

基于动力学同步的复杂网络结构识别速度研究*

杨浦 郑志刚[†]

(北京师范大学物理学系, 北京 - 香港 - 新加坡非线性与复杂性研究联合中心, 北京 100875)

(2012年3月1日收到; 2012年3月16日收到修改稿)

基于动力学同步的复杂网络结构识别是探测复杂网络拓扑性质的重要方面, 其中识别速度是一个重要但鲜有讨论问题。首先对弱耦合条件下耦合非线性振子网络结构识别速度的问题进行了研究。发现识别速度随耦合强度成正比增长。通过解析讨论, 肯定了这一关系是普适的。之后基于我们最近提出的反复驱动识别方法, 将处于同步稳定态的耦合区域也纳入研究范围。在这种情形下存在一个最佳的时间片段的长度使识别速度达到极值。这些结论加深了对时间序列中蕴含的拓扑结构信息量的理解。

关键词: 复杂网络, 拓扑识别, 自适应反馈, 收敛速度

PACS: 05.45.Tp, 89.75.-k, 02.60.Cb

1 引言

复杂网络动力学行为与其拓扑结构之间关系的反问题, 即由已知的网络动力学行为来反推出此网络的拓扑结构, 近几年来逐渐成为复杂网络动力学方面的一个热点, 被广泛深入地讨论^[1], 有很多有效的识别方法被提出。它们总体上可以分为两大类: 一类侧重于识别网络的微观结构, 即识别每条边是否存在及权重大小^[2-5]; 另一类侧重于识别网络结构的宏观性质, 如度分布、聚集系数等^[6-8]。

迄今为止, 大部分方法的研究主要集中于识别方法在理论上的收敛性和普适性等核心问题; 而对识别过程的快慢鲜有讨论。若某一识别方法虽然在理论上肯定其是收敛的但收敛速度过慢, 实际应用上此方法等同于无效。比如网络的拓扑结构是间歇性变化的; 再比如受条件限制, 只能获取有限长度的时间序列。第二种情形主要出现在网络内有稳定同步态的情况。因为网络内部同步是拓扑结构识别的障碍^[9], 这时只有从非同步态向同步态过渡的暂态时间序列才是有效的^[10]。在只有有限长时间序

列可以利用的条件下, 对识别速度的了解就显得尤为重要。知道识别速度依赖的参数, 就可以通过调节此参数来提高识别速度。如果此参数不能调节, 也能知道对给定时间序列长度能达到的识别精度的量级。了解识别速度的依赖条件对提高识别效率也会有很大帮助。所以识别速度是复杂网络拓扑结构识别研究中一个不可忽视的问题。

基于此种考虑, 本文通过自适应反馈识别算法^[2,9]对识别速度在非同步区域和同步区域分别做了探讨。弱耦合的同步态都不稳定, 而稳定同步态的耦合都较强^[11]。为后文叙述方便, 我们约定强耦合即指耦合强度处于同步稳定区域。对于处于非同步区的强耦合情况会在文章最后一节中讨论。

在相对容易处理的弱耦合区域, 我们依据自适应反馈算法的识别误差随时间大体是指数衰减的, 将衰减指数定义为此算法的识别速度。得到识别速度是随耦合强度幂指数增长的一个普适结论, 并在理论上做了分析讨论。基于我们最近提出的反复驱动识别方法可以将识别速度的定义扩展至强耦合区间。发现驱动片段长度与趋向同步的弛豫时间相当时, 识别速度最快。

* 国家自然科学基金(批准号: 10875011, 11075016), 973 项目基金(批准号: 2007CB814805), 教育部博士点基金(批准号: 20100003110007) 和中央高校基本科研业务费专项资金资助的课题。

† E-mail: zgzheng@bnu.edu.cn

2 基于自适应反馈同步的网络结构识别

Yu 等^[2]于 2006 年提出的自适应反馈同步的方法成功地对网络拓扑微观结构进行识别。下面做简要介绍，具体细节请参看文献[9]。假设有一个网络，节点个数为 N ，其动力学演化方程为

$$\dot{\mathbf{x}}_i = \mathbf{F}_i(\mathbf{x}_i) - s \sum_{j=1}^N c_{ij} \mathbf{H}(\mathbf{x}_j). \quad (1)$$

