

用序数网络表征随机时间序列

将时间序列映射到网络的方法已成为处理日益复杂的表征复杂系统数据特征的重要工具。在不同的算法中，最近提出的序数网络因其简单性和计算效率而脱颖而出。但是，序数网络的应用主要集中在非线性动力学系统产生的时间序列上，而与简单随机过程有关的序数网络的基本属性仍然知之甚少。在这里，我们研究了随机时间序列，嘈杂的周期信号，分数布朗运动和地震震级序列中出现的序数网络的几种属性。对于随机序列的序数网络，我们提出了一种建立邻接矩阵的精确形式的方法，该方法反过来对于检测时间序列中的非随机行为以及序数模式之间缺失过渡的存在很有用。我们发现，从有序网络的相邻节点之间的转移概率估计的局部熵的平均值比标准排列熵更能抵抗噪声添加。我们表明，序数网络可用于估算时间序列的赫斯特指数，其准确性可与最新方法相媲美。最后，我们认为序数网络可以检测到由大地震引起的地球地震活动的突然变化。

一、引言

在过去的几十年中，研究人员当前处理的数据量急剧增加。如今，不仅数据量增加了，而且调查的主题和详细程度也是空前的。尽管庞大而详细的数据集使我们能够探讨有关材料科学[1]和艺术[2]等各种主题的定量问题，但这种复杂性通常会挑战可用于执行数据分析的方法和技术。甚至简单数据之间的区分（例如时间序列）也变得具有挑战性，具体取决于任务中涉及的主题和数据量。因此，开发有效且经常跨学科的数据分析方法代表了从复杂数据集中提取隐藏且有意义的模式的重要一步[3]。

自Nicolis等人的开创性工作以来，在已提出的不同方法中，应用复杂的网络工具来研究时间序列的想法引起了科学界的极大兴趣。[4]，张和小[5]，和拉卡萨等。[6]，特别是从事统计物理学的研究人员[7]。这些开创性的工作催生了将时间序列映射到复杂网络中的三种主要方法：分别是过渡网络，邻近网络和可见性图。在可见性图中，时间序列的每个数据点都映射到网络的一个顶点，并且根据成对的数据点之间的可见性条件绘制链接[6,8]。依次，邻近网络假设时间序列段发挥节点的作用，每对段之间估计的相似性度量定义了网络链接。主要是由于其与递归图的紧密联系[7]，递归网络是基于邻近性的方法中的一种重要方法。在这种方法中，将时间序列段视为节点，并使用标准度量标准（例如欧几里得距离）来创建连接。另一方面，过渡网络将时间序列分区上的离散状态视为节点，并且连接表示这些状态之间的过渡概率。说谎的过渡网络下的思想有很长的历史先例，可以追溯到马尔可夫链理论[9]，但是将其用作时间序列分析的工具的确是最近的[7]。

这项工作特别感兴趣的过渡网络是序数网络。此方法由Small [10]提出，并在Bandt和Pompe [11]的工作中直接的启发，该方法由基于有序或置换模式的思想的嵌入方法（称为Bandt-Pompe符号化方法）组成。在顺序网络中，节点表示时间序列分区内的元素之间可能的排序模式，并根据这些模式的时间顺序绘制链接。自从其在文献中首次出现以来，序数网络主要用于研究非线性动力学系统（例如通过迭代混沌图获得的时间序列）产生的时间序列，而很少有作品使用这些网络来表征随机过程或现实世界。时间序列[12]。同样令人惊讶的是，尽管最近对这种方法进行了概括（包括一种处理多元时间序列的方法[13]），但从简单的单变量随机时间序列（例如随机游走）获得的序数网络的基本属性仍然知之甚少。

在这里，我们通过调查从周期，随机和分数布朗运动时间序列获得的序数网络的属性来帮助填补这一空白。我们表明，序数网络的属性可用于高精度估计时间序列的赫斯特指数，其性能优于诸如去趋势波动分析之类的最新技术[14,15]。通过证明序数网络能够识别大地震（主震）发生后地球地震活动的突然变化，我们进一步证明了该算法在研究经验时间序列方面的有用性。

