Python金融计算（第三讲）：基于复杂网络的金融时间序列分析

谢文杰

华东理工大学 金融学系

2022年春季

目录

1 基于复杂网络的时间序列分析背景

2 可视图构建和结构分析

3 四元网络模体结构分析

4 三元网络模体结构分析

目录

1 基于复杂网络的时间序列分析背景

2 可视图构建和结构分析

3 四元网络模体结构分析

4 三元网络模体结构分析

基于复杂网络的时间序列分析

时间序列映射到网络的方法有很多：

- 时序网络
- 周期网络
- 最近邻网络
- n-元组网络
- 循环网络
- 分段相关网络
- 可视图网络
- 平行可视图网络

基于复杂网络的时间序列分析文献

[129] LACASA L, LUQUE B, BALLESTEROS F, et al. From time series to complex networks: The visibility graph[J]. Proc. Natl. Acad. Sci. U.S.A., 2008, 105: 4972-4975.

[130] NI X H, JIANG Z Q, ZHOU W X. Degree distributions of the visibility graphs mapped from fractional Brownian motions and multifractal random walks[J]. Phys. Lett. A, 2009, 373(42): 3822-3826.

[131] QIAN M C, JIANG Z Q, ZHOU W X. Universal and nonuniversal allometric scaling behaviors in the visibility graphs of world stock market indices[J]. J. Phys. A, 2010, 43(33): 335002.

[132] ELSNER J B, JAGGER T H, FOGARTY E A. Visibility network of United States hurricanes[J]. Geophys. Res. Lett., 2009, 36: L16702.

[133] LACASA L, LUQUE B, LUQUE J, et al. The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion[J]. Europhys. Lett., 2009, 86: 30001.

[134] LACASA L, TORAL R. Description of stochastic and chaotic series using visibility graphs[J]. Phys. Rev. E, 2010, 82: 036120.

基于复杂网络的时间序列分析文献

[137] TANG Q, LIU J, LIU H L. Comparison of different daily streamflow series in US and China, under a viewpoint of complex networks[J]. Mod. Phys. Lett. B, 2010, 24: 1541-1547.

[138] XIE W J, HAN R Q, JIANG Z Q, et al. Analytic degree distributions of horizontal visibility graphs mapped from unrelated random series and multifractal binomial measures[J]. EPL, 2017, 119(4): 48008.

[139] XIE W J, HAN R Q, ZHOU W X. Tetradic motif profiles of horizontal visibility graphs[J]. Commun. Nonlinear Sci. Numer. Simul., 2019, 72: 544-551.

[140] XIE W J, HAN R Q, ZHOU W X. Triadic time series motifs[J]. EPL, 2019, 125(1): 18002.

目录

1 基于复杂网络的时间序列分析背景

2 可视图构建和结构分析

3 四元网络模体结构分析

4 三元网络模体结构分析

可视图定义

可视图算法（Visibility graph, VG）由Lacasa等人首先提出，给定时间序列$X = \{x_i : i = 1, 2, ..., L\}$的每个数据点与可视图网络中的节点存在一一对应关系。如果两个数据点x_i 和 x_j之间的所有其他节点x_n满足如下公式：

$$\frac{x_i - x_n}{i - n} > \frac{x_i - x_j}{i - j}, \quad \forall n | i < n < j \tag{1}$$

则数据点x_i 和 x_j是相互可视的，即对应的两个网络节点之间存在连边。将网络中相互可视的节点两两相连，可以形成一个可视图网络。

水平可视图定义

水平可视图算法（Horizontal visibility graph, HVG）将时间序列$X = \{x_i : i = 1, 2, ..., L\}$转化成水平可视图网络$G = \langle V, E \rangle$，其中 $V = \{v_i\}$为水平可视图G的网络节点，与时间序列中的数据点$\{x_i\}$具有一一对应关系。$E = \{e_{ij}\}$为水平可视图G的邻接矩阵，邻接矩阵元素$e_{ij} = 1$表示数据点x_i 和 x_j是水平可视的，即对应的两个网络节点之间存在连边。用数学语言表示为当且仅当$\forall x_n \in X$，其中$i < n < j$，x_n满足如下公式，

$$x_i, x_j > x_n. \tag{2}$$

将网络中相互水平可视的节点两两相连，就形成了一个水平可视图。

水平可视图网络示意图

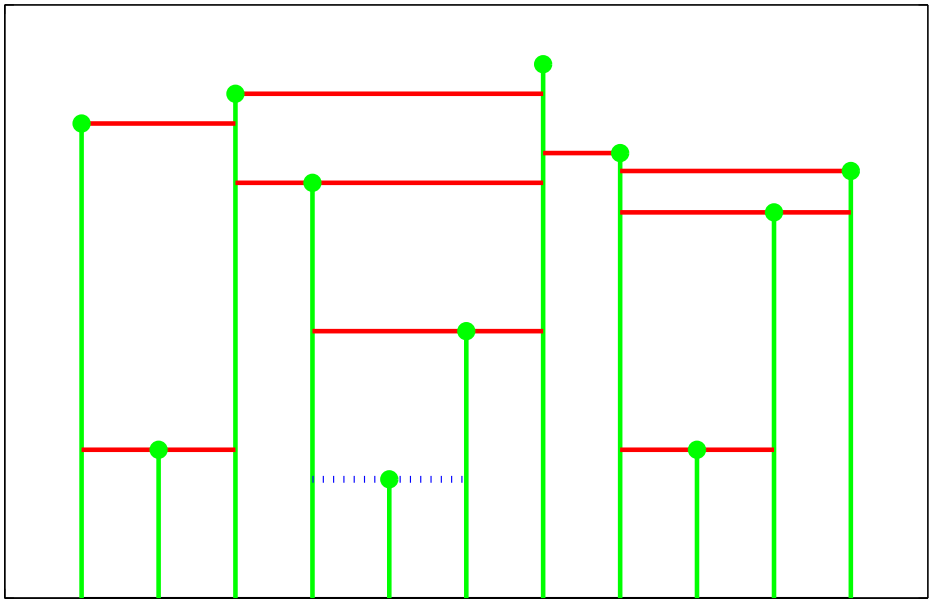


图: 水平可视图网络示意图

Fig.: Illustrative example of the construction of undirected HVGs mapped from time series containing 11 data points.

水平可视图的度分布（随机时间序列）

我们把包含n个节点的水平可视图中节点度数为k的概率定义为，

$$p(k,n) = N(k,n)/n. \tag{3}$$

当n足够大时，$\delta_{k2}/n = 0$，则:

$$p(k,n) \approx \left(1 - \frac{3}{n}\right) p(k,n-1) + \frac{2}{n} p(k-1,n-1), \tag{4}$$

当$n \to \infty$时，我们有

$$\begin{cases} p(k-1,n) = p(k-1,n-1) = p(k-1), \\ p(k,n) = p(k,n-1) = p(k). \end{cases} \tag{5}$$

由式(5)和式(4)可得，

$$p(k) = \frac{2}{3} p(k-1). \tag{6}$$

水平可视图的度分布（随机时间序列）

由式(5)和式(4)可得，

$$p(k) = \frac{2}{3} p(k-1). \tag{7}$$

代入$\sum_{k=2}^{\infty} p(k) = 1$，我们可以求解得到，

$$p(k) = \frac{3}{4} \left(\frac{2}{3} \right)^{k}. \tag{8}$$

有向水平可视图的度分布（随机时间序列）

$$p_d(k,n) = \left(1 - \frac{2}{n}\right) p_d(k,n-1) + \frac{1}{n} p_d(k-1,n-1). \quad (9)$$

当$n \to \infty$时，我们有

$$\begin{cases} p_d(k-1,n) = p_d(k-1,n-1) = p_d(k-1), \\ p_d(k,n) = p_d(k,n-1) = p_d(k). \end{cases} \quad (10)$$

由式(10)和式(9)可得，

$$p_d(k) = \frac{1}{2} p_d(k-1). \quad (11)$$

代入$\sum_{k=1}^{k=\infty} p_d(k) = 1$，我们可以求解得到，

$$p_d(k) = \left(\frac{1}{2}\right)^k. \quad (12)$$

目录

1 基于复杂网络的时间序列分析背景

2 可视图构建和结构分析

3 四元网络模体结构分析

4 三元网络模体结构分析

四元无向网络模体结构示意图

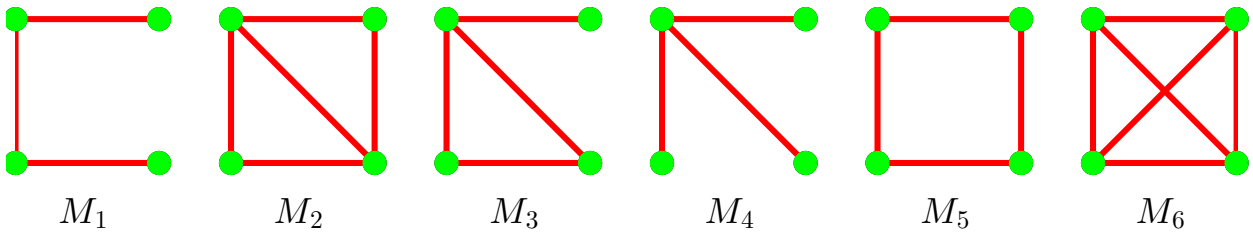


图: 四元无向网络模体结构示意图

Fig.: Plots of undirected tetradic motifs.