其中， $i, j = 1, 2, 3, \dots, N$ ； \mathbf{x}_i 是第 i 个节点的状态； \mathbf{F}_i 是第 i 个节点孤立时的动力学演化方程； \mathbf{H} 为输出函数； s 是耦合强度； $\mathbf{C} = \{c_{ij}\}_{N \times N}$ ，是描述网络结构的拉普拉斯矩阵或称邻接矩阵。我们假设该网络系统的拓扑结构未知；已知条件是节点的动力学 $\mathbf{F}_i(\mathbf{x}_i)$ 和输出函数 \mathbf{H} 。

为了能够在已知时间序列 $\{\mathbf{x}_i(t)\}$ 的情况下反推出网络的邻接矩阵 \mathbf{C} ，我们可以由同样 N 个振子组成响应网络，而将待探测系统(1)作为驱动系统。响应网络中节点之间的连接是随时间变化的，其第 i 个节点的动力学演化方程为

$$\dot{\mathbf{y}}_i = \mathbf{F}_i(\mathbf{x}_i) - s \sum_{j=1}^N d_{ij}(t) \mathbf{H}(\mathbf{y}_j) + \mathbf{u}_i, \quad (2)$$

其中

$$\begin{aligned} \dot{d}_{ij} &= -\mathbf{e}_i^T \mathbf{H}(\mathbf{y}_j), \\ \mathbf{u}_i &= k_i \mathbf{e}_i, \\ \mathbf{e}_i &= \mathbf{y}_i - \mathbf{x}_i, \\ \dot{k}_i &= \frac{1}{2} \|\mathbf{e}_i\|^2. \end{aligned} \quad (3)$$

$d_{ij}(t)$ 或 $\mathbf{D}(t) = \{d_{ij}(t)\}_{N \times N}$ 称为估计子。可以证明，如果 N 个时间序列 $\{\mathbf{H}(\mathbf{x}_i)\}$ 完全线性独立，即有估计子 $\lim_{t \rightarrow \infty} \mathbf{D}(t) = \mathbf{C}$ 。驱动系统的未知网络结构被完全识别。

(2) 式中各变量初始条件的选取并不影响识别的收敛性。为了减少会影响收敛速度的因素，规定初始条件始终选取 $\mathbf{y}_i(0) = \mathbf{x}_i(0)$, $d_{ij}(0) = 0$ 。因为在响应系统的动力学方程(2)中隐含有随时间变化的外加驱动 \mathbf{x}_i ，所以在数值模拟中采用 2 阶龙格库塔算法，时间步长为 0.002。选取一个 4 节点的有向网络作为驱动系统的网络结构，

$$\mathbf{C} = \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & -2 & 4 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix}.$$

不失一般性，在以下数值模拟中节点均用 Lorenz 振子^[12]: $\mathbf{F}_i(\mathbf{x}) = [10(x_2 - x_1), 28x_1 - x_1x_3 - x_2, x_1x_2 - 8/3x_3]$ ，其中 $i = 1, 2, 3, 4$ ；输出函数 $\mathbf{H}(\mathbf{x}) = (x_1, 0, 0)$ 。

3 弱耦合条件下的识别速度

为了衡量识别速度，先定义估计子的平均相对识别误差

$$d_{\text{err}} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \frac{|d_{ij} - c_{ij}|}{c'_{ij}},$$

其中， $c'_{ij} = \max\{|c_{ij}|, \tilde{c}\}$ ， \tilde{c} 等于 $|\mathbf{C}|$ 中最小的非零元。平均是对估计子的所有矩阵元进行平均。在下文中还意味着对不同时间序列的样本做系综平均，仍记作 d_{err} 。

