从数据到结构——动力学网络重构

张朝阳^{1,2}, 陈阳³, 弭元元⁴, 胡岗^{5*}

1. 宁波大学物理科学与技术学院, 宁波 315211

2. 宁波大学商学院, 宁波 315211

3. 中国科学院自动化研究所模式识别国家重点实验室, 北京 100190

4. 重庆大学神经智能研究中心, 重庆 400044

5. 北京师范大学物理学系, 北京 100875

*联系人, E-mail: ganghu@bnu.edu.cn

收稿日期: 2019-04-14; 接受日期: 2019-05-08; 网络出版日期: 2019-11-25

国家自然科学基金(编号: 11835003, 11605098, 31771146, 11734004)、宁波市自然科学基金(编号: 2017A610142)、北京市科学技术委员会类脑计算专项(编号: Z171100000117007)、北京市科技新星项目(编号: Z181100006218118)和宁波大学王宽诚基金资助项目

摘要 大数据是一笔越来越重要并不断快速增长的财富, 合理利用这一财富的关键是有效的分析手段. 大数据中一大类数据是由复杂网络代表的实际动力学系统产生的, 其中网络各个单元的输出数据可以测量, 但产生数据的网络结构却不为所知; 而了解这些网络结构对我们理解、预测和控制实际系统功能极为重要. 因此, 从分析网络数据出发揭示网络结构的重构问题就成为数学物理特别是统计物理以及一系列交叉领域对网络研究的核心问题之一. 网络重构的重要性还来源于解决实际网络重构中所面对的各种困难的理论要求. 网络结构的复杂性、网络节点动力学的非线性、未知噪声对网络动力学演化数据的影响以及测量中有效数据的缺失等都是在实际网络重构中要面对的常见且非常重要的困难. 本综述介绍并讨论了如何有效克服这些困难的方法, 特别是通过数据扩张充分利用数据信息的方法, 针对不同的系统特征和重构任务选择合适的关联量计算方法, 以及利用噪声帮助克服重构困难的方法等. 网络重构研究将逐步解决实际复杂系统重构问题并引起复杂网络相关学者越来越多的关注和研究兴趣.

关键词 网络重构, 非线性动力学, 数据分析

PACS: 89.75.Hc, 05.10.Gg, 05.45.Tp

1 引言

近年来大数据的重要性逐渐被人们所认识和接受^[1-3]. 越来越多的数据积累使大数据成为无与伦比而又增长极快的财富, 但要获取这笔财富正确的

分析手段是关键. 积累和记录的大量数据有各种各样丰富的来源, 而形形色色的动力学网络是产生数据的一大类极为重要的复杂系统. 近几十年来, 网络结构和动力学的研究成为物理、数学和各类交叉学科极其重要并受到广泛关注的重要领域^[4-7]. 许多

引用格式: 张朝阳, 陈阳, 弭元元, 等. 从数据到结构——动力学网络重构. 中国科学: 物理学 力学 天文学, 2020, 50: 010502

Zhang Z Y, Chen Y, Mi Y Y, et al. From data to network structure——Reconstruction of dynamic networks (in Chinese). Sci Sin-Phys Mech Astron, 2020, 50: 010502, doi: [10.1360/SSPMA2019-0127](https://doi.org/10.1360/SSPMA2019-0127)

实际系统(如金融系统^[8]、电力系统^[9]、交通系统、互联网系统等社会系统和神经网络^[10]、基因调控系统^[11-13]、气象系统^[14]等自然系统)大都可以用动力学网络来描述. 在这些网络中人们可以较容易地记录到系统各个单元显示出的各类数据, 但难以获取系统内部各个单元相互影响的结构及系统演化的各种动力学机制. 因此, 从网络中易于收集的各类数据出发揭示隐藏的网络结构和动力学演化规律成为网络数据分析的重要课题, 称为动力学网络重构^[15-20].

网络重构问题在理论发展和实际应用中都有非常重要的意义^[21-25]. 在实际中存在大量的由网络数据分析来探索网络结构的需求, 如大脑功能活动可以被核磁共振、红外血氧探测等多种方法记录成为各类数据^[26], 它们为大脑活动的认知和调控提供了有用的信息. 从数据分析到大脑功能网络的构建就是网络重构的重要任务. 类似的实际网络重构的需求出现在与数据相关的大量实际应用中^[27-30]. 另一方面网络重构研究有重要的理论意义, 这与网络重构所面临的多重困难有关. 网络重构的第一个困难是通常实际网络都非常复杂, 网络常常会包含大量单元(称为网络节点(Nodes)), 单元之间会有复杂的相互影响(称为相互作用或连接(Links)). 要从收集到的数据分析出所有这些复杂的结构信息会面临各种困难^[31,32]. 网络重构的第二个大困难是数据中包含有许多与所研究网络无关的因素. 例如, 在各类基因调控网络中, 基因既会受到分子内部各种未知的物理化学因素的影响, 又会受到网络外部环境的各种生物活动的影响, 所有这些影响在网络重构研究中被认定为未知的噪声^[33]. 噪声的存在是几乎所有实际网络重构问题中都不可回避的因素, 也是网络重构的一大类困难^[34-42]. 网络重构的第三个困难是网络动力学的非线性. 如果上述第一和第二个困难出现在线性系统中, 则可以系统地运用已有的数学和统计方法加以处理. 所有这些困难与非线性行为结合就实质性地增加了问题的难度, 需要发展新的方法和手段. 网络重构的研究中, 还有一个在实际应用中普遍存在的困难是数据的缺失. 数据记录并不总是完备的, 如网络中有些节点的输出数据较容易测量和记录, 而另一些节点的数据却不容易得到

或者根本不为人们所知, 称为隐藏(Hidden)节点. 隐藏节点的存在是许多实际网络重构研究中普遍存在的问题^[43-49], 但这些节点和可测节点在影响网络动力学行为上同等重要, 在所测数据中包含了它们的影响. 如果不有效地处理隐藏节点的问题, 这些影响就会破坏网络重构的正确性.

以上困难只是网络重构所面临诸多困难的一部分, 也是极为重要的一部分. 本文系统地探讨从分析网络输出数据出发重构网络结构的问题, 并重点讨论如何有效地处理上述四种困难.

2 动力学网络重构问题

许多实际系统可以用以下动力学网络描述:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}(\mathbf{x}(t)) + \mathbf{\Gamma}(t), \quad (1)$$

其中

$$\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_N(t)) \quad (2)$$

是网络中 N 个节点的状态变量.

$$\mathbf{F}(\mathbf{x}(t)) = (F_1(\mathbf{x}(t)), F_2(\mathbf{x}(t)), \dots, F_N(\mathbf{x}(t))) \quad (3)$$

描述节点自身和相互作用动力学,

$$\mathbf{\Gamma}(t) = (\Gamma_1(t), \Gamma_2(t), \dots, \Gamma_N(t)) \quad (4)$$

是各个节点所受到的噪声影响. 本文考虑了简单的加性白噪声问题

$$\langle \Gamma_i(t) \rangle = 0, \langle \Gamma_i(t) \Gamma_j(t + t') \rangle = Q_{ij} \delta(t'). \quad (5)$$

所谓网络重构问题就是在已知部分节点的输出数据

$$\mathbf{x}(t)^M = (x_{i_1}(t), x_{i_2}(t), \dots, x_{i_M}(t)), M \leq N \quad (6)$$

的条件下, 尽可能求解未知网络内部隐藏的结构信息, 包括

$$[F_i(\mathbf{x}(t)), Q_{ij}], i, j = 1, 2, \dots, M \quad (7)$$

的行为.

引言中所述的实际网络重构问题中普遍存在的困难(图 1), 在网络(1)式中都可以存在. 首先, 网络规模可能很大 $N \gg 1$, 而节点之间相互作用连接众

多和复杂, 网络的复杂性是绝大多数实际网络都会面临的问题. 同时噪声的存在非常普遍, 其影响的值 $\Gamma_i(t)$ 都是未知的, 甚至其统计特征 $\hat{Q}$ 也常常不为所知, 而这些因素都会极大影响输出信号 $x_i(t)$ 的值, 从而影响重构运算. 特别重要的是网络自身和相互作用动力学 $F_i(\mathbf{x}(t))$ 通常多是非线性而且未知. 这些未知的非线性的行为结合网络的复杂性和噪声作用就会给网络重构带来极大的困难. 网络(1)的所有节点通过相互作用连接成为一个整体, 任意一个节点的数据都会受到网络中其他所有节点动力学的影响. 人们很自然地设想要重构网络(1)中任何部分的结构, 所有节点的输出数据 $x_i(t), i = 1, 2, \dots, N$ 都必须测量和掌握. 但在实际问题中通常不能满足这一要求, 人们常常只能记录网络中部分甚至很少节点的数据($M \ll N$), 另一些节点难以测量, 我们甚至不知道这些节点存在与否. 因此, 网络重构必须解决在 $M \ll N$ (数据缺失)的条件下利用已有的数据并通过有效的分析方法得到尽可能多的关于网络结构的有用信息.

本节最后强调的是, 我们所研究的网络(1)中包含了很多在网络重构中常见的并且起重要作用的

图 1 (网络版彩图) 实际网络重构问题中普遍存在的困难, 如网络结构的复杂性; 网络动力学的非线性; 各种各样来自节点内部或外部环境的未知因素, 即噪声作用; 网络中众多隐藏节点. 上述困难都可以出现在(1)式代表的动力学网络重构问题中

Figure 1 (Color online) Some difficulties extensively exist in the reconstruction of real-world networks such as complexity of network structures, strong nonlinearity of network dynamics, diverse and unknown impacts from the interiors of nodes and externals of networks, i.e., noises, and many hidden nodes in networks. All the above mentioned difficulties are present in our network reconstruction of eq. (1).

困难, 但有一些同样重要的困难并未涉及. 例如, 方程(5)只考虑了关联时间很短的噪声, 即满足白噪声近似的条件. 这些假设的确具有一定的实际意义, 如当噪声来自节点内部微观涨落影响时, 噪声关联时间比起网络中宏观节点动力学时间尺度的确会短很多; 但该假设并不能代表宏观世界和外部环境的影响, 与动力学时间尺度相比, 这些外部噪声的关联时间并不短, 必须用色噪声来描述, 对这样的噪声(5)式的假设就不再成立^[50]. 另外, (1)式中各项都以相同时间 t 参与演化, 特别是任何节点 j 对节点 i 的作用都是等时作用. 在实际系统中, 任何相互作用的影响都要通过物理、化学、生物或其他过程传播, 需要一定的时间, 这造成相互作用常常会有不同程度的时间延迟. 从非线性动力学知识可知时间延迟会使系统动力学及其维度发生重大变化, 会给网络重构带来重大的实质性困难. 所有这些问题在该领域中都有专门的讨论^[51-53], 在本文中我们不涉及这些问题, 而集中讨论以方程(1)代表的网络动力学的重构问题.

