^2}{2}$ 。当所有两体交互都存在时, $s = \frac{n^2}{2}$, 这意味着逐步策略退化为直接策略。对于显著稀疏的两体结构, $s \ll \frac{n^2}{2}$ 。使用已识别的近似低阶结构筛选潜在高阶交互的方法有多种。为了简化, 我们采用了一种典型策略, 即仅当所有相应的低阶交互已经存在时, 才包含高阶交互。因此, 在此处描述的网络重建过程中, 建模三体交互的先决条件是已识别的两体交互矩阵中存在所有三对相应的两体交互。 $G_i^{(3)}$ 的具体例子如图 1(d)-1(e) 所示, 其中仅当 g_{122} 、 g_{132} 和 g_{232} 都被认为存在时, g_{1233} 才被包含在 $G_1^{(3)}$ 中。

需要注意的是, s 的形式可能会根据采用的不同高阶结构筛选规则而略有变化。然而, 一般来说, s 主要受网络稀疏性的影响。网络越稀疏, s 越小, 导致需要建模的交互数量减少。由于现实世界的网络通常是稀疏的, 因此逐步策略建模的交互数量将显著少于直接策略。如图 1 所示:

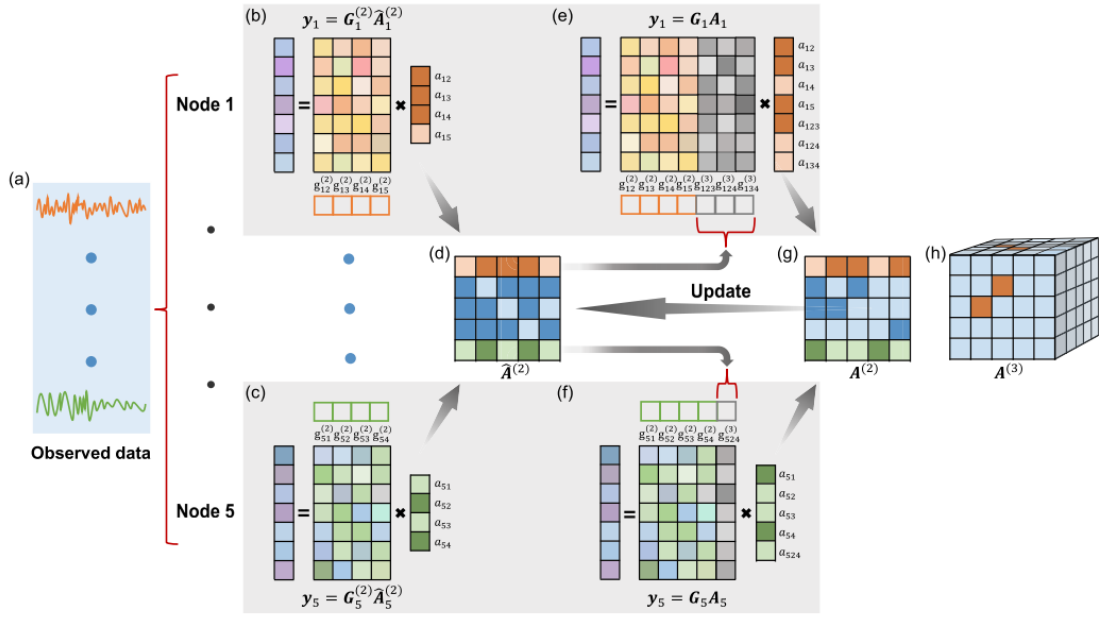


图 1. 方法示意图

3.2 数据生成模块

主要功能是模拟不同动态模型的网络行为，并生成对应的时间序列数据，为后面的网络重建提供了数据支持，可以通过改变不同的模型来验证网络重建方法的准确性。。模块支持多种动态模型（如 Kuramoto 模型、Michaelis-Menten 模型等），通过这些模型模拟网络中节点的相互作用并生成相应的时间序列。具体功能是模拟不同的动态模型，目前支持四种模型：‘kuramoto1’、‘kuramoto2’、‘michaelis menten’ 和 ‘roessler’。这些模型通过不同的数学方程描述节点的动态行为。[1] 网络结构设置：每个模拟的网络由若干个节点组成，每个节点有一定数量的输入连接。通过设置节点的数量（‘N’）、每个节点的输入连接数量（‘NI’）、模拟的时间序列数量（‘S’），以及每个时间序列的时间点数（‘M’），可以定制不同的模拟任务。