由于网络系统内部同步是识别网络结构的障碍，我们先考虑弱耦合这种处于非同步区域的简单情况。通过数值模拟多种耦合方式，我们发现在估计子达到稳定值前平均相对识别误差 d_{err} 都是随时间指数递减的(图 1)。即在识别误差达到极限值之前有关系

$$d_{\text{err}} \propto e^{-vt}, \quad (4)$$

其中 v 是一个参数。在耦合强度相同时，不同线性耦合的 Lorenz 网络收敛速度近似相等；耦合强度不同时，收敛速度差别明显。图 1 中耦合强度 $s = 0.1$ 的识别速度明显比 $s = 0.001$ 高两个数量级。对于其他的耦合方式和不同振子(如 Rössler 振子^[13]、Chen 振子^[14])，平均相对识别误差 d_{err} 随时间变化的行为也都是指数衰减的。这样就可以定义衰减指数 v 为自适应反馈算法的识别速度。

下面我们来考察一下识别速度 v 和耦合强度 s 之间的依赖关系(图 2)。Lorenz 网络中四种线性耦合方式的曲线(a1—a4)十分接近，其中(a2)和(a4)几乎重合在一起。(a5)曲线对应的耦合方式为非线性输出，输出时间序列的数量级高于线性输出，所以远离前四条曲线；而(a6)对应的是 Rössler 振子，

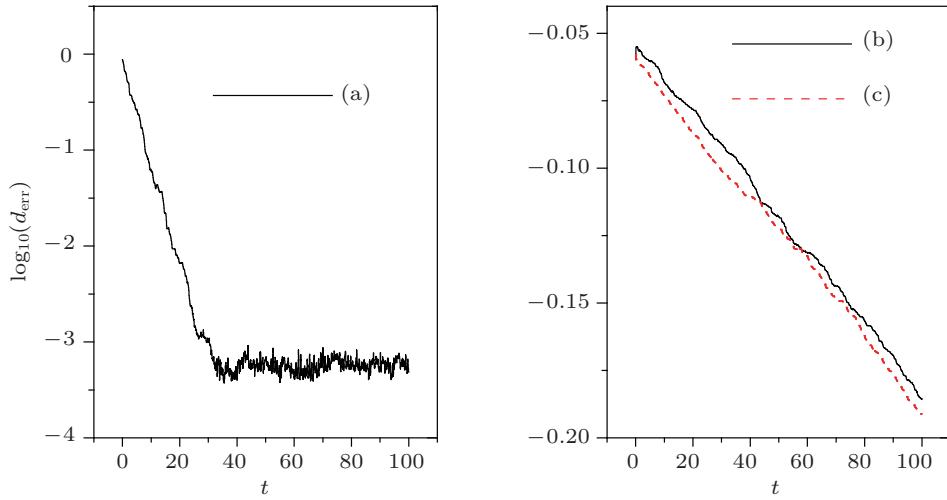


图 1 两图中的三条曲线均为 Lorenz 四节点网络的平均相对误差随时间变化的曲线. 三条曲线对应的耦合强度 s 、输出函数 $\mathbf{H}(\mathbf{x})$ 和斜率分别为 (a) 0.1, $(x_1, 0, 0)$, -0.103 ; (b) 0.001, $(0, 0, x_1)$, -0.00129 ; (c) 0.001, $(x_1, 0, 0)$, -0.00128

也远离前四条曲线. 但是这六条曲线都显示识别速度随耦合强度的增加而加快, 而且都服从的幂律关系

$$v(s) \propto s^\gamma. \quad (5)$$

这间接说明在弱耦合条件下, 单位时间序列中所包含的网络拓扑结构的信息量与耦合强度正相关. 我们还对 Chen 振子等其他振子网络也进行了数值验证, 全都满足幂律关系.