3 线性网络重构

我们首先研究(1)式的最简单的线性近似

$$\frac{d\mathbf{x}(t)}{dt} = \hat{\mathbf{A}}\mathbf{x}(t) + \boldsymbol{\Gamma}(t), \quad (8)$$

其中 $\hat{\mathbf{A}}, \hat{\mathbf{Q}}$ 为常数矩阵,

$$\hat{\mathbf{A}} = [\mathbf{A}_{ij}], \quad \hat{\mathbf{Q}} = [\mathbf{Q}_{ij}]. \quad (9)$$

现在考虑从所有节点的输出数据

$$\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_N(t)) \quad (10)$$

出发重构网络连接结构与噪声关联信息. 方程(8)的性质在统计物理中作为一个基本问题受到广泛关注与讨论. 其中与我们重构问题直接相关的一个关系式是^[54]

$$\hat{\mathbf{A}}\hat{\mathbf{C}} + \hat{\mathbf{C}}\hat{\mathbf{A}}^T = -\hat{\mathbf{Q}}, \quad (11)$$

其中 $\hat{\mathbf{C}}$ 是一个关联矩阵,

$$\hat{\mathbf{C}} = [\mathbf{C}_{ij}], \quad \mathbf{C}_{ij} = \langle x_i(t)x_j(t) \rangle = \frac{1}{T} \int_0^T x_i(t)x_j(t)dt,$$

其中 T 取足够长时间 ($T \gg 1$) 的数据采集长度. 所谓足够长是指增加 T 的长度不会显著改变 C_{ij} 的数值. 对重构而言, 式(11)的问题是已知 $\mathbf{x}(t)$ 的时间序列后不能直接求出 $\hat{\mathbf{A}}, \hat{\mathbf{Q}}$ 的值.

现在我们从可测数据开始来讨论重构式(8)的问题. 实际问题中 $x_i(t)$ 不可能连续测量, 所以只能假设可测得离散的数据

$$x_i(t_1), x_i(t_2), \dots, x_i(t_L); i = 1, 2, \dots, N, L \gg 1. \quad (12)$$

现代测量技术可以做到很高的测量频率

$$0 < t_{k+1} - t_k \ll 1, k = 1, 2, \dots, L-1. \quad (13)$$

简单起见(不失普遍性)我们假设以固定频率测量变量数据, 得到数据输出式(12)

$$\Delta t = t_{k+1} - t_k, k = 1, 2, \dots, L-1. \quad (14)$$

值得指出的是式(12)中包含大量信息, 远远超过了仅仅 L 个数字的信息. 例如, 它们可以给出变量变化速率

$$\dot{x}_i(t_k) = \frac{x_i(t_{k+1}) - x_i(t_k)}{\Delta t}, k = 1, 2, \dots, L-1, \quad (15)$$

这个速率信息可以完全独立于式(12)的变量数值的信 息, 为重构增加了极大潜力. 类似的从式(12)中还可以挖掘出大量式(15)以外的其他有用信息.

现在我们就利用式(12)和(15)的信息以及对噪声(5)的认识进行网络重构, 即同时求解矩阵 $\hat{\mathbf{A}}, \hat{\mathbf{Q}}$ 的值. 用式(15)代替方程(8)的左边, 可以得到

$$\dot{\mathbf{x}}(t_k) = \hat{\mathbf{A}}\mathbf{x}(t_k) + \mathbf{\Gamma}(t'_k), \quad (16)$$

并用 $\mathbf{x}(t)$ 的转置 $\mathbf{x}(t)^T$ 右乘式(16)的两边并计算相应的关联, 可得矩阵等式

$$\hat{\mathbf{B}} = \hat{\mathbf{A}}\hat{\mathbf{C}} + \langle \mathbf{\Gamma}(t')\mathbf{x}(t)^T \rangle, \quad (17)$$

其中 $\mathbf{\Gamma}(t')$ 的含义为 t 到 $t + \Delta t$ 一段时间内的噪声影响, 矩阵元的定义和计算为

$$\hat{\mathbf{B}} = [B_{ij}], B_{ij} = \frac{1}{L-1} \sum_{k=1}^{L-1} \dot{x}_i(t_k)x_j(t_k), \quad (18)$$

$$\hat{\mathbf{C}} = [C_{ij}], C_{ij} = \frac{1}{L-1} \sum_{k=1}^{L-1} x_i(t_k)x_j(t_k), \quad (19)$$

$$\langle \mathbf{\Gamma}(t')\mathbf{x}(t)^T \rangle = [\langle \Gamma_i(t')x_j(t) \rangle],$$

$$\langle \Gamma_i(t')x_j(t) \rangle = \frac{1}{L-1} \sum_{k=1}^{L-1} \Gamma_i(t'_k)x_j(t_k). \quad (20)$$

在式(17)中, $\hat{\mathbf{B}}, \hat{\mathbf{C}}$ 矩阵均可直接从测量数据进行计算得到, 但方程右面的第二项似乎不能求得, 因为 $\Gamma_i(t'_k)$ 是完全未知的噪声序列. 幸运的是 $\mathbf{\Gamma}(t)$ 的关联时间很短, 远小于测量间隔 Δt . 由定义式(15)可见, 在式(16)中左边 $\dot{x}_i(t_k)$ 只受 t_k 到 t_{k+1} 之间的噪声 $\mathbf{\Gamma}(t')$ 影响, $x_i(t_k)$ 是 t_k 时间点上的节点输出数据, 因而 $x_i(t_k)$ 只与 t_k 之前的噪声相关而与 t_k 之后的噪声无关, 所以 $\mathbf{\Gamma}(t')$ 与 $x_i(t_k)$ 统计不相关, 则有

$$\langle \mathbf{\Gamma}(t')\mathbf{x}(t)^T \rangle = 0. \quad (21)$$

这就是我们在白噪声近似下的去噪声操作. 将式(21)代入(17)可得

$$\hat{\mathbf{B}} = \hat{\mathbf{A}}\hat{\mathbf{C}}, \quad (22)$$

其中 $\hat{\mathbf{B}}, \hat{\mathbf{C}}$ 矩阵均可以从可测数据求得, 因而网络连接 $\hat{\mathbf{A}}$ 可以解析求出. 与式(11)不同的是在式(22)中噪声的影响已被有效去除, 利用可测数据可以计算 $\hat{\mathbf{B}}, \hat{\mathbf{C}}$ 两个矩阵, 则 $\hat{\mathbf{A}}$ 矩阵可被独立求得,

$$\hat{\mathbf{A}} = \hat{\mathbf{B}}\hat{\mathbf{C}}^{-1}. \quad (23)$$

从数据式(12)及其拓展式(15)还可以独立地求出式(9)中的噪声关联矩阵. 将 $\mathbf{x}(t_{k+1})^T$ 乘以式(16)两边并对 k 求和, 可以得到另一个方程

$$\hat{\mathbf{B}}^+ = \hat{\mathbf{A}}\hat{\mathbf{C}}^+ + \langle \mathbf{\Gamma}(t')\mathbf{x}^+(t)^T \rangle, \quad (24)$$

其中矩阵元的定义和计算为

$$\hat{\mathbf{B}}^+ = [B_{ij}^+], B_{ij}^+ = \frac{1}{L-1} \sum_{k=1}^{L-1} \dot{x}_i(t_k)x_j(t_{k+1}), \quad (25)$$

$$\hat{\mathbf{C}}^+ = [C_{ij}^+], C_{ij}^+ = \frac{1}{L-1} \sum_{k=1}^{L-1} x_i(t_k)x_j(t_{k+1}), \quad (26)$$

$$\langle \mathbf{\Gamma}(t')\mathbf{x}^+(t)^T \rangle = [\langle \Gamma_i(t')x_j^+(t) \rangle],$$

$$\langle \Gamma_i(t')x_j(t) \rangle = \frac{1}{L-1} \sum_{k=1}^{L-1} \Gamma_i(t'_k)x_j(t_{k+1}). \quad (27)$$

与式(17)不同的是, 在式(24)中噪声的影响不能消除, 因为 $\Gamma_i(t'_k)$ 是在 t_k 与 t_{k+1} 之间的时间段上的噪声,

而 $x_i(t_{k+1})$ 是 t_{k+1} 时间点上的节点输出数据, 它直接受到噪声 $\Gamma_i(t'_k)$ 的影响, 其影响值可以解析计算为

$$\langle \Gamma(t') \mathbf{x}^+(t)^T \rangle = e^{\hat{A}t} \int_t^{t+0} e^{-\hat{A}t'} \hat{\mathbf{Q}} \delta(t-t') dt' = \hat{\mathbf{Q}}. \quad (28)$$

从式(24)和(28)可以得到

$$\hat{\mathbf{B}}^+ = \hat{\mathbf{A}}\hat{\mathbf{C}}^+ + \hat{\mathbf{Q}}, \quad (29)$$

其中除了 $\hat{\mathbf{A}}, \hat{\mathbf{Q}}$ 以外的各项都可以从数据中求出. 用式(29)减去式(22), 可以进一步求出 $\hat{\mathbf{Q}}$

$$\begin{aligned} \hat{\mathbf{Q}} &= \hat{\mathbf{B}}^+ - \hat{\mathbf{A}}\hat{\mathbf{C}}^+ - (\hat{\mathbf{B}} - \hat{\mathbf{A}}\hat{\mathbf{C}}) \\ &= \hat{\mathbf{B}}^+ - \hat{\mathbf{B}} \\ &= -\hat{\mathbf{B}}^T - \hat{\mathbf{B}}. \end{aligned} \quad (30)$$

第二个等式是因为 $\hat{\mathbf{C}}^+ \doteq \hat{\mathbf{C}}$, 一个小的 $\mathbf{x}(t)$ 时间错位不会改变 $\mathbf{x}(t)$ 与 $\mathbf{x}^+(t)^T$ 的关联值, 但会显著改变 $\dot{\mathbf{x}}(t)$ 与 $\mathbf{x}^+(t)^T$ 的关联值. 当然如果 $\hat{\mathbf{A}}$ 已通过式(23)计算得到, 也可以将式(23)代入式(11), 求出 $\hat{\mathbf{Q}}$, 得到式(30)的第三个等式.

由式(23)和(30)可以仅由 $\mathbf{x}(t)$ 的数据出发分别重构网络连接 $\hat{\mathbf{A}}$ 和噪声统计矩阵 $\hat{\mathbf{Q}}$, 比通过传统的关系式(11)实质性地前进了一大步, 这个成功的关键是拓宽数据式(12)的应用(即同时应用数据式(12)和(15)). 可以看到在噪声条件下, 各种统计关联的计算是从可测数据到网络重构的核心, 而选择计算合适的关联矩阵是重中之重, 这一要点将在本文以下讨论中反复出现.

