使用优化理论进行网络重建的通用方法研究

姓名: 郭剑华　　学号: 2410103039

摘要

复杂网络的结构重建是理解和控制复杂系统的重要前提。本文基于 Li 等人在《Physical Review E》（2017）中提出的通用数据驱动方法，系统研究了具有二元状态动态的复杂网络重建问题。该方法通过线性化技术和凸优化，将复杂的非线性、分段或随机动态转化为稀疏信号重建问题，无需依赖具体的动态函数形式。研究方法包括基于归一化汉明距离的二元状态数据处理、数据驱动的线性化技术以及基于 Lasso 的网络拓扑重构。实验在 Erdős Rényi 随机网络、无标度网络、小世界网络及多个真实网络上验证了该方法的性能，表明其在噪声和数据缺失条件下仍具有高准确性和鲁棒性。本报告详细阐述了研究背景、方法、理论验证、实验结果及局限性，并探讨了未来研究方向，为复杂网络分析提供了理论与实践参考。

关键词：凸优化；复杂网络；优化理论；网络重构

1　引言

复杂网络在自然、技术和社会系统中广泛存在，其结构和动态特性对系统的功能和行为起着关键作用。从生物体内的基因调控网络、神经元网络，到社会中的社交网络、经济网络，再到技术领域的互联网、电力网络等。这些系统的一个共同特点是它们都由大量的个体（节点）组成，这些个体之间通过复杂的相互作用（边）形成一个整体。理解这些复杂系统的结构和动态特性对于揭示系统的功能、预测其行为以及设计有效的控制策略至关重要。

Li 等人（2017）提出了一种通用的数据驱动方法，专注于具有二元状态动态的复杂网络重建。该方法不依赖于动态函数的先验知识，仅利用节点的时间序列数据，通过线性化技术和凸优化技术（如 Lasso）将网络重建问题转化为稀疏信号重建问题。该方法的创新性在于其通用性，能够处理线性、非线性、分段或随机动态，适用于多种复杂系统，如神经网络、传染病传播、进化博弈和社交观点传播等。

本报告基于 Li 等人的研究，全面总结该方法的理论框架、实现步骤、实验验证及局限性。报告结构如下：第 2 节介绍研究背景，第 3 节详细阐述研究方法，第 4 节分析实验结果，第 5 节讨论方法优势与局限性，第 6 节总结并展望未来研究方向。

2　研究背景

复杂网络作为一种由大量节点和边构成的系统，其拓扑结构与动态特性对系统的整体行为具有决定性影响。在生物系统中，基因调控网络通过基因间的复杂相互作用调控生命活动，如细胞分化与代谢过程；在社会系统中，观点传播网络塑造群体行为与舆论动态，影响社会

决策与文化演变；在技术系统中，电力网络或通信网络的拓扑结构直接决定了系统的稳定性与抗干扰能力。二元状态动态是复杂网络研究中的核心模型之一，其特点是每个节点可在两种状态间切换，例如神经网络中的活跃与非活跃状态、进化博弈中的合作与背叛策略、传染病传播中的易感与感染状态，以及社交网络中的意见采纳或拒绝等 [1]。这种动态模型广泛应用于描述神经网络活动、流行病扩散、群体合作行为及社会信息传播等现象，为理解复杂系统的涌现行为提供了重要视角 [2]。由于二元状态动态的复杂性和多样性，准确重建网络结构成为复杂性科学与网络科学领域的一个关键挑战。

网络重建的目标是从可观测的时间序列数据中推断节点间的连接关系，即重构网络的邻接矩阵 [3]。这一过程对于揭示网络的拓扑特性、理解动态行为的机制以及设计有效的系统控制策略至关重要。然而，二元状态动态的网络重建面临多重困难。首先，节点的切换概率通常依赖于其邻居的状态，这种依赖关系可能表现为线性、非线性、分段甚至随机函数，且在实际应用中这些函数的具体形式往往未知，极大地增加了重建的复杂性。其次，网络的结构信息通常以隐含的方式嵌入在二元状态数据中，而解空间的维度极高，使得传统的枚举方法在计算上不可行 [4]。此外，现实数据常伴随测量噪声、数据缺失或动态过程中的随机效应，这些因素进一步加剧了重建的难度。已有研究中，尽管一些方法（如基于压缩感知的网络重建技术）在特定场景下取得了一定进展，但这些方法通常假设动态函数已知或仅适用于特定类型的动态过程，缺乏针对二元状态动态的通用性和鲁棒性 [5]。