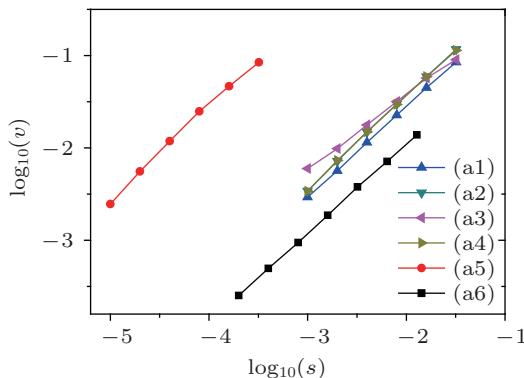


图 2 识别速度 v 与耦合强度 s 的依赖关系 (a1)–(a5) 都是由 Lorenz 振子组成的网络, 但耦合方式 $\mathbf{H}(\mathbf{x})$ 不同, 依次为 $(x_1, 0, 0)$; $(x_2, 0, 0)$; $(x_3, 0, 0)$; $(0, 0, x_2)$ 和 $(x_1 * x_2, 0, 0)$; (a6) 是由 Rössler 振子组成的网络, 输出函数 $\mathbf{H}(\mathbf{x}) = (x_1, 0, 0)$. 曲线 (a3) 的斜率 $\gamma \approx 0.82$, 其他曲线的斜率 $\gamma = 1 \pm 0.04$

4 对识别速度与耦合强度之间关系的分析

下面对弱耦合的数值结果的做一些定性的理论分析, 来进一步理解得到的数值结果. 为分析方

便, 考虑振子为一维的情形. 响应系统与驱动系统的差异演化方程为

$$\dot{e}_i = F_i(x_i) - F_i(y_i) + s \sum_{j=1}^N [c_{ij} H(x_j) - d_{ij}(t) H(y_j)] - k_i e_i. \quad (6)$$

若 k_i 初值取得充分大, k_i 近似为常数^[2,9].

由于平均相对误差 d_{err} 的收敛行为和估计子中最慢的收敛行为等价, 假设 d_{mn} 是收敛最慢的估计子且满足

$$d_{mn}(t) = c_{mn} + b_{mn} \exp[-v_{mn}(s)t],$$

其中 b_{mn} 为常数, $v_{mn}(s)$ 是估计子 d_{mn} 依赖于耦合强度 s 的收敛速度. 令上式对时间 t 求导, 有

$$\dot{d}_{mn} = -v_{mn}(s)b_{mn} \exp[-v_{mn}(s)t]. \quad (7)$$

而 \dot{d}_{mn} 在 (3) 式中有原始定义. 对 (6) 式做一阶近似并积分后, 将 e_m 代入原始定义并与 (7) 式联立, 可以得到关于识别速度的二次方程

$$v_{mn}^2(s) - \alpha_m(s)v_{mn}(s) - s\beta_n^2 = 0,$$

其中, $\alpha_m(s)$ 是 s 的一次式, 由 $F'(x_m), H'(x_m), k_m$ 和 c_{mm} 决定; β_n 不依赖于 s , 由 $H(y_n)$ 决定. 在系统平均的意义下, 所有变量都与时间 t 无关. $\alpha_m(s)$ 与 β_n 的具体表达形式不影响后面的定性讨论. 因为耦合强度 $s > 0$, 所以此方程一定存在正负两个实根. 而 $v_{mn} > 0$, 仅保留正根

$$v_{mn}(s) = 1/2[\alpha_m(s) + \sqrt{\alpha_m^2(s) + 4\beta_n^2 s}]. \quad (8)$$

识别速度 v_{mn} 与耦合强度 s 近似有正比关系。这与数值模拟结果符合。在误差是指数收敛的前提下，关系(8)是普适的，不依赖于具体的振子动力学和输出函数的具体形式。

5 反复驱动识别方法的识别速度

如果用来驱动的时间序列是从非同步态演化到同步态的过渡暂态，一般来说平均识别误差随时间的演化过程是复杂的，而且会存在多个非零极限稳定值（例如图3(b)）。这些因素都给我们定义识别速度带来困难。我们在文献^{*}中提出可以反复利用同一段过渡暂态时间序列进行识别，并在理论和数值模拟中都肯定了这种反复驱动识别方法的有效性。