本文的组织如下。我们回顾了序数网络如何将时间序列映射到复杂的网络，并在第II节中讨论了其一般的结构约束。接下来，在第III节，我们研究了从周期性（III A节），完全随机（III B节），噪声周期（III C节），分数布朗运动（III D节）和经验地震活动中出现的序数网络的性质 时间序列（第III E节）。我们在第IV节中介绍我们的结论和最后的评论。

二、方法

如前所述，序数网络与Bandt和Shiha的符号化方法[16]以及Bandt和Pompe的置换熵框架[11]有直接关系。为了更好地理解这种联系，我们首先回顾一下这些想法。在所谓的Bandt-Pompe方法中，根据长度为N的给定时间序列 $\{x_t\}_{t=1, \dots, N}$ ，我们构造 $n = N - d + 1$ 个长度为d的重叠分区（嵌入尺寸），由 $w_s = (x_s, x_{s+1}, \dots, x_{s+d-2}, x_{s+d-1})$ 表示，其中 $s = 1, \dots, n$ 是分区索引。接下来，我们评估 $(0, 1, \dots, d-1)$ 的置换 $\pi_s = (r_0, r_1, \dots, r_{d-1})$ ，将 w_s 的元素排序（按升序排列），是，由 $x_{s-r_{d-1}} \leq x_{s-r_{d-2}} \leq \dots \leq x_{s-r_0}$

r_0 定义的置换。我们进一步假设在分区内绘制的情况下，如果 $x_{s-r_i} < x_{s-r_{i-1}}$ ，则 $r_i < r_{i-1}$ 保持出现的顺序[17,18]。经过这两个步骤，我们获得了符号序列 $\{\pi_s\}_{s=1, \dots, n}$ 。获得所有排列后，我们可以计算出每个 $d!$ 的相对频率 $p_i(\pi_i)$ 。符号 $(0, 1, \dots, d-1)$ 的可能置换 π_i ,

$$p_i(\pi_i) = \frac{\text{number of partitions of type } \pi_i \text{ in } \{\pi_s\}}{n}, \quad (1)$$

从中我们估计序数概率分布 $P = \{p_i(\pi_i)\}_{i=1, \dots, d!}$ 。置换熵就是序数概率分布的香农熵[19]，即

$$H = - \sum_{i=1}^{d!} p_i(\pi_i) \log_2 p_i(\pi_i), \quad (2)$$

$\log_2(\dots)$ 代表以2为底的对数。嵌入维数 $d > 1$ 是该方法的唯一参数，并且由于它定义了可能排列的数量，因此条件必须满足 $d! \leq N$ ，为了可靠地估计分布 $P = \{p_i(\pi_i)\}_{i=1, \dots, d!}$ 。H的值量化了 x_t 的局部排序模式中的随机性： $H \approx \log_2 d!$ 表示 x_t 的元素在本地是随机排序的，而 $H \approx 0$ 表示 x_t 的元素很可能以特定顺序出现。