因此，开发一种不依赖动态函数先验知识、能够从有限且可能存在噪声的二元状态时间序列中高效、鲁棒地重建复杂网络结构的通用方法，具有重要的理论与实际意义。

3 研究方法

Li 等人提出了一种基于数据的通用方法，通过线性化技术和凸优化解决二元状态动态网络的重建问题。以下为方法的核心步骤和理论基础。

3.1 问题描述

该论文针对复杂网络中二值状态动态的逆问题，即从可观测的二值状态时间序列数据中重建网络结构和动态过程，提出了一种通用的数据驱动方法。复杂网络的节点具有二值状态（如 0 表示非活跃，1 表示活跃，例如神经元网络中的活跃/非活跃、博弈网络中的合作/背叛、或传染病网络中的感染/健康等）。

考虑一个由 N 个节点组成的复杂网络，节点 i 在时间 t 的状态为 $s_i(t) \in \{0,1\}$。节点的状态转换概率依赖于其度 k_i（邻居数量）和活跃邻居数量 $m_i(t) = \sum_{j=1,j\neq i}^{N} a_{ij}s_j(t)$，其中 a_{ij} 为邻接矩阵元素（$a_{ij}=1$ 表示节点 i 和 j 存在连接，否则 $a_{ij}=0$）。节点状态的切换概率由两个函数 $F(m,k)$ 和 $R(m,k)$ 决定，其中 k 为节点度数，m 为活跃邻居数量，这两个函数分别表示节点从状态 0 切换到 1 以及从 1 切换到 0 的概率。这些函数可能呈现线性、非线性、分段或随机特性，涵盖了如投票模型（Voter Model，$F(m,k)=\frac{m}{k}$，$R(m,k)=\frac{k-m}{k}$）、Ising 模型、SIS 传染病模型、博弈模型等多种动态过程。

然而重建网络结构的挑战在于以下几点：首先，切换概率函数 F 和 R 的具体形式通常未知，导致难以直接推导网络拓扑；其次，网络结构信息隐藏在高维二值状态数据中，解空间维度极高（例如，网络邻接矩阵 $\{a_{ij}\}$，其中 $a_{ij}=1$ 表示节点 i 和 j 相连，否则为 0），使

得暴力枚举所有可能网络配置在计算上不可行；此外，数据可能受到测量噪声、缺失数据或随机效应的干扰，进一步增加重建难度。具体来说，节点 i 在时间 t 从状态 0 切换到 1 的概率为

$$P_i^{01}(t) = F\left(m_i(t), k_i\right) = F\left(\sum_{j=1, j\neq i}^{N} a_{ij} s_j(t), k_i\right),$$

其中 $s_j(t)$ 为节点 j 在时间 t 的状态，$m_i(t) = \sum_{j=1,j\neq i}^{N} a_{ij} s_j(t)$ 为活跃邻居数量，但 k_i、$P_i^{01}(t)$ 和 F 的形式均未知，阻碍了邻接矩阵 $\{a_{ij}\}$ 的推导。因此论文的目标是开发一种无需预先了解切换函数形式、仅依赖二值时间序列数据、能够高效且鲁棒地应对噪声和数据缺失的重建框架，通过将网络重建问题转化为稀疏信号重建问题，利用凸优化方法（如 lasso）实现高精度网络拓扑推断。