图中网络模体M_i代表了时间序列中数据之间的特定数量关系，模体M_i的分布f_i从微观层面反映了时间序列特征规律。基于网络模体分析方法对模拟生成的时间序列特征进行分析，并且运用到现实系统中的时间序列实证研究。

四元无向网络模体分布

水平可视图算法将时间序列特征信息转化为对应的可视图网络结构，如网络模体结构。网络模体分析是从微观层次对复杂网络结构进行分析，通过研究四元无向网络模体来挖掘不同时间序列的特征性质。

四元无向网络模体共有6种结构，六种模体结构分别表示为M_1，M_2，M_3，M_4，M_5 以及 M_6。水平可视图网络中模体M_i出现的频数表示为N_i，所占比例为

$$f_i = N_i / \sum_{i=1,\ldots,6} N_i. \tag{13}$$

多重分形二项测度时间序列分析

- 采用p-model算法构造多重分形二项测度时间序列，并构建水平可视图进行网络模体分析。
- 选取的参数为$p_1 = 0.25$ 和 $p_2 = 1 - p_1 = 0.75$。
- 对于多重分形二项测度，水平可视图中各个数据点之间相互可见的网络结构特征是不变的，因而参数p_1的选择不会影响结果。
- 将两个长度为2^{m-1}的多重分形二项测度序列生成的可视图拼接起来，构造包含2^m个数据点的水平可视图网络。
- 图给出了不同m对应的水平可视图网络结构，其中m的取值分别为$m = 2, 3, 4, 5$。

多重分形二项测度的水平可视图网络构

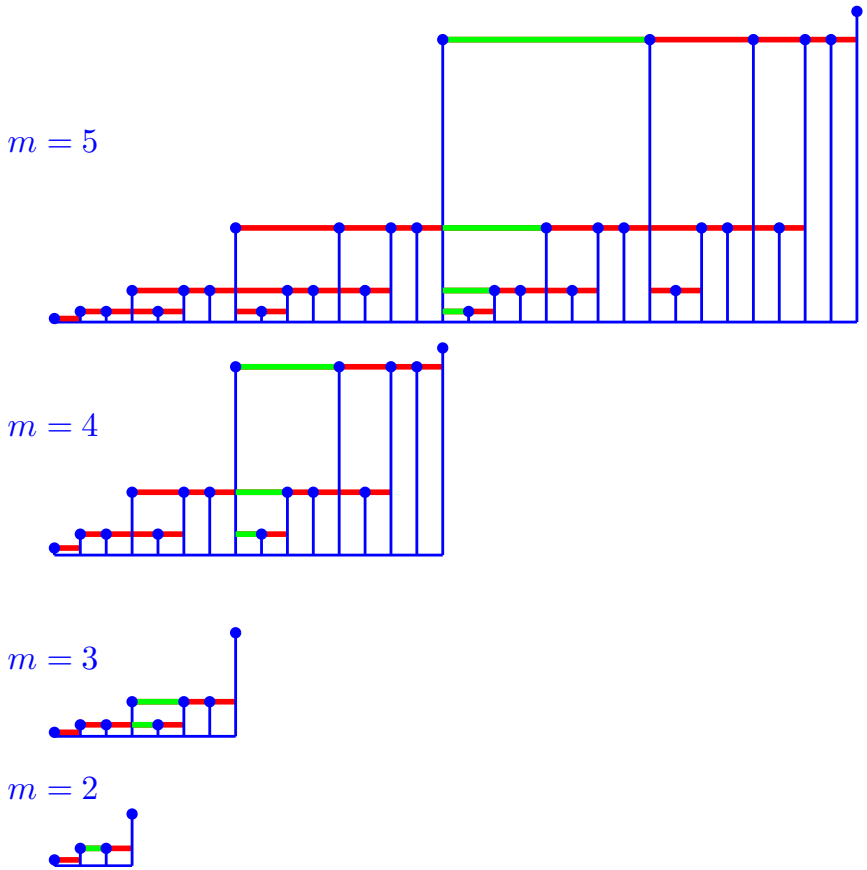


图: 多重分形二项测度的水平可视图网络构建示意图

多重分形二项测度的水平可视图网络构

- 将长度为2^m的多重分形二项测度序列对应的水平可视图中网络模体M_i的数量表示为$N_i(m)$。
- 通过模体识别算法，计算得到包含2^m个节点的水平可视图中网络模体M_i 的数量$N_i(m)$如下页表格所示，其中m的取值分别为$m = 1, 2, \cdots, 8$。
- 下页表格中可以看到$N_2 = N_6 = 0$。
- 事实上，$N_6 = 0$对于任何时间序列生成的水平可视图都成立。

多重分形二项测度水平可视图四元网络模体

表: 长度为 2^m 的多重分形二项测度水平可视图四元网络模体出现次数

m	N_1	N_2	N_3	N_4	N_5	N_6
1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	6	0	2	0	0	0
4	25	0	8	4	1	0
5	89	0	22	22	3	0
6	260	0	52	76	7	0
7	666	0	114	212	15	0
8	1567	0	240	524	31	0
9	3487	0	494	1202	63	0
10	7478	0	1004	2628	127	0

多重分形二项测度水平可视图四元网络模体

定义水平可视图中四元模体M_i的数量表达式$N_i(m)$。包含2^m个节点的水平可视图网络可以看成由两个分别包含2^{m-1}个节点的水平可视图拼接所得。在此过程中，先前的网络模体数量不变，添加$m-1$条连边后网络结构发生了改变。为了简化表达，用多项式函数$\eta_i(m)$表示新增模体M_i的数量，那么网络模体M_i的数量表达式可以写为

$$N_i(m) = 2N_i(m-1) + \eta_i(m), \quad i = 1, ..., 6. \tag{14}$$

新增$m-1$条边会导致网络模体数量增加，不同模体M_i的数量变化不一样，对应的多项式函数 $\eta_i(m)$也有所不同。此处无法给出多项式函数 $\eta_i(m)$的严格数学表达形式，结合文献中的理论推导，假设网络模体M_i的数量$N_i(m)$满足如下数学表达式:

多重分形二项测度水平可视图四元网络模体

$$N_i(m) = a_i 2^m + b_i m^3 + c_i m^2 + d_i m + e_i, \quad i = 1, ..., 6. \qquad (15)$$

其中参数a_i, b_i, c_i, d_i, e_i为待定系数。将表 1中的m和$N_i(m)$分别代入方程(15)可以求解得到参数a_i, b_i, c_i, d_i, e_i的值以及$N_i(m)$的数学表达式。为了拟合参数a_i, b_i, c_i, d_i, e_i，我们将表 1中模体M_i对应不同m的数量$N_i(4)$, $N_i(5)$, $N_i(6)$, $N_i(7)$, $N_i(8)$分别代入方程(15)。通过求解多元线性方程组，得到模体M_i的数量表达式(15)中参数a_i, b_i, c_i, d_i, e_i的值，将求解所得的参数值再代入式(15)可得:

多重分形二项测度水平可视图四元网络模体

$$N_i(m) = \begin{cases} \frac{33}{4}2^m - \frac{2}{3}m^3 - \frac{5}{2}m^2 - \frac{29}{6}m - 5, & i = 1 \\ 0, & i = 2 \\ 2^m - 2m, & i = 3 \\ 3 \times 2^m - \frac{1}{3}m^3 - m^2 - \frac{2}{3}m - 4, & i = 4 \\ \frac{1}{8}2^m - 1, & i = 5 \\ 0, & i = 6 \end{cases} \tag{16}$$

多重分形二项测度水平可视图四元网络模体

随着时间序列长度的增加，网络模体数量呈指数增长。我们定义模体M_i的出现频率为$f_i = N_i(m)/N(m)$，其中：

$$N(m) = \sum_{i=1}^{6} N_i(m), \tag{17}$$

及

$$N(m) = \frac{99}{8}2^m - m^3 - \frac{7}{2}m^2 - \frac{45}{6}m - 10 \tag{18}$$

将其代入，可得f_i的具体表达式为：

多重分形二项测度水平可视图四元网络模体

$$
\begin{cases}
\dfrac{\frac{33}{4}2^m-\frac{2}{3}m^3-\frac{5}{2}m^2-\frac{29}{6}m-5}{\frac{99}{8}2^m-m^3-\frac{7}{2}m^2-\frac{45}{6}m-10}, & i=1\\
0, & i=2\\
\dfrac{2^m-2m}{\frac{99}{8}2^m-m^3-\frac{7}{2}m^2-\frac{45}{6}m-10}, & i=3\\
\dfrac{3\times 2^m-\frac{1}{3}m^3-m^2-\frac{2}{3}m-4}{\frac{99}{8}2^m-m^3-\frac{7}{2}m^2-\frac{45}{6}m-10}, & i=4\\
\dfrac{\frac{1}{8}2^m-1}{\frac{99}{8}2^m-m^3-\frac{7}{2}m^2-\frac{45}{6}m-10}, & i=5\\
0, & i=6
\end{cases}
\tag{19}
$$

多重分形二项测度水平可视图四元网络模体

随着m的增加，数学表达式(15)的指数部分 $a_i 2^m$会远远大于多项式部分$b_i m^3 + c_i m^2 + d_i m + e_i$。因此当$m \to \infty$时，多项式部分可以忽略，各个模体的出现频率将收敛至常数，如下式(20)所示，

$$\lim_{m\to\infty} f_i(m) = \begin{cases} \dfrac{a_1}{a_1 + a_3 + a_4 + a_5} = 2/3, & i = 1 \\ 0, & i = 2 \\ \dfrac{a_3}{a_1 + a_3 + a_4 + a_5} = 8/99, & i = 3 \\ \dfrac{a_4}{a_1 + a_3 + a_4 + a_5} = 8/33, & i = 4 \\ \dfrac{a_5}{a_1 + a_3 + a_4 + a_5} = 1/99, & i = 5 \\ 0, & i = 6 \end{cases} \tag{20}$$

多重分形二项测度水平可视图四元网络模体

简化后可得，

$$\lim_{m\to\infty} \mathrm{f} = \lim_{m\to\infty} \{f_i(m) : i = 1, 2, \cdots, 6\} = \left(\frac{2}{3}, 0, \frac{8}{99}, \frac{8}{33}, \frac{1}{99}, 0\right) \tag{21}$$

