

基于社区划分和关键节点的高阶网络重构

王建波^{1,2,3*}, 蒋睿¹, 杜占玮^{4,3}, 李平¹, 许小可^{5,6*}

1. 西南石油大学计算机与软件学院, 成都 610500

2. 四川警察学院智慧警务与国家安全风险治理重点实验室, 泸州 646000

3. 香港大学公共卫生学院, 香港 999077

4. 南方科技大学公共卫生及应急管理学院, 深圳 518055

5. 北京师范大学计算传播学研究中心, 珠海 519087

6. 北京师范大学新闻传播学院, 北京 100875

* 通信作者. E-mail: jianbow2021@gmail.com, xuxiaoke@foxmail.com

国家自然科学基金 (批准号: 62173065)、智慧警务与国家安全风险治理重点实验室 2025 年度开放课题 (批准号: ZHKFYB2503) 资助项目、2024 年度开放课题重点项目 (批准号: ZHKFZD2401) 资助项目和北京市自然科学基金 (批准号: 4242040)。本论文的数值计算得到了西南石油大学高性能计算中心的计算支持和帮助

摘要 高阶网络重构是理解复杂系统群体交互机制的关键问题。然而, 现有方法通常依赖全局搜索策略, 导致组合爆炸问题, 进而引发高时间复杂度, 且重构性能仍存在提升空间。为此, 本文提出了一种基于社区划分与关键节点识别的高阶网络重构框架。该框架基于高阶交互的稀疏性与局部性先验, 将全局重构问题转化为局部社区重构与关键节点全局重构相结合的计算范式, 从而显著缩小了解空间的规模。该框架通过设计一种高效的分阶段重构算法来实现。首先, 基于压缩感知理论重构成对相互作用网络。接着, 通过社区划分对每个社区进行局部重构, 并结合多维权值函数精确识别关键节点。然后, 仅对关键节点进行全局补充重构, 显著提高计算效率, 同时保证重构的完整性。最后, 将局部重构和全局重构结果整合并二值化, 得到最终的高阶网络重构结果。理论分析表明, 与现有先进方法相比, 所提框架显著降低了计算复杂度。实验结果表明, 该框架在多种真实与合成的高阶网络上, 相比现有先进方法, 不仅大幅减少了运行时间, 还提升了重构性能。本文提出的框架为大规模高阶网络的高效重构提供了一种兼顾性能与效率的可行方案。

关键词 高阶网络, 社区划分, 压缩感知, 关键节点, 网络重构

1 引言

复杂网络建模是对现实世界系统中个体交互的有效分析工具。在过去的数十年间, 借助于数学工具和统计数据, 重构复杂网络在低阶网络领域已经取得了巨大的成功。成对网络重构的核心思想是基于可观测的时间序列数据或统计关联信息, 推断出节点间是否存在边连接, 从而重建出网络的

拓扑结构。典型的方法包括基于相关系数（如皮尔逊相关、斯皮尔曼等级相关）[1]、互信息 [2]、最大似然估计 [3]、格兰杰因果分析 [4]、压缩感知 [5~7] 以及转移熵 [8] 等。这些方法在多个领域得到广泛应用并取得显著成效：在神经科学中，通过 fMRI 或 EEG 时间序列重构功能脑网络 [9]；在生态学中，利用物种丰度数据推断物种间的竞争或共生关系 [10]；在社交网络分析中，基于交互频率或信息流重建用户间的联系 [11]。这些成功使得科学家对现实世界中的众多系统从结构到动力学进行了完整的刻画——从交通网络到基因调控网络，现实世界中的交互大多以图结构的形式表示，并使用节点和边分别描述实体及其二元关系。

然而，新的实证数据表明，在多数现实世界网络中，交互并不局限于成对作用，而是经常涉及三个及以上节点之间的群体性、协同性相互作用。传统网络模型局限于两两节点间的二元关系，难以描述真实系统中广泛存在的多体协同效应。相比传统图模型中仅建模二元关系的边结构，高阶结构可更准确地刻画诸如神经网络中三元突触同步 [12]、群聊中的共同暴露、社交网络中的小团体传播结构 [13] 等现象。例如，在社交关系网络中，个体可能同时接收到来自多个他人的共同影响；在生态系统中，两物种之间的相互作用可被第三物种调节（如中介或抑制）[10]；在生物分子网络中，蛋白质往往以复合物形式协作完成功能 [14]。高阶相互作用不仅深刻影响系统的结构组织，也显著制约多节点集体动力学行为，包括传播过程、群体共识、同步现象与合作进化等 [12]。

近年来，高阶网络的重构方法主要遵循两大范式：其一是基于统计推断的模型，如利用最大熵原理 [8] 或张量分解 [15]；其二是基于深度学习的模型，如超图神经网络 [16]。统计推断方法基于概率建模与观测数据，利用参数估计 [17] 或稀疏推断 [15] 方法，确定最可能存在的高阶相互作用。深度学习方法则依托于表示学习 [18] 和非线性函数逼近 [19]，从数据中直接学习节点集合之间的高阶交互概率分布，把高阶相互作用的重构任务转化为一种端到端的预测问题。然而，无论是基于统计推断还是深度学习的方法，它们的一个共同核心局限在于：无法克服固有的组合爆炸问题与现有方法所依赖的全局搜索策略之间的根本性矛盾。具体而言，这两类方法均严格遵循了高阶网络的全局耦合设定 [20]，即高阶相互作用可能存在于任意节点组之间，从而驱使算法对网络中的所有潜在交互进行无差别的全局搜索与建模。这直接导致了计算空间的超指数增长，例如，所有可能的三元交互组合数即高达 $O(N^3)$ 。此种全局性策略必然使算法面临严峻的时间与存储瓶颈，致使现有方法复杂度过高。

幸运的是，近年来的实证研究揭示，真实复杂系统中的高阶相互作用并非均匀或随机分布，而是普遍呈现出两种显著的结构先验：其一为稀疏性，即高阶交互的数量远少于成对交互，且超边规模分布往往随阶数升高而急剧衰减 [21]；其二为局部性，即高阶结构往往在功能社区内部高度集中，而跨社区的高阶连接相对稀少 [16]。这种结构性规律为突破高阶重构的计算瓶颈提供了新的思路：若重构策略能够有效利用此类先验，则可避免对组合空间的全局搜索，从而大幅降低计算开销。为此，本工作利用稀疏性与局部性先验，提出了一种基于社区划分和关键节点的高阶网络重构策略，将全局重构问题以分治策略分解为各个社区内的局部重构和跨社区的关键节点的全局重构，其核心贡献如下：

- (1) 利用结构先验，提出了一种新的高阶网络重构框架。该框架利用高阶交互固有的稀疏性与局部性先验，将原本面向全局组合空间的重构问题，转化为“社区内局部重构为主体、跨社区全局重构为补充”的新型计算范式，从而为缓解计算瓶颈提供了一种有效途径。

- (2) 进一步设计了一种高效的分阶段重构算法来实现该框架。该算法融合社区划分与关键节点识别机制，通过两阶段执行重构：首先在社区内部进行局部重构，随后仅对识别出的关键节点实施全局补充重构。该策略在显著提升计算效率的同时，有效保证了重构的完整性与准确性。
- (3) 理论分析表明，本文所提方法将计算复杂度从现有先进方法的主导项 $O(N^{D+1})$ [15] 显著降低至 $O(|C_{\max}|^{D+1}) + O(|V_{\text{critical}}| \cdot N^D)$ 。在多个真实及合成的高阶网络上进行的实验结果进一步验证，该框架不仅显著降低了运行时间，还在重构性能（以 F1-Score 衡量）方面优于现有先进方法。

本文的其余部分安排如下：第二节回顾了网络重构与高阶交互建模的相关工作；第三节形式化定义高阶网络重构问题；第四节详细阐述本文提出的分阶段重构框架及其核心算法；第五节展示在真实与合成网络上的实验结果与分析；最后第六节总结全文并展望未来研究方向。

2 相关工作

复杂网络作为刻画现实世界系统的强大建模工具，其以拓扑结构反映真实交互，使用节点代表实体，边则描述了实体间的各种交互关系（如社交关系 [11]、生物相互作用 [14]、信息流通 [22]）。然而，在绝大多数实际场景中，这些复杂的拓扑结构往往是无法直接观测的。我们能够获取的通常是节点在时间演化过程中所产生的状态数据（如社交媒体用户的发帖记录、神经元的活动电位、传感器的读数）或统计关联信息（如物种共现数据）。如何从这些可观测的、通常是动态的数据中，反向推断出潜在的、静态的网络连接结构，构成了网络重构领域的核心任务。其最终目标在于揭示系统底层的作用骨架，为理解其上的动力学行为（如扩散 [23]、同步 [12]、控制 [24]）奠定结构基础。网络重构研究的兴起，源于多重驱动因素。首先，从技术层面看，许多重要网络（如大脑功能连接、基因调控网络）由于其规模庞大或探测技术的局限，难以进行全域的直接测量。其次，从科学层面看，结构决定功能是复杂系统研究的基本共识 [25]。唯有准确重构出网络结构，才能解释乃至预测系统的集体行为与功能涌现。最后，从数据层面看，随着监测技术的发展，高维时间序列数据变得日益易得，这为从动力学反推结构提供了数据基础。形式化地讲，网络重构问题可表述为：给定一组节点 V 在多个时间点或多次观测下的状态数据 X ，设计算法 F 以估计出邻接矩阵 \hat{A} ，使其尽可能逼近真实的、未知的邻接矩阵 A ，即 $\hat{A} = F(X)$ ，该问题本质上是一个高维统计推断问题，其面临的核心挑战包括但不限于：计算复杂度（随节点数呈超线性增长）、数据稀疏性与噪声、以及如何区分直接交互与间接关联以控制误判率。

在网络重构的诸多分支中，旨在恢复节点间两两交互（Pairwise Interactions）的成对网络重构理论 [26] 最为成熟，已发展出多种基于不同原理的计算范式。基于相关性的方法 [27] 是最直观的一类：其核心假设是节点行为的时间序列之间的统计相关性暗示着它们之间存在连接。例如，皮尔逊相关系数用于衡量线性关联 [28]，而斯皮尔曼等级相关则用于捕捉单调关系 [1]。这类方法计算高效、实现简单，但缺陷在于“相关不等于因果”，容易产生虚假连接。为捕捉非线性依赖关系，基于信息论的方法被广泛应用：如互信息（Mutual Information）[2] 和转移熵（Transfer Entropy）[8]，其分别能够捕捉非线性依赖和时间方向性，揭示复杂的统计依赖，但计算成本较高。基于动力系统模型的方法为网络重构 [17] 提供了坚实的理论基础。该类方法假设时间序列由特定动力学模型生成，并将重构转化为参数估计问题。其中，压缩感知（Compressed Sensing）[5~7] 利用网络稀疏性，将

重构转化为稀疏回归问题,效果显著。此外,基于概率图模型的方法:如最大熵原理 [8]、贝叶斯推断 [29],通过概率分布建模来推断网络结构,灵活但计算复杂度高。这些方法已在神经科学 [9]、生态学 [30]、社交网络分析和基因调控网络中取得了显著成效。然而,其本质局限在于只考虑二元交互,忽略了真实世界中普遍存在的群体性作用机制。