具体的反复驱动识别方法如下：设已有长度为 Δt 的时间序列 $\{\mathbf{x}_i(t)\}$ 。响应系统的构造不变。在第 l 次驱动的初始时刻，令 $\mathbf{y}_i^l(0) = \mathbf{x}_i(0)$ ， $d_{ij}^{l+1}(0) = d_{ij}^l(\Delta t)$ 。当 l 趋于无穷时，有 $\mathbf{D}^l(\Delta t) \rightarrow$

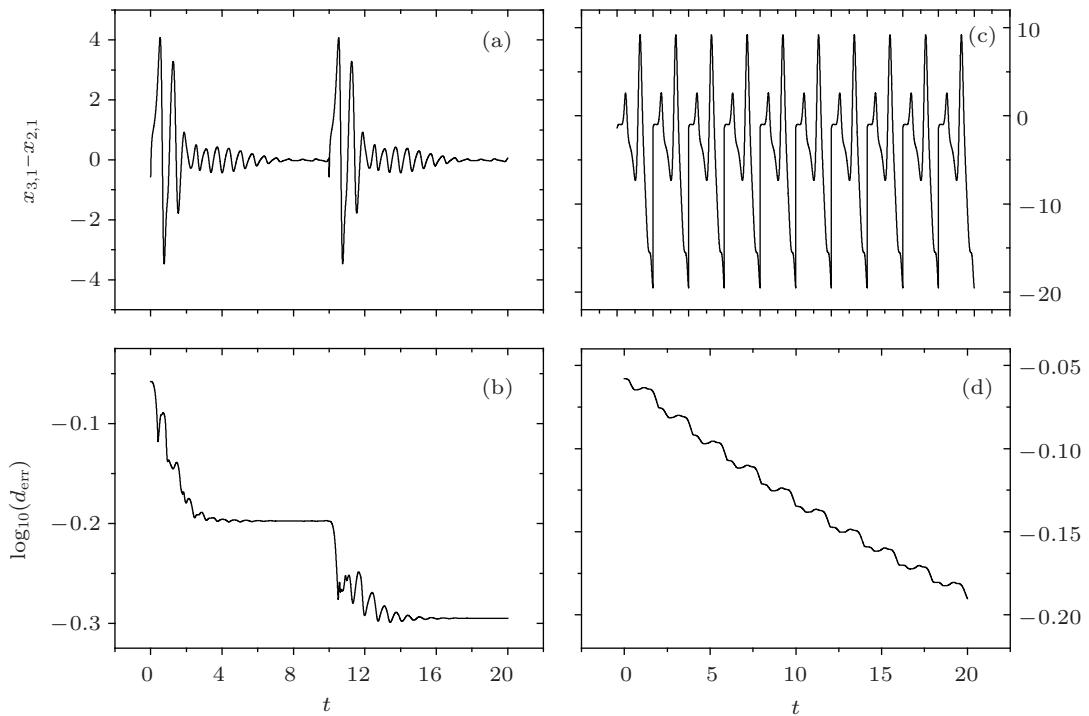


图3 用反复驱动识别方法识别非同步态的时间序列。在(a)和(b)中 $s = 4$ ，驱动网络处于同步稳定态。在(c)和(d)中 $s = 0.01$ ，驱动网络处于非同步态。(a)和(c)用驱动网络中两节点的状态差示意反复用来驱动的时间序列 $\{\mathbf{x}_i\}$ 的长度和同步情况。 $x_{i,j}$ 表示第 i 个节点的第 j 个分量。从(b)可以看出，在(a)的驱动下误差 $\log_{10}(d_{\text{err}})$ 在驱动网络同步后停止演化，但再次驱动后会下降相同的幅度。从(d)可以看出，在(c)每次驱动后误差 $\log_{10}(d_{\text{err}})$ 也会下降相同的幅度

C. 如果驱动网络虽然始终处于非同步态但只能取得的时间序列长度有限，识别速度又不能提高，估计子不能够达到我们要求的精度。这时利用同一段时间序列进行重复驱动的方法可以延长有效时间序列提高识别精度。总之，若有效的时间序列不是充分长而又不对识别用时做要求，就可以利用反复驱动方法成功识别。