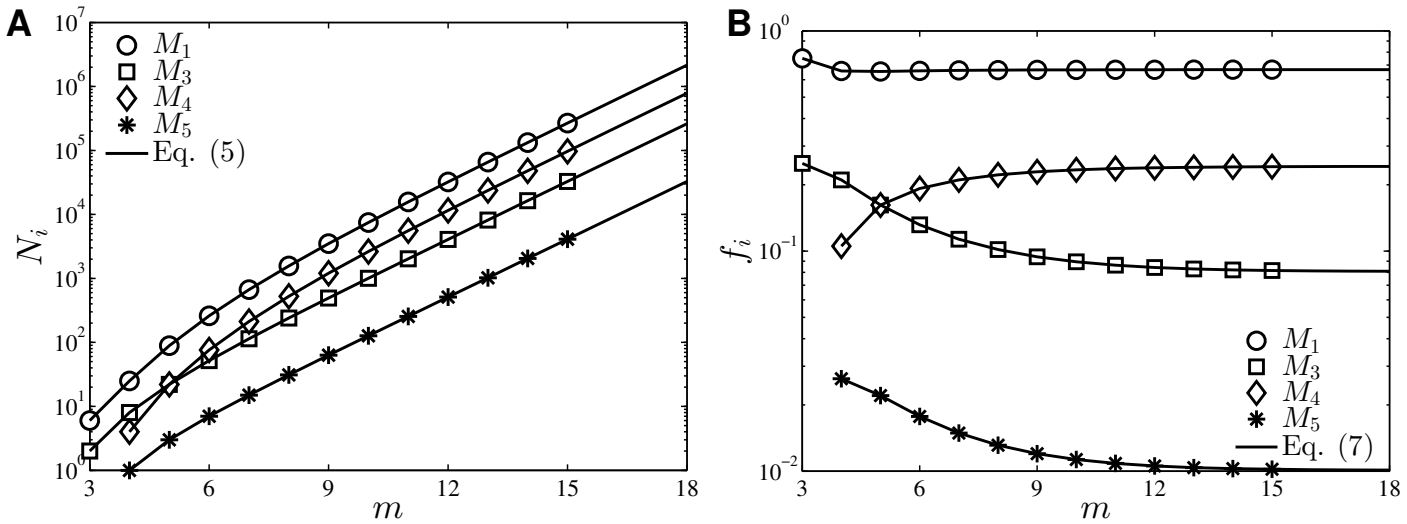


图: (A) 多重分形二项测度水平可视图中四元模体M_i的出现次数N_i；
(B) 多重分形二项测度水平可视图中四元模体M_i的出现频率f_i。

心率间隔时间序列

- 现实世界中时间序列表现出了复杂的动力学特征，水平可视图网络模体结构能够刻画时间序列特征。
- 心电信号是最常见的也是最能直观反映人体机能的生理信号，研究心率时间序列可视图网络的拓扑结构对于揭示生理信号的内在动力学特征具有重要意义。
- 根据上述分析结果，我们进一步运用网络模体分析方法来研究心率间隔时间序列，揭示不同类型心率数据的特征规律。
- 研究样本数据分别来自5名健康受试者、5名充血性心力衰竭患者和5名心房颤动患者，心率间隔时间序列长度N在70,000到150,000之间，这些数据来自PhysioNet数据库（http://physionet.org/physiobank/database/）。

心率间隔时间序列

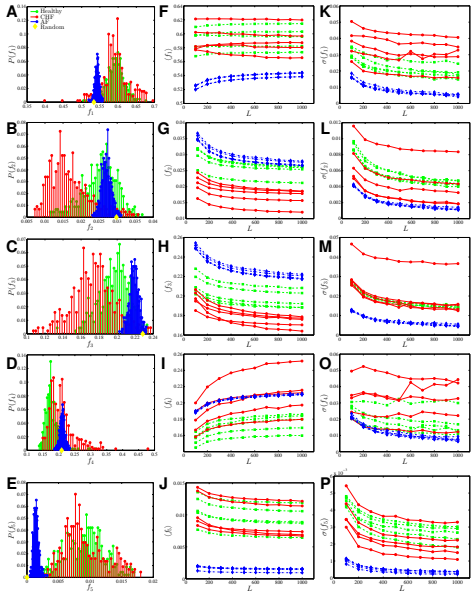
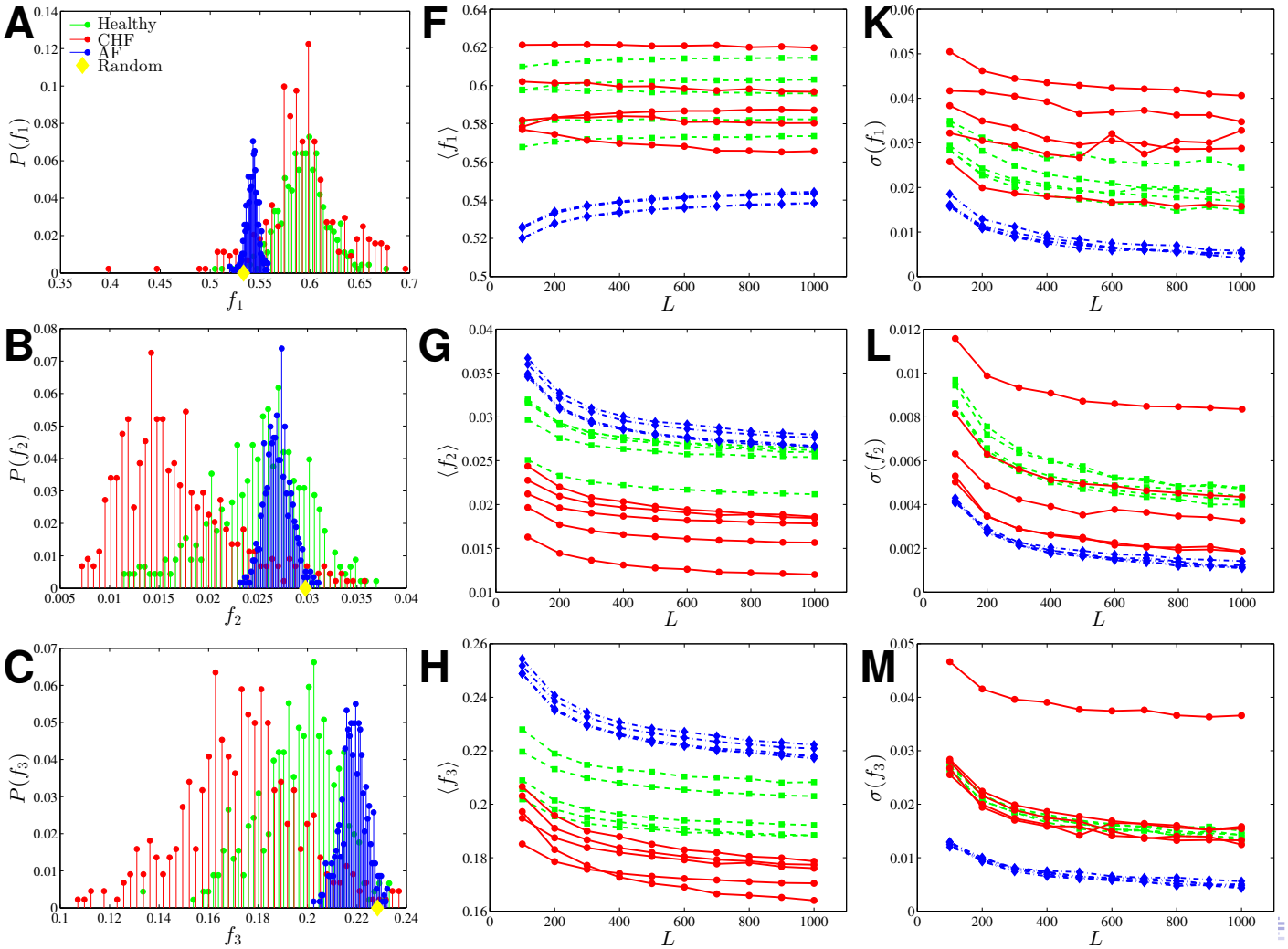


图: (A-E) 健康受试者、充血性心力衰竭（CHF）患者以及心房颤动（AF）患者的心率时间序列中模体M_i的出现频率分布$P(f_i)$；(F-J) 模体M_i的平均出现频率$\langle f_i \rangle$与时间序列长度L的关系图；(K-P) 模体M_i出现频率的均方差$\sigma(f_i)$与时间序列长度L的关系图。

心率间隔时间序列


A
Healthy
CHF
AF
Random
$P(f_1)$
f_1
F
$\langle f_1 \rangle$
L
K
$\sigma(f_1)$
L
B
$P(f_2)$
f_2
G
$\langle f_2 \rangle$
L
L
$\sigma(f_2)$
L
C
$P(f_3)$
f_3
H
$\langle f_3 \rangle$
L
M
$\sigma(f_3)$
L

目录

1 基于复杂网络的时间序列分析背景

2 可视图构建和结构分析

3 四元网络模体结构分析

4 三元网络模体结构分析

三元时间序列模体的定义

定义二元时间序列模体，任意两个数据点x_i和 x_j $(i < j)$可以构成二元时间序列模体，当且仅当两者之间所有数据点x_n的取值小于x_i和 x_j，

$$x_i > x_n \quad 且 \quad x_j > x_n, \tag{22}$$

其中$i < n < j$。如图6所示，根据数据点x_i和x_j的取值大小存在两种不同的二元时间序列模体

$$x_i > x_j \quad 或 \quad x_i \leq x_j. \tag{23}$$

可以将对应的二元时间序列模体分别表示为$(1, 2)$和$(2, 1)$。注意到单调递增的时间序列只存在一种二元模体$(1, 2)$，例如 Devil's Staircase；而单调递减的时间序列只存在另一种二元模体$(2, 1)$。