为突破上述局限,研究者们逐渐将目光投向高阶相互作用 (Higher-order Interaction) 的建模与重构。高阶建模范式如超图 (Hypergraph) [19]、单纯复形 (Simplicial Complex) [31] 和 k -交互网络 [20],能够通过超边直接表示多个节点的群体关系,更好地刻画社交小团体、蛋白质复合物 [32]、多物种共生等现象 [33]。现有高阶网络重构主要分为两类方法:其一是基于统计推断的方法:如最大熵建模、互信息验证、低秩张量恢复等。这类方法利用观测数据的统计特性,恢复潜在的高阶结构,但在大规模网络上会遇到维度灾难。二是基于深度学习的方法:如超图神经网络 (HGNN) [19],通过神经网络自动提取时间序列特征,学习高阶交互的隐式映射。然而,这类方法计算开销巨大且可解释性不足。但无论是统计推断还是深度学习,当前方法普遍依赖“全局搜索” [34] 范式,隐含假设高阶相互作用可能存在于任意节点组之间,使得重构时需面对超指数增长的参数空间,尽管诸如稀疏正则化等技术被用于从该巨大空间中筛选真实交互,但其底层的优化过程仍无法避免对该庞大组合空间的全局性遍历与评估,无法克服固有的组合爆炸问题与现有方法所依赖的全局搜索策略之间的根本性矛盾。这导致传统方法在大规模网络上陷入计算瓶颈。

值得注意的是,这一计算瓶颈与真实高阶网络所普遍呈现的结构性规律形成了鲜明对比。大量实证研究指出,真实高阶网络并非均匀分布,而是呈现稀疏性与局部性。例如,在野外节肢动物群落的研究中,虽然存在高阶调节作用,但由三个或更多物种共同参与的交互观测次数远少于简单的消费者-资源对 [35],证明高阶网络的数量稀疏性:高阶交互数量远低于成对交互,且超边阶数分布快速衰减。Feng 等人 [16] 通过三物种共培养实验验证了局部 HOI(Higher-Order Interactions) 形成的拓扑空洞占全局高阶连接的 70% 以上,而跨群落的高阶交互仅占不足 10%,发现局部社区的拓扑空洞具有更高的持久性,而跨社区的高阶连接则随时间快速消解。同时,基于同源性理论的超网模型 [36] 研究证明了社区内高阶相互作用的局部性与稳定性,这共同证明了高阶网络的空间局部性:高阶交互多集中在社区内部,而跨社区交互稀疏且不稳定。此外,Francesco 等人 [21] 聚焦于多层物理与电影合作网络,发现超边阶数的分布呈现明显衰减:小群体 (阶数 2-3) 占据绝大多数,而更高阶群体 (阶数 > 5) 极为少见,表现出“小团体偏好”特征,证明了高阶网络的阶数衰减性:三阶及以下的低阶相互作用占据主导地位,其中三阶相互作用形成的三角结构在真实网络中具有高度聚集性。然而,现有主流方法往往忽视了这些结构性先验,导致冗余计算与泛化误差。基于此,与其在全局空间中进行昂贵的筛选,不如将上述结构性规律直接作为算法设计的基本先验。通过结构性引导,将计算资源集中于最可能产生高阶交互的、规模大幅减小的子空间内,首先在社区内部进行高效重构;再针对跨社区的关键节点实施全局补充重构。这样既能规避全局搜索带来的计算灾难,又能在保持性能的同时显著降低时间复杂度。

在本文中,我们提出了一种基于社区划分和关键节点的高阶网络重构框架 (Higher-Order Network Reconstruction Based on Community Partitioning and Key Nodes, HRCK),旨在高效重构大规模高阶网络。该框架首先通过社区划分将网络划分为若干内部连接紧密的子系统 (社区),随后分两阶段实施重构:第一阶段:在社区内部采用压缩感知与张量分解方法,仅重构社区内节点的高阶作用,降低计算复杂度;第二阶段:基于节点权值函数识别潜在的跨社区关键节点,仅对这些节点

实施全局范围的精细重构；最终整合局部与全局重构结果，得到完整的高阶相互作用矩阵。本文的核心创新在于：通过结构化稀疏先验与重点采样机制，将全局组合爆炸式的高阶重构问题转化为以社区内局部重构为主体，配合关键节点全局重构进行补充的混合策略，从而在保留高阶拓扑信息的同时，显著降低时间复杂度。理论分析与实验结果表明，该框架能够在保证性能的前提下，有效减少计算与存储开销，适用于大规模高阶网络的重构场景。

3 问题描述

3.1 网络结构形式化定义

考虑一个包含 N 个节点的无权无向网络，其底层拓扑结构（包括成对与高阶相互作用）是未知的。我们旨在解决的问题是，如何根据有限的观测数据，重构出该网络的结构。

考虑一种常用于社会系统和自然系统建模的二元状态动力学过程，网络的二元连接结构用邻接矩阵 $A \in \{0, 1\}^{N \times N}$ 表示，其中元素 A_{ij} 定义为：

$$A_{ij} = \begin{cases} 1, & \text{若节点 } i \text{ 与节点 } j \text{ 之间存在直接连接;} \\ 0, & \text{若节点 } i \text{ 与节点 } j \text{ 之间不存在连接。} \end{cases} \quad (1)$$

超越成对的高阶交互用超图 $H = (V, \mathcal{E})$ 表示，其中 V 是节点集合，节点总数为 N ， $\mathcal{E} \subseteq 2^V \setminus \emptyset$ 是超边集合。每条超边 e 是节点的非空子集，即 $e \subseteq V$ 且 $|e| \geq 2$ ，用于表示节点之间的协同作用。超图也可用一个 N 维张量 \mathcal{A} 表示，其中元素 $\mathcal{A}_{i_1, i_2, \dots, i_k}$ 表示节点集合 $\{i_1, i_2, \dots, i_k\}$ 是否存在 k 阶交互。

值得注意的是，高阶结构也可表示为单纯复形 [37]，即每条高阶边不仅包括所有参与节点，还要求其所有低维子集（例如三元组的所有二元子集）也必须同时存在于结构中。这种表示在处理带有边界性质的拓扑分析时具有优势，尤其适用于代数拓扑方法 [20]。基于其强大的表示能力，本文采用超图模型，使用张量恢复方法重构高阶网络，该方法具有表示自由度高、结构约束少的优势，特别适用于动力学驱动下的结构重构任务。

3.2 高阶网络重构问题的数学描述

在成对作用网络上，节点动力学遵循一个受网络结构驱动的二元状态过程。 $x_i(t)$ 表示节点 i 在时刻 t 的状态（如感染/易感、活跃/不活跃），节点的状态演化由其邻居（包括成对和高阶邻居）的联合状态驱动。通过观测节点状态的时间序列 $\{x_i(t)\}$ ，可以构造出一个回归形式：

$$Y_i = \Phi_i a_i + \xi_i, \quad (2)$$

其中 $Y_i \in \mathbb{R}^T$ 表示节点 i 的观测向量（如状态变化量或差分信号）； $\Phi_i \in \mathbb{R}^{T \times N}$ 是由其他节点观测信号构造的基函数矩阵； $a_i \in \mathbb{R}^N$ 是邻接矩阵第 i 行（即节点 i 的邻接向量）； ξ_i 是噪声项。

由于真实网络通常是稀疏的， a_i 也是稀疏向量。于是成对重构问题等价于：

$$\min_{a_i} \|a_i\|_1 \quad \text{s.t.} \quad Y_i \approx \Phi_i a_i. \quad (3)$$

该优化可通过 Lasso 或其他稀疏回归方法求解，从而恢复邻接矩阵 A 。

高阶重构问题可统一写为：

$$\min_{\theta_i} \|\theta_i\|_1 \quad \text{s.t.} \quad Y_i \approx \sum_{k=2}^K \Phi_i^{(k)} \theta_i^{(k)}, \quad (4)$$

其中 $\theta_i^{(k)}$ 对应高阶张量 $\mathcal{A}^{(k)}$ 的一部分。

3.3 动力学模型

本文所述框架基于张量恢复框架 [15]，其核心思想是：将网络中所有潜在的 k 阶交互表示为一个 k 阶张量，其中张量元素代表超边，将重构问题形式化为一个稀疏张量恢复问题。该框架通过建立节点状态导数与拓扑结构间的非线性关系，并求解稀疏优化问题，从时间序列中识别潜在的高阶相互作用。相较于仅恢复成对结构的方法，该框架能够有效识别三元及以上的群体协同作用，是高阶网络重构领域的代表性方案之一。我们将该框架推广应用于合成网络（SIS 模型、Ising 模型驱动下的超图）与经验网络（如社交或生态系统中的高阶结构），以验证其在不同场景中的有效性与泛化能力。

为验证高阶网络重构框架的有效性，须基于已知真实结构的网络及其产生的动力学数据对算法进行测试。本研究选用两种在物理学和社会学中广泛使用的、能够产生丰富动力学行为的模型作为数据生成器：高阶 SIS 流行病模型与高阶 Ising 模型。前者模拟了非平衡态的传播过程，后者则模拟了平衡态的统计物理相变行为，二者结合可以全面检验重构算法在不同场景下的普适性与鲁棒性。

3.3.1 高阶 SIS 流行病模型

在 SIS 流行病传播模型中，易感个体的感染过程可视为病原体侵蚀的结果。这一过程不仅源于成对的接触传播（即两两交互），也可能来自多个感染源的协同效应（即高阶交互）。为刻画此类高阶相互作用的影响，本文在经典的基于两两交互的 SIS 模型基础上，引入三体相互作用项，构建了一个高阶流行病传播模型。

在该模型中，我们设定 $S_i(t)$ 为节点 i 在 t 时刻的状态，每个节点仅可能处于两种状态之一：

$$S_i(t) = \begin{cases} 0, & \text{节点未被感染, 处于易感态} \\ 1, & \text{节点已被感染, 处于感染态} \end{cases} \quad (5)$$

易感节点可以通过两种途径被感染：一是通过成对相互作用 (i, j) 被邻居节点 j 感染，其对应的基础感染概率为 γ_1 。其次是通过与两个已感染邻居节点 j 和 k 的协同交互 (i, j, k) 被感染，其中 j 和 k 在 t 时刻都处于感染态，其对应的协同感染率为 γ_2 。

因此，任意节点 i 在 t 时刻被感染的概率可以表示为：

$$P_i^\gamma(t) = \min \left(1, \gamma_1 \sum_{j=1, j \neq i}^N A_{ij} S_j(t) + \gamma_2 \sum_{j=1, j \neq i}^N \sum_{k=j+1, k \neq i}^N A_{ij} A_{ik} S_j(t) S_k(t) \right), \quad (6)$$

其中 A 为邻接矩阵， $S_j(t)$ 和 $S_k(t)$ 分别代表节点 j 和 k 在 t 时刻的状态。

我们设定 $\gamma_1 = \frac{\alpha}{\kappa_1}$, $\gamma_2 = \frac{\beta}{\kappa_2}$, 其中 α 、 β 为非零常量, κ_1 、 κ_2 分别为网络的两体和三体相互作用的占比。在本文工作中, 将 α 、 β 选取在各自流行病传播阈值附近, 以确保动力学过程对网络结构的高阶特征保持敏感, 从而实现网络的高性能重构。对于稀疏网络, 一般约束 $\alpha < \beta$, 以保证两两相互作用在传播动力学中占据主导地位。另一方面, 每个感染节点有 σ 的概率恢复到易感状态, 节点 i 在 t 时刻的恢复率表示为: $P_i = \sigma$ 。