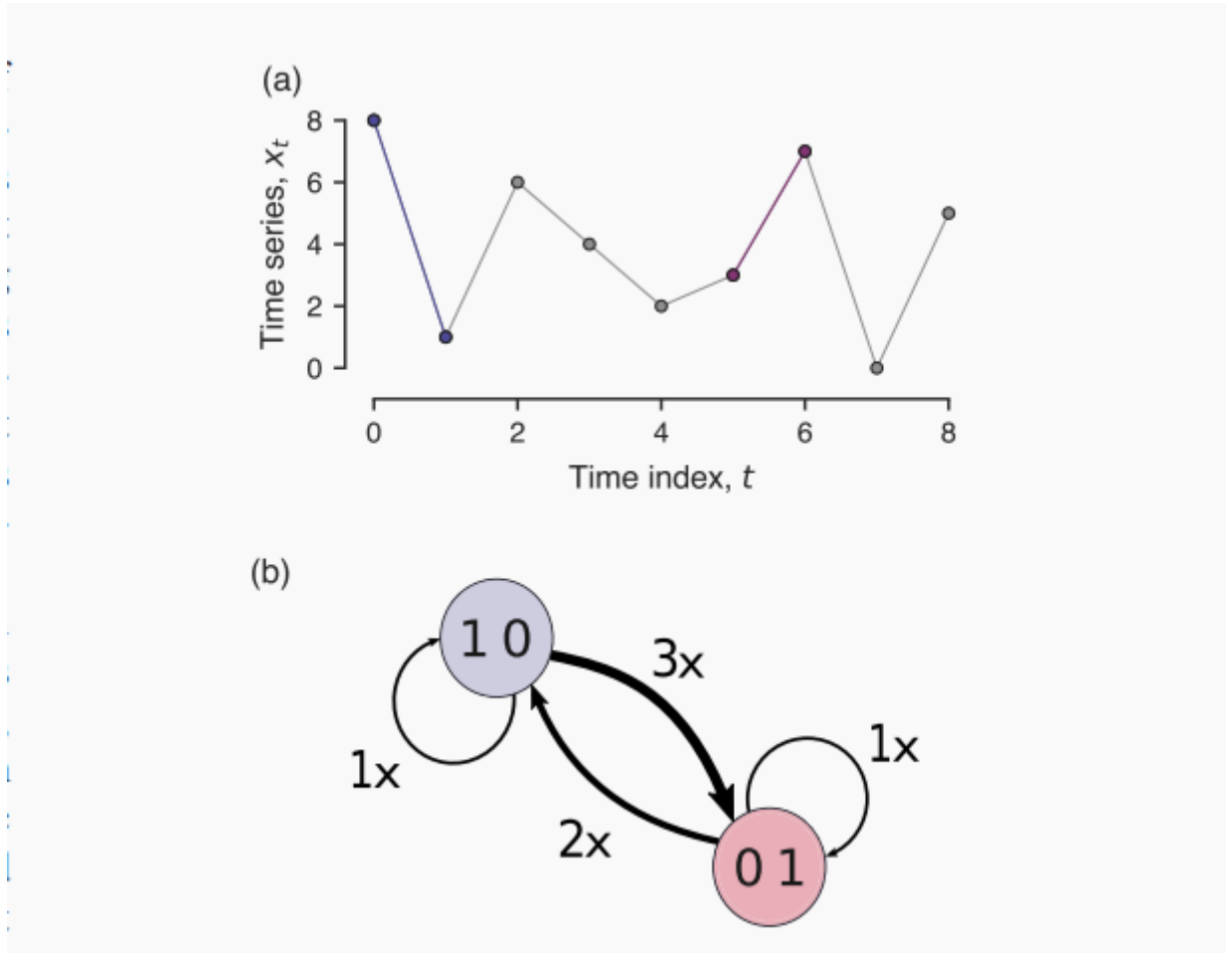


图 1.用于将时间序列映射到网络的序数网络方法。(a) 时间序列 $x_t = \{8, 1, 6, 4, 2, 3, 7, 0, 5\}$ 和相应的符号序列 $\{(1, 0), (0, 1), (1, 0), (1, 0), (0, 1), (0, 1), (1, 0), (0, 1)\}$ 是从Bandt-Pompe方法获得的，嵌入尺寸为 $d=2$ 。(b) 与时间序列 x_t 相关的序数网络的可视化。节点“01”(红色)和“10”(蓝色)分别对应于排列 $(0,1)$ 和 $(1,0)$ 。基于符号序列中排列的时间顺序，在网络节点之间绘制有向边，并且它们的权重反映每个可能的顺序的相对频率。当自身在符号序列中跟随置换时，就会出现自环。