3.3 数据评估指标

在模拟中，逐步重建方法假设了对节点度的先验知识，并仅在所有相应的低阶交互存在时才纳入高阶交互。例如，仅当认为存在 $g(122)$ 、 $g(132)$ 和 $g(232)$ 时， $g(1233)$ 才会被包含在 $G1(3)$ 中。此外，考虑到使用逐步策略进行网络重建时，识别出的近似低阶交互矩阵可能表现出非对称性。为了提高重建的准确性和鲁棒性，在筛选需要建模的高阶交互之前，会对近似低阶交互矩阵进行对称化处理，尽可能地在搜索空间中包含存在和不存在的交互。逐步策略的终止条件，如算法 1 中所描述，设置为达到最多 30 次迭代。重建性能使用 F1 分数进行评估，计算公式为 $F1 = 2PR/(P + R)$ ，其中 $P = TP/(TP + FP)$ ， $R = TP/(TP + FN)$ 。[7] TP、FP 和 FN 分别代表真正例、假正例和假负例的数量。F1 分数的范围从 0 到 1，值越高表示准确性越好。F1 分数为 1 表示完美重建了原始网络。

4 复现细节

4.1 复现代码功能分析

在此申明，本文没有参考任何相关源代码。下面为复现代码的功能描述以及运行流程

计算时间序列的导数

为了推断网络中节点的交互，首先需要计算时间序列数据的导数，使用五点法对每个时间序列进行平滑和导数计算。`calculate_derivative` 函数通过对每个时间序列应用五点差分法，计算出每个节点在各个时间步的状态变化率。

推断二体交互

通过函数 `infer_three_body_interactions`，基于已知的二体交互矩阵，推断出可能的三体交互。此过程基于对二体交互强度的评估，并根据预设的阈值 `epsilon` 和交互强度参数 `sigma2`、`sigma3` 进行筛选。函数会返回符合条件的三体交互节点三元组列表。

构建三体交互矩阵

在 `calculate_three_body_interactions_per_node` 函数中，计算了每个节点的二体交互矩阵。该矩阵表示节点之间通过三个节点交互而影响彼此的程度。通过使用 `g2` 函数，为每个时间点计算了与其他节点间的二体交互影响，结果被存储在矩阵 `g_list_3` 中。

合并二体和三体交互矩阵

使用 `merge_interaction_matrices` 函数，将二体交互矩阵 `g_list_1` 和三体交互矩阵 `g_list_3` 合并为一个综合矩阵。该矩阵包含了所有可能的交互信息，能够为后续的交互矩阵求解提供输入。

使用正则化方法求解交互矩阵

- **二体交互矩阵求解：**`solve_interaction_matrix` 函数使用 Ridge 回归方法来求解每对节点间的二体交互矩阵。通过最小化正则化误差，得到节点间的交互系数。
- **三体交互矩阵求解：**在 `solve_interaction_matrices` 函数中，不仅求解了二体交互矩阵 `a2_matrix`，还通过考虑三体交互关系，求解了三体交互矩阵 `a3_matrix`。这两个矩阵结合起来，帮助我们准确建模网络中节点间的相互作用。

输出结果

- 最终，得到的二体交互矩阵和三体交互矩阵被存储在 `Data/a2_matrix.dat` 和 `Data/a3_matrix.dat` 文件中。每个矩阵的元素代表节点之间的交互强度。
- **实验参数：**`Data/ts_param.dat` 文件包含了时间序列数据的参数（例如时间步数和时间序列的数量），这些参数可以帮助我们在后续分析中提取和使用不同的时间序列数据。

4.2 实验环境搭建

安装所需的 Python 库：

将所需的 Python 库安装到虚拟环境中。首先创建一个 ‘requirements.txt’ 文件，内容如下：

```
numpy==1.21.2
scipy==1.7.1
scikit-learn==0.24.2
tomlkit==0.7.1
matplotlib==3.4.3
```