3.3.2 高阶 Ising 模型

Ising 模型最初用于描述铁磁体中的自旋相互作用, 因其在统计物理相变中的重要作用而被广泛应用于许多领域 [38]。该模型在临界温度附近对网络结构异常敏感, 为测试重构算法提供了理想的高阶动力学数据。我们选用 Ising 模型以验证方法在平衡态统计物理系统中的适用性。

在该模型中, 每个节点具有两个可能状态, 自旋向上 ($S_i(t) = +1$) 或自旋向下 ($S_i(t) = -1$)。在传统的二体相互作用下, 节点 i 的能量 E_i (哈密顿量贡献) 由其与邻居的耦合决定:

$$E_i = -J_1 \sum_{j=1, j \neq i}^N S_i(t) S_j(t), \quad (7)$$

其中 N 为节点个数, J_1 表示两两相互作用的强度, $S_i(t)$ 和 $S_j(t)$ 分别表示节点 i 和 j 在 t 时刻的状态。

当引入三体相互作用时, 其能量表达式拓展为:

$$E_i = -J_1 \sum_{j=1, j \neq i}^N S_i(t) S_j(t) - J_2 \sum_{j=1, j \neq i}^N \sum_{k=j+1, k \neq i}^N S_i(t) S_j(t) S_k(t), \quad (8)$$

其中第一项为两体相互作用, 第二项为三体相互作用。在每个时间步, 每个节点 i 的自旋以如下概率发生翻转:

$$P_i^{E_i}(t) = \frac{1}{1 + e^{\Delta E_i/T}}, \quad (9)$$

其中 T 为温度, ΔE_i 为在 t 时刻翻转节点 i 自旋所引起的能量变化:

$$\Delta E_i^t = 2J_1 \sum_{j \in \xi_1} S_i(t) S_j(t) + 2J_2 \sum_{(j,k) \in \xi_2} S_i(t) S_j(t) S_k(t), \quad (10)$$

其中 ξ_1 为节点 i 的邻居集合, ξ_2 为所有包含节点 i 的节点对集合, 即所有满足 $j \neq i, k \neq i, j \neq k$ 的有序对。

4 算法设计与分析

本文提出了基于社区划分和关键节点的高阶网络重构框架 (HRCK), 旨在高效破解大规模高阶网络重构的计算瓶颈。该方法将全局组合爆炸问题分解为局部重构与全局补充两个高效阶段, 其核心流程主要包括以下四个步骤。第一阶段, 成对相互作用重构: 首先, 从观测到的时间序列数据中, 利用压缩感知技术重构出成对相互作用网络, 建立一个可靠的拓扑骨架。第二阶段, 社区划分: 基

于此骨架, 采用基于概率生成模型的社区发现算法 PCDMS [39], 划分出以三角形模体为结构基础的多个社区, 为后续的局部重构划定边界。第三阶段, 关键节点识别: 为捕捉稀疏的跨社区高阶相互作用, 本阶段设计了一种多维度权值函数, 从拓扑连接、动力学活跃度与模体参与性三个层面评估节点的跨社区交互潜力, 从而精准筛选出少数需要进行全局重构的关键节点。第四阶段, 高阶网络重构: 此为算法核心, 在各社区内部并行地执行高阶张量恢复, 解决大部分社区内的高阶连接。然后仅对已识别的关键节点, 执行面向全局的、精细的高阶相互作用搜索, 以查漏补缺。最后整合局部与全局的重构结果, 通过统计准则确定最终的二值化超图结构。HRCK 框架的重构流程如图 1 所示。

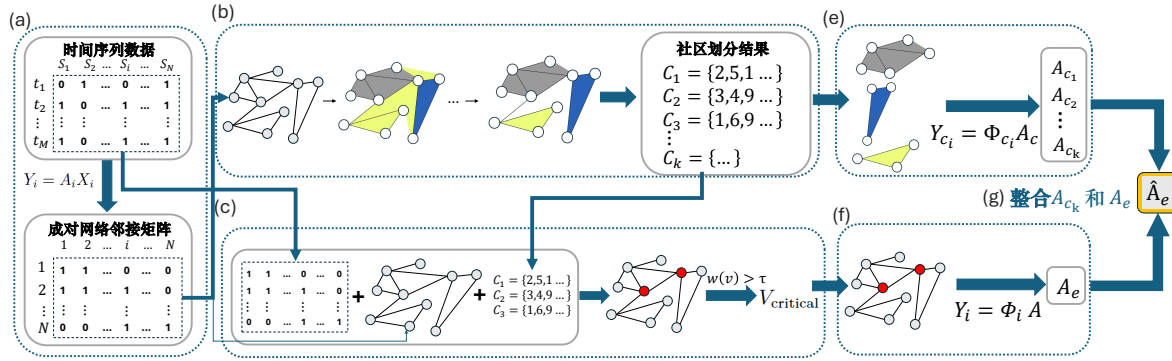


图 1 HRCK 框架重构高阶网络流程图。(a) 基于时间序列数据进行成对网络重构, 获得成对网络邻接矩阵; (b) 基于成对网络邻接矩阵得出成对网络拓扑, 使用 PCDMS 算法进行社区划分, 多次迭代后获得社区划分后的各个社区 c_k ; (c) 基于时间序列数据、成对网络邻接矩阵、社区划分结果进行关键节点识别, 求出关键节点集 $V_{critical}$; (d) 在社区内进行局部重构, 得出社区内的张量矩阵 A_{c_k} ; (e) 整合 A_{c_k} 和 A_e , 得到完整的全局张量矩阵 \hat{A}_e ; (f) 对关键节点进行全局重构, 求解全局张量矩阵 A_e 。

4.1 成对相互作用网络重构

在高级相互作用结构重构之前, 识别节点间的成对作用关系是构建复杂网络拓扑结构的重要前提。在多数现实系统中, 两两相互作用不仅构成网络的基础连接骨架, 也对高级相互作用的形成和传播具有决定性影响。因此, 在本文中, 我们首先从观测到的时间序列数据出发, 采用稀疏性驱动的压缩感知 (Compressed Sensing, CS) 算法 [40] 重构节点间的成对连接关系, 为后续高级结构的重构提供初始支撑。我们考虑一个由 N 个节点组成的动力系统, 其中每个节点 i 的状态记作 $x_i(t)$ 。节点之间的成对相互作用服从如下耦合动力学模型:

$$\dot{x}_i = f_i(x_i) + \sum_{j=1, j \neq i}^N A_{ij} g(x_j, x_i), \quad i = 1, \dots, N, \quad (11)$$

其中 $x_i \in \mathbb{R}^N$ 是节点 i 的状态向量, $f_i: \mathbb{R}^N \rightarrow \mathbb{R}^N$ 是节点 i 的自身动力学函数 (函数形式未知), $g(x_j, x_i)$ 为节点 i 与节点 j 的耦合函数, 本研究中取线性差分形式: $g(x_j, x_i) = x_j - x_i$, $A_{ij} \in \{0, 1\}$ 表示节点之间是否存在成对连接, $A_{ij} = 1$ 表示节点 i 和节点 j 之间存在连接, 否则不存在。我们假设观测到每个节点在时间区间 $[0, T]$ 上的状态时间序列 $x_i(t)$, 采样时间长度为 M , 并通过数值差

分近似估计其导数 $\dot{x}_i(t)$ 。邻接矩阵 A 可表示为：

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ \vdots & \ddots & & \vdots \\ a_{N,1} & a_{N,2} & \cdots & a_{N,N} \end{bmatrix}. \quad (12)$$

由于系统中真实的邻接矩阵 A 是未知的，我们将网络结构识别问题转化为一个稀疏回归问题。对每个节点 i ，将其在 M 个时间点上的状态导数向量记作观测向量 $Y_i \in \mathbb{R}^M$ ，其他节点在对应时间点的状态组成测量矩阵 $X_i \in \mathbb{R}^{M \times N}$ （对于 $j = i$ 的对角线元素，通常置为零或用于处理自身动力学），因此节点 i 的动力学方程可写为：

$$Y_i = A_i X_i + \varepsilon, \quad (13)$$

其中 $A_i \in \mathbb{R}^N$ 是待求解的、稀疏的节点 i 的连接权重向量， ε 为测量噪声， $i = 1, \dots, N$ 。求解邻接矩阵 A 的问题转化为对每个节点 i 独立求解一个稀疏回归问题，其优化目标函数为：

$$\min_A \|Y - AX\|_2^2 + \lambda \|A\|_1, \quad (14)$$

其中第一项为最小二乘拟合误差，第二项为 L_1 正则项，促进系数向量 A_i 的稀疏性； $\lambda > 0$ 为正则化参数，控制解的稀疏程度。根据最大化似然函数的方法可得两两拓扑矩阵，将其编码在矩阵 A 中。该问题属于标准的 LASSO 回归问题，可使用坐标下降法 [41] 或 FISTA [42] 等高效优化算法求解。

需要强调的是，压缩感知算法的有效性依赖于待重构信号的稀疏性前提。这一条件在复杂网络研究中具有坚实的理论与现实基础：大多数真实世界的网络（包括社交、生物、技术网络）都呈现出显著的稀疏特性，即节点的平均度远小于网络规模。在本文的合成网络实验中，我们通过控制网络生成参数明确保障了该条件：在合成数据实验中，我们使用稀疏生成模型（如 ER 模型中的低连接概率 p ，BA 模型中的低初始度 m_0 ），以确保生成网络的稀疏性满足压缩感知的理论要求。这一设计使我们能够在满足方法前提的条件下，有效验证所提重构框架的性能。除此之外，多数现实网络本身即表现出低连接密度、幂律度分布等稀疏特征。因此，压缩感知方法在这类系统中具有良好适应性和解释力。

4.2 基于社区划分的局部重构

在完成成对网络重构后，本文将社区划分思想引入高阶网络重构领域，首先将得到的成对网络划分为多个社区，并在各社区内部独立进行高阶重构，从而将全局组合爆炸问题分解为多个可处理的局部问题。为最大限度地保留成对网络中的高阶结构信息，本文利用基于概率生成模型的 PCDMS 算法 [39] 进行社区划分，该算法能够通过三元组模体捕捉社区内的高阶聚集特性。

目前成对网络上的主流社区划分方法包括基于模块度优化的方法 [3]、基于图划分的方法 [43]、基于随机游走的方法 [44]、基于标签传播的方法 [45]、基于统计推断的方法 [46]、基于深度学习的方法 [47]。然而，这些传统方法仅基于成对连接进行划分，忽略了高阶相互作用所承载的社群信息。高阶相互作用（如三元组）在真实网络中并非随机分布，而是高度聚集，形成“小团体偏好” [48]。因此，直接依据高阶结构的分布模式进行社区划分，有望得到更精确、更具功能意义的结果。由于高

阶网络存在“小团体偏好”、局部性等特性，其在成对网络上表现为节点以多元组的形式构成模体。在进行社区划分时可以识别这些模体结构，从而提取高阶信息。本文采用的社区发现算法 PCDMS，通过构建基于三角形基序（三阶超边）的概率生成模型，利用社区内三角结构的密集性作为划分依据。对比传统社区检测算法，该方法将基于二元连接的密度假设推广到高阶相互作用场景，这与真实网络中高阶相互作用（HOI）的“小团体偏好”特性（即低阶超边占比高）一致。通过三阶超边的概率建模，将三阶相互作用考虑到社区划分中，很大程度上能够保留部分高阶信息，避免仅依赖二元连接的拓扑损失，从而更有效地捕捉高阶相互作用在局部社区中的聚集特性。根据成对网络重构得到的拓扑结构 $G = (V, E)$ ，其中 V 和 E 分别代表节点和边，生成节点 u 和社区 C 之间的强度联系作为非负值 M_{uc} ， $M_{uc} = 0$ 表示 u 不是 C 的成员。 M 矩阵显示了每个节点和每个社区之间的依赖程度。随机初始化每个节点的社区归属矩阵 $M \in \mathbb{R}^{|V| \times k}$ ，每个元素 $M_{uc} > 0$ 表示节点 u 与社区 c 之间的关联程度。