受Bandt-Pompe方法的启发，Small [10]建议使用符号序列 $\{\pi_s\}_{s=1, \dots, n}$ 创建序数网络，即时间序列 $\{x_t\}$ 的图形表示 $\{t=1, \dots, N\}$ 。该方法包括将所有可能的置换 $\{\pi_i\}_{i=1, \dots, d!}$ 视为网络节点，并在序列 $\{\pi_s\}_{s=1, \dots, n}$ 内连续发生的每一对置换之间绘制连接。 $\{s=1, \dots, n\}$ 。边根据排列的时间顺序来定向（例如，对于符号序列 $\{\pi_1, \pi_2\}$ ，链接为 $\pi_1 \rightarrow \pi_2$ ），并且还按相对权重进行加权连续 (π_i, π_j) 发生的频率为 $\{\pi_s\}_{s=1, \dots, n}$ 。该网络的加权邻接矩阵的元素为

$$p_{i,j} = \frac{\text{number of times } \pi_i \text{ is followed by } \pi_j \text{ in } \{\pi_s\}}{n-1}, \quad (3)$$

其中 $i, j=1, 2, \dots, d!$ 分母 $n-1$ 表示排列转变的总数。

为了阐明有序网络方法，让我们考虑一个简单的时间序列 $x_t = \{8, 1, 6, 4, 2, 3, 7, 0, 5\}$ ，其中 $N=9$ 个元素，并且嵌入维数 $d=2$ [图10]。1 (a)]。我们的网络由与置换 $\pi_1 = (0, 1)$ （标记为“01”）和 $\pi_2 = (1, 0)$ （标记为“10”）相关的两个节点组成。第一个分区为 $w_1 = (8, 1)$ ，并且对应于置换 $(1,0)$ ，因为 w_1 的元素按降序排列。由于 w_2 的元素按升序排列，因此第二个分区为 $w_2 = (1, 6)$ ，并且对应于置换 $(0,1)$ 。通过对所有 $n=8$ 个分区重复此过程，我们获得符号序列 $\{(1, 0), (0, 1), (1, 0), (1, 0), (0, 1), (0, 1), (1, 0), (0, 1)\}$ 。接下来，通过分析排列对之间的所有连续演替，我们发现 $(0, 1) \rightarrow (0, 1)$ ， $(0, 1) \rightarrow (1, 0)$ ， $(1, 0) \rightarrow (0, 1)$ 和 $(1, 0) \rightarrow (1, 0)$ 分别出现1、2、3和1次。因此，与此特定时间序列关联的加权邻接矩阵为

$$\begin{array}{cc} & \begin{matrix} (0, 1) & (1, 0) \end{matrix} \\ \begin{matrix} (0, 1) \\ (1, 0) \end{matrix} & \begin{bmatrix} 1/7 & 2/7 \\ 3/7 & 1/7 \end{bmatrix} \end{array}.$$

图1 (b) 显示了生成的网络的可视化。对于这个说明性的时间序列，由于 $p_1(\pi_1) = 1/2$ 和 $p_2(\pi_2) = 1/2$ ，所以置换熵为 $H=1$ 位。

通过与排列熵进一步并行，McCullough等人。[20]提出了基于过渡概率来估计另一个熵度量的方法（另见Unakafov和Keller [21]）。这个想法是通过计算节点 i 的局部熵

$$h_i = - \sum_{j \in O_i} p'_{i,j} \log_2 p'_{i,j}, \quad (4)$$

其中 $p'_{i,j} = p_{i,j} / \sum_{k \in O_i} p_{i,k}$ 是从节点 i 到节点 j 的重新归一化的转移概率，而 O_i 代表节点 i 的输出邻域（全部离开节点 i 的边）。因此， h_i 的值量化了与从置换 π_i 开始的转变相关的随机度。例如， $h_i = \log_2 |O_i|$ （ $| \dots |$ 代表集合基数，即 i 的可能输出邻居的数目），如果所有可能的离开 π_i 的跃迁都是等概率的，而 $h_i = 0$ ，当仅存在一个离开 π_i 的跃迁时。对于图1中的网络， $h_1 = h(01) = 0.918$ ， $h_2 = h(10) = 0.811$ 。