Model	$F(m,k)$	$R(m,k)$
Voter [13]	$\frac{m}{k}$	$\frac{k-m}{k}$
Kirman [49]	$c_1 + dm$	$c_2 + d(k-m)$
Ising Glauber [17,50]	$\frac{1}{1+e^{\frac{\beta}{k}(k-2m)}}$	$\frac{e^{\frac{\beta}{k}(k-2m)}}{1+e^{\frac{\beta}{k}(k-2m)}}$
SIS [14]	$1-(1-\lambda)^m$	μ
Game [3]	$\frac{1}{\alpha+e^{\frac{\beta}{k}[(a-c)(k-m)+(b-d)m]}}$	$\frac{1}{\alpha+e^{\frac{\beta}{k}[(c-a)(k-m)+(d-b)m]}}$
Language [51]	$s(\frac{m}{k})^\alpha$	$(1-s)(\frac{k-m}{k})^\alpha$
Threshold [52]	$\begin{cases}0 & \text{if } m \leqslant M_k \\ 1 & \text{if } m > M_k\end{cases}$	0
Majority vote [53]	$\begin{cases}Q & \text{if } m < k/2 \\ 1/2 & \text{if } m = k/2 \\ 1-Q & \text{if } m > k/2\end{cases}$	$\begin{cases}1-Q & \text{if } m < k/2 \\ 1/2 & \text{if } m = k/2 \\ Q & \text{if } m > k/2\end{cases}$

图 1. 不同二元状态动态模型的状态转换函数

3.2 数据驱动的线性化

为克服 F 未知的难题，该论文提出了一种数据驱动的线性化方法，用于从复杂网络的二值状态时间序列中重建网络结构，核心在于将未知的切换概率函数线性化，从而将网络重建问题转化为可通过凸优化解决的稀疏信号重建问题。具体过程如下：对于任意节点 i，其在时间 t 从状态 0 切换到 1 的概率为

$$P_i^{01}(t) = F\left(m_i(t), k_i\right) = F\left(\sum_{j=1, j\neq i}^{N} a_{ij} s_j(t), k_i\right),$$

其中 $m_i(t) = \sum_{j=1,j\neq i}^{N} a_{ij} s_j(t)$ 表示活跃邻居数量，k_i 为节点度数，$s_j(t)$ 为节点 j 在时间 t 的状态，a_{ij} 为邻接矩阵元素。由于 F 的形式未知，论文通过合并过程（merging process）将 F 线性化为

$$F \approx c_i \sum_{j=1, j\neq i}^{N} a_{ij} s_j(t) + d_i,$$

其中 c_i 和 d_i 为与节点 i 相关的常数。为实现这一线性化，方法首先筛选出所有 $s_i(t)=0$ 的时间步，因为只有这些时刻包含关于 $P_i^{01}(t)$ 的信息。然后，通过构造一个网络，节点表示 $s_i(t)=0$ 时其他节点的二值状态序列 $s_{-i}(t)$，边权为序列间的归一化汉明距离（Hamming distance），并设定阈值 Δ 删除小权边，保留高连接度的顶点（通过阈值 σ 筛选），再从中选取度数最小的 M 个顶点作为基础字符串（base strings），确保基础字符串既有足够的相似字符串，又彼此差异较大。对于每个基础字符串，收集与其汉明距离小于 Δ 的从属字符串，计算从属字符串中节点状态的平均值 $\langle s_j(t)\rangle$ 和下一时刻节点 i 状态的平均值 $\langle s_i(t+1)\rangle$，利用大数定律近似 $P_i^{01}(t)\approx\langle s_i(t+1)\rangle$。由此得到线性关系

$$\langle s_i(t+1)\rangle \approx c_i \sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(t)\rangle + d_i.$$