图2是对线性网络(8)重构的例子. 我们考虑了 $N = 100$ 两组随机线性网络, 两组网络分别具有固定和分布的连接强度. 图2(a)和(c)分别画出了两组网络中的两个节点数据在相平面的投影. 运用矩阵代数运算式(23)和(30)可以重构线性网络(8)连接(图2(b)和(d))和噪声统计情况(图2(e)). 在噪声提高10倍的情况下, 我们仍可以重构原网络, 图2(f)和(g)画出了对应分布网络的数据和重构结果. 我们可以定义重构精度 D

$$D(L) = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^N (A_{ij} - A'_{ij})^2}{N^2}}, \quad (31)$$

用以定量刻画重构结果的优劣程度, 其中 A'_{ij} 为应用式(23)计算出的 A_{ij} 对应矩阵元. 图2(h)中画出了两种情况下重构精度随数据量的变化关系.

4 非线性网络重构

线性网络式(8)的重构可以选取恰当的关联计算并过去噪声而建立简单的线性矩阵代数运算式(23)实现, 而非线性网络重构就要复杂得多^[39].

现在我们来考虑式(1)一般形式的重构问题. 在数学物理和非线性科学中我们早已熟悉了任意非线性函数用基函数进行线性展开, 对函数 $F_i(\mathbf{x})$ 可以展开为

$$F_i(\mathbf{x}) = \sum_{\mu=1}^{\infty} \mathfrak{Y}_{i,\mu} Y_{\mu}(\mathbf{x}), \quad (32)$$

展开式中 $\mathbf{Y}(\mathbf{x}) = [Y_{\mu}(\mathbf{x})]$ 为已知的一组非线性基函数, 而 $\mathfrak{Y} = [\mathfrak{Y}_{i,\mu}]$ 为线性矩阵

$$\mathfrak{Y} = [\mathfrak{Y}_{i,\mu}], \quad i = 1, 2, \dots, N; \mu = 1, 2, \dots, \infty. \quad (33)$$

在这个框架下上一节的关联运算技术可以完全应用于非线性网络重构. 具体地运用一组对耦基 $\mathbf{Y}^*(\mathbf{x})$

$$\mathbf{Y}^*(\mathbf{x}) = (Y_1^*(\mathbf{x}), Y_2^*(\mathbf{x}), \dots, Y_{\infty}^*(\mathbf{x})) \quad (34)$$

的转置乘以方程(1)两边并进行关联运算.

由于我们采用有限频率收集数据

$$\mathbf{x}(t_1), \mathbf{x}(t_2), \dots, \mathbf{x}(t_L), \quad (35)$$

方程(1)可以离散化为

$$\dot{\mathbf{x}}(t_k) = \mathbf{F}(\mathbf{x}(t_k)) + \Gamma(t'_k), \quad k = 1, 2, \dots, L-1. \quad (36)$$

将函数 $\mathbf{F}(\mathbf{x})$ 进行非线性基展开, 方程转化为

$$\dot{\mathbf{x}}(t_k) = \mathfrak{Y} \mathbf{Y}(\mathbf{x}(t_k)) + \Gamma(t'_k), \quad k = 1, 2, \dots, L-1. \quad (37)$$

将 $\mathbf{Y}^*(\mathbf{x}(t_k))^T$ 乘以方程(37)两边, 进行类似式(17)的关联计算, 并考虑和式(21)完全同样的去噪声条件, 可以得到与式(22)同样的简单线性矩阵方程

$$\mathfrak{B} = \mathfrak{Y} \hat{\mathbf{C}}, \quad (38)$$

只是这里 $\mathfrak{B}$ 和 $\hat{\mathbf{C}}$ 要用非线性基计算

$$\mathfrak{B} = [\mathfrak{B}_{i,\mu}], \quad \mathfrak{B}_{i,\mu} = \frac{1}{L-1} \sum_{k=1}^{L-1} \dot{x}_i(t_k) Y_{\mu}^*(\mathbf{x}(t_k)), \quad (39)$$

图 2 (网络版彩图)运用矩阵代数运算式(23)和(30)重构线性网络式(8). 我们考虑了 $N = 100$ 两组随机线性网络, 网络具有 5% 正相互作用和 5% 负相互作用. 两组网络分别具有固定连接强度 (± 1) (Case 1) 和均匀分布的连接强度 $[(-2, -1), (1, 2)]$ (Case 2). (a) 和 (c) 分别画出了两组网络中的两个节点数据在相平面的投影. (b) 和 (d) 分别用 (23) 重构 (a) 和 (c) 中的网络连接, 可以看到在两种情况下重构结果与真实网络连接非常接近. (e) 两种情况下式 (30) 推断的噪声结果与真实情况对比. (f) 和 (g) 具有分布连接强度 (Case 2) 的网络, 在噪声提高 10 倍的情况下的两节点数据在相平面投影 (f) 与重构结果 (g). (h) 两种情况下重构精度 D (31) 式随数据量的变化关系. 引自文献 [38]

Figure 2 (Color online) Solving the inverse problem of linear system eq. (8) with algorithms eqs. (23) and (30). $N = 100$. 500 active links and 500 repressive ones are randomly chosen, and $A_{ij}(0) = 0$ for all other off-diagonal matrix elements. Time step for simulations of eq. (8) is $dt = 5 \times 10^{-4}$ and time interval for velocity computation is $\Delta t = 10^{-3}$. $L = 5 \times 10^5$ samples are yielded for inferences. (a) and (b) Case 1: Jacobian matrix $\hat{A}$ is given as: active interactions $A(a)_{ij} = 1$; repressive interactions $A(r)_{ij} = -1$; the diagonal terms are set to $A_{ii} = -3$; and noise correlation matrix is given as $Q_{ij} = \sigma_i \delta_{ij}$, $\sigma_i = 0.01$. (c) and (d) Case 2: the same as (a) and (b) with heterogeneous $\hat{A}$ and $\hat{Q}$, i.e., $A_{ij}(a) \in (1.0, 2.0)$ and $A_{ij}(r) \in (-2.0, -1.0)$ with uniform distributions, $A_{ii} = -5$, and $\sigma_i \in (0.005, 0.015)$ with uniform distribution. (a) and (c) The trajectories in a 2D phase plane for the two sets of networks and parameters. (b) The distribution of off-diagonal elements of $\hat{A} = \hat{B}\hat{C}^{-1}$ (blue triangles for $A_{ij}(a)$, red squares for $A_{ij}(r)$, and green circles for $A_{ij}(0)$) calculated by eq. (23) from data of (a). The computational elements of matrix $\hat{B}\hat{C}^{-1}$ coincide very well with the actual $\hat{A}$ with certain fluctuated errors, and all active, repressive and null interaction natures can be accurately predicted. (d) Matrix elements computed by $\hat{B}\hat{C}^{-1}$ are plotted against true Jacobian matrix elements A_{ij} for Case 2 (all colors have the same meanings as (b)). All dots locate closely near the diagonal line, indicating correct inferences of both interaction natures and coupling strengths. (e) Element values of $\hat{B}^+ - \hat{B}$ plotted versus element values of actual $\hat{Q}$ for both the noises of (a) and (c). The agreements between $\hat{B}^+ - \hat{B}$ and $\hat{Q}$ convincingly show that our method can depict not only Jacobian matrix $\hat{A}$ but also noise correlation matrix $\hat{Q}$. (f) and (g) Exactly the same as (c) and (d), respectively, with noise intensities increasing 100 times ($\sigma_i \in (0.5, 1.5)$ with uniform distribution). Matrix $\hat{A}$ can be retrieved equally well. (h) Standard deviation (D) of eq. (31) plotted against the number of samples L . D s decay as $\frac{1}{\sqrt{L}}$. Quoted from ref. [38].

$$\hat{\mathbf{C}} = [\mathbb{C}_{\mu\nu}], \quad \mathbb{C}_{\mu\nu} = \frac{1}{L-1} \sum_{k=1}^{L-1} Y_{\mu}(\mathbf{x}(t_k)) Y_{\nu}^*(\mathbf{x}(t_k)), \quad (40)$$

$$i = 1, 2, \dots, N; \quad \mu, \nu = 1, 2, \dots, \infty.$$

值得指出的是, 虽然上述计算采用了非线性基 $\mathbf{Y}(\mathbf{x})$ 和对偶基 $\mathbf{Y}^*(\mathbf{x})$, 但这些基函数都已知, 所以都可以用采集到的数据来计算. 这样 $\hat{\mathbf{B}}$ 和 $\hat{\mathbf{C}}$ 中的所有元素都可以从数据中显式算出, 理论上可以用它们来求出常数矩阵 $\hat{\mathbf{A}}$, 进而完成重构非线性网络式 (1) 的任务. 但这些矩阵 $\hat{\mathbf{C}}$ 的维数是无穷大的, 在运算中必须采用截断近似

$$\hat{\mathbf{C}} = [C_{\mu\nu}], \quad \mu, \nu = 1, 2, \dots, q; \quad q \geq N. \quad (41)$$

要这样做首先 (32) 式基展开必须收敛, 在收敛条件下采用适当的截断 q . 对于收敛条件问题, 本文暂无统一的理论, 但会根据实际问题提供一些经验考虑. 对于采用自恰的截断条件选择合适截断点的方法, 将在下面讨论. 在具有确定的截断值后, (38) 式成为可解问题

$$\hat{\mathbf{A}} = \hat{\mathbf{B}}\hat{\mathbf{C}}^{-1}. \quad (42)$$

令人惊奇的是该式和线性网络算式 (23) 完全一致. 只不过线性网络的矩阵元 $\hat{\mathbf{B}}$ 和 $\hat{\mathbf{C}}$ 分别由式 (18) 和 (19) 计算, 而非线性网络重构的矩阵元 $\hat{\mathbf{B}}$ 和 $\hat{\mathbf{C}}$ 由式 (39) 和 (40) 的非线性基之间的关联计算给出. 另外动力

学方程(1)中的非线性项并不影响噪声的计算, 与式(30)完全同样的分析和计算也可以由可测数据重构作用在非线性网络节点上的白噪声的统计关联矩阵 $\hat{Q}$.