随后构建基于模体的概率生成模型，该模型包含两个核心假设：一是一个三角形模体 (u, v_1, v_2) 在社区 c 中生成的概率，等于其两条边在给定社区条件下条件独立地生成的概率的乘积： $P_c(u, v_1, v_2) = P_c(u, v_1) \cdot P_c(u, v_2)$ ，其中 $(v_1, v_2) \in N(u)$ ， $N(u)$ 是节点 u 的邻居集合， P_c 为对应的概率模型。其次是一条边 (u, v) 存在于社区 c 中的概率，由两节点在该社区的归属强度决定，使用基于社区归属强度 M_{uc} 的 softmax 函数建模： $P_c(u, v) = \exp(-M_{uc}M_{vc}^T) / \sum_{v_i \in N(u)} \exp(-M_{uc}M_{v_i c}^T)$ ，我们的目标是找到社区归属矩阵 M ，以最大化图中所有观测到的三角形模体被生成的概率。定义图中存在的模体为 $L(M)$ ，通过最小化负对数似然函数，实现最大化网络中实际存在模体的对数似然概率：

$$L(M) = - \sum_{(v_1, v_2) \in N(u)} \log P(u, v_1, v_2). \quad (15)$$

对于单个节点 u ，通过最小化负对数似然函数，得到最终优化目标：

$$L(M_u) = M_u M_{v_1}^T + M_u M_{v_2}^T + 2 \sum_{v_i \in N(u)} \exp(-M_{uc} M_{v_i c}^T). \quad (4)$$

潜在变量 M 包含在负非线性似然函数中。为优化此非凸目标函数，使用块坐标下降（Block Coordinate Descent）[49]，依次将每个节点的社区归属向量 M_u 视为独立优化块。具体而言，每轮迭代依次遍历所有节点 u ，保持其邻居节点 v 的归属向量 M_v 固定，然后通过梯度下降（Gradient Descent）[50] 更新 M_u ：

$$\frac{\partial L(M_u)}{\partial M_u} = M_{v_1} + M_{v_2} - \frac{\sum_{v_i} M_{v_i} \exp(-M_u M_{v_i}^T)}{\sum_{v_i} \exp(-M_u M_{v_i}^T)}. \quad (16)$$

根据梯度更新规则：

$$M_u^{(t+1)} = \max \left(0, M_u^{(t)} - \eta \frac{\partial L(M_u)}{\partial M_u} \right), \quad (17)$$

在基于社区归属矩阵 M 的迭代优化过程中，节点的社区归属向量会随着迭代逐步调整，直至达到稳定状态。迭代次数直接影响算法的收敛性与社区划分的清晰度：迭代次数不足可能导致节点尚未找到最合适的社区归属，而过多迭代则可能引入噪声，使社区划分出现碎片化。为了量化迭代过程对社区划分效果的影响，我们将迭代次数记作 W ，并将其作为评估算法收敛性与划分质量的重要指

标。在迭代过程中， η 为学习率参数， \max 操作确保 M_u 的非负性。当所有节点的社区归属向量变化小于阈值 Δ 或达到最大迭代次数 T_{\max} 时，迭代停止。若 M_{uc} 超过阈值，则节点 u 被分配到社区 c 。值得注意的是，迭代次数 W 会显著影响算法的收敛性和最终社区划分质量：迭代过少可能导致社区结构模糊未收敛，而迭代过多则可能增加计算开销并产生过拟合，降低社区划分的泛化能力及后续高阶网络重构性能。

在完成基于 PCDMS 的社区划分后，在每个社区内独立进行高阶相互作用的局部重构。设划分得到的社区集合为 c_1, c_2, \dots, c_k ，在每个社区内基于时间序列数据与节点导数信息，构建社区内的局部张量矩阵：

$$Y_{c_i} = \Phi_{c_i} \mathcal{A}_c, \quad (18)$$

其中 Y_{c_i} 为社区 c 内节点 i 的时间序列数据向量， Φ_{c_i} 是社区内节点 i 的动力学在采样时间点上的离散表示，用于从数据中反推出交互结构， \mathcal{A}_c 为社区 c 的待求解张量矩阵，具体求解过程见 4.4 节。

4.3 基于关键节点的全局重构

对社区内的节点进行社区内的高阶重构具有局部性，会忽略真实网络中可能跨社区进行高阶交互的关键节点。为在大规模网络中有效捕捉潜在的跨社区高阶相互作用，我们设计了一种基于拓扑结构与动态活跃度联合建模的节点评分函数。该函数综合考虑了节点的跨社区连接潜力（PC）、动态活跃性（AC）与高阶模体参与性（MPS），分别从结构性参数、节点活跃度、模体参与数方面分析节点权值，形成一种可扩展且具普适性的高阶交互潜力评估机制。该机制不仅提升了高阶重构的性能，还有效降低了其计算复杂度。结构性参数用于刻画节点在成对网络拓扑结构中所处的位置及其“连接策略”的复杂性。其结构越重要、越靠近网络骨架，通常越可能成为高阶交互的“连接枢纽”。在网络结构性研究上，核心-边缘模型（Core-Periphery Model）[35] 和 k-core 分解理论 [51] 被广泛用于刻画网络中信息扩散和功能组织的中心结构。特别地，Chen 等 [52] 在高阶网络传播模型中发现，核心区域的节点在多节点同时激活模型（如同步刺激下的三节点激活）中具有更高的参与概率；Jin 等 [48] 在高阶社交网络中使用模拟实验指出，高阶集体事件（如会议共现）常围绕度大、核心度高的节点组织发生。节点活跃度衡量的是一个节点在时间序列中参与活动的频繁程度，其反映节点是否经常处于“动态活跃状态”，具备更高的参与高阶相互作用的机会。根据活跃度驱动的超图模型（Activity-driven Hypergraph Models）[53]，节点的时序活跃状态决定其参与高阶事件的概率。Cencetti 等 [21] 利用真实教学场景下的佩戴传感器数据发现，学生的“活跃度分布”高度相关于其在三人、四人同时互动中的出现频率。我们在进行社区划分时考虑了节点模体的影响，节点参与的三元、四元团（完全模体）反映其历史上所处的“高阶结构环境”。三角形或四元团越多，节点越可能是高阶交互的组成部分，丘成桐团队 [16] 的研究指出完全模体是高阶交互的主要构成。综上，节点参与高阶相互作用的可能性，既依赖其在网络结构中的地位，也与其动态活跃度密切相关，还受到其在模体结构中的参与程度的影响。因此，我们构造如下节点权值函数，用于量化节点的高阶潜力：

$$\mathbf{w}(v) = \alpha \cdot \text{PC}(v) + \beta \cdot \text{AC}(v) + \sum_{k=3}^K \gamma_k \cdot \text{MPS}_k(v), \quad (19)$$

其中 $PC(v)$ 为参与系数，代表结构性参数，可预测节点的跨社区潜能，即高参与系数的节点往往连接多个社区，具有潜在的跨团体交互属性； $AC(v)$ 为节点 v 的动力学活跃度，衡量节点在时间序列中参与活动的频繁程度，其反映节点是否经常处于“动态活跃状态”，其值越大，节点参与高阶相互作用的机会越高； $MPS_k(v)$ 为模体参与度，可以理解为节点 v 在 k 阶完全联通模体中的“存在密度”，反映节点潜在的高阶依赖结构，理论上可推广至任意 k （三元至 n 元模体）。各个参数的具体定义如下：

$$PC(v) = \frac{c_v}{C}, \quad (20)$$

其中 c_v 为节点 v 参与的社区数， C 为社区总数， $PC(v)$ 越接近 1 表明节点 v 连接越分散于多个社区，更可能出现在跨社区的高阶交互中。

$$AC(v) = \frac{\sum_{t=1}^M x_v(t)}{\sum_{v'=1}^N \sum_{t=1}^M x_{v'}(t)}, \quad (21)$$

其中 $x_v(t)$ 为节点 v 在时间步 t 的活跃值， M 为总时间步数。

$$MPS_K(v) = \frac{\text{节点 } v \text{ 所在的 } k\text{-团数量}}{\text{网络中 } k\text{-团总数}}. \quad (22)$$

可展开为：

$$MPS_K(v) = \sum_{r=3}^n \gamma_r \cdot M_v^{(r)}(v), \quad (23)$$

其中 $M_v^{(r)}(v)$ 为节点参与的 r 元完全子图数， n 为最大模体阶。在本文中取 $n = 3$ ，最大考虑到 3 阶模体，而在稀疏网络中 n 可提升至 4、5 甚至 k 元，取决于网络稀疏程度。因此，本文节点权值函数可表示为：

$$w(v) = \alpha \cdot PC(v) + \beta \cdot AC(v) + \sum_{r=3}^3 \gamma_r \cdot M_v^{(r)}(v). \quad (24)$$

在本工作中，我们设定 $\alpha = \beta = 0.3$ ， $\gamma_3 = 0.4$ ，以均衡各维度贡献。基于已有时间序列数据和划分出的社区结构及成对拓扑网络结构，我们计算节点 $i \in \mathbb{R}^N$ 的节点权值 $w(i)$ ，定义阈值参数 τ ，仅当 $w(i) > \tau$ 时，将节点标记为关键节点并执行全局补充性重构，以确保跨社区的高阶相互作用不被遗漏，实现大规模网络中兼顾性能与效率的高阶建模。

通过节点权值函数识别出关键节点后，对关键节点进行全局重构，恢复跨社区潜在的高阶相互作用结构。我们仅针对关键节点集合 V_{critical} 构建张量模型，以全局范围的节点状态数据为输入，识别关键节点与其跨社区邻域节点之间可能存在的高阶相互作用。对于每个关键节点 $i \in V_{\text{critical}}$ ，我们构建其对应的全局张量矩阵：

$$Y_i = \Phi_i \mathcal{A}. \quad (25)$$

其中 Y_i 为节点 i 的时间序列数据向量， Φ_i 是节点 i 的动力学在采样时间点上的离散表示， \mathcal{A} 为待求解的全局张量，由此重构出具有跨社区相互作用的关键节点的高阶作用，具体求解过程见 4.4 节。

4.4 高阶网络重构

在完成社区划分并识别关键节点后，本文分别在社区内进行局部高阶网络重构，并在关键节点上执行全局高阶相互作用重构。所使用的高阶网络重构算法为基于数据驱动的模式识别算法，其建立在稀疏回归模型上。在合成网络中，我们通过稀疏生成模型保证该性质，而在真实网络中，其低连接密度等稀疏特性同样使该算法适用。