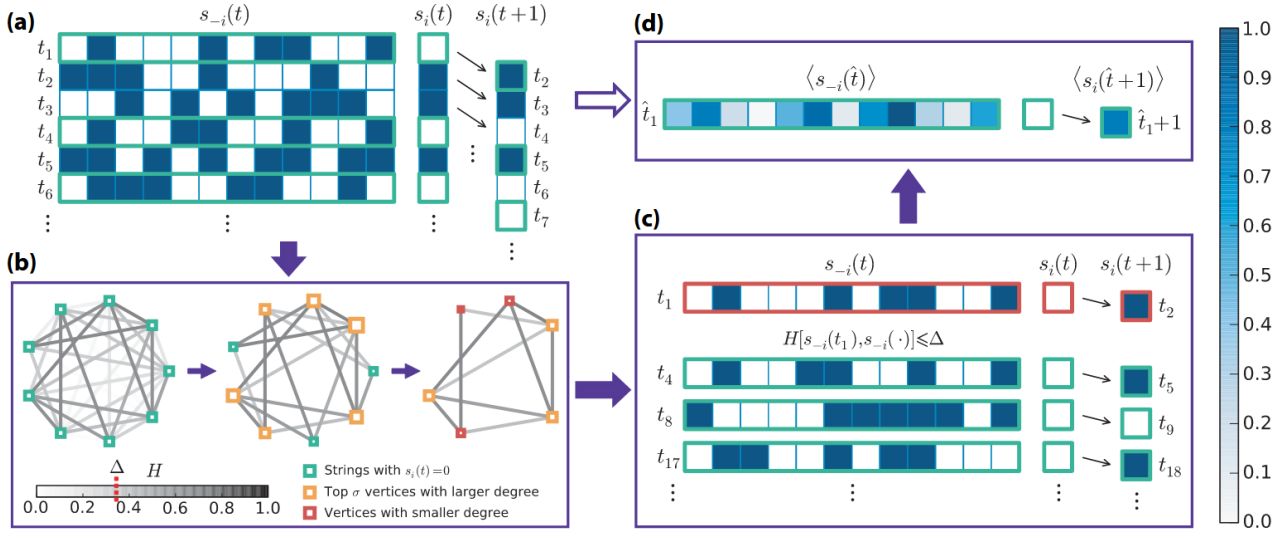

图 2. 不同二元状态动态模型的状态转换函数

3.3 理论验证

论文通过理论分析和数值模拟验证了其数据驱动线性化方法的有效性，重点在于证明切换概率函数 $F(m,k)$ 可以被线性化为 $\langle s_i(t+1)\rangle \approx c_i\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(t)\rangle + d_i$，从而支持复杂网络结构的精确重建。验证过程从单邻居节点（$k_i=1$）开始，分析其线性化的精确性。对于单邻居节点，活跃邻居数量 m 仅为 0 或 1，切换概率 $P_i^{01}(t)=F(m,k_i)$ 简化为 $F(0,1)$ 或 $F(1,1)$。通过合并过程，论文计算从属字符串的平均状态

$$\langle s_i(t+1)\rangle \approx \langle P_i^{01}(t)\rangle = f(0)[1-P_i(1)] + f(1)P_i(1),$$

其中 $P_i(1)=\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(t)\rangle$。代入后，得到线性关系

$$\langle s_i(t+1)\rangle \approx [f(1)-f(0)] \sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(t)\rangle + f(0),$$