现在整个非线性网络重构式(42)的计算中还有未定的参数是式(41)级数展开截断值 q . 在式(32)展开收敛的条件下, 可以用如下自恰验证的方法来确定 q . 首先取任意 $q \geq N$, 采用此 q 可以用式(42)计算 $\hat{\mathbf{y}}(q)$; 然后再任取 $q' > q$ 的基, 而且 q' 中包含了 q 中的全部基. 由 q' 也可计算矩阵 $\hat{\mathbf{y}}(q')$, 矩阵 $\hat{\mathbf{y}}(q')$ 中包含了 $\hat{\mathbf{y}}(q)$ 中所有矩阵元, 同时还增加了一部分 $q < \mu < q'$ 矩阵元. 如果

$$\begin{cases} \hat{\mathbf{y}}_{i,\mu}(q') \simeq \hat{\mathbf{y}}_{i,\mu}(q), & 1 \leq \mu \leq q; \\ \hat{\mathbf{y}}_{i,\mu}(q') \simeq 0, & q < \mu \leq q', \end{cases} \quad (43)$$

则截断 q 就被自恰地验证正确, 因为从 q 出发再增加展开项的个数并不会显著改变重构结果. 这里 q 的选择显然并不唯一, 取决于我们对重构计算精度的要求. 在实际操作中我们会选取在满足重构精度条件下尽可能小的 q 值, 这样既简化计算过程, 又在物理上更好地理解实际问题中关键的非线性特征及重要相互作用. (38)–(43)式提供了完备的算式从数据式(35)出发重构非线性网络式(1)的运算. 下面用几个例子来验证整个重构过程.

首先, 我们研究洛伦兹方程(Lorenz)的重构问题.

$$\begin{cases} \frac{dx(t)}{dt} = \sigma(y - x) + \Gamma_1, \\ \frac{dy(t)}{dt} = x(\rho - z) - y + \Gamma_2, \\ \frac{dz(t)}{dt} = xy - \beta z + \Gamma_3. \end{cases} \quad (44)$$

该方程可以看作是 $N=3$ 的非线性网络. 我们可以测量节点的输出时间序列 $x(t_k), y(t_k), z(t_k)$; $k = 1, 2, \dots, L, L \gg 1$, 任务是从这些可测数据重构整个动力学方程(44). 图3(a)和(b)中画出 $x(t), y(t)$ 的时间序列及其在相平面的投影. 非线性网络重构首先要选择(32)式的基函数. 在无法获取网络动力学特征相关先验知识的情况下, 常常可以用幂级数展

开作为初步尝试. 我们选择基函数

$$\mathbf{Y} = (1, x, y, z, xy, xz, yx, x^2, y^2, z^2, xyz, x^2y, \dots). \quad (45)$$

利用(43)式来寻找恰当的截断同时求解 $\hat{\mathbf{y}}$ 及重构方程(44). 在图3(c)和(d)中, 利用幂级数展开时分别比较了变量的不同幂次展开的重构结果, 发现幂级数展开至二次所得矩阵元 $\hat{\mathbf{y}}_{i,\mu}(q' = 10)$ 与幂级数展开至一次矩阵元 $\hat{\mathbf{y}}(q = 4)$ 数值差别较大(图3(c)), 但通

图 3 (网络版彩图) 重构洛伦兹(Lorenz)方程(44). 参数值分别为 $\sigma=10, \rho=50, \beta=8/3, Q_{ij}=50\delta_{ij}$. (a), (b) 方程(44)的时间序列与相平面投影. (c), (d) 利用式(43)自恰检验. (c) 式(44)中级数展开至一次幂($q=4$)与级数展开至二次幂($q=10$)的对比. (d) (44)式中级数展开至二次幂($q=10$)与级数展开至三次幂($q=20$)的对比. 从中看出级数展开至少至二次幂才满足式(43)自恰检验的条件. (e), (f) 利用 $q=10$ 重构式(44)后, 具有噪声与滤去噪声后的重构结果在相平面上的投影. 引自文献[39]

Figure 3 (Color online) Reconstruction of Lorenz equation (44). Parameters are taken as $\sigma=10, \rho=50, \beta=8/3, Q_{ij}=50\delta_{ij}$. (a) and (b) Time sequences and trajectories of noise-driven Lorenz system. Data are chaotic and strongly random. (c) and (d) Demonstration of self-consistent checking of our method, and its application in inference computations of eq. (44). (c) Reconstruction results obtained by eq. (43) with the truncated bases of the zero and the first orders of powers ($q=4$) plotted against those with power bases truncated up to the second order ($q=10$). (d) The same as (c) with power bases truncated at the third order ($q=20$) plotted against those truncated at the second order ($q=10$). The truncation at $q=10$ is self-consistently confirmed to be suitable. (e) and (f) The reconstructed trajectories of noise-driven Lorenz system (e) and the noise-free Lorenz system (f). Quoted from ref. [39].

过比较幂级数展开至三次所得矩阵元 $\hat{\mathbf{A}}_{i,\mu}(q'=20)$ 与展开至二次幂级数所得矩阵元 $\hat{\mathbf{A}}_{i,\mu}(q=10)$ 数值, 发现两个矩阵对应元素值差别很小(图3(d)), 另外 $\hat{\mathbf{A}}_{i,\mu}(q'=20)$ 中 $q>10$ 的其余元素均接近于零. 由(43)式的自恰截断条件判断 $q=10$ 二次展开是合适的截断位置, 利用 $q=10$ 截断可近似重构方程(44), 图3(e)和(f)分别给出了重构系统在具有噪声与去除噪声后情况下的相平面投影图, 原测量数据(图3(b))和Lorenz吸引子的特征得到很好的重现.

我们进一步研究高维的非线性网络.

$$\frac{dx_i(t)}{dt} = \Phi_i(x_i(t)) + \sum_{j=1, j \neq i}^N W_{ij} x_j(t) + \Gamma_i(t), \quad i = 1, 2, \dots, N, \quad (46)$$

其中

$$\Phi_i(x_i) = a_i x_i + b_i x_i^3 + c_i e^{-x_i}.$$

这里假设相互作用具有线性的特点, 但节点动力学是非线性的. 同样采用幂级数展开,

$$\begin{aligned} Y_i(t) &= (Y_1(t), Y_2(t), \dots, Y_{M_i}(t)) \\ &= (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_N, 1, x_i, x_i^2, \dots, x_i^{m_i}). \end{aligned} \quad (47)$$

不同于(44)式的是, (46)式中幂级数展开有无穷多项, 所以截断成为必要操作. 我们在图4(a)–(c)中画出了当幂级数展开至不同幂次 m_i 时, 利用(43)式自恰验证的过程, 图4(d)给出了 $m_i = 3$ 截断下的重构结果. 整个数值模拟很好地验证了从式(35)到(42)重构运算的正确性.

上面两个模型中都采用了幂级数展开作为基函数, 但在实际问题中我们常常也会采用不同的非线性基. 例如, 在以下例子中

$$\frac{d\theta_i}{dt} = \omega_i + \Phi_i(\theta_i) + \sum_{j \neq i}^N \Psi_{ij}(\theta_j - \theta_i) + \Gamma_i(t), \quad i = 1, 2, \dots, N, \quad (48)$$

其中

$$\theta_i + 2k\pi = \theta_i, \quad k = \pm 1, \pm 2, \dots, \pm p, \dots, \quad i = 1, 2, \dots, N.$$

Φ_i 与 Ψ_{ij} 为未知的非线性函数

图 4 (网络版彩图)利用(43)式自恰验证重构高维非线性网络式(46)的结果. 参数值为 $N = 50, a_i \in (-1, 1), b_i \in (2, 5), c_i \in (0, 2), W_{ij} \in (0.5, 1), Q_{ij} \in \delta_{ij}(0.5, 1)$. (a)–(c) 当基函数(47)分别展开至 $m_i = 1, 2, 3$ 时与展开至 $m_i + 1$ 高次幂重构结果的对比. 可以发现只有基函数展开至三次幂时才满足(43)式自恰验证条件. (d) 基函数展开至三次幂时重构结果与真实网络连接的对比, 此时可以很好地重构网络连接. 引自文献[39]

Figure 4 (Color online) The same as Figure 3 with network (46) of $N = 50$ considered. Parameters are taken with uniform distributions in the intervals $a_i \in (-1, 1), b_i \in (2, 5)$ and $c_i \in (0, 2)$ and white noises $Q_{ij} \in \delta_{ij}(0.5, 1)$. The interaction intensities are $W_{ij} \in (0.5, 1)$ with 10% probability for positive ones; $W_{ij} \in (-1, -0.5)$ also with 10% for negative ones; and $W_{ij} = 0$ otherwise. Power bases are organized from low-order to higher-order ones as described in eq. (47). (a) Results for $m_i = 2$ plotted vs. those for $m_i = 1$. Triangles are considerably away from the diagonal line, representing incorrect truncation at $m_i = 1$. (b) The same as (a) with results of $m_i = 3$ plotted with those of $m_i = 2$. The truncation at $m_i = 2$ is not suitable either. (c) Results for $m_i = 4$ plotted vs. those for $m_i = 3$. Both results coincide with each other fairly well, that self-consistently confirms the correctness of our method for sufficiently large $m_i, m_i \geq 3$. (d) The results for $m_i = 3$ plotted vs. the actual values of $A_{i,\mu}$. With suitable nonlinearity considered, all nonlinear and interacting structures are inferred correctly with certain fluctuations. Quoted from ref. [39].

$$\begin{aligned} \Phi_i(\theta_i + 2k\pi) &= \Phi_i(\theta_i); \quad \Psi_{ij}(\theta_j - \theta_i + 2k\pi) \\ &= \Psi_{ij}(\theta_j - \theta_i). \end{aligned} \quad (49)$$

这里取

$$\Phi_i(\theta_i) = \sum_{k=1}^3 [\alpha_{i,k} \sin(k\theta_i) + \alpha'_{i,k} \cos(k\theta_i)], \quad (50)$$

$$\begin{aligned} \Psi_{ij}(\theta_j - \theta_i) &= \sum_{k=1}^3 [\beta_{ij,k} \sin(k(\theta_j - \theta_i)) \\ &\quad + \beta'_{ij,k} \cos(k(\theta_j - \theta_i))]. \end{aligned} \quad (51)$$

研究角度相关的网络系统, 对其动力学方程同样可以采用幂级数展开但收敛很慢. 直接的傅里叶基展

开常常更有效, 采用基展开

$$\begin{aligned}
 Y_i = & (1, \sin(\theta_i), \cos(\theta_i), \dots, \sin(m_i\theta_i), \cos(m_i\theta_i), \dots, \\
 & \sin(\varphi_1), \cos(\varphi_1), \dots, \sin(m_i\varphi_1), \cos(m_i\varphi_1), \\
 & \sin(\varphi_2), \cos(\varphi_2), \dots, \sin(m_i\varphi_2), \cos(m_i\varphi_2), \dots, \\
 & \sin(\varphi_{i-1}), \cos(\varphi_{i-1}), \dots, \sin(m_i\varphi_{i-1}), \cos(m_i\varphi_{i-1}), \\
 & \sin(\varphi_{i+1}), \cos(\varphi_{i+1}), \dots, \sin(m_i\varphi_{i+1}), \cos(m_i\varphi_{i+1}), \dots, \\
 & \sin(\varphi_N), \cos(\varphi_N), \dots, \sin(m_i\varphi_N), \cos(m_i\varphi_N)),
 \end{aligned} \quad (52)$$

其中,

$$\varphi_j = \theta_j - \theta_i, j = 1, 2, \dots, N; j \neq i.$$

在图5(a)–(c)中我们采用(43)式自恰验证的方法确定 $m_i = 3$ 是很好的截断, 在图5(d)中利用 $m_i = 3$ 截断并计算式(42)中相应矩阵元的方法能够很好地重构高维非线性网络(48).