三元时间序列模体结构示意图

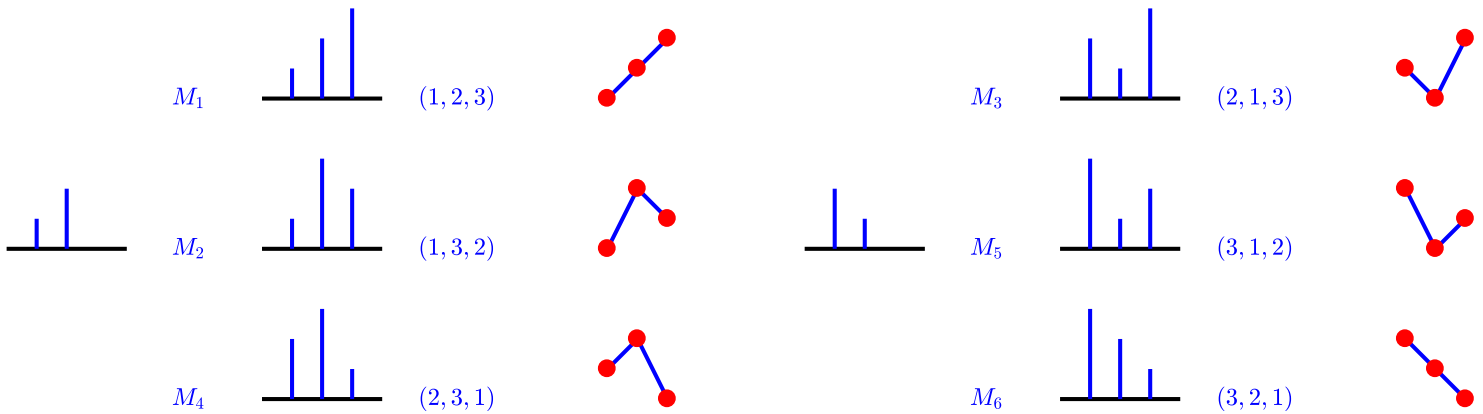


图: 三元时间序列模体结构示意图

Fig.: Illustrative example showing the six types of triadic time series motifs.

三元时间序列模体

对于给定的时间序列$\{x_n\}_{n=1}^{L}$，任意三个数据点$\{x_i, x_j, x_k\}(i < j < k)$可以构成三元时间序列模体当且仅当以下条件成立：

$$\begin{cases} x_i, x_j > x_n, & \forall n \in (i,j), \quad j-i \geq 1 \\ x_j, x_k > x_m, & \forall m \in (j,k), \quad k-j \geq 1 \end{cases} \tag{24}$$

可以得到 6 种不同类型的三元时间序列模体，如图6所示，分别记为$M_1 = (1,2,3)$, $M_2 = (1,3,2)$, $M_3 = (2,1,3)$, $M_4 = (2,3,1)$, $M_5 = (3,1,2)$和$M_6 = (3,2,1)$。其中M_1, M_2, M_4和M_6为非闭合三元模体，而M_3和M_5属于闭合三元模体。

三元时间序列模体的构建过程

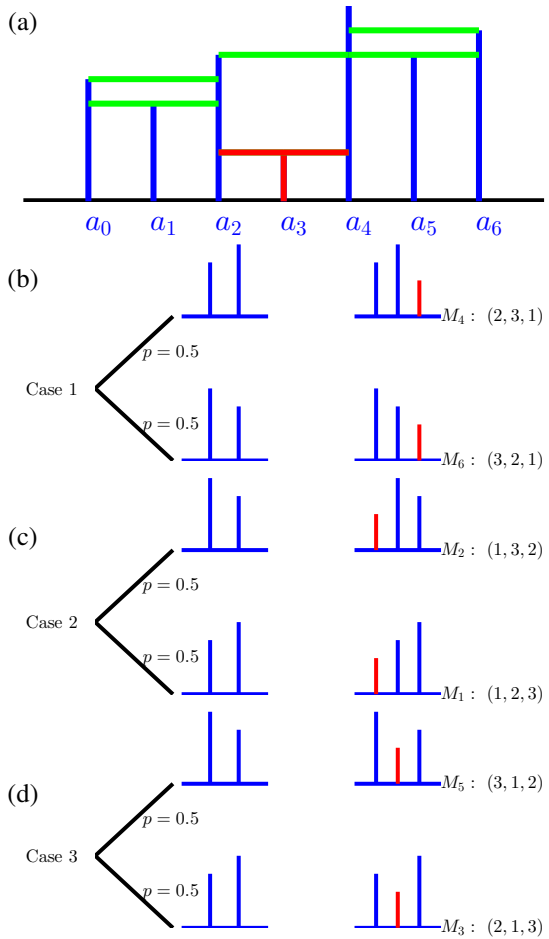


图: 三元时间序列模体的构建过程示意图

三元时间序列模体分布

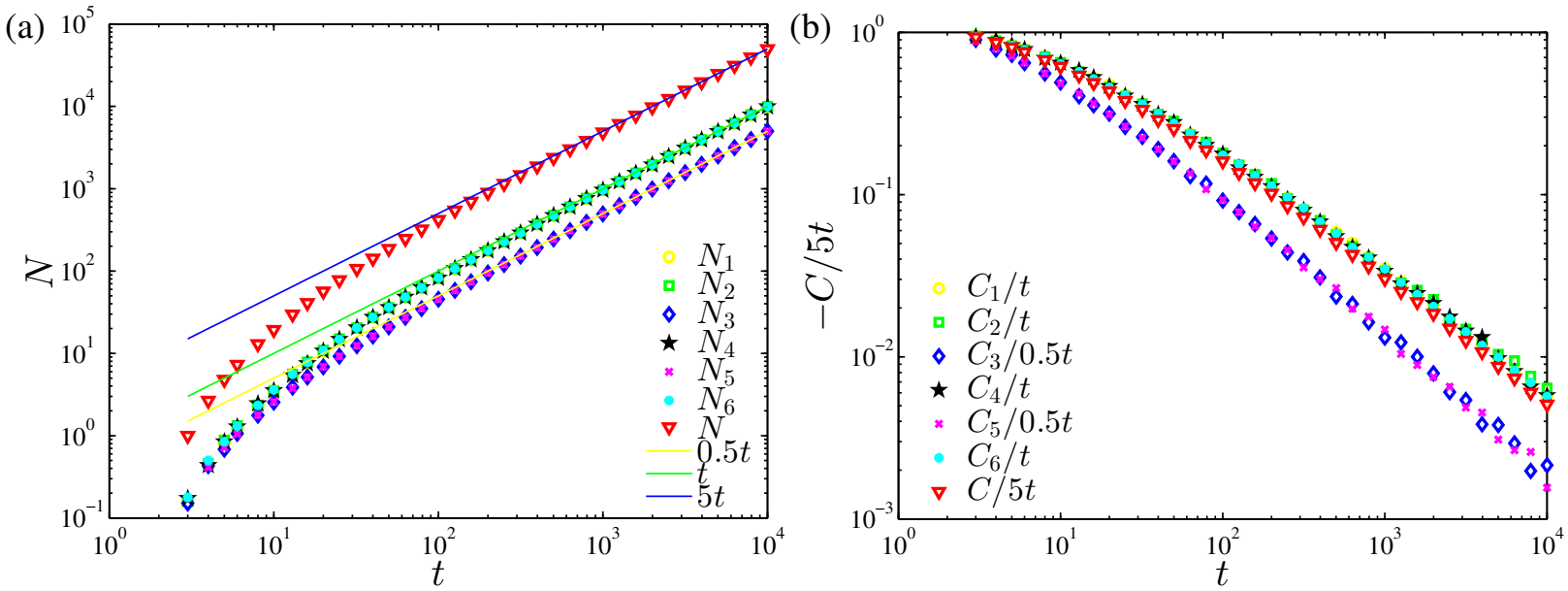


图: 三元时间序列模体分布与随机时间序列长度 t 之间的关系图。(a) 模体出现次数的均值；(b) 模体出现次数的平均增量。

三元时间序列模体分布

定义$N_i(t)$为第t步中时间序列模体M_i的数量。对于非闭合的三元时间序列模体M_i $(i \in \{1,2,4,6\})$，得到模体数量$N_i(t)$的迭代方程为，

$$N_i(t) = N_i(t-1) + 1 + \varepsilon_i(t), \quad i \in \{1,2,4,6\}, \tag{25}$$

其中$\varepsilon_i(t)$为随机变量。求解可得其近似解为，

$$N_i(t) = t + C_i(t), \quad i \in \{1,2,4,6\}, \tag{26}$$

其中$C_i(t)$为非闭合三元模体M_i的实际数量与理论近似值之间的差值。当t取值较小时，有$N_i(1) = 0$, $N_i(2) = 0$。非闭合三元模体M_i的实际增量小于1，即$C_i(t) < 0$。如图所示，通过大量的数值模拟实验进一步验证了上述理论分析。

三元时间序列模体分布

对于闭合的三元时间序列模体M_i $(i \in \{3,5\})$，得到模体数量$N_i(t)$的迭代方程为，

$$N_i(t) = N_i(t-1) + 0.5 + \varepsilon_i(t), \quad i \in \{3,5\}, \tag{27}$$

其中$\varepsilon_i(t)$为随机变量。求解可得其近似解为，

$$N_i(t) = 0.5t + C_i(t), \quad i \in \{3,5\}, \tag{28}$$

其中$C_i(t)$为闭合三元模体M_i的实际数量与理论近似值之间的差值。当t取值较小时，我们有$N_i(1) = 0$，$N_i(2) = 0$。闭合三元模体M_i的实际增量小于0.5，即$C_i(t) < 0$。如图所示，我们通过大量的数值模拟实验进一步验证了上述理论分析。