对于通过成对和高阶相互作用耦合的 N 个节点动力系统，其重构技术工作原理如下：设系统包含 N 个节点，以固定间隔 Δt 采样，对系统进行持续时间 $T = m\Delta t$ 的观察，采集 $M + 1$ 个时间点的状态数据。在每个时间点，记录节点状态 $x_1(t), x_2(t), \dots, x_N(t)$, $x_i(t) \in \{0, 1\}$ ，形成时间序列数据矩阵。根据该数据矩阵，得出成对网络重构结果和社区划分结果。在划分得出 k 个社区后，对每个社区 c 内的节点进行局部高阶重构时，从数据矩阵中提取社区 c 内节点 i 的列向量组成子矩阵 Φ_c ，基于该子矩阵，以社区内 N_c 个节点构建张量 \mathcal{A}_c ，并通过求解张量以重构高阶相互作用。节点动力学方程表示为：

$$\dot{x}_i = f_i(x_i) + \sum_{d=1}^{D_c} \sum_{j_1, \dots, j_d=1}^{N_c} a_{ij_1 \dots j_d}^{(d_c)} g^{(d_c)}(x_i, x_{j_1}, \dots, x_{j_d}), \quad (26)$$

其中 $i = 1, \dots, N_c$, $x_i \in \mathbb{R}^{N_c}$ 为节点 i 的状态向量，其时间序列构成重构数据基础； $f_i: \mathbb{R}^{N_c} \rightarrow \mathbb{R}^{N_c}$ 是社区内节点的局部非线性动力学函数； $g^{(d_c)}: \mathbb{R}^{N_c \times (d_c+1)} \rightarrow \mathbb{R}^{N_c}$ 建模社区内 $d_c + 1$ 阶高阶相互作用； $\mathcal{A}_c^{(d_c)}$ 为编码 $d_c + 1$ 阶相互作用拓扑的张量，其分量 $a_{ij_1 \dots j_{d_c}}^{(d_c)}$ 仅在索引互异时非零，以避免自相互作用并保证张量唯一性。假设函数 f_i 和 $g^{(d_c)}$ 是已知的，可以给出以下等式：

$$z_i^m = f_i(x_i^m) + \sum_{d_c=1}^{D_c} \sum_{j_1, \dots, j_{d_c}=1}^{N_c} a_{ij_1 \dots j_{d_c}}^{(d_c)} g^{(d_c)}(x_i^m, x_{j_1}^m, \dots, x_{j_{d_c}}^m), \quad (27)$$

其中 $i = 1, \dots, N_c$, $m = m_{\min}, \dots, m_{\max}$, $m_{\min} \geq 0$, $m_{\max} \leq M$, x_i^m 是 $x_i(t_m)$ 的简写， z_i^m 是时间 t_m 处 x_i 的导数测量值， m_{\min} 和 m_{\max} 的取值决定重构性能。由此，定义残差矩阵：

$$y_i^m = z_i^m - f_i(x_i^m), \quad (28)$$

其中 $Y_{c_i} = [y_i(m_{\min})^T, \dots, y_i(m_{\max})^T]^T$ ，定义社区 c 内的张量矩阵 \mathcal{A}_c ，用于编码存储社区内的 d_c 阶相互作用。接着，利用 (18) 式构造社区内张量矩阵，即 $Y_{c_i} = \Phi_{c_i} \mathcal{A}_c$ 。这里 Φ_{c_i} 是社区内节点 i 的动力学在采样时间点上的离散表示，用于从数据中反推出交互结构，可由下式计算：

$$\Phi_{c_i} = \begin{bmatrix} g_{i,1}^{(1)}(m_{\min}) & \cdots & g_{i,i-1}^{(1)}(m_{\min}) & g_{i,i+1}^{(1)}(m_{\min}) & \cdots & g_{i,N_c}^{(1)}(m_{\min}) & g_{i,1,2}^{(2)}(m_{\min}) & \cdots & g_{i,N_c-D_c+1,\dots,N_c}^{(D_c)}(m_{\min}) \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ g_{i,1}^{(1)}(m_{\max}) & \cdots & g_{i,i-1}^{(1)}(m_{\max}) & g_{i,i+1}^{(1)}(m_{\max}) & \cdots & g_{i,N_c}^{(1)}(m_{\max}) & g_{i,1,2}^{(2)}(m_{\max}) & \cdots & g_{i,N_c-D_c+1,\dots,N_c}^{(D_c)}(m_{\max}) \end{bmatrix}, \quad (29)$$

其中 $g_{i,j_1,\dots,j_{d_c}}^{(d_c)}(m)$ 是对 $g^{(d_c)}(x_i^m, x_{j_1}^m, \dots, x_{j_{d_c}}^m)$ 的简写。为保证 Φ_{c_i} 满秩，需选择 $\widetilde{M} = (m_{\max} - m_{\min} + 1)n_{c_g} \geq \mathcal{H}_c$ ，其中 $\mathcal{H}_c = N_c - 1 + (N_c - 1)(N_c - 2) + \cdots + (N_c - 1) \cdots (N_c - D_c)$ 为非平凡耦合函数分量总数， n_{c_g} 为非零分量个数。此时可唯一确定 \mathcal{A}_c 。对每个社区执行上述操作，得到各社区的局部高阶重构结果： $\mathcal{A}_{c_1}, \mathcal{A}_{c_2}, \dots, \mathcal{A}_{c_k}$ 。

考虑跨社区相互作用时, 对所有节点的权值函数 $w(i)$ 进行判断, 当 $w(i) > \tau$ 时, 节点 i 被标记为关键节点, 并执行全局重构。此时: $\Phi_{c_i} \rightarrow \Phi$, $N_c \rightarrow N$, $\mathcal{A}_c \rightarrow \mathcal{A}$, $d_c \rightarrow d$, $D_c \rightarrow D$, $Y_{c_i} \rightarrow Y_i$, 并通过全局等式 (25) 式, 即 $Y_i = \Phi_i \mathcal{A}$, 来求解全局张量 \mathcal{A} 。由此重构出具有跨社区相互作用的关键节点的高阶作用。由于关键节点的全局重构使用了全部的时间序列数据, 其重构结果相对于仅使用部分数据的单个社区重构结果更为全面和可靠, 因此直接将关键节点的重构结果编码在全局张量矩阵 \mathcal{A} 中。

为整合局部社区重构与关键节点全局重构结果, 采用全局信息驱动的补丁整合策略。该策略优先采纳关键节点的全局重构结果, 以确保跨社区相互作用的准确性。需要特别说明的是, 本文所采用的 PCDMS 社区划分算法允许节点以不同强度属于多个社区, 即其输出结果本质上是具有重叠性的社区结构。这意味着, 同一个高阶相互作用 (如一个三元组) 可能被多个重叠社区所包含, 从而在多个社区的重构结果 \mathcal{A}_c 中出现。如果一个高阶相互作用在多个重叠社区中被独立地重构出来, 那么它存在的可信度更高。通过取平均, 我们获得了一个基于“社区共识 [54]”的稳健估计。因此, 全局张量矩阵 \mathcal{A} 的更新规则为:

$$\mathcal{A}_e = \begin{cases} \mathcal{A}_e, & \text{if } \mathcal{A}_e > 0, \\ \frac{1}{n} \sum_{k=1}^n \mathcal{A}_{c_{k_e}}, & \text{if } \mathcal{A}_e = 0 \text{ 且 } \mathcal{A}_{c_{k_e}} > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (30)$$

其中 e 表示超边 (例如三元组 (i, j, k)), \mathcal{A}_e 为该超边在全局重构结果中的置信度, $\mathcal{A}_{c_{k_e}}$ 为在社区 c_k 中包含的 (i, j, k) 三元组的局部高阶相互作用, $\mathcal{A}_e = 0$ 表示全局重构算法未能检测到该超边的存在证据, $\mathcal{A}_{c_{k_e}} \neq 0$ 表示社区 c_k 内局部重构中检测到了该超边。因此, 当 $\mathcal{A}_e = 0$ 且 $\mathcal{A}_{c_{k_e}} > 0$ 时, 我们使用所有检测到该超边的社区的重构结果的平均值 (即基于社区共识的概率估计) 来更新 \mathcal{A}_e 的值。完成概率张量 \mathcal{A} 的更新后, 需要将其转换为二值的高阶邻接矩阵 $\hat{\mathcal{A}}$, 以明确判定每条超边是否存在。为此, 我们通过优化贝叶斯信息准则 (BIC) 确定全局最优阈值 θ , 并采用如下二值化规则:

$$\hat{\mathcal{A}}_e = \begin{cases} 1, & \text{if } \mathcal{A}_e > \theta, \\ 0, & \text{otherwise,} \end{cases} \quad (31)$$

当 $\hat{\mathcal{A}}_e = 1$ 时, 表示超边 e 存在于最终重构出的超图中; 反之则不存在。

综上, 本文高阶网络重构框架包括四个主要步骤: (1) 基于压缩感知的成对网络重构; (2) 基于 PCDMS 的社区划分; (3) 基于多维权值的关键节点识别; (4) 分阶段的高阶相互作用重构与整合: 包括社区内局域重构和关键节点全局重构。前两个步骤为重构提供基础网络骨架与社区结构; 最后两个步骤为本文核心创新, 旨在解决全局重构计算瓶颈。算法 1 给出了 HRCK 框架包含的分阶段高阶网络重构算法从时间序列到最终高阶网络的逐步重构过程。

算法 1 分阶段高阶网络重构

Require: 状态时间序列矩阵 $\mathbf{S} \in \mathbb{R}^{M \times N}$, 最大交互阶数 D

Ensure: 重构的高阶邻接张量 $\hat{\mathbf{A}}$

```
1: 阶段 1: 成对相互作用重构
2:  $\mathbf{A} \leftarrow \text{压缩感知重构}(\mathbf{S})$  {建立拓扑骨架}
3: 阶段 2: 社区划分
4:  $\{C_1, C_2, \dots, C_k\} \leftarrow \text{PCDMS 社区发现}(\mathbf{A})$  {基于三角形模体的社区划分}
5: 阶段 3: 关键节点识别
6: for 每个节点  $v \in V$  do
7:    $PC(v) \leftarrow \frac{\text{包含节点 } v \text{ 的社区数量}}{k}$ 
8:    $AC(v) \leftarrow \frac{\sum_{t=1}^M S_v(t)}{\sum_{u=1}^N \sum_{t=1}^M S_u(t)}$ 
9:    $MPS_3(v) \leftarrow \frac{\text{包含节点 } v \text{ 的三角形数量}}{\text{网络中三角形总数}}$ 
10:   $w[v] \leftarrow \alpha PC(v) + \beta AC(v) + \gamma_3 MPS_3(v)$ 
11: end for
12:  $V_{\text{critical}} \leftarrow \text{TopK 节点}(W, k)$  {筛选跨社区交互潜力最高的  $k$  个节点}
13: 阶段 4: 高阶网络重构
14: 4.1 社区内局部高阶重构
15: for 每个社区  $C_i \in \{C_1, \dots, C_k\}$  do
16:    $\mathbf{S}_{C_i} \leftarrow \text{提取社区子序列}(\mathbf{S}, C_i)$ 
17:    $\mathcal{A}_{C_i} \leftarrow \text{高阶张量恢复}(\mathbf{S}_{C_i}, D)$  {并行执行}
18: end for
19: 4.2 关键节点全局高阶重构搜索
20: for 每个关键节点  $v \in V_{\text{critical}}$  do
21:    $\mathcal{A}_v \leftarrow \text{全局高阶搜索}(\mathbf{S}, v, D)$  {查漏补缺跨社区交互}
22: end for
23: 4.3 结果整合与二值化
24: for 每个潜在高阶交互  $e$  do
25:   if  $e$  包含关键节点 且在  $\mathcal{A}_v$  中被检测到 then
26:      $\mathcal{A}(e) \leftarrow \mathcal{A}_v(e)$  {优先采用全局重构结果}
27:   else if  $e$  完全位于某个社区  $C_i$  内 then
28:      $\mathcal{A}(e) \leftarrow \mathcal{A}_{C_i}(e)$  {采用局部重构结果}
29:   else
30:      $\mathcal{A}(e) \leftarrow 0$  {忽略其他情况}
31:   end if
32: end for
33:  $\hat{\mathbf{A}} \leftarrow \text{统计二值化}(\mathcal{A}, \theta)$  {基于统计准则确定最终超图结构}
34: return  $\hat{\mathbf{A}}$ 
```