文献[20]提出了计算全局节点熵的方法。

$$H_{GN} = \sum_{i=1}^{d!} p'_i h_i, \quad (5)$$

其中 $p'_i = \sum_{j \in I_i} p_{j,i}$ 是从其传入邻域 I_i （节点 i 的强度）到达节点 i 的概率。值得注意的是，对于长时间序列， $p'_i \approx p_i(\pi_i)$ ，因此置换熵是与序数网络节点强度相关的香农熵。（5）表示局部加权平均值节点熵，并量化在时间序列 x_t 中发生的所有（一阶）置换过渡的全局随机性程度。与与随机过程相关的时间序列相比，更规则的时间序列具有较小的HGN值。对于图1所示的示例，我们发现 $HGN = 0.872$ 。文献[20]进一步提出在估计全局节点熵（HGN）时忽略自动循环。正如他们所论证的那样，在处理过采样的时间序列时，自动循环可能会变得过于激烈，从而可能使HGN的值产生偏差。在我们的工作中，由于要研究的时间序列的性质，我们选择考虑自动循环。

为了更精确地定义术语，公式 (5) 通常被称为条件置换熵[21]。还值得一提的是，我们的序数网络的熵定义和其他拓扑性质与图的单纯形复形[22]的概念不同，后者使用代数拓扑方法定义拓扑信息量度[23]，例如图的熵。Andjelković等人使用的拓扑级别。 [24]用于调查交通拥堵情况。

除了全局节点熵 H_{GN} 以外，基于有序网络的时间序列研究还探索了可用于表征复杂网络的众多网络指标。但是，序数网络的重要特征自然会被继承，并受到产生网络的符号序列的限制。序数网络的这些属性中的某些已经在以前的工作中隐式讨论过[10,25,26]，但它们仍然缺乏关注。一个重要的限制与节点的最大连接数有关。当 $d = 2$ 时，这种约束并不明显，因为这种情况显示了所有可能的连接[图 1 (b)]。另一方面，当 $d = 3$ 时，存在六个可能的排列（即 $d!$ ），因此在有序网络中有六个节点，但是每个节点仅与其他三个节点连接。这种限制源自以下事实：对分区 w_s 进行排序的排列 π_s 只能跟随与分区 w_{s+1} 关联的三个顺序模式。

为了说明这种约束，让我们考虑具有置换 $\pi_1 = (0, 1, 2)$ 的分区 $w_1 = (x_1, x_2, x_3)$ ，也就是说，其元素的排序应使 $x_1 < x_2 < x_3$ 。现在假设下一个分区是 $w_2 = (x_2, x_3, x_4)$ ，其中 x_2 和 x_3 是 w_1 中包含的相同元素。因为 $\pi_1 = (0, 1, 2)$ ，我们知道条件 $x_2 < x_3$ 应该在 w_2 中成立，因此，在与 w_2 相关的置换中，数字0应该在1之前。在 $d = 3$ 的所有可能置换中， $\{(0,1,2), (0,2,1), (1,0,2), (1,2,0), (2,0,1), (2,1,0)\}$ ，只有 $(0,1,2)$ ， $(0,2,1)$ 和 $(2,0,1)$ 满足先前条件。因此，只能出现三个排列 π_2 在 π_1 之后，这又将节点 $(0,1,2)$ 的传入连接数限制为3。通过使用相同的参数，我们可以验证传入的连接数也限制为3，并且这些约束保留所有节点。

这些想法推广到嵌入维数 d 的所有值，以便排列网络中的所有节点的入度和出度限制为0到 d 之间的数字。因此，序数网络的边不能超过 $d \times (d!)$ 。这种限制是由于以下事实造成的：分区 w_s 中的元素的排序部分地执行到下一个分区 w_{s+1} ，因为 w_{s+1} 的元素包括 w_s 的 $d-1$ 个元素。此外，嵌入维数越小，时间序列分区之间共享的元素越少，因此，过去的序列对于确定将来的排列迅速变得不重要。