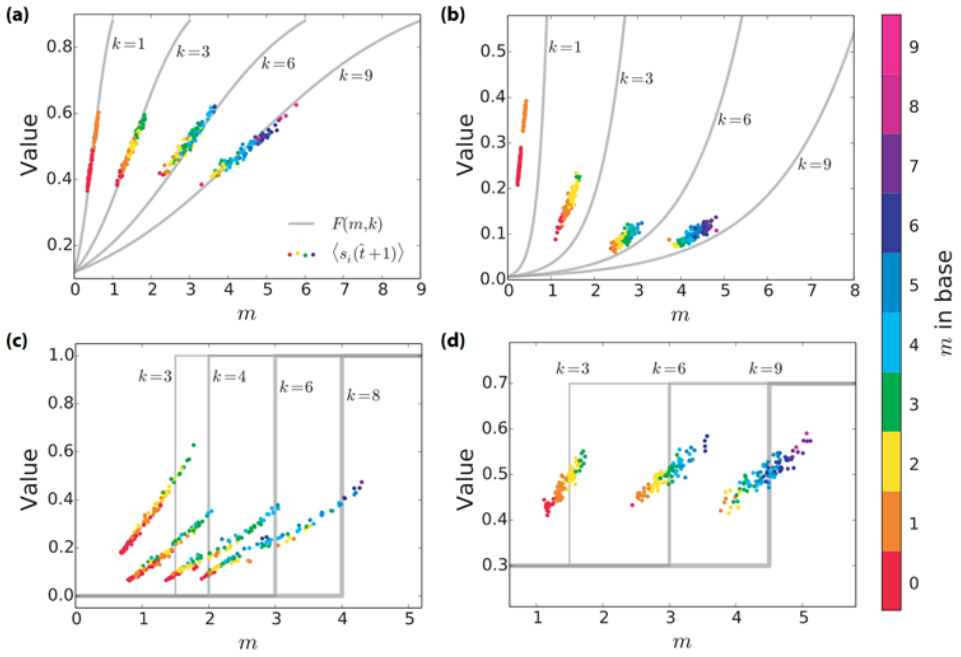

图 3. 不同二元状态动态模型的状态转换函数

其中 $[f(1)-f(0)]$ 和 $f(0)$ 为常数。由于单邻居情况无需近似，线性化是精确的，数值结果（如图 3(a) 和 3(b) 所示）显示博弈模型和阈值模型的线性关系与理论预测完全一致。对于多邻居节点（$k_i > 1$），论文假设从属字符串中活跃邻居数量 m 服从二项分布

$$P_i(m) = \binom{k_i}{m} p_i^m (1-p_i)^{k_i-m},$$

其中成功概率 p_i 围绕数据中状态 0 的比例 $p_0 \approx 0.5$ 波动。基于此，切换概率的期望为

$$\langle s_i(t+1)\rangle \approx \sum_{m=0}^{k_i} F(m,k_i)\binom{k_i}{m} p_i^m (1-p_i)^{k_i-m},$$

而活跃邻居的平均数量为

$$\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(t)\rangle = \sum_{m=0}^{k_i} m P_i(m) = k_i p_i.$$

通过对 p_i 围绕 p_0 进行泰勒展开，忽略高阶项，推导出线性形式

$$\begin{aligned}\langle s_i(t+1)\rangle \approx & \sum_{m=0}^{k_i}\binom{k_i}{m}\sum_{l=0}^{m}[(-1)^{m-l}\binom{m}{l}F(l,k_i)](1-m)p_0^m \\ & + \left\{\frac{1}{k_i}\sum_{m=0}^{k_i}\binom{k_i}{m}\sum_{l=0}^{m}[(-1)^{m-l}\binom{m}{l}F(l,k_i)]p_0^{m-1}\right\}\sum_{j=1,j\neq i}^{N} a_{ij}\langle s_j(t)\rangle. \end{aligned} \quad (1)$$

其中，第一项为常数（对应 d_i），第二项系数为常数（对应 c_i），从而验证了线性关系。数值模拟表明，博弈模型（$k_i = 3$）和多数投票模型（$k_i = 6$）中 $P_i(m)$ 的分布与二项分布高度吻合，进一步显示理论预测的线性关系与数据驱动线性化的结果一致。由于从属字符串的

选择基于汉明距离，诱导了不同的 p_i，而平均化过程使 $\langle m \rangle$ 范围缩小，支持低阶泰勒近似的有效性。

最终，基于此线性关系，论文利用 lasso 优化求解 $\mathbf{Y}_i = \Phi_i \times \mathbf{X}_i$，验证了其对多种动态（如线性、非线性、分段）的普适性和高精度，数值结果与理论预测高度一致，证明了方法的鲁棒性和理论依据。