三元时间序列模体分布

综合上述三种情况可知，当把第t个最大的数放入序列时，三元时间序列模体的预期增量为5。定义$N(t)=\sum_{i=1}^{6}N_i(t)$为第$t$步中三元时间序列模体的数量，可以得到以下迭代方程，

$$N(t)=N(t-1)+5+\varepsilon(t), \tag{29}$$

其中$\varepsilon(t)$为随机变量。求解可得其近似解为，

$$N(t)=5t+C(t), \tag{30}$$

其中$C(t)$为模体的实际数量与理论近似值之间的差值。当t取值较小时，我们有$N(1)=0$，$N(2)=0$，$N(3)=1$，$N(4)<5$。三元时间序列模体的实际增量小于5，即$C(t)<0$。如图所示，通过大量的数值模拟实验进一步验证了上述理论分析。从图中可以看出，三元时间序列模体的数量与随机时间序列的长度近似存在线性增长关系。

三元时间序列模体分布

可以定义时间序列模体M_i的出现频率为，

$$f_i(t) = N_i(t)/N(t). \tag{31}$$

结合式(30)、式(26)以及式(28)可知，当t足够大时，模体M_i在随机时间序列中的出现频率为$\mathrm{f} = (0.2, 0.2, 0.1, 0.2, 0.1, 0.2)$。
通过大量的数值模拟实验进一步验证了上述理论分析。对于给定的时间序列长度t，重复进行1000次数值模拟得到模体M_i出现次数的均值与时间序列长度t之间的关系如图(a)所示。当t足够大时，数值模拟结果与理论分析结果是相吻合的。图(b)给出了实际模体数量与理论近似值之间的相对差值$C(t)/5t$, $C_i(t)/t$, $C_i(t)/0.5t$ $(C(t) < 0)$与时间序列长度t的关系图。从图中可以看出，实际模体数量与理论近似值之间相对差值的绝对值$-C(t)/5t$, $-C_i(t)/t$, $-C_i(t)/0.5t$随着t的增加而减小。

心率间隔时间序列

- 首先考察不同健康状况的个体心率间隔时间序列，研究样本数据分别来自5名健康受试者、5名充血性心力衰竭患者和5名心房颤动患者。
- 这些数据来自PhysioNet数据库（http://www.physionet.org/challenge/chaos/）。
- 不同研究对象的心率间隔时间序列长度T在70,000到150,000之间，其中5名健康受试者的心率间隔时间序列长度分别为 99762、86925、101523、86822、81280，
- 5名充血性心力衰竭患者的时间序列长度分别为 74496、76948、88501、88499、115062，
- 5名心房颤动患者的时间序列长度分别为 101145、117100、85304、138209、141628。