与置换之间的转换相关联的约束的另一个有趣的结果与网络中的自边缘有关。通过分析所有转换，我们得出结论，自环仅对于序数网络的两个特定节点存在，而与嵌入维数 d 无关。这两个节点仅与升序或降序关联，也就是说，与元素在所有元素中连续增加或减少的分区相关的置换。例如，在 $d = 3$ 的情况下，这些节点对应于排列 $(0,1,2)$ 和 $(2,1,0)$ ，而在 $(0,1,2,3)$ 和 $(3,2,1, 0)$ $d = 4$ 。

值得注意的是，已经提出并应用了从时间序列构建序数网络的其他通用算法。这些不同的数字配方包括使用非重叠分区[10,27]，具有时间滞后元素的分区[25]以及包含有关时间序列元素的幅度信息[26]。自然，我们在这里讨论的约束对于这些通用算法不成立。例如，在非重叠分区的情况下，给定的置换 π 可以跟随任何可能的置换，因为连续的分​​区不共享任何时间序列元素。对于时间滞后的元素，对于高阶过渡，排列之间也出现了类似的限制。这些通用算法也很有趣，可能值得进一步研究。在这里，我们将重点放在由Bandt和Shiha [16]和Bandt和Pompe [11]的开创性工作直接启发的序数网络中，如图1所示。

三、结果

A.简单时间序列的序数网络

我们通过分析基本时间序列产生的序数网络的结构来开始实证研究。可能要考虑的最简单的时间序列是单调（递增或递减）序列。在这种情况下，无论嵌入维数 d 是多少，有序网络都由一个节点（代表唯一的递增或递减置换）组成，并且只有一个自动循环（表示置换始终由其自身进行）。因此，单调时间序列的网络指标都是微不足道的。

周期序列是简单但有趣的信号的另一个示例。简单周期序列的序数网络形成循环结构，其中节点和边的排列暗示了序列本身。如果在一个时间序列的周期内数据点的数量为 T 并且嵌入维数为 $d = T$ ，则由于时间序列重复自身，因此有序网络是具有 T 个节点（不同排列的数量）的循环图（无自动循环）在 T 个数据点之后。此外，如果我们考虑更大的嵌入尺寸（ $d > T$ ），则此循环结构不会改变；仅修改与网络节点关联的排列符号[图 2 (a)]。由于这种行为，周期信号的序数网络显示出恒定的平均度数（对于输入和输出连接，图2 (c)），并且直径随嵌入尺寸线性增长[图 2 (d)]。我们还可以证明，从一个节点到所有其他节点的平均加权距离为 $1/d \sum_{i=1}^{d-1} i$ （ i/d ），这与我们由于图对称而选择的节点无关。因此，整个网络的平均加权最短路径也为

$$\langle l \rangle_{\text{per}} = \frac{1}{d} \sum_{i=1}^{d-1} \left(\frac{i}{d} \right) = \frac{d-1}{2d}, \quad (6)$$

其中， $\langle l \rangle$ 表示从特定节点到所有其他节点的平均加权距离，并且因子 $1/d$ 占网络中所有节点的平均值[图 2 (e)]。当所有现有边具有相同权重时，此表达式有效，该权重适用于长时间序列或由整数周期组成的序列。

B. 随机序数网络

我们进一步研究了随机信号产生的序数网络。与单调和周期性信号不同，从随机序列出现的序数网络（以下称为随机序数网络）被期望显示所有可能的连接，因为所有排列都同样可能发生在足够长的随机信号中。图 2 (b) 显示了由高斯白噪声产生的具有不同嵌入维数的有序网络的示例。我们注意到这些网络是由 d 组成的！节点，每个节点都有 d 个传出链接和 d 个传入链接。由于这些特性，随机序数网络的平均度和直径随嵌入尺寸 d 线性增加。 2 (c) 和 2 (d)]。