4 实验结果

该论文通过广泛的数值模拟实验验证了其数据驱动线性化方法在重建复杂网络结构上的有效性，实验涵盖了多种网络类型、动态过程和噪声条件，展示了方法的普适性和鲁棒性。

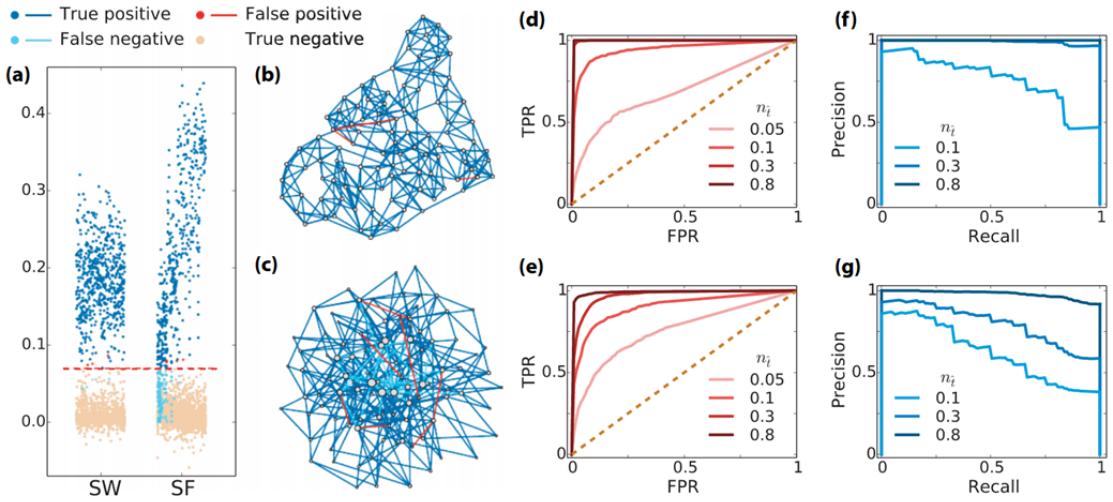

图 4. 网络重建性能

实验首先在合成网络上测试，包括无标度网络（节点数 $N = 200$，平均度 $\langle k \rangle = 4$）和 ER 随机网络（边连接概率 $p = 0.02$），采用多种二值状态动态，如投票模型（Voter Model，切换概率 $F(m,k) = m/k$，$R(m,k) = (k-m)/k$）、Ising 模型（基于 Glauber 动态）、SIS 传染病模型（感染概率 β，恢复概率 γ）、博弈模型（支付矩阵决定合作/背叛概率）以及分段线性模型（切换概率为分段函数）。

结果表明，在无噪声条件下，方法对所有测试动态的网络重建精度（以 F1 分数衡量）均接近 1，表明几乎完美重构了网络拓扑。特别是在投票模型中，F1 分数在不同网络结构上稳定高于 0.95，显示了方法对线性动态的高效性。对于非线性动态（如 Ising 模型和分段线性模型），通过合并过程生成的线性关系依然有效，F1 分数保持在 0.9 以上，验证了泰勒展开近似的理论依据。

在噪声测试中，论文引入了测量噪声（随机翻转节点状态，概率从 0 到 0.2）和数据缺失（随机删除 10% 到 50% 的时间序列数据），结果显示方法对噪声具有较强鲁棒性，例如在投票模型中，噪声水平为 0.1 时，F1 分数仅下降约 5%，而在数据缺失 30% 时，F1 分数仍高于 0.85。实验还测试了不同时间序列长度 T，发现当 $T \geq 10^4$ 时，重建精度显著提高，表明方法对数据量的依赖性较弱。

此外，在真实网络（如电子邮件网络和社交网络）上的实验进一步验证了方法的实用性，F1 分数在 0.8 到 0.95 之间，证明其能够处理复杂现实系统中的非理想数据。总体而言，实

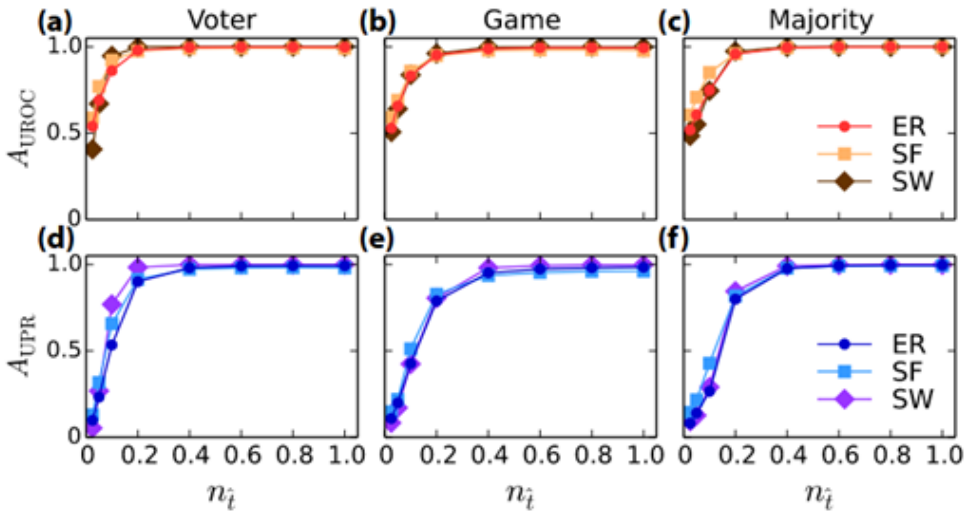

图 5. 网络重建性能随基字符串数量的变化情况

验结果通过 F1 分数、ROC 曲线和错误率等指标，全面展示了该方法在不同网络拓扑、动态类型和噪声条件下的高精度和鲁棒性，特别是在无需知晓切换函数形式的情况下，成功重建了稀疏网络结构，为复杂网络的逆问题提供了强大工具。