4.5 复杂度分析

本小节从理论与实现层面, 对所提 HRCK 框架的时间与空间复杂度进行分析, 并与先进的全局重构方法进行对比, 以量化其效率优势。

4.5.1 时间复杂度

HRCK 框架的整体时间复杂度由若干顺序执行的阶段构成, 其主导项可分解如下:

基础预处理阶段: 包括成对网络重构 ($O(MN^2)$) 与社区检测 ($O(N\langle k \rangle^3)$)。鉴于真实网络的稀疏性 ($\langle k \rangle \ll N$), 此阶段为次主导项。

核心重构阶段: 此为算法效率的关键。相比传统全局重构方法必须处理的 $O(N^{D+1})$ 复杂度,

HRCK 通过分治策略实现了根本性降低:社区内局部重构:将全局问题分解为 k 个规模为 $O(|C_i|^{D+1})$ 的独立子问题。由于最大社区规模 $|C_{\max}| \ll N$, 此阶段复杂度远低于全局基准。关键节点全局补充:仅对稀疏的关键节点子集 V_{critical} 执行全局重构, 贡献 $O(|V_{\text{critical}}| \cdot N^D)$ 的复杂度。

因此, HRCK 的总时间复杂度可表述为:

$$O(MN^2 + N\langle k \rangle^3 + |C_{\max}|^{D+1} + |V_{\text{critical}}| \cdot N^D), \quad (32)$$

其中, $|C_{\max}|^{D+1}$ 与 $|V_{\text{critical}}| \cdot N^D$ 共同取代了传统的 $O(N^{D+1})$ 主导项。

结果整合与优化: 采用哈希表存储非零交互, 将整合复杂度从 $O(N^D)$ 降至与实际检测到的交互数量 $|\mathcal{H}|$ 成正比, 即 $O(|\mathcal{H}|)$, 在稀疏网络中此项可忽略。

加速比分析: HRCK 的理论加速比可近似为:

$$\frac{O(N^{D+1})}{O(|C_{\max}|^{D+1}) + O(|V_{\text{critical}}| \cdot N^D)}. \quad (33)$$

由于 $|C_{\max}| \ll N$ 且 $|V_{\text{critical}}| \ll N$, 该加速比是超线性的。此外, 各社区的局部重构及关键节点的全局重构任务彼此独立, 框架天然支持并行计算, 可进一步将阶段耗时降低至 $O(|C_{\max}|^{D+1})$ 与 $O(N^D)$ 。

4.5.2 空间复杂度

空间复杂度: 全局方法需要存储规模为 $O(N^D)$ 的稠密张量, 这在大规模场景下不可行。HRCK 通过以下方式显著降低了空间开销: 仅在社区内存储规模为 $|C_i|$ 的张量; 同时使用全局哈希表仅存储被检测到的非零高阶交互。因此, HRCK 的有效空间复杂度与实际检测到的交互数量 $|\mathcal{H}|$ 成正比, 即

$$O\left(\sum_{i=1}^k |C_i|^D\right) + O(|\mathcal{H}|), \quad (34)$$

这意味着 HRCK 在运行过程中对内存的需求远低于必须一次性处理整个全局张量的传统方法。

5 实验结果与分析

为了验证所提框架的重构性能, 本文在多个真实和合成的高阶网络上进行了系统性的仿真实验, 并与现有的先进重构方法进行了对比。

5.1 合成网络和真实网络的构造

本文采用 4 个真实高阶网络数据集和 2 个合成网络数据集进行实验, 真实网络数据集参数如表 1 所示:

表 1 真实网络数据集参数

网络	N	K_3	描述
Hypertext2009	85	33	ACM 超文本 2009 会议参与者面对面接触网络
Thiers12	156	154	Thiers 高中学生友谊网络
InVS15	211	64	研究机构工作场所接触网络
LyonSchool	222	62	小学学生日常接触网络

其中, N 为节点个数, K_3 (Number of 3-hyperedges) 代表网络中的三体连接个数。

人工合成的 2 个高阶网络为: ER 网络, 节点个数 $N = 300$, 连接概率 $p = 0.02$, 其上运行高阶 SIS 流行病模型。BA 网络, 节点个数 $N = 300$, 初始节点数为 3, 其上运行高阶 Ising 流行病模型。

我们使用 F1 分数来量化方法的重构性能, 其定义如下:

$$F_1 = \frac{2P \cdot R}{P + R},$$

其中

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}.$$

TP 、 FP 、 TN 、 FN 分别是真阳性、假阳性、真阴性和假阴性类别的数量。F1 分数越大, 重构性能越好, 当 $F_1 = 1$ 时, 代表重构结果与原始高阶网络结构完全相同, 误差为 0。

5.2 运行时间对比分析

计算效率是本工作的核心贡献之一。为了定量评估 HRCK 在实际运行时间上的优势及其可扩展性, 我们首先生成了五组不同规模的人工网络 (节点数 N 分别为 25、50、100、150、200、250)。在完全相同的硬件环境下, 我们将本文框架与两种最先进的高阶重构方法进行比较: 统计推断重构法 (HOR) [15] 和动力系统重构法 (DSR) [37], 实验结果如图 2 所示。结果表明, HRCK 在所有网络规模下均展现出显著的效率优势, 且随着网络规模的扩大, 其优势呈指数级增长。具体而言, 在 $N = 100$ 的中等规模网络中, HRCK 的加速比超过了 5 倍; 而当网络规模增至 $N = 250$ 时, HRCK 方法仅需 HOR 方法 15% 的运行时间。这一趋势与我们的理论分析完全一致。

效率提升机制分析: HRCK 的效率优势直接源自其分治策略有效解决了组合爆炸问题。社区内局部重构将原本昂贵的全局 $O(N^{D+1})$ 问题, 分解为多个规模仅为 $O(|C_i|^{D+1})$ 的、可并行处理的子问题。由于最大社区规模 $|C_{\max}| \ll N$, 此步骤实现了超线性的计算增益。同时, 通过精确采样对关键节点进行全局补充, 将全局搜索的范围从所有 N 个节点压缩至一个极小的关键节点子集 V_{critical} , 从而将该部分的计算开销限制在 $O(|V_{\text{critical}}| \cdot N^D)$ 。正是通过这种分阶段的重构策略, HRCK 成功规避了现有方法需要处理的随节点数超指数增长的组合空间, 从而实现了运行时间的大幅减少。实验结果表明, HRCK 框架是一种行之有效的解决大规模高阶网络重构可扩展性瓶颈的路径。

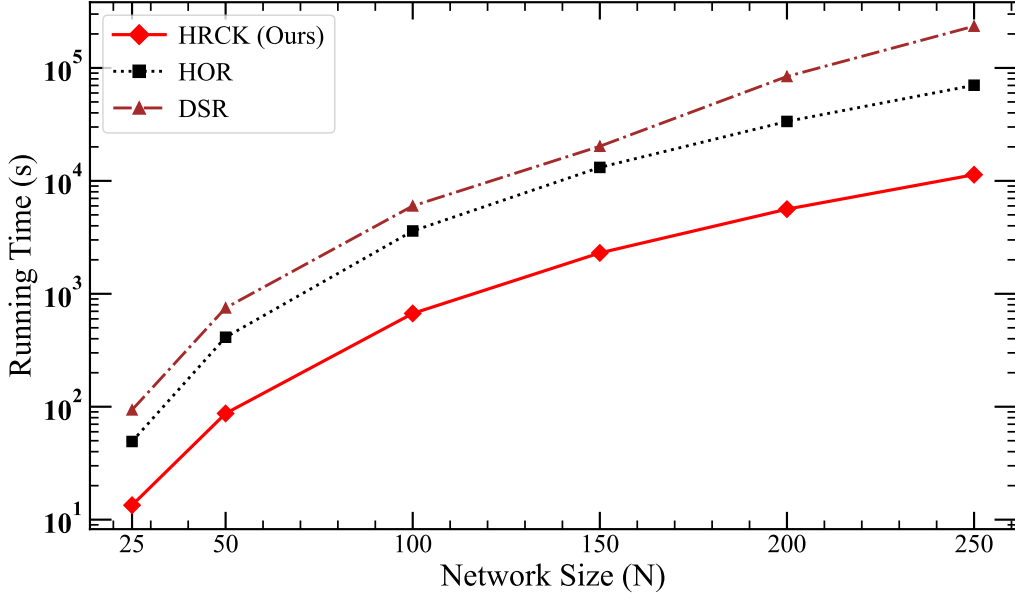


图 2 HRCK、HOR、DSR 在不同网络规模下的运行时间折线对比图。横坐标为节点个数，纵坐标为运行时间（单位为 10 的幂次方）。

5.3 重构性能对比实验

为评估所提框架的性能，我们在四个真实高阶网络（Hypertext2009、Thiers12、InVS15、Lyon-School）上进行了重构性能对比实验。如图 3(e)-(h) 所示，展示了 HRCK 与两种现有最先进方法——HOR [37] 和 DSR [15]——的对比。

在图 3(a)-(d) 中，可视化了每个网络的结构，其中蓝色边表示成对连接，黄色部分表示高阶相互作用。正如图中所示，Thiers12 和 LyonSchool 网络中的高阶连接相对集中，且在社区划分时能够保留更多的高阶信息，从而提升重构性能。对于图 3(a)-(d) 中的四个真实网络，我们分别选取了迭代次数为 70、60、80 和 50 次进行对比。我们重点分析了时间序列数据长度 T 对重构性能 (F1-Score) 的影响。实验结果显示，尽管在 T 较小的情况下，HRCK 的性能略低于 HOR 和 DSR 方法（见图 3(e)-(h) 左部），但当可获得的时间序列数据足够长时 ($T > 17000$)，HRCK 在所有测试的真实网络上均能达到与两种最先进方法相当甚至更优的重构性能 (F1-Score > 0.75)。这一发现有力地验证了 HRCK 在充分数据条件下所具备的重构性能优势。值得注意的是，在社区结构清晰、组织化程度较高的真实网络中（如 Thiers12），HRCK 展现出了独特的优势，其最终重构性能稳定地优于对比的两种现有方法。这一结果印证了本文方法的核心设计：通过利用网络固有的结构性先验，HRCK 能够更精准地捕捉高阶交互的局部聚集模式。考虑到大多数真实世界网络普遍具有一定的结构化特性，这一优势表明 HRCK 尤其适用于实际应用场景中的网络重构任务。