图3中给出强、弱耦合两种情况下用反复驱动方法识别的效果。单独一次驱动时，误差 d_{err} 的演化并不规则。但是从每次驱动的误差末值看，或者从多次驱动的整体效果看，误差 d_{err} 是随时间指数递减的，类似图1中 d_{err} 的行为。这就解决了在驱动网络处于同步稳定态时定义识别速度 v 的两个困难。1) 通过反复驱动方法解决了识别能否成功的问题；2) 从远大于时间片段长度的宏观时间尺度来看平均相对误差随时间演化的行为是指数衰减的。这样，在反复驱动的方式下仍然可以用(4)式定义识别速度 v 。

下面考察对于固定的耦合强度，不同长度的驱

*Yang Pu and Zheng Zhi-Gang Topological identification of networks by using asynchronous transient(unpublished).

动时间序列对于识别速度的影响。驱动系统处于弱耦合强度条件时的情况比较简单。在图 4(a) 中, 曲线斜率的绝对值, 即识别速度, 始终随驱动时间片段长度的增加而增加。这是因为耦合较弱且时间序列始终不同步, 片段越长则包含的拓扑结构信息越多, 识别的速度就越快。片段趋于无穷时, 识别速度趋于极限。

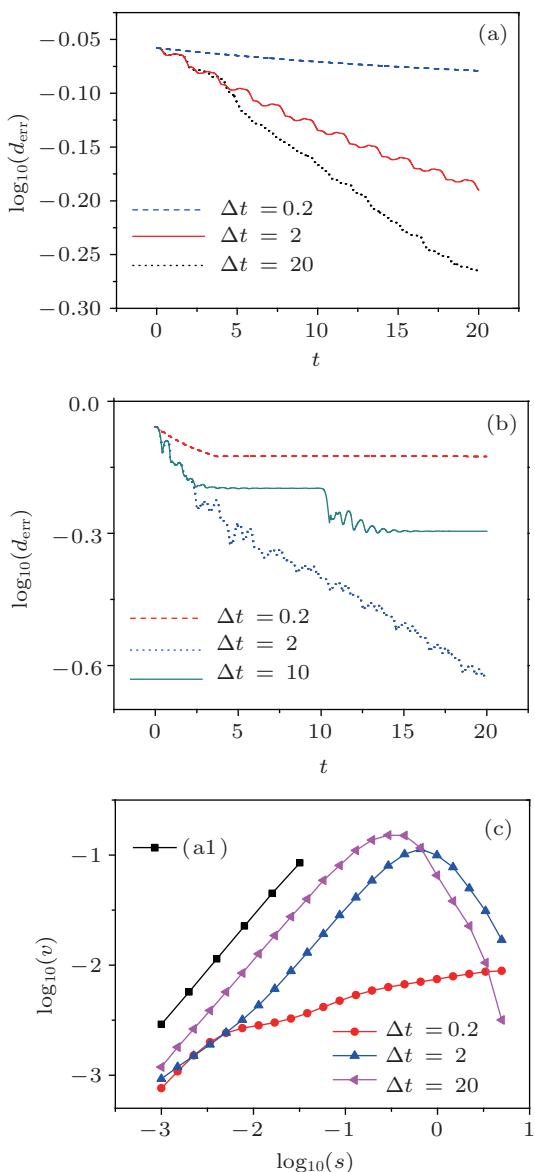


图 4 考查不同驱动时间序列长度 Δt 及不同耦合强度 s 对识别速度的影响 (a) 耦合强度 $s = 0.01$, 识别速度随驱动时间序列的长度增加; (b) 耦合强度 $s = 4$, 存在一个适中的驱动时间序列长度使识别速度最快; (c) 给出用不同长度 ($\Delta t = 0.2, 2, 20$) 的时间序列片段驱动时, 识别速度随耦合强度变化图, 并将图 2 中的曲线 (a1) 附在图中以供比较