心率间隔时间序列

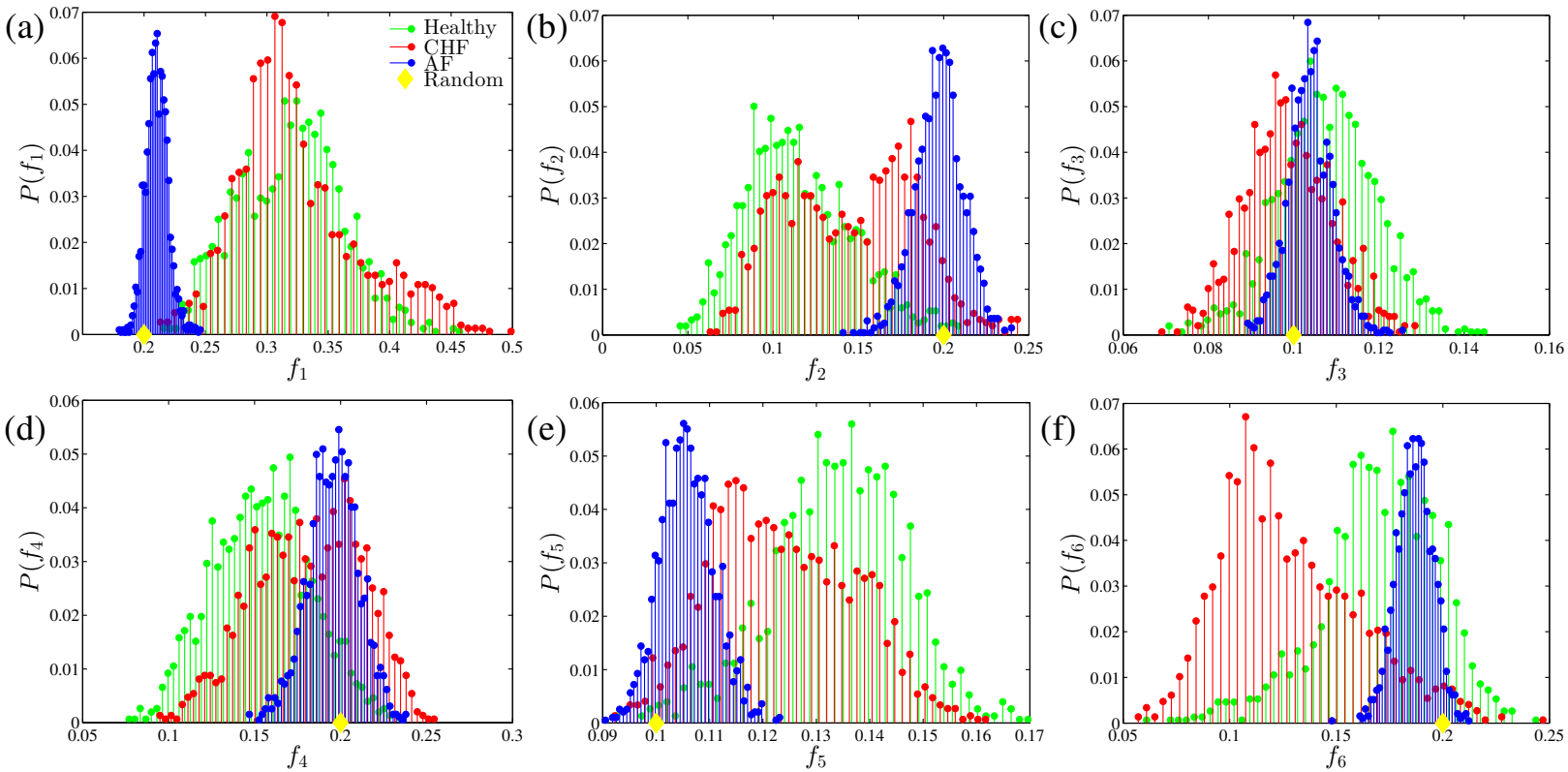


图: 健康受试者、充血性心力衰竭患者（CHF）以及心房颤动患者（AF）的心率间隔时间序列中模体 M_i 的出现频率分布 $P(f_i)$。

心率间隔时间序列

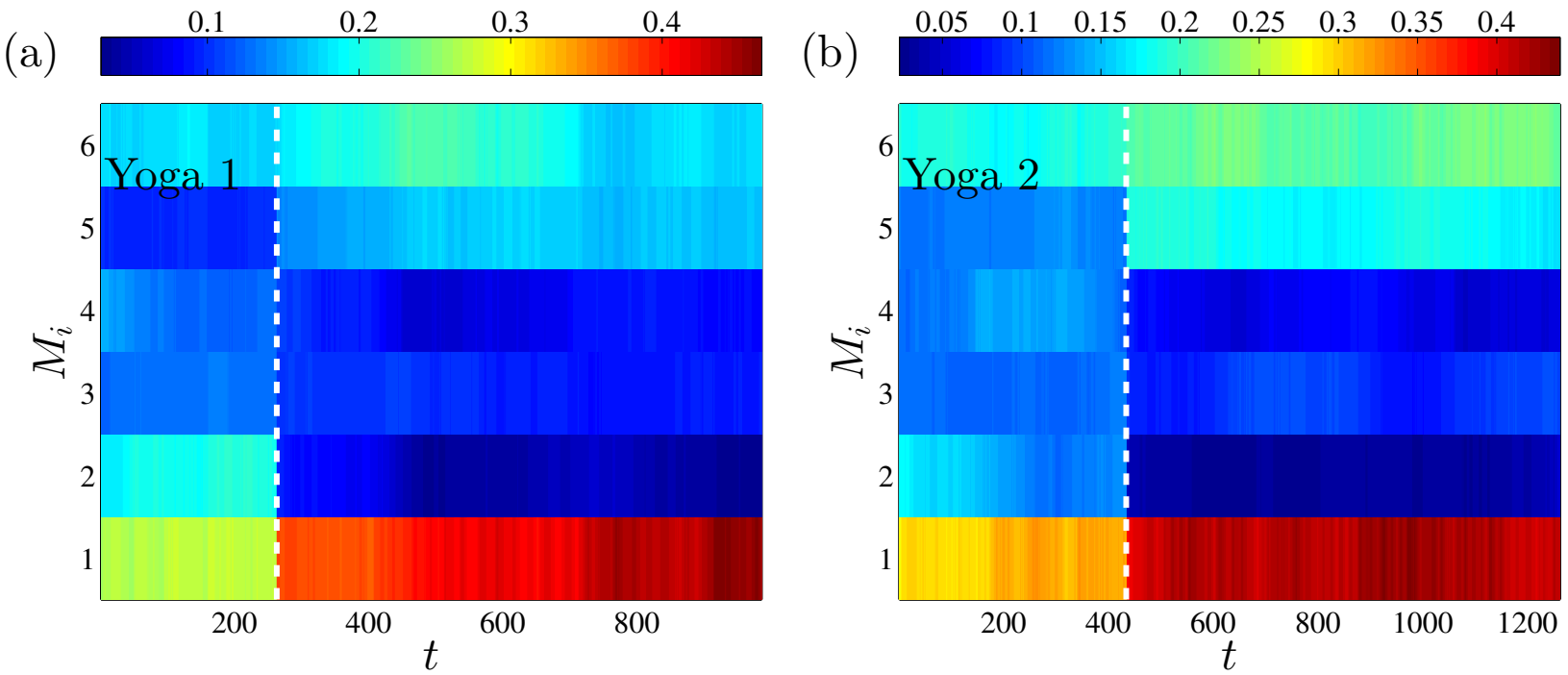


图: 瑜伽练习前后心率时间序列的模体分布情况。

心率间隔时间序列

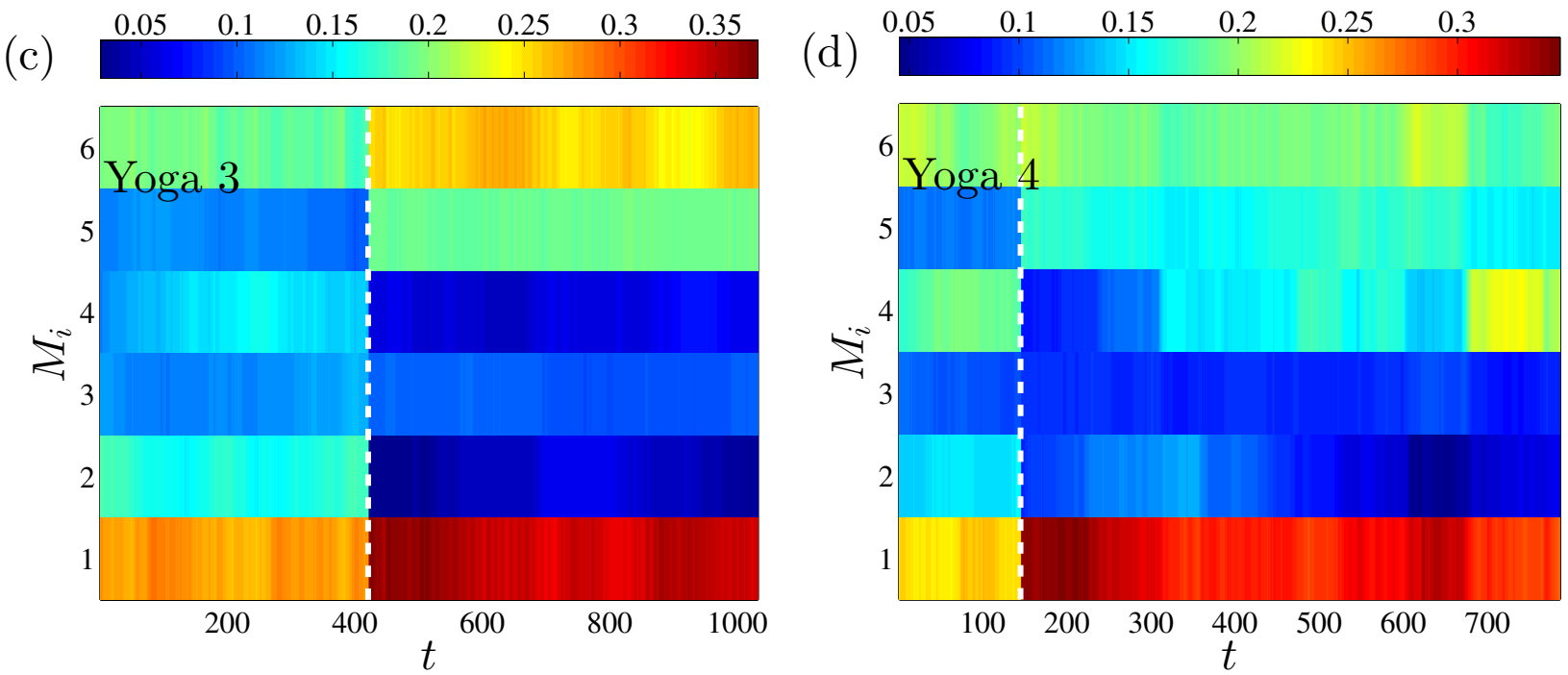
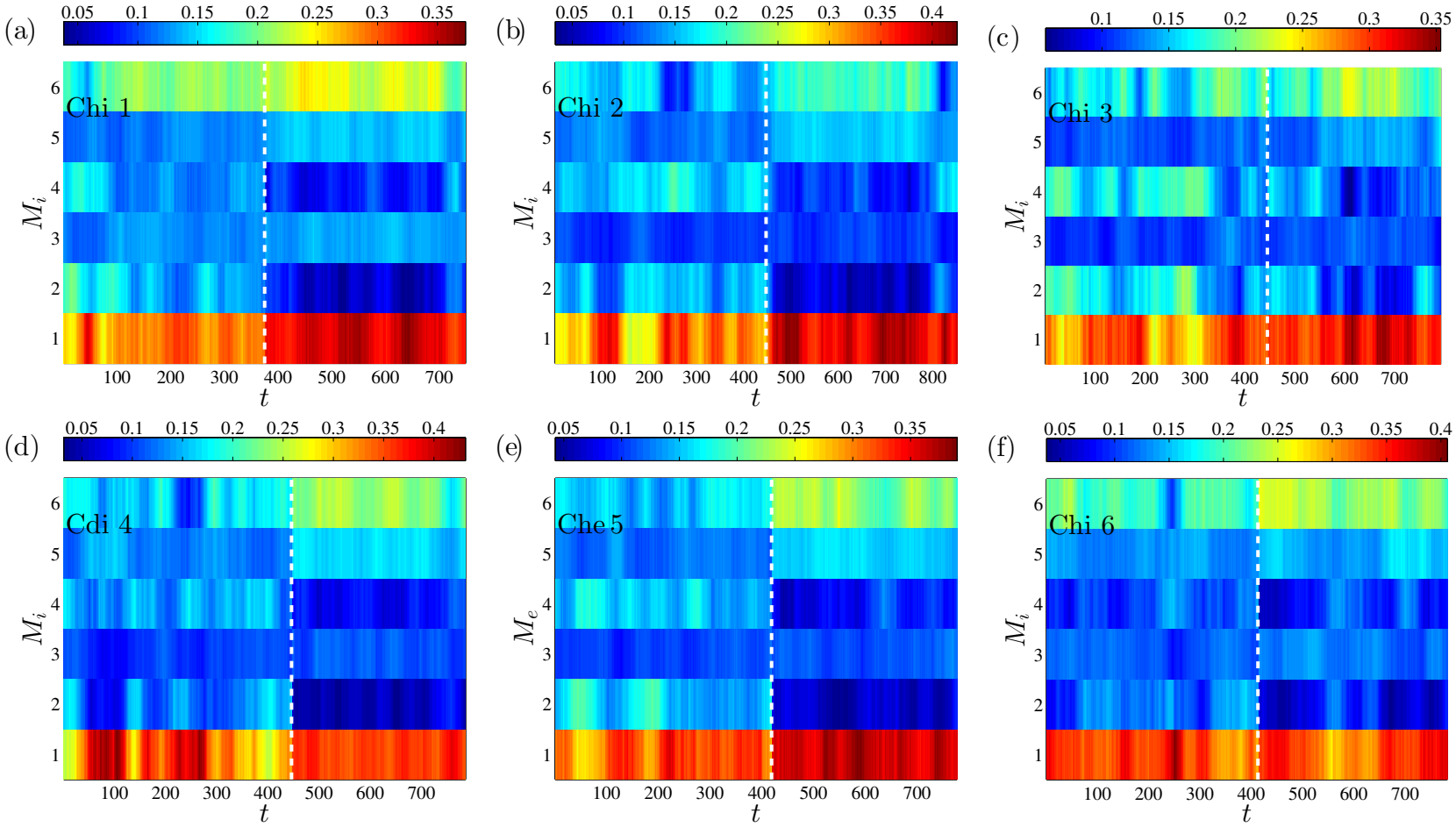


图: 瑜伽练习前后心率时间序列的模体分布情况。

心率间隔时间序列



图· 气功练习前后心率时间序列的模体分布情况。

金融时间序列

- 进一步应用时间序列模体分析方法来研究金融时间序列。首先选取美国股市的道琼斯工业平均指数（DJIA）作为研究样本，时间跨度从2005年6月2日至2010 年8月3日。
- 采用移动时间窗口的方法研究DJIA指数日度收益率$r(t)$的三元时间序列模体分布情况，选取的移动窗口长度为$l = 400$个交易日，以1天为移动步长。
- 在每个移动时间窗口中，将收益率时间序列$\{r(t-l+1), r(t-l+2), \cdots, r(t)\}$转化为水平可视图网络，通过模体识别算法得到时间序列模体M_i的数量。
- 以2007年12月31日为分界点把DJIA指数序列划分成两个子序列，便于比较2008年金融危机前后牛市和熊市对时间序列模体分布的影响。
- 对于每个子序列，我们计算模体M_i在移动窗口中的出现频率f_i，得到相应的概率分布$P(f_i)$。