5 数据缺失条件下的网络重构

在以上两节的讨论中我们都假设网络中所有节点的输出数据都可以测量. 这个假设似乎是必要的, 因为相互连通的网络中所有节点通过连接互相影响, 缺少了部分节点的数据会严重阻碍对已测节点在网络中的动力学和连接结构的重构, 但这一困难并不是不可克服的. 数据(6)加上(12)的时间标记包含了极大的信息量, 尽管数据(6)在 $M < N$ 时有缺失, 但在充分发掘可测数据的信息后, 这一缺失可以补充, 从而网络结构和动力学可以在数据缺失的条件下得到准确重构. 例如我们可以利用式(12)的时间标记对式(6)进行数据扩张.

$$\begin{cases} (x_{i_1}(t_k), x_{i_2}(t_k), \dots, x_{i_M}(t_k)), \\ (x_{i_1}^{(1)}(t_k), x_{i_2}^{(1)}(t_k), \dots, x_{i_M}^{(1)}(t_k)), \\ \vdots \\ (x_{i_1}^{(\gamma)}(t_k), x_{i_2}^{(\gamma)}(t_k), \dots, x_{i_M}^{(\gamma)}(t_k)), \\ \vdots \\ (x_{i_1}^{(\nu)}(t_k), x_{i_2}^{(\nu)}(t_k), \dots, x_{i_M}^{(\nu)}(t_k)), \\ k = 1, 2, \dots, L - \nu, \end{cases} \quad (53)$$

图 5 (网络版彩图)与图4做法相同, (a)–(c)利用(43)式自恰验证合适的截断值, 并在(d)中画出重构高维非线性网络(48)的结果. 具体参数值为 $N = 10, \omega_i \in (1, 2), [\alpha_{i,k}, \alpha'_{i,k}, \beta_{i,j,k}, \beta'_{i,j,k}] \in (-1, 1), i, j = 1, 2, \dots, N; k = 1, 2, 3$. 引自文献[39]

Figure 5 (Color online) The same as Figure 4 with network (48) reconstructed by using Fourier basis set. $N = 10$, and other parameters are uniformly distributed on $\omega_i \in (1, 2), [\alpha_{i,k}, \alpha'_{i,k}, \beta_{i,j,k}, \beta'_{i,j,k}] \in (-1, 1), i, j = 1, 2, \dots, N; k = 1, 2, 3$. Truncations are made from low-order harmonics to higher-order ones. (a), (b) The same as Figure 4(a) with phase dynamics eq. (48) considered, and the results of $m_i = 2$ plotted vs. those of $m_i = 1$ in (a); $m_i = 3$ vs. $m_i = 2$ in (b). Without high-order harmonics, many dots distribute away from the diagonal lines. (c) The results of $m_i = 4$ plotted vs. those of $m_i = 3$. All dots are distributed around the diagonal line, indicating self-consistently the suitable truncation at $m_i = 3$. (d) The same as Figure 4(d) with eq. (48) computed. All computed results at $m_i = 3$ coincide with those of the actual values with certain fluctuations. Quoted from ref. [39].

其中

$$x_{i_\mu}^{(\gamma)}(t_k) = \frac{x_{i_\mu}^{(\gamma-1)}(t_{k+1}) - x_{i_\mu}^{(\gamma-1)}(t_k)}{\Delta t}, \quad (54)$$

$$\gamma = 1, 2, \dots, \nu; \mu = 1, 2, \dots, M.$$

这里要强调的是尽管数据(53)都来自数据(6)和(12), 但由于时间标签的使用, 所有各阶导数(包括零阶导数为原数据本身)都包含互相独立的信息, 用来有效弥补缺失信息对网络重构的影响. 由式(53)和(54)可以把式(1)网络中可测量的 M 个节点的数据扩展到 νM 个相互独立的方程($\nu M > N$)

$$\begin{cases} x_{i_1}^{(\mu)}(t) = \Phi_{i_1}(\mathbf{x}(t)^M, \mathbf{x}^{(1)}(t)^M, \dots, \mathbf{x}^{(\nu-1)}(t)^M) + \Gamma_{i_1}^{(\mu)}, \\ x_{i_2}^{(\mu)}(t) = \Phi_{i_2}(\mathbf{x}(t)^M, \mathbf{x}^{(1)}(t)^M, \dots, \mathbf{x}^{(\nu-1)}(t)^M) + \Gamma_{i_2}^{(\mu)}, \\ \vdots \\ x_{i_M}^{(\mu)}(t) = \Phi_{i_M}(\mathbf{x}(t)^M, \mathbf{x}^{(1)}(t)^M, \dots, \mathbf{x}^{(\nu-1)}(t)^M) + \Gamma_{i_M}^{(\mu)}, \\ \mu = 1, 2, \dots, \nu. \end{cases} \quad (55)$$

在这个方程中左边的所有项和右边的所有变量均可以从已测数据中得到或算出, 而所有函数关系均未知并且是重构的目标. 式(55)右边的噪声项具有奇异结构, 但均有很短的关联时间的特征, 前面的去噪声关联计算(21)的方法都可以适用. 另外可测数据变量中包含的噪声行为仍可以用 M 维的(30)式算出.

综上所述, (55)式满足(1)式网络重构的全部要求, 只要数据扩张的程度足够, 即满足

$$vM \geq N \quad (56)$$

的条件即可. 由于 N 是未知的变量, 该条件满足与否也要依赖自恰性检验来确定, 其操作与上一节确定截断近似的(43)式的作法类似, 这里就不再重复. 以下我们简单假设已找到合适的扩张量 v (即 v 足够小), 但又满足完备性条件(56)式. (55)式的网络结构可以解决缺失部分节点的输出数据对可测节点影响的问题, 从而准确重构可测节点的网络结构和动力学行为. 但这个重构当然不能也不可能定义和重构隐藏节点本身的网络行为, 这一点在以下不再强调.

我们考虑由FitzHugh-Nagumo (FHN)方程描述的神经元耦合的复杂网络[55,56]

$$\frac{dv_i}{dt} = v_i - \frac{1}{3}v_i^3 - \omega_i + c_1 + \sum_{j=1}^N W_{ij}v_j + \Gamma_{i,1}, \quad (57)$$

$$\frac{d\omega_i}{dt} = \delta(v_i - b\omega_i + c_2) + \Gamma_{i,2}, \quad (58)$$

其中 v_i, ω_i 分别为第 i 个神经元的膜电压变量和对应的恢复变量, FHN模型为Hodgkin-Huxley模型的简化版本. 在式(57)中我们考虑电突触耦合的神经网络, 因此不同神经元 v_i 变量通过扩散效应彼此耦合. 这些 v_i 变量在实验中容易测得, 而恢复变量 ω_i 的数值难以获取. 现在的问题是能否在这种情况下准确重构网络中可测节点相关的参数以及更为重要的网络连接作用 W_{ij} .

应用式(55)的方法, 对(57)式求导并代换 ω_i 变量, 可以得到

$$\begin{aligned} \frac{d^2v_i}{dt^2} = & (1 - \delta b)\dot{v}_i - v_i^2\dot{v}_i + \delta(b - 1)v_i - \frac{1}{3}\delta b v_i^3 \\ & + \delta(bc_1 - c_2) + \delta b \sum_{j=1}^N W_{ij}v_j + \sum_{j=1}^N W_{ij}\dot{v}_j \end{aligned}$$

$$+ \delta b \Gamma_{i,1} - \delta \Gamma_{i,2} + \Gamma'_{i,1}. \quad (59)$$

现在上式看作非线性展开的话, 我们的基矢量为

$$Y_i = (1, \dot{v}_1, \dots, \dot{v}_i, \dots, \dot{v}_N, v_1, \dots, v_i, \dots, v_N, \dot{v}_i^2, v_i^3), \quad (60)$$

对应的待定系数 $\mathfrak{A}_{ij}$ 为

$$\begin{aligned} (\mathfrak{A}_{i,1:2N+3}) = & (\delta(bc_1 - c_2), W_{i1}, \dots, W_{ii} + (1 - \delta b), \\ & \dots, W_{iN}, \delta b W_{i1}, \dots, \delta b W_{ii} + \delta(b - 1), \\ & \dots, \delta b W_{iN}, -1, -\frac{1}{3}\delta b). \end{aligned} \quad (61)$$

这里我们指出, 原网络(57)的连接结构其实是包含在我们所重构的 $\mathfrak{A}$ 矩阵(61)式之中, 即

$$W_{ij} = \mathfrak{A}_{i,j+1}, i \neq j. \quad (62)$$

图6应用(53)式数据扩展的方式重构噪声驱动

图 6 (网络版彩图)应用(53)式数据扩展的方式重构噪声驱动的FHN神经元耦合网络(57). 我们在数据扩展中只使用了 v_i 变量, 另外的 ω_i 变量并未涉及. 所有节点参数值为 $b = 0.8, \delta = 0.08, c_1 = 0, c_2 = 0.7, W_{ij} \in [0, 1)$. (a)–(c)网络规模 $N = 10$. (a)网络中单个节点时间序列. (b)和(c)分别不应用和应用平滑技术(63)时重构结果与真实结果对比. 对于类似(a)中强噪声影响的数据, 只有应用平滑技术(63) (平滑窗口 $\Omega = 40$)才可以很好地重构(57)网络. (d)应用平滑技术的重构网络(57) ($N = 100$). 引自文献[48]

Figure 6 (Color online) Reconstruction of white-noise-driven FHN neural network (57) using the algorithm (53) with all ω_i variables hidden. We use $b = 0.8, \delta = 0.08, c_1 = 0, c_2 = 0.7, W_{ij} \in [0, 1)$ for all nodes. (a) The time sequence for one node in the FHN network; (b) and (c) comparison of the parameters $\mathfrak{A}_{ij}$ reconstructed using eq. (53) with the actual parameters $\mathfrak{A}_{ij}$ without smoothing and with smoothing under eq. (63), respectively, the network size is $N = 10$, the smoothing window size in (c) is $\Omega = 40$, the results in (b) are not satisfactory, while in (c) the reconstruction accuracy is considerably better; (d) the same as (c) for a larger network with $N = 100$. Quoted from ref. [48].