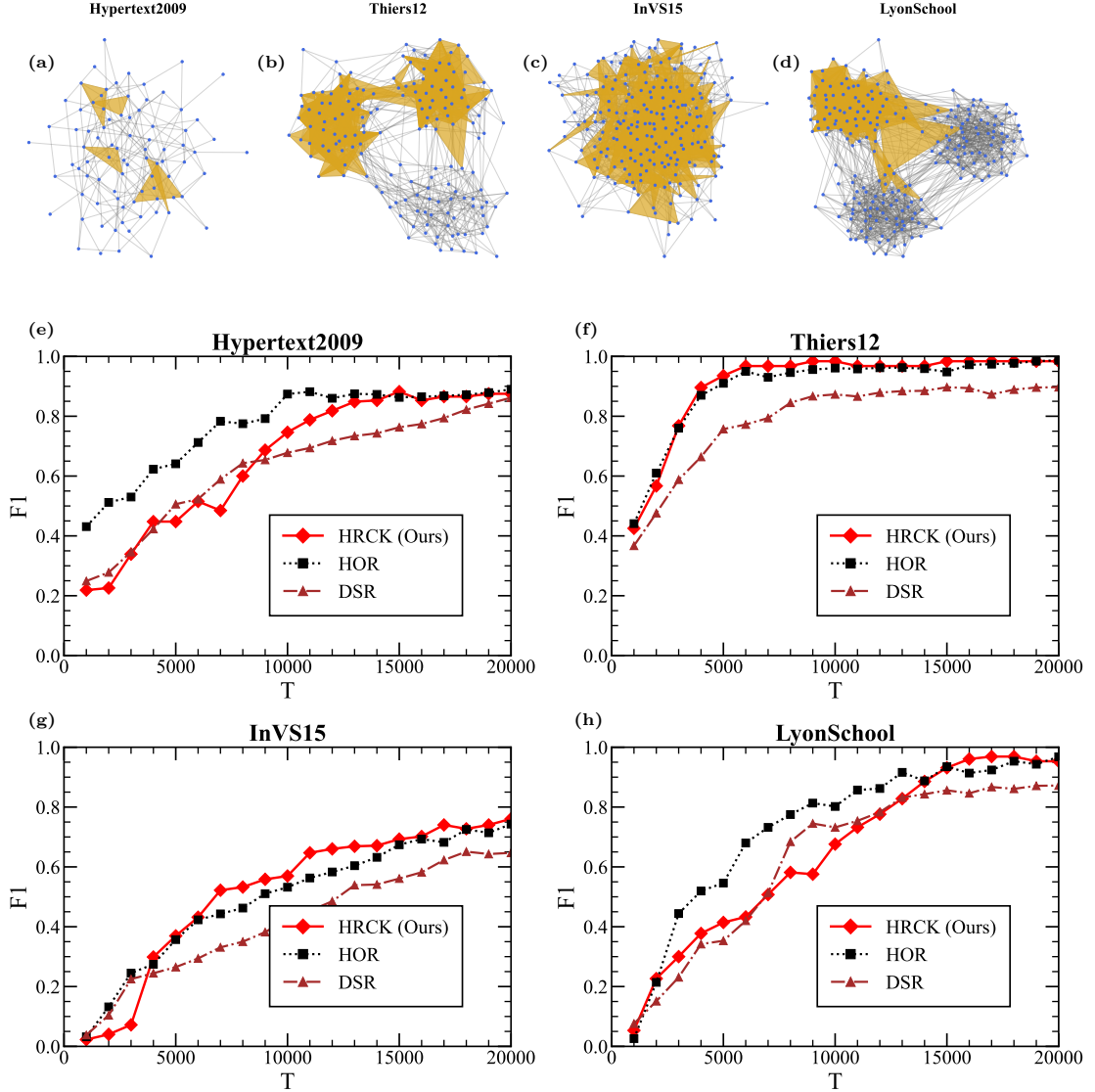


图 3 真实网络的可视化结构: (a): Hyperext2009, (b): Thiers12, (c): InVS15, (d) LyonSchool; (e)-(h) 为在四个真实世界网络 (Hyperext2009, Thiers12, InVS15, LyonSchool) 上, 三种高阶重构方法 HRCK、HOR 和 DSR 的重构性能 (F1-Score) 随时间序列数据长度 T 增加的变化趋势。

5.4 参数对重构性能的影响

为了确保公平对比, 上一小节中我们为每种网络选择了最优参数 (对于 HRCK, 即选择了最优的社区划分迭代次数)。为进一步考察 HRCK 框架在不同实验条件下的表现, 我们系统分析了时间序列长度和迭代次数对重构性能的影响, 并展示了在合成网络与真实网络中, 针对不同时间序列长度和迭代次数下的重构结果。图 4 给出了 4 种不同的真实网络和 2 种合成网络, 在不同时间步和迭代次数下的重构性能对比。根据图 4, 我们可以观察到以下特征: 随着时间序列长度 T 的增加, 重构性能逐渐提升, 且当 $T > 14000$ 时, 重构性能趋于稳定。不同的迭代次数对重构性能也有显著影响。当迭代次数过少时, 社区划分无法充分保存高阶信息, 导致跨社区的高阶交互增多, 社区内的

高阶相互作用也未能完整保留，进而使得 F1 分数较低（见图 4(a)-(f)）。随着迭代次数的增加，高阶信息得到了更完整的利用，但过多的迭代可能导致过拟合，社区划分过于细化，忽略了跨度较大的超边，进而导致重构性能下降。因此，横向像素点呈现出由蓝到黄再到蓝的变化，F1 分数先增大后减小。对于 4 个真实网络 Hypertext2009、InVS15、LyonSchool 和 Thiers12，重构结果在迭代次数分别为 70、80、50、60 次时趋于稳定；对于 SIS 流行病传播模型的 ER 网络和 Ising 动力学模型的 BA 网络，其最优迭代次数分别为 80 次和 80 次。

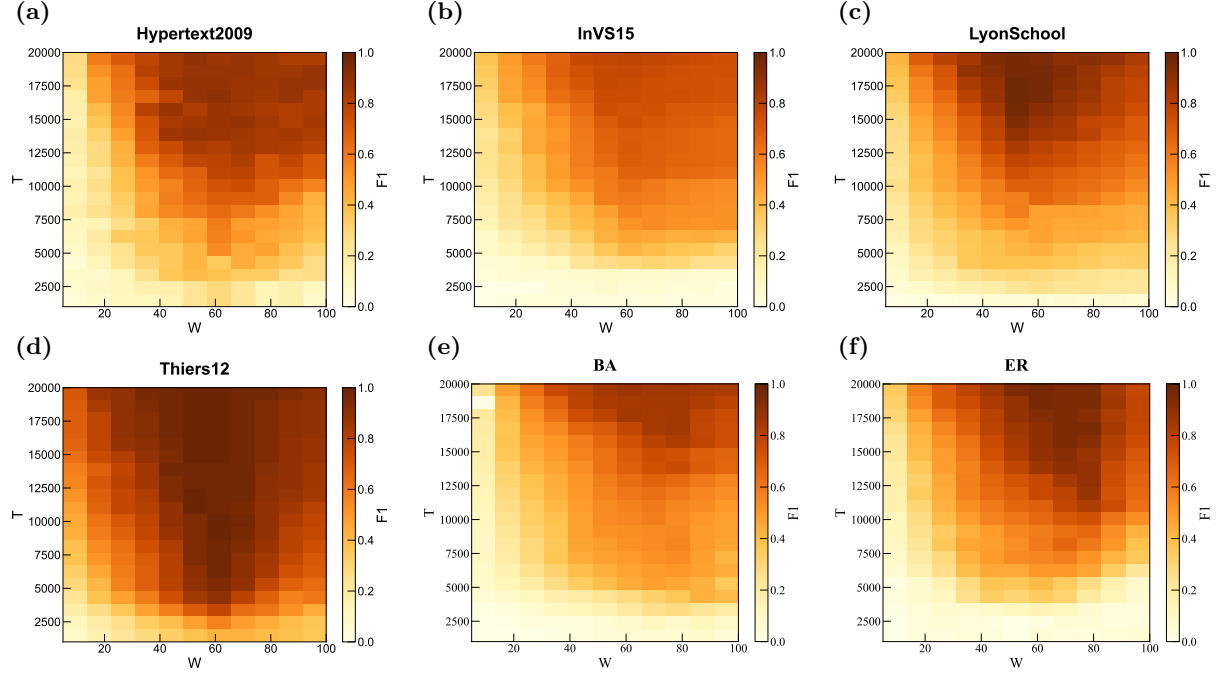


图 4 不同时间步和迭代次数下的重构性能对比。横坐标为迭代次数，纵坐标为时间序列数据长度，图中颜色深度代表 F1 分数。图 (a)-(d) 分别代表 Hypertext2009、InVS15、LyonSchool、Thiers12 网络；图 (e)-(f) 分别代表使用 SIS 和 Ising 动力学生成的 ER 网络和 BA 网络。

5.5 鲁棒性分析

为验证所提重构框架的鲁棒性，我们进行了抗干扰实验，以评估框架在面对随机扰动时的稳健性。具体而言，我们对时间序列数据矩阵 S 中的感染状态与易感状态值进行随机翻转，翻转的比例为 M 。我们研究了不同翻转比例 M 对重构性能的影响，实验结果如图 5 所示，涵盖了合成网络和真实网络。

从结果中可以看出，随着翻转比例 M 的增加，各个网络的 F1 分数均有所下降，且较低迭代次数的重构实验表现出更为明显的下降趋势。同时，在相同翻转比例 M 下，翻转感染状态对实验结果的影响显著大于翻转易感状态的影响。这是由于网络的稀疏性，在相同的翻转概率下，感染状态的翻转更可能影响社区划分结果，从而导致低阶信息利用率较低的划分结果。然而，即使在 30% 的感染状态翻转的情况下，高迭代次数的重构实验的 F1 分数仍能维持在 50% 以上，这充分证明了所提框架在应对数据中的随机波动时具备较强的鲁棒性。

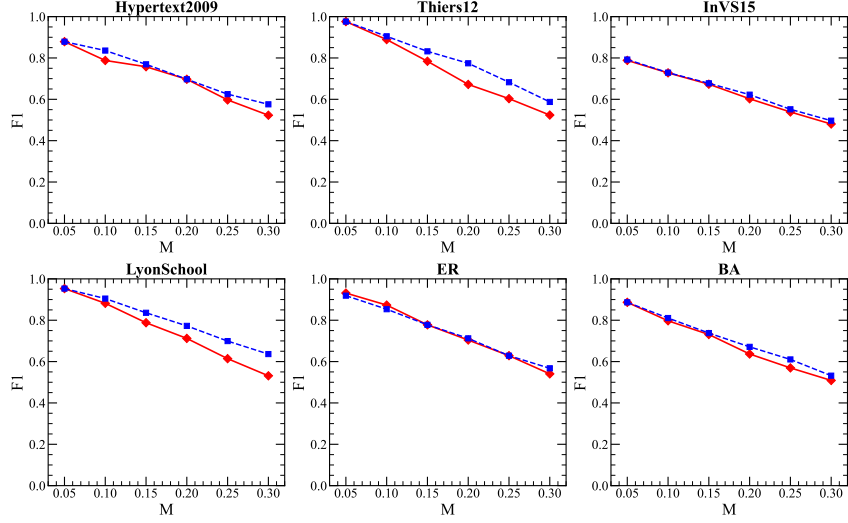


图 5 随机翻转对 F1 的影响。在 6 种网络上随机翻转 0 节点或 1 节点对重构性能的影响，其中 M 为翻转比例，红线为翻转 1 节点，蓝线为翻转 0 节点。