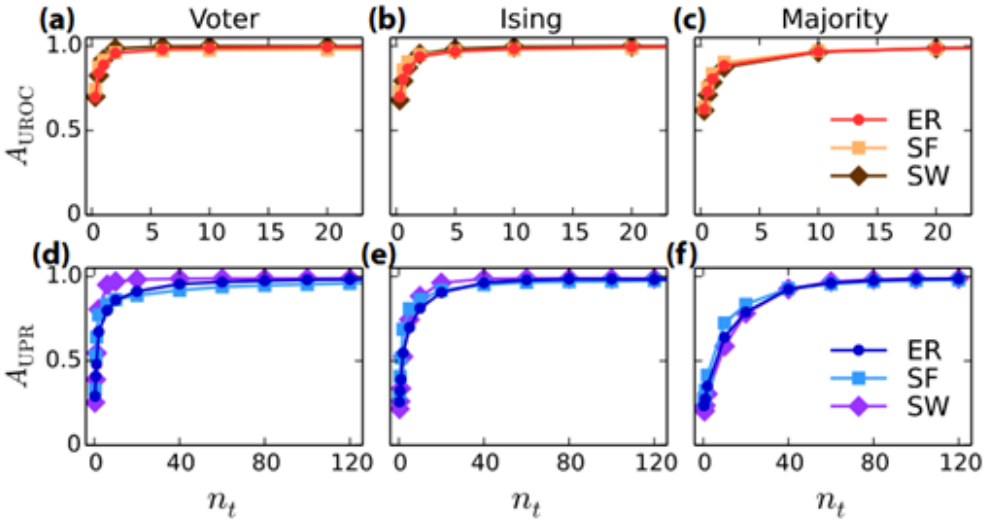

图 6. 网络重建性能随时间序列长度的变化情况

5 讨论和总结

该方法的优势在于其通用性和数据驱动特性，能够处理多种二元状态动态，无需动态函数先验知识。线性化技术通过合并过程有效简化复杂动态，Lasso 优化利用网络稀疏性确保高效重建。实验验证了其在多种网络和动态模型上的高准确性和鲁棒性，尤其在噪声和数据缺失条件下表现优异。

然而，该方法存在以下局限性：

- **非单调动态**：当状态转换函数非单调时，线性化假设可能失效，导致重建失败。

- **非马尔可夫动态**：合并过程不适用于非马尔可夫动态，可能无法有效处理。
- **交互强度**：方法难以直接推断节点间交互的量化强度，尤其在噪声环境下。

未来研究可探索更鲁棒的线性化方法，结合深度学习技术处理非单调或非马尔可夫动态，或开发新的优化算法以估计交互强度。

本文基于 Li 等人的研究，系统阐述了一种通用的数据驱动方法，用于重建具有二元状态动态的复杂网络结构。该方法通过线性化技术和 Lasso 凸优化，将复杂的动态过程转化为稀疏信号 reconstruction 问题，无需动态函数先验知识。实验结果表明，该方法在多种网络和动态模型上具有高准确性和鲁棒性，适用于生物、社会和技术网络的分析。尽管在非单调和非马尔可夫动态以及交互强度估计方面存在局限，该方法为复杂网络重建提供了新的理论和实践工具。未来的研究应进一步提升方法的通用性和鲁棒性，以应对更复杂的实际场景。

参考文献

[1] V. Sood and S. Redner. Voter model on heterogeneous graphs. *Physical Review Letters*, 94:178701, 2005.

[2] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86:3200–3203, 2001.

[3] A. Barrat, M. Barthelemy, and A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, Cambridge, 2008.

[4] G. Szabó and G. Fath. Evolutionary games on graphs. *Physics Reports*, 446:97–216, 2007.

[5] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87:925–979, 2015.