然后通过以下命令安装：

```
bash
pip install -r requirements.txt
或者可以手动安装每个库：
bash
pip install numpy scipy scikit-learn tomlkit matplotlib
```

4.3 界面分析与使用说明

准备环境：

- 安装必要的 Python 库：numpy 和 matplotlib。

运行代码：

- 在终端中运行以下命令：python3 main.py
- 这将执行 plot_fig() 函数，生成并显示包含逐步重建和直接重建方法的 F1 分数与节点数关系的图表。

5 实验结果分析

重建方法

- **逐步重建方法：**使用 stepwise_reconstruction() 函数，在每个实验中构建二体交互矩阵，并通过引入三体交互项逐步更新交互矩阵。
- **直接重建方法：**使用 direct_reconstruction() 函数，通过最小二乘法直接计算二体交互矩阵。

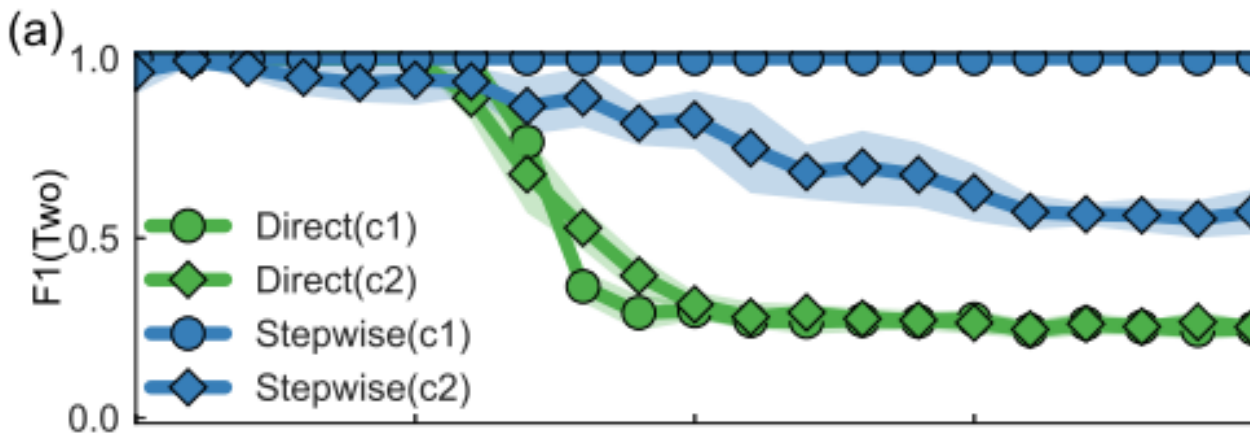
结果可视化

绘制 F1 分数对比图：使用 plot_fig() 函数生成图表，比较两种重建方法在不同节点数下的 F1 分数。

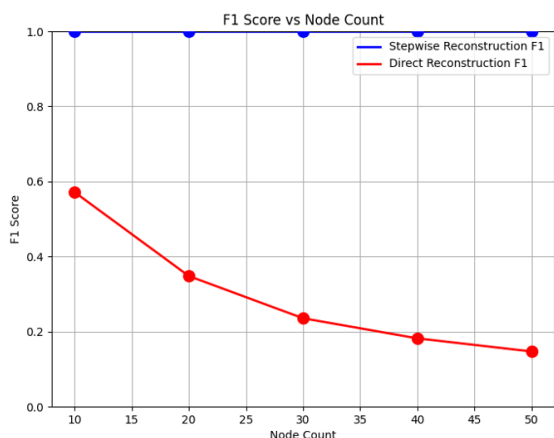
- 横轴：节点数量 (20, 30, 40, 50, 60)。

- 纵轴：F1 分数。
- 两条曲线：一条代表逐步重建方法，另一条代表直接重建方法。

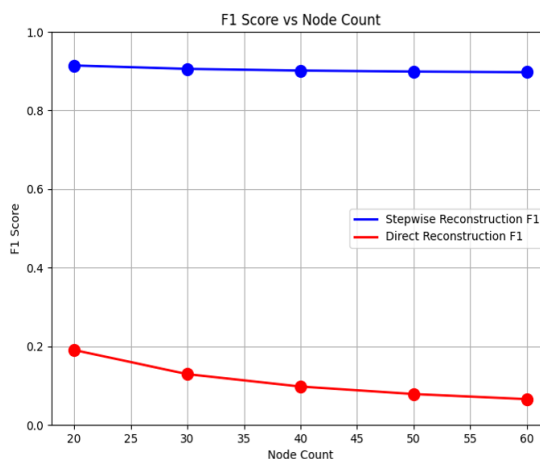
F1 分数越高代表网络重建效果越好，随着节点数量的增加，逐步重建与直接重建效果均有下降，但逐步重建效果下降较为不明显。如图2所示



(a) 论文结果



(b) 复现结果



(c) 复现结果

图 2. 论文结果与复现结果对比

6 总结与展望

本文围绕复杂网络中高阶相互作用的重建问题，提出了一种结合逐步策略与优化技术的新方法。通过从时间序列数据中提取低阶交互信息，并逐步筛选高阶交互，本文的方法有效地降低了维度诅咒的影响，显著提升了重建效率和鲁棒性。在模拟实验中，该方法在不同网络结构和动力系统中的应用均优于传统直接策略。逐步策略的应用表明，高阶网络中交互的稀疏性与低阶依赖性为网络重建提供了新的突破口。尽管本文的方法在性能和计算效率上取得了一定的优势，但仍存在一些值得进一步研究的方向：

1. 方法适用性：当前方法主要针对稀疏网络，对于密集网络或高维度复杂系统的适用性尚需进一步验证。

2. 参数敏感性：算法中涉及的超参数（如阈值设定）可能对结果有显著影响，未来研究可探索自适应参数优化策略以提高鲁棒性。

3. 动态特性建模：现有方法在处理动态网络或时变交互时存在一定局限，扩展方法以适应动态网络的特性将是未来的重要方向。

4. 高阶交互形式的扩展：除本文研究的三体交互外，现实网络中可能存在更复杂的高阶交互形式，对此进行建模和分析将进一步丰富高阶网络的研究体系。

总之，高阶网络重建的研究正处于快速发展阶段。随着计算能力的提升和数据挖掘技术的进步，未来的研究将不仅限于结构重建，还可以结合功能分析，为复杂系统的理论研究与实际应用提供更有力的工具支持。

参考文献