强耦合时情况则不同。时间序列从非同步态经过依赖耦合强度和网络动力学的弛豫时间达到同步态。而同步态中不包含有关的驱动网络拓扑结构

的信息。用来驱动的时间片段如果远大于弛豫时间, 则多出来的时间就是无效的, 会减慢识别速度; 如果远小于弛豫时间, 片段中包含的网络拓扑结构信息就少, 也会减慢识别速度。两方面竞争的结果就是对于固定的耦合强度存在一个最优的时间序列片段长度, 它的量级与网络内达到同步的弛豫时间相同。

在图 4 的系统中, $s = 0.3$ 是网络出现同步的临界值。从图 4(b) 可以明显看出, 当 $\Delta t = 10$ 时, 在每次驱动的后半段估计子停止演化; 当 $\Delta t = 0.2$ 时, 由于估计子随时间演化不规则导致识别速度慢, 甚至识别失败; 当 $\Delta t = 2$ 时, 既充分利用了有效时间片段又没有包含多余的同步态时间序列, 识别速度最快。在图 3(a) 中显示在相同的动力学网络中 $s = 4$ 时, 弛豫时间约为 2, 与识别速度最快的时间片段长度相当。

对于固定长度的时间序列 (如 $\Delta t = 20$), 识别速度与耦合强度在弱耦合区域有良好的正幂指数关系 (见图 4(c)), 和图 2 中 (a1) 的行为类似; 在强耦合区域 ($s > 0.5$), 虽然识别速度与耦合强度仍有良好的幂指数关系, 但是负幂指数, 单调下降。随着耦合强度的增大, 在此网络系统中逐渐出现同步, 并且弛豫时间逐渐缩短。在中间的过渡区域, 同步时间长度逐渐取代单位时间内时间序列包含拓扑结构信息量成为影响识别速度的主要因素。识别速度由上升变为下降, 曲线出现拐点。

随着驱动时间序列长度的减小, 曲线的峰值逐渐向右移动。当 $\Delta t = 0.2$ 时, 曲线在整个图示区间呈严格单调上升趋势。这说明单位时间序列包含的网络拓扑结构信息与耦合强度正相关不仅在弱耦合范围内适用, 在强耦合区域同样适用。

6 结 论

收敛速度是一个重要的问题, 但很少有人讨论。原因是很难就此进行理论上普适的探讨。它对具体情况的依赖性较大。我们通过对多种振子, 多种耦合方式做大量数值模拟, 对自适应反馈识别算法识别的收敛速度进行了细致的研究, 对估计子的收敛情况得到一些新的认识。

在弱耦合条件下, 识别平均相对误差随时间指数衰减; 衰减指数 (即识别速度) 与耦合强度有正比

的普适关系. 这间接说明单位时间序列内所含的网络拓扑结构的信息量随耦合强度的增加而增多. 我们用反复驱动识别方法将识别速度的定义拓展到强耦合区间. 验证了在强耦合区间非同步的单位时间序列内含有网络拓扑结构信息量与耦合强度正相关的关系也成立.

由于单位时间序列的信息量和网络内趋向同步的弛豫时间之间的竞争, 导致识别速度与耦合强度在整个区间的函数关系为单峰映射, 且峰值位置随驱动片段的长度向耦合强度正方向推移. 在强耦合区域存在一个驱动片段的最优长度, 使识别速度最快. 最后, 我们用随机初始条件取代了固定初始条件 $\mathbf{y}_i(0) = \mathbf{x}_i(0)$, $d_{ij}(0) = 0$. 发现初始条件只对开始的极短时间内的识别有影响. 随着响应网络与驱动网络趋向同步, 初始条件对识别速度的影响逐渐消失.