的FHN神经元耦合网络(57). 我们在数据扩展中只使用了 v_i 变量, 另外的 ω_i 变量并未涉及. 图6(a)画出了网络中单个节点时间序列. 图6(b)应用式(42)重构结果与真实结果对比. 可以发现重构结果与真实网络连接差别较大, 这是因为(59)式的求导操作使得噪声强度提高了 $1/\Delta t$ 倍. 因此对于有限的数据量, 式(21)的去噪声关联操作可能显现出较大的波动. 可以对窗口 Ω 内的测量数据 $D(t)$, 采用平滑技术

$$\bar{D}(t_k) = \frac{1}{\Omega} \sum_{l=1}^{\Omega} D(t_{k+l-1}) \quad (63)$$

降低噪声的影响, 这里数据 $D(t)$ 可以是计算(39)和(40)式中 x_i , $\dot{x}_i$, $\ddot{x}_i$ 及其他任何可测数据. 对于类似图6(a)中强噪声影响的数据, 合理应用平滑技术(63)可以很好地重构(57)网络(图6(c)). 图6(d)应用平滑技术(63)重构 $N=100$ 网络(57), 同样得到很好的结果.

由(55)式可见, 为了解决隐藏节点的数据缺失问题, 引入高频采样的变量数据, 并利用各阶导数进行有效数据的扩张, 其导数阶数 ν 取决于不可测量节点数目在网络总节点数目中所占的比例($\nu > N/M$). 这个比例越小, 所要求计算的导数阶数越高. 由于采样频率总是有限的, ν 越大高阶导数的计算误差就越大, 所以(55)式网络重构不能应用于隐藏节点数目过多情况下的网络重构问题.

例如在图7(a)的网络中只能测量两个节点的数据($x_A(t)$ 和 $x_B(t)$), 而其他的($N-2 \gg 1$)所有网络节点均被隐藏. (55)式要求计算 $N/2 \gg 1$ 阶的导数, 这实际上不可能也并不可行. 现在我们讨论在图7(a)仅仅测量A和B节点数据($x_A(t)$ 和 $x_B(t)$)的条件下, 重构A对B(或B对A)的连接作用. 这时(1)式的 N 个方程中只有A和B节点的两个方程

$$\begin{aligned} \frac{dx_A(t)}{dt} &= \Phi_A(x_A(t)) + W_{AB}(x_A(t), x_B(t)) \\ &+ \sum_{j \neq A, B}^N W_{Aj}(x_A(t), x_j(t)), \end{aligned} \quad (64a)$$

$$\begin{aligned} \frac{dx_B(t)}{dt} &= \Phi_B(x_B(t)) + W_{BA}(x_B(t), x_A(t)) \\ &+ \sum_{j \neq A, B}^N W_{Bj}(x_B(t), x_j(t)), \end{aligned} \quad (64b)$$

图7 (网络版彩图)噪声的影响提高目标相互作用示意图. (a) 在连通的网络中所有节点互相影响, 如果仅能测量A和B节点的数据而网络中所有其他节点均为隐藏节点, 此时无法仅仅用测得的这两个节点的数据来重构这两个节点之间的作用(如节点A对节点B的作用). (b) 网络(a)中子网络的放大图示, 显示了对B节点作用的所有节点. 可以清楚地看到从A到B的作用掩盖在众多隐藏节点对B节点的作用之中. (c) 如果对A节点施加一定的外界驱动, A对其他节点的作用就可能在众多未知的作用中突显出来. (d) 如果A节点受到的驱动作用使得A对B的作用大大增强从而在众多隐藏作用中显示出来, 在该情形下仅获取A和B这两个节点的数据就可能推断A对B的作用. 快速变化的噪声能够提供这样的驱动, 使我们在数据极端缺失的条件下重构两个节点之间的相互作用. 引自文献[47]

Figure 7 (Color online) Schematic figures on noise enhancement of certain targeted interactions. (a) In connected networks all nodes influence each other, and the data of nodes A and B are measurable while all other nodes (black nodes) are hidden. Interaction from A to B can never be inferred with data of nodes A and B only ((A, B) can be any pair of nodes in the network). (b) An amplified figure of subnetwork of (a) with all interactions to node B shown. Interaction from A to B submerges in many hidden (black) interactions. (c) All interactions from node A are visible in the black dark interaction sea with certain external driving acting to node A. (d) Interaction from A to B in (b) is enhanced by the driving at node A, so that the contribution of the interaction from A to B is greatly larger than those of all other interactions from the hidden nodes and thus computable with data of A and B only. Fast-varying noises can serve as such driving, called fast-varying noise driving method. Quoted from ref. [47].

可以用来分析重构问题. 这里我们只考虑了两两节点之间的非线性相互作用(如节点A对节点B的作用 $W_{BA}(x_B, x_A)$). 注意该方程中除了A和B节点的数据起作用外, 网络中大量不可测节点的数据也都参与到方程的演化之中.

我们的问题是在方程(64)式(图7(a))中, 已知 $x_A(t)$ 和 $x_B(t)$ 高质量的数据(充分长和足够高的测量频率), 我们能否重构A节点对B节点的作用. 答案是否定的, 因为B节点还受到其他大量隐藏节点的作用. 无法获取它们的数据, 这些隐藏节点对B节点的作用就不能确定(见图7(b)), 从而A对B的作用也不能独立确定. 但对可测节点动力学施加一定的外部注入影响, 则可以完成这一似乎不能完成的任务, 图7(c)和(d)说明了其中合理的逻辑.

下面我们将看到在具有一定性质的噪声的帮助下我们就可能在(64)式中仅从可测的数据 $x_A(t)$ 和 $x_B(t)$ (即不知道所有隐藏节点的数据($x_i(t), i \neq A, B$))出发确定A对B的作用. 在网络重构中我们熟悉许多处理噪声的方法, 绝大多数方法都致力于“去噪声”(如(21)式), 尽可能地减少和消除噪声对重构网络结构的负面影响. 但以下部分我们将看到噪声也可能对重构起及其重要的积极作用, 在给定节点上的噪声有可能帮助重构与该节点相关的相互作用^[57]. 给节点A加上快变噪声(近似为白噪声), (64)式变为

$$\frac{dx_A(t)}{dt} = \Phi_A(x_A(t)) + W_{AB}(x_A(t), x_B(t)) + \sum_{j \neq A, B}^N W_{Aj}(x_A(t), x_j(t)) + \Gamma_A(t), \quad (65a)$$

$$\frac{dx_B(t)}{dt} = \Phi_B(x_B(t)) + W_{BA}(x_B(t), x_A(t)) + \sum_{j \neq A, B}^N W_{Bj}(x_B(t), x_j(t)), \quad (65b)$$

$$\langle \Gamma_A(t) \rangle = 0, \langle \Gamma_A(t) \Gamma_A(t') \rangle = Q_A \delta(t - t'). \quad (65c)$$

对(65b)式求导可以得到

$$\begin{aligned} \frac{d^2 x_B(t)}{dt^2} &= \frac{d\Phi_B(x_B(t))}{dt} + \frac{dW_{BA}(x_B(t), x_A(t))}{dt} \\ &+ \sum_{j \neq A, B}^N \frac{dW_{Bj}(x_B(t), x_j(t))}{dt} \\ &= \frac{\partial \Phi_B(x_B(t))}{\partial x_B(t)} \dot{x}_B(t) + \frac{\partial W_{BA}(x_B(t), x_A(t))}{\partial x_B(t)} \dot{x}_B(t) \\ &+ \frac{\partial W_{BA}(x_B(t), x_A(t))}{\partial x_A(t)} \dot{x}_A(t) \end{aligned}$$

$$\begin{aligned} &+ \sum_{j \neq A, B}^N \frac{\partial W_{Bj}(x_B(t), x_j(t))}{\partial x_B(t)} \dot{x}_B(t) \\ &+ \sum_{j \neq A, B}^N \frac{\partial W_{Bj}(x_B(t), x_j(t))}{\partial x_j(t)} \dot{x}_j(t). \end{aligned} \quad (66)$$

应用测得的数据(12), (66)式可以实际计算如下

$$\begin{aligned} \ddot{x}_B(t_k) &= \frac{1}{2} \left[\frac{\partial \Phi_B(x_B(t_k))}{\partial x_B(t_k)} + \frac{\partial W_{BA}(x_B(t_k), x_A(t_k))}{\partial x_B(t_k)} \right] \dot{x}_B(t_k) \\ &+ \frac{1}{2} \left[\frac{\partial \Phi_B(x_B(t_k))}{\partial x_B(t_k)} + \frac{\partial W_{BA}(x_B(t_k), x_A(t_k))}{\partial x_B(t_k)} \right] \\ &\times \dot{x}_B(t_{k+1}) + \frac{1}{2} \frac{\partial W_{BA}(x_B(t_k), x_A(t_k))}{\partial x_A(t_k)} \dot{x}_A(t_k) \\ &+ \frac{1}{2} \frac{\partial W_{BA}(x_B(t_k), x_A(t_k))}{\partial x_A(t_k)} \dot{x}_A(t_{k+1}) \\ &+ \sum_{j \neq A, B}^N \frac{1}{2} \frac{\partial W_{Bj}(x_B(t_k), x_j(t_k))}{\partial x_B(t_k)} (\dot{x}_B(t_k) + \dot{x}_B(t_{k+1})) \\ &+ \sum_{j \neq A, B}^N \frac{1}{2} \frac{\partial W_{Bj}(x_B(t_k), x_j(t_k))}{\partial x_j(t_k)} (\dot{x}_j(t_k) + \dot{x}_j(t_{k+1})). \end{aligned} \quad (67)$$