- [1] J. Casadiego, M. Nitzan, S. Hallerberg, and M. Timme. Model-free inference of direct network interactions from nonlinear collective dynamics. *Nature Communications*, 8:2192, 2017.
- [2] Ting-Ting Gao and Gang Yan. Autonomous inference of complex network dynamics from incomplete and noisy data. *Nature Computational Science*, 2(3):160–168, March 2022.
- [3] X. Han, Z. Shen, W.-X. Wang, and Z. Di. Robust reconstruction of complex networks from sparse data. *Physical Review Letters*, 114:028701, 2015.
- [4] G. Li, N. Li, S. Liu, and X. Wu. Compressive sensing-based topology identification of multilayer networks. *Chaos*, 29:053117, 2019.
- [5] C. Ma, H.-S. Chen, X. Li, Y.-C. Lai, and H.-F. Zhang. Data based reconstruction of duplex networks. *SIAM Journal on Applied Dynamical Systems*, 19:124–150, 2020.
- [6] F. Malizia, A. Corso, L. V. Gambuzza, G. Russo, V. Latora, and M. Frasca. Reconstructing higher-order interactions in coupled dynamical systems. *Nature Communications*, 15(1):5184, 2024.
- [7] D. M. W. Powers. Evaluation: From precision, recall and f-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2:37–63, 2011.
- [8] Yingbang Zang, Ziyi Fan, Zixi Wang, Yi Zheng, Li Ding, and Xiaoqun Wu. Stepwise reconstruction of higher-order networks from dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34(7):073156, 07 2024.
- [9] Y. Zhang, Y. Guo, Z. Zhang, M. Chen, S. Wang, and J. Zhang. Universal framework for reconstructing complex networks and node dynamics from discrete or continuous dynamics data. *Physical Review E*, 106:034315, 2022.