金融时间序列

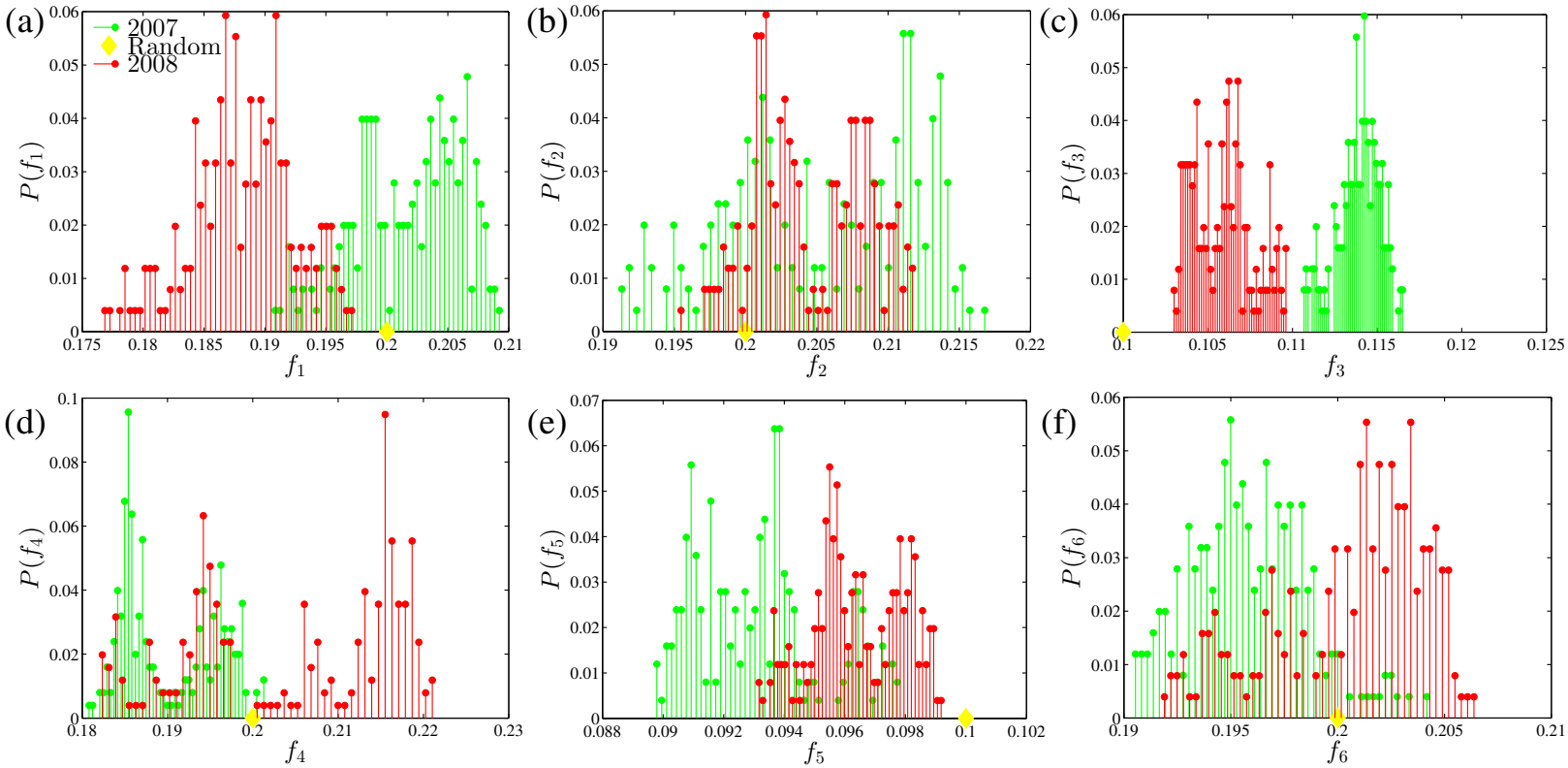


图: DJIA指数以2007年12月31日为分界点的收益率时间序列中模体 M_i 的出现频率分布 $P(f_i)$。

金融时间序列

- 图给出了金融危机前后DJIA指数收益率序列对应模体M_i的出现频率分布$P(f_i)$。
- DJIA指数以2007年12月31日为分界点的两个子序列中模体M_i的出现频率分布$P(f_i)$ 具有显著差异，
- 特别是模体M_3在两个子序列中的出现频率分布$P(f_3)$ 图像没有重叠部分，
- 这表明三元时间序列模体分布可以识别不同市场状态下股指回报率的波动情况。

金融时间序列

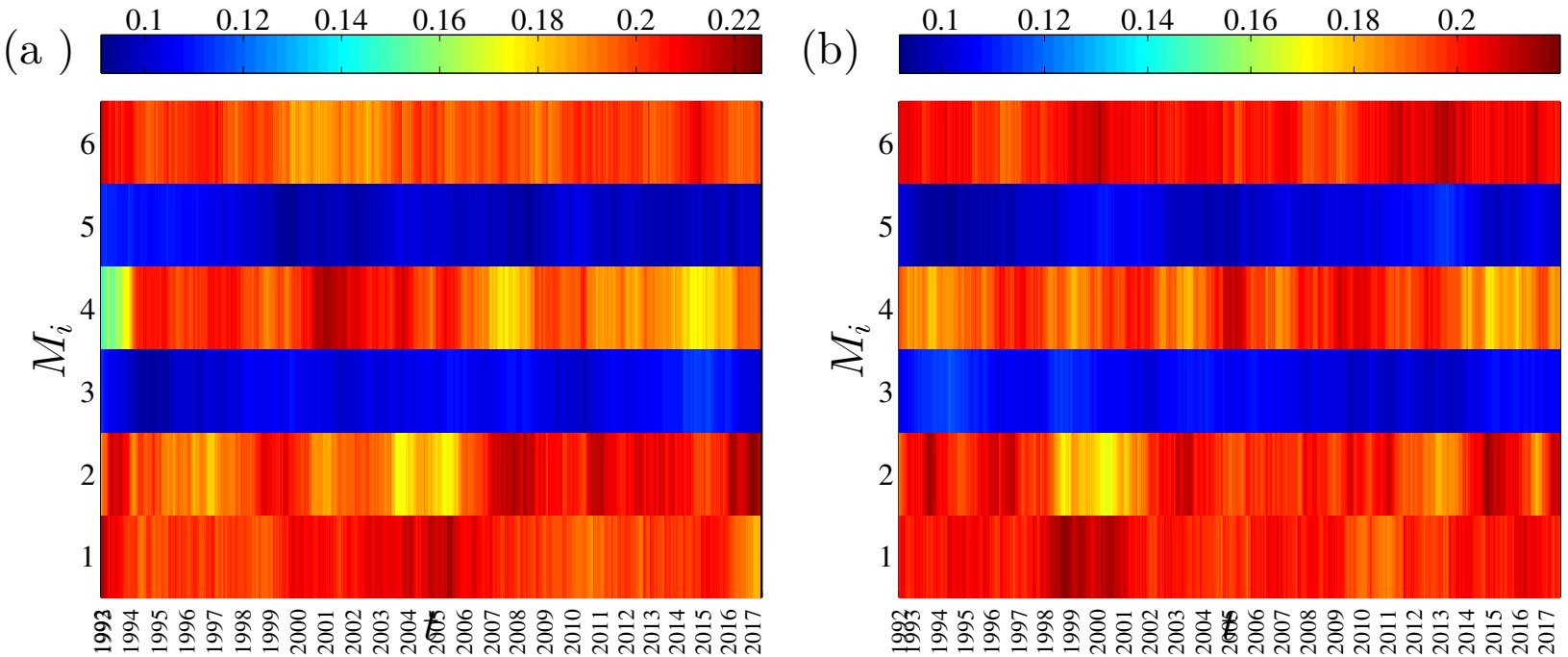


图: 金融时间序列的模体分布情况。(a) 上证指数；(b) 杜邦股票；(c) 埃克森美孚股票；(d) CHF-EUR汇率。

金融时间序列

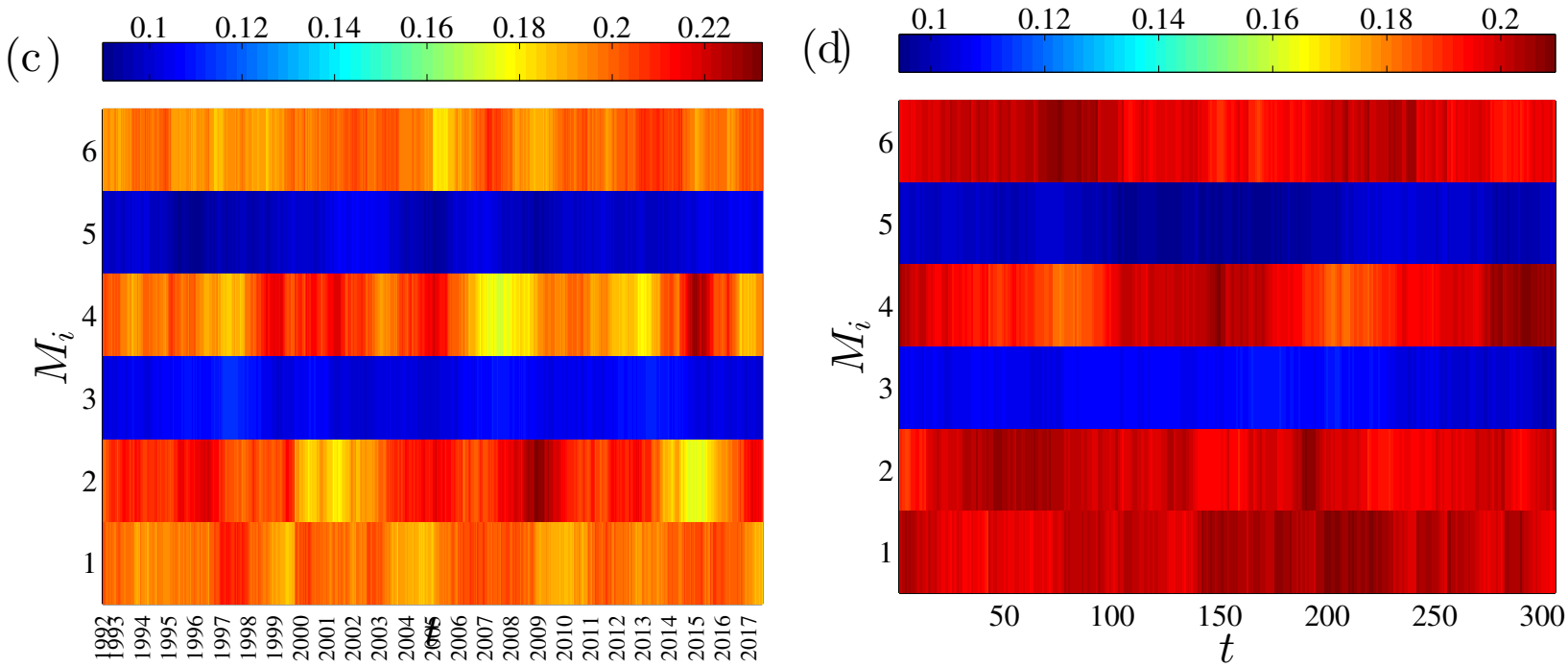


图: 金融时间序列的模体分布情况。(a) 上证指数；(b) 杜邦股票；(c) 埃克森美孚股票；(d) CHF-EUR汇率。

金融时间序列

- 图给出了上证指数收益率时间序列的模体分布情况。
- 可以看到不同模体M_i的出现频率存在显著差异，其中非闭合的三元模体M_1，M_2，M_4和M_6的出现频率较高，闭合的三元模体M_3和M_5的出现频率相对较低。
- 每个模体M_i在不同时期的分布情况没有明显的差异，因此我们无法通过上证指数收益率的模体分布来区分不同市场态势下的股市结构变化。
- (b-c)分别给出了纽约证券交易所上市股票杜邦（NYSE:DD）和埃克森美孚（NYSE:XOM）日度收益率的模体分布情况。
- 图中可以看出，模体M_i在两只股票收益率时间序列中的出现频率与上证指数的模体分布基本一致。
- 同时研究了外汇市场CHF-EUR以小时为单位的收益率时间序列，样本的时间跨度从2001年09月01日至2001年12月31日，相应的模体分布情况如图 (d)所示。