从文献 [11] 可以看到存在非同步态的强耦合情形. 若网络仍能保持非同步的混沌行为, 关于

弱耦合网络的结论这时同样适用. 比如, Rössler 网络 $\mathbf{H}(\mathbf{x}) = (x_2, 0, 0)$ 在 $s < 6$ 时, 识别速度始终随耦合强度线性增加. 但在耦合强度特别大时可能会出现多种复杂的情况. 比如, Lorenz 网络 $\mathbf{H}(\mathbf{x}) = (x_3, 0, 0)$ 在 $s = 10$ 时会出现阵发同步; Lorenz 网络 $\mathbf{H}(\mathbf{x}) = (x_2, 0, 0)$ 在 $s = 10$ 时会出现非完全同步的严格周期态; Rössler 网络 $\mathbf{H}(\mathbf{x}) = (x_2, 0, 0)$ 在 $s = 10$ 时振子状态会发散; 还有振幅死亡等等. 以上情况虽然不是严格的同步稳定态, 但是都需要用反复驱动识别法来分析, 并可得到与处于强耦合同步稳定态一致的结论.

本文虽然是基于自适应反馈识别算法, 但通过它得到的有关识别速度的知识可以帮助我们来定量理解不同时间序列中蕴含的有关网络拓扑结构的信息量. 这种理解不局限于某一种具体方法. 当然, 本文还有一些需要日后完善的地方. 比如怎样直接衡量时间序列中包含的有关网络拓扑结构的信息量等等.

-
- [1] Lu J A 2010 *Complex Systems and Complexity Science* **7** 19 (in Chinese) [陆君安 2010 复杂系统与复杂性科学 7 19]
 - [2] Yu D, Righero M, Vicente P 2006 *Phys. Rev. Lett.* **96** 114102
 - [3] Timme M 2007 *Phys. Rev. Lett.* **98** 224101
 - [4] Ren j, Wang W X, Li B, Lai Y C 2010 *Phys. Rev. Lett.* **104** 058701
 - [5] Liang X M, Liu Z H, Li B W 2009 *Phys. Rev. E* **80** 046102
 - [6] Bu S L, Jiang I M 2008 *Europhys. Lett.* **82** 68001
 - [7] Liang X, Liu Z, Li B 2009 *Phys. Rev. E* **80** 046102
 - [8] Shen Y, Hou Z, Xin H 2010 *Chaos* **20** 013110
 - [9] Chen L, Lu J A, Tse C K 2009 *IEEE Trans. Circuits Syst. II-Express Briefs* **56** 310
 - [10] Sun F, Peng H P, Xiao J H 2012 *Nonlinear Dyn.* **67** 1457
 - [11] Huang L, Chen Q F, Lai Y C, Pecora L M 2009 *Phys. Rev. E* **80** 036204
 - [12] Lorenz E N, Atmos J 1963 *Sci.* **20** 130
 - [13] Rössler O E 1976 *Phys. Lett. A* **57** 397
 - [14] Chen G, Ueta T 1999 *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **9** 1465

Analysis the convergency speed of estimating the network topology based on the dynamical synchronization*

Yang Pu Zheng Zhi-Gang[†]

(Beijing - Hong Kong - Singapore Joint Center for Nonlinear and Complex Studies, Department of Physics, Beijing Normal University, Beijing, 100875, China)

(Received 1 March 2012; revised manuscript received 16 March 2012)

Abstract

Identifying convergent speed is an important but rarely discussed problem in estimating topologies of complex networks. In this paper, we discuss this problem mainly in both weakly and strongly coupled conditions. In the weakly coupled conditions, the convergent speed we defined increases linearly with coupling strength increasing. After analyzing the dynamics, we find that this relation is universal. In light of the repeatedly driving method we proposed recently, we generalize the definition of the convergent speed into the area of synchronization. In this case, there is a best length of the driving time series to maximize the convergent speed. The knowledge of convergent speed helps us understand the topological information embedded in the time series.

Keywords: complex network, estimate topology, adaptive-feedback, convergent speed

PACS: 05.45.Tp, 89.75.-k, 02.60.Cb

* Project supported by the National Natural Science Foundation of China (Grant Nos. 10875011, 11075016), the 973 Program(Grant No. 2007CB814805), the Foundation for Doctoral Training from MOE (Grant No. 20100003110007), and the Fundamental Research Funds for the Central Universities.

† E-mail: zgzheng@bnu.edu.cn