6 结论与未来展望

本文针对高阶网络重构中的计算瓶颈和性能问题，提出了一种基于社区划分和关键节点的高阶网络重构框架 HRCK。该框架的核心思想是将高阶交互的稀疏性与局部性特征转化为有效的分阶段算法设计原则，从而将复杂的全局重构问题有效地分解为一系列可处理的局部子问题。基于该框架的核心原则，我们设计了一种融合社区划分与多维权值函数关键节点识别的两阶段算法。该算法通过社区内局部重构与关键节点全局重构的协同机制，确保重构完整性的同时，显著提高了计算效率。理论分析和实验结果一致表明，本框架将计算复杂度从传统全局方法的 $O(N^{D+1})$ 主导项显著降低至 $O(|C_{\max}|^{D+1}) + O(|V_{\text{critical}}| \cdot N^D)$ 。在多个真实与合成网络上的实验结果表明，HRCK 显著降低了运行时间，并取得了比现有先进方法更优的重构性能，特别是在具有明确社区结构的真实网络中。综上所述，本研究构建了一个高效且精准的分阶段重构框架，通过社区内局部重构与关键节点全局重构的协同机制有效解决了高阶网络重构中的可扩展性难题，并同步优化了重构性能，为分析大规模复杂系统中的高阶相互作用提供了可行的思路。

尽管本研究提出的分阶段重构框架在计算效率方面取得了进展，但仍存在以下局限性，这些局限性为未来的研究指明了方向。首先，社区划分质量的依赖性：分阶段策略的有效性在启动阶段依赖于社区划分的质量。当前框架在社区结构明确且局部稠密的网络中表现尤为出色，但对于社区边界模糊或具有多重层次结构的复杂网络，其鲁棒性可能下降。开发自适应多尺度社区发现或重叠社区检测算法，以根据网络固有特性动态调整划分粒度，是提升框架普适性的关键。其次，对数据质量与完备性的敏感度：作为动力学驱动的重构框架，其性能与时间序列的长度 T 、分辨率以及信噪比密切相关。尽管该框架降低了对数据总量的需求，但在数据极度稀疏或含有高强度噪声的现场景下，仍可提升性能。未来可考虑开发针对不完整与噪声数据的鲁棒推断模型，或引入主动学习机制以智能地优先采集信息量最大的数据。

参考文献

- 1 Singh H P, Singh H T, Earnest V J. Correlation networks from random walk time series. *Physical Review E*, 2018, 98: 032311
- 2 Faith J J, Hayete B, Thaden J T, et al. Large-scale mapping and validation of *escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biology*, 2007, 5: e8
- 3 Clauset A, Moore C, Newman M E. Hierarchical structure and the prediction of missing links in networks. *Nature*, 2008, 453: 98–101
- 4 Jiang C, Wu W. Anomaly detection in distributed systems based on spatio-temporal causal inference. *Engineering Applications of Artificial Intelligence*, 2025, 160: 111857. URL <https://www.sciencedirect.com/science/article/pii/S0952197625018597>
- 5 Fan Z, Wu X. Identifying partial topology of simplicial complexes. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2022, 32: 113128
- 6 Shen Z, Wang W X, Fan Y, et al. Reconstructing propagation networks with natural diversity and identifying hidden sources. *Nature Communications*, 2014, 5: 4323
- 7 Liu Y, Huang K, Yang C, et al. Distributed network reconstruction based on binary compressed sensing via admm. *IEEE Transactions on Network Science and Engineering*, 2023, 10: 2141–2153
- 8 Schreiber T. Measuring information transfer. *Physical Review Letters*, 2000, 85: 461–464
- 9 Stam C J, Van Straaten E C W. The organization of physiological brain networks. *Clinical Neurophysiology*, 2012, 123: 1067–1087
- 10 Faust K, Raes J. Microbial interactions: from networks to models. *Nature Reviews Microbiology*, 2012, 10: 538–550
- 11 Eagle N, Pentland A S, Lazer D. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 2009, 106: 15274–15278
- 12 Arenas A, Díaz-Guilera A, Kurths J, et al. Synchronization in complex networks. *Physics Reports*, 2008, 469: 93–153
- 13 Iacopini I, Petri G, Barrat A, et al. Simplicial models of social contagion. *Nature Communications*, 2019, 10: 2485
- 14 Gavin A C, Aloy P, Grandi P, et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 2006, 440: 631–636
- 15 Monti F, Calder A, Lombardi V G, et al. Reconstructing higher-order interactions in coupled dynamical systems. *Nature Communications*, 2024, 15: 5184
- 16 Feng L, Gong H, Zhang S, et al. Hypernetwork modeling and topology of high-order interactions for complex systems. *Proceedings of the National Academy of Sciences*, 2024, 121: e2412220121
- 17 Timme M, Casadiego J. Revealing networks from dynamics: an introduction. *Journal of Physics A: Mathematical and Theoretical*, 2018, 51: 124001
- 18 Fu S, Liu W, Tao D, et al. Hsegcn: Hessian graph convolutional networks for semi-supervised classification. *Information Sciences*, 2020, 514: 484–498
- 19 Feng Y, You H, Zhang Z, et al. Hypergraph neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33: 3558–3565
- 20 Battiston F, Cencetti G, Iacopini I, et al. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 2020, 874: 1–92
- 21 Lotito Q F, Montresor A, Battiston F. Multiplex measures for higher-order networks. *Applied Network Science*, 2024, 9: 55
- 22 Paninski L. Estimation of entropy and mutual information. *Neural Computation*, 2003, 15: 1191–1253
- 23 Pastor-Satorras R, Castellano C, Van Mieghem P, et al. Epidemic processes in complex networks. *Reviews of Modern Physics*, 2015, 87: 925–979
- 24 Liu Y Y, Barabási A L. Control principles of complex systems. *Reviews of Modern Physics*, 2016, 88: 035006
- 25 Bassett D S, Sporns O. Network neuroscience. *Nature Neuroscience*, 2017, 20: 353–364
- 26 Han X P, Shen Z X, Wang W X, et al. Robust reconstruction of complex networks from sparse data. *Physical Review Letters*, 2015, 114: 028701
- 27 Palla G, Barabási A L, Vicsek T. Quantifying social group evolution. *Nature*, 2007, 446: 664–667
- 28 Pearson K. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 1895, 58: 240–242
- 29 Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*, 1978, 6: 461–464
- 30 Montoya J M, Pimm S L, Solé R V. Ecological networks and their fragility. *Nature*, 2006, 442: 259–264
- 31 Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks. *Nature Physics*, 2010, 6: 888–893
- 32 Benson A R, Gleich D F, Leskovec J. Higher-order organization of complex networks. *Science*, 2016, 353: 163–166

- 33 Grilli J, Barabás G, Michalska-Smith M J, et al. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, 2017, 548: 210–213
- 34 Zang Y, Fan Z, Wang Z, et al. Stepwise reconstruction of higher-order networks from dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2024, 34: 073156
- 35 Terry J C D, Bonsall M B, Morris R J. The impact of structured higher-order interactions on ecological network stability. *Theoretical Ecology*, 2025, 18: 9
- 36 Grigor'yan A, Lin Y, Muranov Y, et al. Homotopy theory for digraphs. *Pure and Applied Mathematics Quarterly*, 2014, 10: 619–674
- 37 Wang H, Ma C, Chen H S, et al. Full reconstruction of simplicial complexes from binary contagion and ising data. *Nature Communications*, 2022, 13: 3043
- 38 Stauffer D. Social applications of two-dimensional ising models. *American Journal of Physics*, 2008, 76: 470–473
- 39 Hajibabaei H, Seydi V, Koochari A. A motif-based probabilistic approach for community detection in complex networks. *Journal of Intelligent Information Systems*, 2024, 62: 1285–1303
- 40 Yang J, Yang G. Network reconstruction based on compressive sensing. In: 2015 34th Chinese Control Conference (CCC). Hangzhou, China 2015. 2123–2128
- 41 Lin Y, Shames I, Nešić D. Online convex optimization using coordinate descent algorithms. *Automatica*, 2024, 165: 111681
- 42 Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2009, 2: 183–202
- 43 Fiedler M. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 1973, 23: 298–305
- 44 Pons P, Latapy M. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 2006, 10: 191–218
- 45 Raghavan U N, Albert R, Kumara S. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 2007, 76: 036106
- 46 Ma C, Chen H S, Lai Y C, et al. A statistical inference approach to structural reconstruction of complex networks from binary time series. *Physical Review E*, 2018, 97: 022301
- 47 Ding X, Kong L W, Zhang H F, et al. Deep-learning reconstruction of complex dynamical networks from incomplete data. *Chaos*, 2024, 34: 043115
- 48 Benson A R, Gleich D F, Leskovec J. Higher-order organization of complex networks. *Science*, 2016, 353: 163–166
- 49 Xu Y, Yin W. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 2013, 6: 1758–1789
- 50 Robbins H, Monro S. A stochastic approximation method. *The Annals of Mathematical Statistics*, 1951, 22: 400–407
- 51 Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks. *Nature Physics*, 2010, 6: 888–893
- 52 Chen W, Teng S H, Zhang H. On the equivalence between high-order network-influence frameworks: General-threshold, hypergraph-triggering, and logic-triggering models. *arXiv preprint arXiv:2012.00355*, 2020. URL <https://arxiv.org/abs/2012.00355>
- 53 Petri G, Barrat A. Simplicial activity driven model. *Physical Review Letters*, 2018, 121: 228301
- 54 Kittler J, Hatef M, Duin R P, et al. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20: 226–239

High-order network reconstruction based on community division and key nodes

Jianbo Wang^{1,2,3*}, Jiang Rui¹, Zhanwei Du^{4,3}, Ping Li¹ & Xiaoke Xu^{5,6*}

1. *School of Computer Science and Software Engineering, Southwest Petroleum University, Chengdu 610500, China;*
2. *The Intelligent Policing and National Security Risk Management Laboratory, Sichuan Police College, Luzhou 646000, China;*
3. *School of Public Health, The University of Hong Kong, Hong Kong 999077, China;*
4. *School of Public Health and Emergency Management, Southern University of Science and Technology, Shenzhen*

518055, China;

5. *Computational Communication Research Center, Beijing Normal University, Zhuhai 519087, China;*

6. *School of Journalism and Communication, Beijing Normal University, Beijing 100875, China*

* Corresponding author. E-mail: jianbow2021@gmail.com, xuxiaoke@foxmail.com

Abstract Higher-order network reconstruction is a key problem in understanding the interaction mechanisms of complex system populations. However, existing methods often rely on global search strategies, leading to combinatorial explosion, which in turn results in high time complexity, while there is still room for improvement in reconstruction performance. To address this, this paper proposes a higher-order network reconstruction framework based on community partitioning and key node identification. The framework leverages the sparsity and locality priors of higher-order interactions to transform the global reconstruction problem into a computational paradigm that combines local community reconstruction with global key node reconstruction, significantly reducing the size of the search space. This framework is realized through the design of an efficient staged reconstruction algorithm. First, the pairwise interaction network is reconstructed based on compressive sensing theory. Then, local reconstruction is performed for each community through community partitioning, and critical nodes are accurately identified using a multidimensional weight function. Next, only the critical nodes undergo global supplementary reconstruction, significantly improving computational efficiency while ensuring reconstruction completeness. Finally, the local and global reconstruction results are integrated and binarized to obtain the final higher-order network reconstruction. Theoretical analysis demonstrates that, compared to existing state-of-the-art methods, the proposed framework significantly reduces computational complexity. Experimental validation shows that the framework not only reduces runtime substantially but also improves reconstruction performance on a variety of real-world and synthetic higher-order networks, compared to current state-of-the-art methods. The proposed framework offers a feasible solution for the efficient reconstruction of large-scale higher-order networks, balancing both performance and efficiency.

Keywords High order networks, network reconstruction, community division, compressive sensing, key nodes