注意在(65)到(67)式中我们仅仅知道可测量的 $x_A(t)$ 和 $x_B(t)$ 的数据, 当然由于时间标签同时也知道了它们对 t 的一阶和二阶导数. 但我们不知道所有隐藏节点的数据($x_i(t), i \neq A, B$), 当然也不知道所有动力学函数 $\Phi_i(x_i)$, $W_{ij}(x_i, x_j)$ 的形式. 用 $\dot{x}_A(t_k)$ 乘以(67)式的右边并计算对应的关联, 可以得到

$$\begin{aligned} \langle \ddot{x}_B(t_k) \dot{x}_A(t_k) \rangle &= \frac{1}{2} \left\langle \frac{\partial W_{BA}(x_B(t_k), x_A(t_k))}{\partial x_A(t_k)} \dot{x}_A(t_k) \dot{x}_A(t_k) \right\rangle \\ &+ \frac{1}{2} \left\langle \frac{\partial W_{BA}(x_B(t_k), x_A(t_k))}{\partial x_A(t_k)} \dot{x}_A(t_{k+1}) \dot{x}_A(t_k) \right\rangle \\ &+ \frac{1}{2} \left\langle \frac{\partial \Phi_B(x_B(t_k))}{\partial x_B(t_k)} \dot{x}_B(t_k) \dot{x}_A(t_k) \right\rangle \\ &+ \frac{1}{2} \left\langle \frac{\partial W_{BA}(x_B(t_k), x_A(t_k))}{\partial x_B(t_k)} \dot{x}_B(t_k) \dot{x}_A(t_k) \right\rangle \\ &+ \frac{1}{2} \left\langle \frac{\partial \Phi_B(x_B(t_k))}{\partial x_B(t_k)} \dot{x}_B(t_{k+1}) \dot{x}_A(t_k) \right\rangle \\ &+ \frac{1}{2} \left\langle \frac{\partial W_{BA}(x_B(t_k), x_A(t_k))}{\partial x_B(t_k)} \dot{x}_B(t_{k+1}) \dot{x}_A(t_k) \right\rangle \\ &+ \sum_{j \neq A, B}^N \frac{1}{2} \left\langle \frac{\partial W_{Bj}(x_B(t_k), x_j(t_k))}{\partial x_B(t_k)} (\dot{x}_B(t_k) + \dot{x}_B(t_{k+1})) \dot{x}_A(t_k) \right\rangle \end{aligned}$$

$$+ \sum_{j \neq A, B}^N \frac{1}{2} \left\langle \frac{\partial W_{Bj}(x_B(t_k), x_j(t_k))}{\partial x_j(t_k)} (\dot{x}_j(t_k) + \dot{x}_j(t_{k+1})) \dot{x}_A(t_k) \right\rangle. \quad (68)$$

现在可以分析上式中不同项的量级. 所有变量被限制在有限的区域内, 除A节点外所有节点的变化速率($\dot{x}_i(t)$, $i \neq A$)也都是有限的, 在白噪声的驱动下A节点输出数据变化很快并能够到达很高的绝对值. 因而, 上式中第一个关联项比其余各项要大得多, 由(30)式可以得到

$$\langle \dot{x}_A(t_k) \dot{x}_A(t_k) \rangle = \frac{Q_A}{\Delta t} = O\left(\frac{1}{\Delta t}\right), \quad (69)$$

而(68)式中其余各关联项的量级为1或 Q_A 量级(Q_A 的量级为 $O(1)$). 在高频测量($\Delta t \ll 1$)和白噪声作用的情况下, 第一项的量级比其余各项要大得多. 取(68)式中主要量级的项, 可以由式(68)得到

$$\langle \ddot{x}_B(t_k) \dot{x}_A(t_k) \rangle \doteq \frac{1}{2} \left\langle \frac{\partial W_{BA}(x_B(t_k), x_A(t_k))}{\partial x_A(t_k)} \dot{x}_A(t_k) \dot{x}_A(t_k) \right\rangle. \quad (70)$$

对于非线性相互作用我们可以用式(32)基函数展开的方法由(70)式确定中非线性函数 $W_{BA}(x_B(t_k), x_A(t_k))$ 中各个展开项的系数, 而对于简单的线性作用 $W_{BA}(x_B(t_k), x_A(t_k)) = W_{BA} x_A(t_k)$ 可以得到

$$W_{BA} = 2 \frac{\langle \ddot{x}_B(t_k) \dot{x}_A(t_k) \rangle}{\langle \dot{x}_A(t_k) \dot{x}_A(t_k) \rangle}. \quad (71)$$

上述推导的最关键之处是通过作用在A节点上的快变噪声显著增强了A节点对其直接作用的所有节点的连接强度(图7(c)); 另一个关键步骤是选择合适关联计算式($\ddot{x}_B(t_k) \dot{x}_A(t_k)$)来突出A对B的相互作用贡献(图7(d)), 从而在只获取A和B节点数据而缺失所有隐藏节点数据的条件下重构了A对B的作用.

我们考虑如下耦合Rössler振子方程

$$\begin{cases} \dot{x}_i(t) = -\omega_i y_i(t) - z_i(t) + \sum_{j=1}^N W_{ij}(x_j(t) - x_i(t)), \\ \dot{y}_i(t) = \omega_i x_i(t) + a_i y_i(t) + \sum_{j=1}^N W_{ij}(y_j(t) - y_i(t)), \\ \dot{z}_i(t) = b_i + z_i(t)(x_i(t) - c_i) + \sum_{j=1}^N W_{ij}(z_j(t) - z_i(t)), \\ i = 1, 2, \dots, N, \end{cases} \quad (72)$$

其中方程参数在图8中给出. 对于给定的参数, 网络动力学是非常高维的非同步混沌运动. 假定在(72)式中我们仅能测得两组时间序列($x_A(t)$ 和 $x_B(t)$)并且能够对数据的产生影响, 比如能够通过A节点上施加快变的白噪声而影响 $x_A(t)$ 的数据.

$$\dot{x}_A(t) = -\omega_A y_A(t) - z_A(t) + \sum_{j=1}^N W_{Aj}(x_j(t) - x_A(t)) + \Gamma(t), \quad (73a)$$

$$\dot{x}_B(t) = -\omega_B y_B(t) - z_B(t) + \sum_{j=1}^N W_{Bj}(x_j(t) - x_B(t)), \quad (73b)$$

$$\langle \Gamma(t) \rangle = 0, \langle \Gamma(t) \Gamma(t') \rangle = Q_A \delta(t - t'). \quad (73c)$$

利用噪声驱动的动力学(73)式替代(72)式中 $x_A(t)$ 的时间演化规律并模拟网络动力学演化. 在仅获取A和B节点数据的条件下, 我们可以用式(71)推断网络结构. 我们考虑网络中不同组合的(A, B)节点组, 并对所有节点组只获取这两个节点的数据, 并且在A节点施加噪声影响, 就可以推断出整个网络的连接作用.

图8(a)和(b)分别为在有噪声驱动和无噪声情况下的推断结果(W_{ij} 是实际网络连接强度, W'_{ij} 是用(71)式重构的计算). 在A节点有快变噪声 $Q_A=50$ 注入时, (71)式可以很好重构网络(73b). 图8(c)和(d)数据量与噪声强度对重构结果的影响, 其中D为(31)定义的重构精度.

6 总结

本文综述了通过分析动力学网络产生的数据重构网络连接结构和系统演化规律的问题. 在如何处理网络结构复杂性, 噪声影响, 动力学非线性及测量数据缺失等常见困难方面提出了具体的处理方法.

作为总结首先要强调的是在所有这些重构问题中统计关联计算的重要性, 选取恰当的关联量是关乎重构成功与否的关键. 例如, (19)式的变量与变量之间以及(18)式变量与速度之间的关联计算使式(23)的简捷重构计算成为可能; 式(18)与(25)两种稍微不同时差的关联计算实现去噪声(式(21))和计

图 8 (网络版彩图)耦合混沌非同步Rössler 振子方程(式(72)和(73)). 参数选取值为 $N = 100$, $\omega_i \in (0.50, 1.50)$, $a_i \in (0.10, 0.30)$, $b_i \in (0.10, 0.30)$, $c_i \in (5.0, 15.0)$, $W_{ij} \in (0.05, 0.10)$, 网络连接概率为 $p = 0.30$. 在网络的所有变量中, 我们仅能获取A和B节点数据, 并利用(71)式重构A对B的连接作用. (a), (b) 分别在有噪声驱动和无噪声情况下的推断结果(W_{ij} 是实际网络连接强度, W'_{ij} 是用(71)式重构的计算). (a) 在A节点有快变噪声 $Q_A = 50$ 注入时, (71)式可以很好地重构网络相互作用. (c), (d) 数据量与噪声强度对重构结果的影响, 其中 D 为(31)定义的重构精度. 引自文献[47]

Figure 8 (Color online) Interaction inference in coupled chaotic and nonsynchronous Rössler oscillators ($N = 100$) with all nodes hidden except two nodes, driving node A and driven node B of targeted interactions. Parameters are $\omega_i \in (0.50, 1.50)$, $a_i \in (0.10, 0.30)$, $b_i \in (0.10, 0.30)$, $c_i \in (5.0, 15.0)$, $W_{ij} \in (0.05, 0.10)$, link probability $p = 0.30$. (a) Interactions inferred by eq. (71) W'_{ij} plotted against actual interaction values W_{ij} . Noise driving in eq. (73) is used with $Q_A = 50.0$ ($\Delta t = 0.001$ and $L = 20000$). All interactions are reconstructed fairly accurately though only two variables $x_A(t)$ and $x_B(t)$ are measured in the 300-dimensional and densely connected network of eq. (72). (b) The same as (a) without noise driving ($Q_A = 0$). Same computations as (a) give results far away from the diagonal line, and the reconstructed links are independent of actual interactions. (c) Divergence error D defined in eq. (31) plotted against L with all other parameters given in (a), where D decreases as L sharply for small L . (d) D plotted against Q_A with other parameters given in (a). A valley of divergence error given by optimal noise intensity can be observed. Quoted from ref. [47].

算噪声的运算(式(30)); 非线性动力学关系的基展开(式(32))及基函数之间的高阶关联计算(式(39)和(40))将非线性网络的重构简化为线性代数运

算(42). 另外对加速度和更高阶时间导数的关联计算实现了数据缺失(式(55))或严重缺失(式(70))条件下对目标结构的准确重构等.

总结要强调的第二个问题是对数据内包含信息的充分发掘和利用. 例如, (12)式不仅给出了 NL 个数据, 而且给出了这些数据的时间排列. 这些数据和时间标签的结合包含了比 NL 个数大得多的信息, 可以产生出新的独立的数据. (55)式正是利用这种可能性通过计算 ν 阶导数将 ML ($M < N$)个数据扩展到 νML 个独立并且完备的数据从而在未知隐藏节点信息的条件下准确重构 M 个节点之间的相互作用连接结构.

总结的第三个要点是噪声的作用与利用. 通常人们认为噪声因为其未知并且对动力学演化产生影响, 是重构中要利用各种手段消除的负面因素. 这种观点并不总是正确的, 有时具有特定性质的噪声能作为十分积极的因素用来帮助对网络的准确重构. 本文讨论了其中一类问题, 即当网络中只有少量(例如只有两个)可测节点时, 加在节点A上的快变噪声(可以用(5)式白噪声近似)可以极大放大该节点对其直接作用的任一相邻节点(如B节点)的有向作用, 从而在B节点所受的众多隐藏节点的作用中把A对B的作用准确重构出来. 值得指出的是, 作用在A上的快变噪声可以在原网络动力学中自然存在, 也可以在实际系统中较容易和有目的地注入. 类似利用噪声来帮助克服动力学网络重构困难的方法和技术有广阔的应用前景.

网络重构问题在自然科学和社会科学的广阔领域具有重大意义, 同时由于其面临的诸多困难在理论研究中又极具吸引力, 它是当前大数据分析的一个重要领域. 本文只是讨论了如何处理网络重构中几个常见的困难, 而向更深层次的拓展和将这些方法应用于实际数据分析和真实系统重构将是我们面临的重大挑战.