金融时间序列

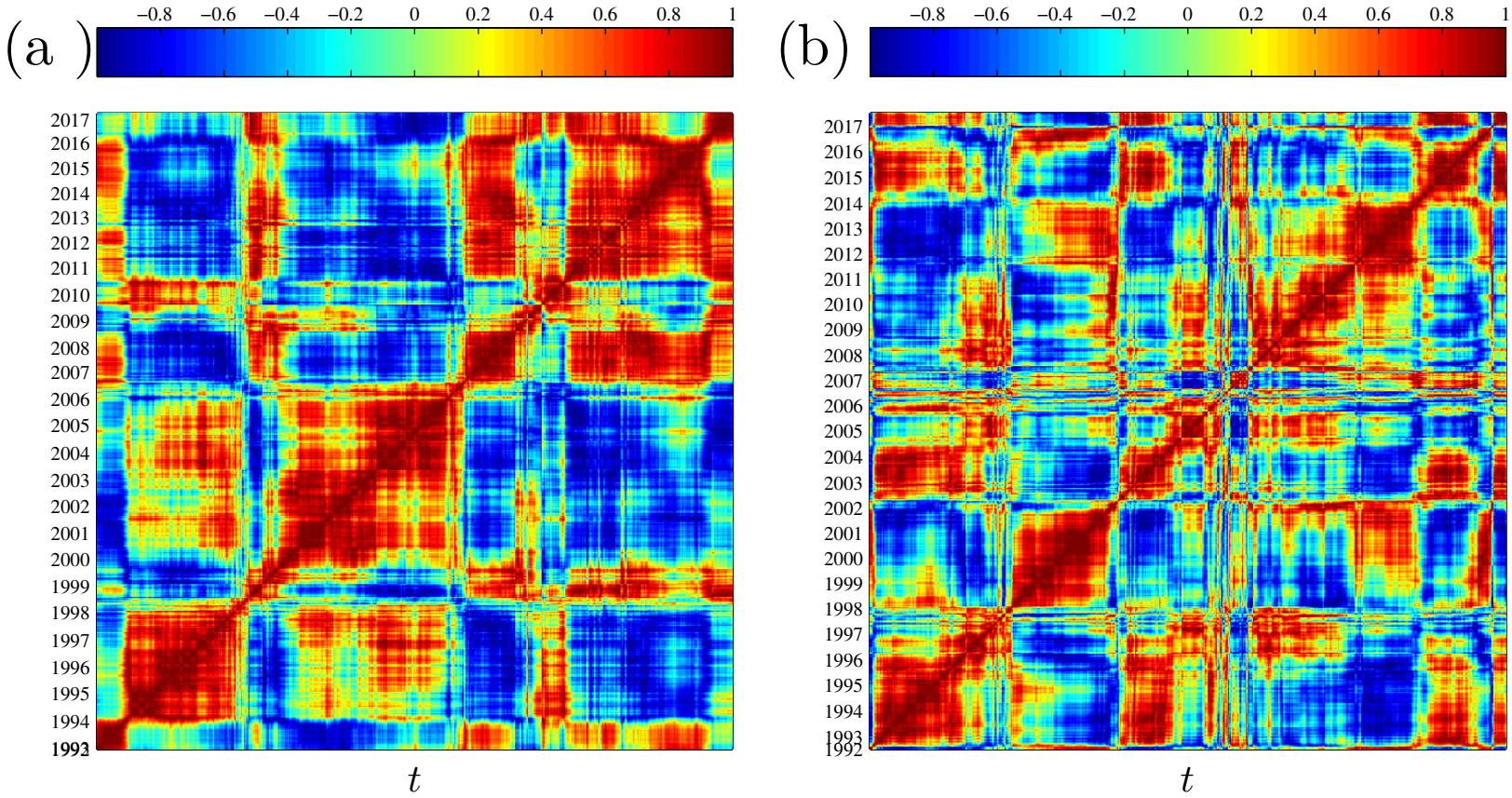


图: 金融时间序列模体分布向量的相关系数矩阵。(a) 上证指数；(b) 杜邦股票。

金融时间序列

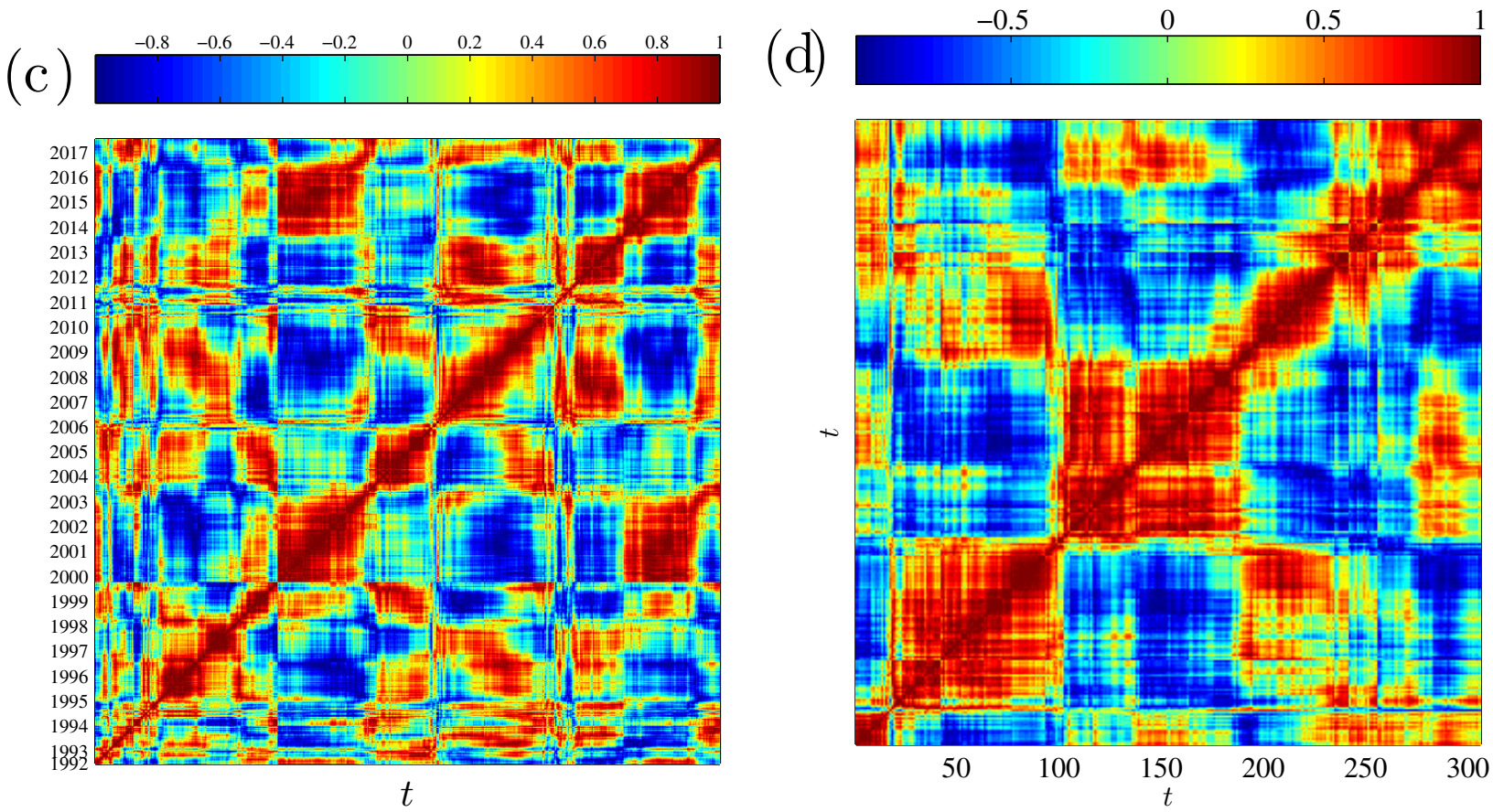


图: 金融时间序列模体分布向量的相关系数矩阵。(c) 埃克森美孚股票; (d) CHF_EUR汇率

金融时间序列

- 将每个模体M_i在移动窗口$[t-l+1,t]$中的出现频率$f_i(t)$进行z-score标准化，就可以得到对应的模体分布向量$z(t)$，计算模体分布向量$z(t_i)$和$z(t_j)$之间的相关系数$s(t_i,t_j)$，得到上证指数收益率时间序列对应模体分布向量的相关系数矩阵。
- 图(a)所示。根据收益率时间序列中模体结构的相似性，我们可以将上证指数划分为几个时期：1997-1998 年亚洲金融危机，2001年互联网泡沫崩溃，2007-2008年的全球金融危机，2015年中国股市危机等。
- 图(b-c)分别给出了纽约证券交易所上市股票杜邦（NYSE:DD）和埃克森美孚（NYSE:XOM）的日度收益率对应模体分布向量的相关系数矩阵。图中的对角线分块表明模体结构的相似性可以识别股票市场状态的变化情况。