参考文献

- Covert M W, Knight E M, Reed J L, et al. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 2004, 429: 92–96
- Feist A M, Herrg M J, Thiele I, et al. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*, 2009, 7: 129–143

- 3 Bar-Joseph Z, Gitter A, Simon I. Studying and modelling dynamic biological processes using time-series gene expression data. *Nat Rev Genet*, 2012, 13: 552–564
- 4 Watts D J, Strogatz S H. Collective dynamics of “small-world” networks. *Nature*, 1998, 393: 440–442
- 5 Barabatz A. Emergence of scaling in random networks. *Science*, 1999, 286: 509–512
- 6 Strogatz S H. Exploring complex networks. *Nature*, 2001, 410: 268–276
- 7 Boccaletti S, Latora V, Moreno Y, et al. Complex networks: Structure and dynamics. *Phys Rep*, 2006, 424: 175–308
- 8 Caldarelli G, Chessa A, Pammolli F, et al. Reconstructing a credit network. *Nat Phys*, 2013, 9: 125–126
- 9 Chen G R. The China power grid: A network science perspective. *Nat Sci Rev*, 2014, 1: 368–370
- 10 Bullmore E, Sporns O. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, 2009, 10: 186–198
- 11 Glass L, Mackey M C. From Clocks to Chaos: The Rhythms of Life. Princeton: Princeton University Press, 1988
- 12 Goldbeter A. Biochemical Oscillations and Cellular Rhythms. Cambridge: Cambridge University Press, 1996
- 13 Tsai T Y C, Choi Y S, Ma W, et al. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science*, 2008, 321: 126–129
- 14 Tsonis A A, Deyle E R, May R M, et al. Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature. *Proc Natl Acad Sci USA*, 2015, 112: 3253–3256
- 15 Bongard J, Lipson H. From the cover: Automated reverse engineering of nonlinear dynamical systems. *Proc Natl Acad Sci USA*, 2007, 104: 9943–9948
- 16 Sugihara G, May R, Ye H, et al. Detecting causality in complex ecosystems. *Science*, 2012, 338: 496–500
- 17 Marbach D, Costello J C, Kuffner R, et al. Wisdom of crowds for robust gene network inference. *Nat Methods*, 2012, 9: 796–804
- 18 Wang W X, Lai Y C, Grebogi C. Data based identification and prediction of nonlinear and complex dynamical systems. *Phys Rep*, 2016, 644: 1–76
- 19 Casadiego J, Nitzan M, Hallerberg S, et al. Model-free inference of direct network interactions from nonlinear collective dynamics. *Nat Commun*, 2017, 8: 2192
- 20 Nitzan M, Casadiego J, Timme M. Revealing physical interaction networks from statistics of collective dynamics. *Sci Adv*, 2017, 3: e1600396
- 21 Yeung M K S, Tegner J, Collins J J. Reverse engineering gene networks using singular value decomposition and robust regression. *Proc Natl Acad Sci USA*, 2002, 99: 6163–6168
- 22 Timme M. Revealing network connectivity from response dynamics. *Phys Rev Lett*, 2007, 98: 224101
- 23 Wang W X, Yang R, Lai Y C, et al. Predicting catastrophes in nonlinear dynamical systems by compressive sensing. *Phys Rev Lett*, 2011, 106: 154101
- 24 Villaverde A, Ross J, Banga J. Reverse engineering cellular networks with information theoretic methods. *Cells*, 2013, 2: 306–329
- 25 Brunton S L, Proctor J L, Kutz J N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc Natl Acad Sci USA*, 2016, 113: 3932–3937
- 26 Roebroeck A, Formisano E, Goebel R. Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage*, 2005, 25: 230–242
- 27 Stuart J M. A gene-coexpression network for global discovery of conserved genetic modules. *Science*, 2003, 302: 249–255
- 28 Vicente R, Wibral M, Lindner M, et al. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J Comput Neurosci*, 2011, 30: 45–67
- 29 Cestnik R, Rosenblum M. Reconstructing networks of pulse-coupled oscillators from spike trains. *Phys Rev E*, 2017, 96: 012209
- 30 Pikovsky A. Reconstruction of a neural network from a time series of firing rates. *Phys Rev E*, 2016, 93: 062313
- 31 Napoletani D, Sauer T D. Reconstructing the topology of sparsely connected dynamical networks. *Phys Rev E*, 2008, 77: 026103
- 32 Timme M, Casadiego J. Revealing networks from dynamics: An introduction. *J Phys A-Math Theor*, 2014, 47: 343001
- 33 Elowitz M B, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*, 2000, 403: 335–338
- 34 Wang W X, Chen Q, Huang L, et al. Scaling of noisy fluctuations in complex networks and applications to network prediction. *Phys Rev E*, 2009, 80: 016116
- 35 Ren J, Wang W X, Li B, et al. Noise bridges dynamical correlation and topology in coupled oscillator networks. *Phys Rev Lett*, 2010, 104: 058701
- 36 Ching E S C, Lai P Y, Leung C Y. Extracting connectivity from dynamics of networks with uniform bidirectional coupling. *Phys Rev E*, 2013, 88: 042817
- 37 Ching E S C, Lai P Y, Leung C Y. Reconstructing weighted networks from dynamics. *Phys Rev E*, 2015, 91: 030801

- 38 Zhang Z, Zheng Z, Niu H, et al. Solving the inverse problem of noise-driven dynamic networks. [Phys Rev E](#), 2015, 91: 012814
- 39 Chen Y, Zhang Z, Chen T, et al. Reconstruction of noise-driven nonlinear networks from node outputs by using high-order correlations. [Sci Rep](#), 2017, 7: 44639
- 40 Ching E S C, Tam H C. Reconstructing links in directed networks from noisy dynamics. [Phys Rev E](#), 2017, 95: 010301
- 41 Lai P Y. Reconstructing network topology and coupling strengths in directed networks of discrete-time dynamics. [Phys Rev E](#), 2017, 95: 022311
- 42 Newman M E J. Network structure from rich but noisy data. [Nat Phys](#), 2018, 14: 542–545
- 43 Wu X, Wang W, Zheng W X. Inferring topologies of complex networks with hidden variables. [Phys Rev E](#), 2012, 86: 046106
- 44 Su R Q, Wang W X, Lai Y C. Detecting hidden nodes in complex networks from time series. [Phys Rev E](#), 2012, 85: 065201
- 45 Su R Q, Lai Y C, Wang X, et al. Uncovering hidden nodes in complex networks in the presence of noise. [Sci Rep](#), 2015, 4: 3944
- 46 Zhang C, Chen Y, Hu G. Network reconstructions with partially available data. [Front Phys](#), 2017, 12: 128906
- 47 Zhang C, Chen Y, Hu G. Inference of targeted interactions of networks with data of driving and driven nodes only by applying fast-varying noise signals. [Phys Lett A](#), 2017, 381: 2502–2509
- 48 Chen Y, Zhang C Y, Chen T Y, et al. Reconstruction of noise-driven nonlinear dynamic networks with some hidden nodes. [Sci China-Phys Mech Astron](#), 2017, 60: 070511
- 49 Ching E S C, Tam P H. Effects of hidden nodes on the reconstruction of bidirectional networks. [Phys Rev E](#), 2018, 98: 062318
- 50 Chen Y, Wang S, Zheng Z, et al. Depicting network structures from variable data produced by unknown colored-noise driven dynamics. [Europhy Lett](#), 2016, 113: 18005
- 51 Wang W X, Ren J, Lai Y C, et al. Reverse engineering of complex dynamical networks in the presence of time-delayed interactions based on noisy time series. [Chaos](#), 2012, 22: 033131
- 52 Sysoev I V, Prokhorov M D, Ponomarenko V I, et al. Reconstruction of ensembles of coupled time-delay systems from time series. [Phys Rev E](#), 2014, 89: 062911
- 53 Sysoev I V, Ponomarenko V I, Kulminskiy D D, et al. Recovery of couplings and parameters of elements in networks of time-delay systems from time series. [Phys Rev E](#), 2016, 94: 052207
- 54 Risken H. The Fokker-Planck Equation. Heidelberg: Springer-Verlag, 1984
- 55 FitzHugh R. Impulses and physiological states in theoretical models of nerve membrane. [Biophys J](#), 1961, 1: 445–466
- 56 Mi Y, Liao X, Huang X, et al. Long-period rhythmic synchronous firing in a scale-free network. [Proc Natl Acad Sci USA](#), 2013, 110: E4931–E4936
- 57 Jiang J J, Huang Z G, Huang L, et al. Directed dynamical influence is more detectable with noise. [Sci Rep](#), 2016, 6: 24088

From data to network structure——Reconstruction of dynamic networks

ZHANG ZhaoYang^{1,2}, CHEN Yang³, MI YuanYuan⁴ & HU Gang^{5*}

¹*Department of Physics, School of Physical Science and Technology, Ningbo University, Ningbo 315211, China;*

²*Business School, Ningbo University, Ningbo 315211, China;*

³*Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China;*

⁴*Center for Neurointelligence, Chongqing University, Chongqing 400044, China;*

⁵*Department of Physics, Beijing Normal University, Beijing 100875, China*

In recent decades, large scale of data sets have been accumulated in various and wide fields, in particular in social and biological systems. There are massive amounts of data available for utilization; and for really using these data it is crucial to find effective methods of analyses. Many practical systems can be described by dynamic networks, for which modern technique can measure their outputs, and accumulate extremely rich data. Nevertheless, the network structures producing these data are often deeply hidden. The structure and intensity of couplings between the network elements (say, network nodes) play key roles in determining the features of network dynamics. Therefore, the problem of network reconstruction, i.e., exploring unknown network structures by analyzing available output network data, has attracted significant interest in many interdisciplinary fields in recent decades. The significance of network reconstruction is due to not only the extreme importance of black-box manifestation in practice, but also serious difficulties, which are theoretically challenging, and appearing in various aspects of real-world systems, such as complexity of network structures, strong nonlinearity of network dynamics, diverse and unknown impacts from the interiors of nodes and the externals of networks, i.e., presence of noises, many hidden and not measurable nodes in networks. In this review, we introduce some effective methods dealing with the above difficulties, such as exploring information of available data by searching rich and diverse data features; seeking different types of correlations aiming at treating different difficulties; finding intelligent ways to remove negative and utilize active effects of noises, and so on. Our approach offers some possibilities towards understanding of complex networked systems in more complicated and difficult conditions and has important implications for the reconstruction tasks of many practical networks.

network reconstruction, nonlinear dynamics, data analysis

PACS: 89.75.Hc, 05.10.Gg, 05.45.Tp

doi: [10.1360/SSPMA2019-0127](https://doi.org/10.1360/SSPMA2019-0127)