这些随机序数网络的一个有趣的方面是，尽管所有排列在随机时间序列中都是等概率的，但边缘权重并非都相同。实际上，所有节点的 d 个向外边缘之一的权重是其余 $d-1$ 个边缘的权重的两倍[图 2 (b) 中的黑色链接]，这一结果适用于从任何连续概率分布中得出的长随机数序列。为了说明这是如何发生的，让我们考虑与随机序列相关联的第一个 $d=3$ 的分区是 $w_1 = (x_1, x_2, x_3)$ ，相应的排列是 $\pi_1 = (0, 1, 2)$ （即， $x_1 < x_2 < x_3$ ）。下一个分区是 $w_2 = (x_2, x_3, x_4)$ ，新元素 x_4 符合以下条件之一：(i) $x_4 < x_1 < x_2 < x_3$ ，(ii) $x_1 < x_4 < x_2 < x_3$ ，(iii) $x_1 < x_2 < x_4 < x_3$ 和 (iv) $x_1 < x_2 < x_3 < x_4$ 。条件 (i) 和 (ii) 得出 $\pi_2 = (2, 0, 1)$ ，而 (iii) 得出 $\pi_2 = (0, 2, 1)$ 和 (iv) 得出 $\pi_2 = (0, 1, 2)$ 。因此，有两种可能性可以找到 $\pi_2 = (2, 0, 1)$ ，这使转换 $(0, 1, 2) \rightarrow (2, 0, 1)$ 的可能性是 $(0, 1, 2) \rightarrow (0, 1, 2)$ 的两倍， $(2, 1)$ 或 $(0, 1, 2) \rightarrow (0, 1, 2)$ 。还值得注意的是，如果我们绘制大量样本 (x_1, x_2, x_3) ，以使 $x_1 < x_2 < x_3$ ，则 x_1, x_2 和 x_3 的平均值收敛到概率分布的四分位数时间序列，因此条件 (i) - (iv) 是等概率的。当 $d=3$ 时，对于所有其他一阶跃迁以及任何其他嵌入维，都适用相同的想法。

确定更可能遵循 π_s 的排列 π_{s+1} 的一般经验法则是选择 π_{s+1} ，其中等于“ $d-1$ ”的符号适合符号“0”在 π_s 中的位置。例如，如果 $\pi_s = (3, 2, 1, 0)$ ，则 $\pi_{s+1} = (2, 1, 0, 3)$ 的可能性是 $\pi_{s+1} \in \{(3, 2, 1, 0), (2, 1, 3, 0), (2, 3, 1, 0)\}$ ，因为 π_{s+1} 中的符号“3”位于同一位置，而符号“0”位于 π_s 中。这个结果使我们能够为任何嵌入维 d 建立随机有序网络的加权邻接矩阵。为此，我们从具有 $d!$ 的网络开始。节点（每个节点代表一个特定的置换 π_i ），并在所有节点对 π_i 和 π_j 之间画出具有单位权重的有向连接，对于这些过渡而言， $\pi_i \rightarrow \pi_j$ 可能是过渡的。接下来，我们将所有可能连接的权重从 1 更新为 2。最后，将所得邻接矩阵 $p_{i,j}$ 的元素除以因子 $(d+1)!$ ，该因子代表所有单位权重的总和 $[d! \cdot (d-1)]$ 加上两倍权重为的边数。通过使用该加权邻接矩阵，我们可以在数值上评估具有任意嵌入维 d 的随机有序网络的任何网络度量。图 2 (e) 显示了根据这些理论网络估计的平均加权最短路径随 d 的变化，其中我们观察到，对于大 d 值，此度量接近零，因为边缘权重随 d 的增加而减小。

我们可以分析估计的随机有序网络的另一个有趣特性是全局节点熵[公式 (5)]。为此，我们首先注意到节点 i 的局部熵为

$$\begin{aligned} h_i &= -\tilde{p}'_{i,j} \log_2 \tilde{p}'_{i,j} - \sum_{j=1}^{d-1} p'_{i,j} \log_2 p'_{i,j} \\ &= -\frac{2}{d+1} \log_2 \left(\frac{2}{d+1} \right) - \left(\frac{d-1}{d+1} \right) \log_2 \left(\frac{1}{d+1} \right), \end{aligned} \quad (7)$$

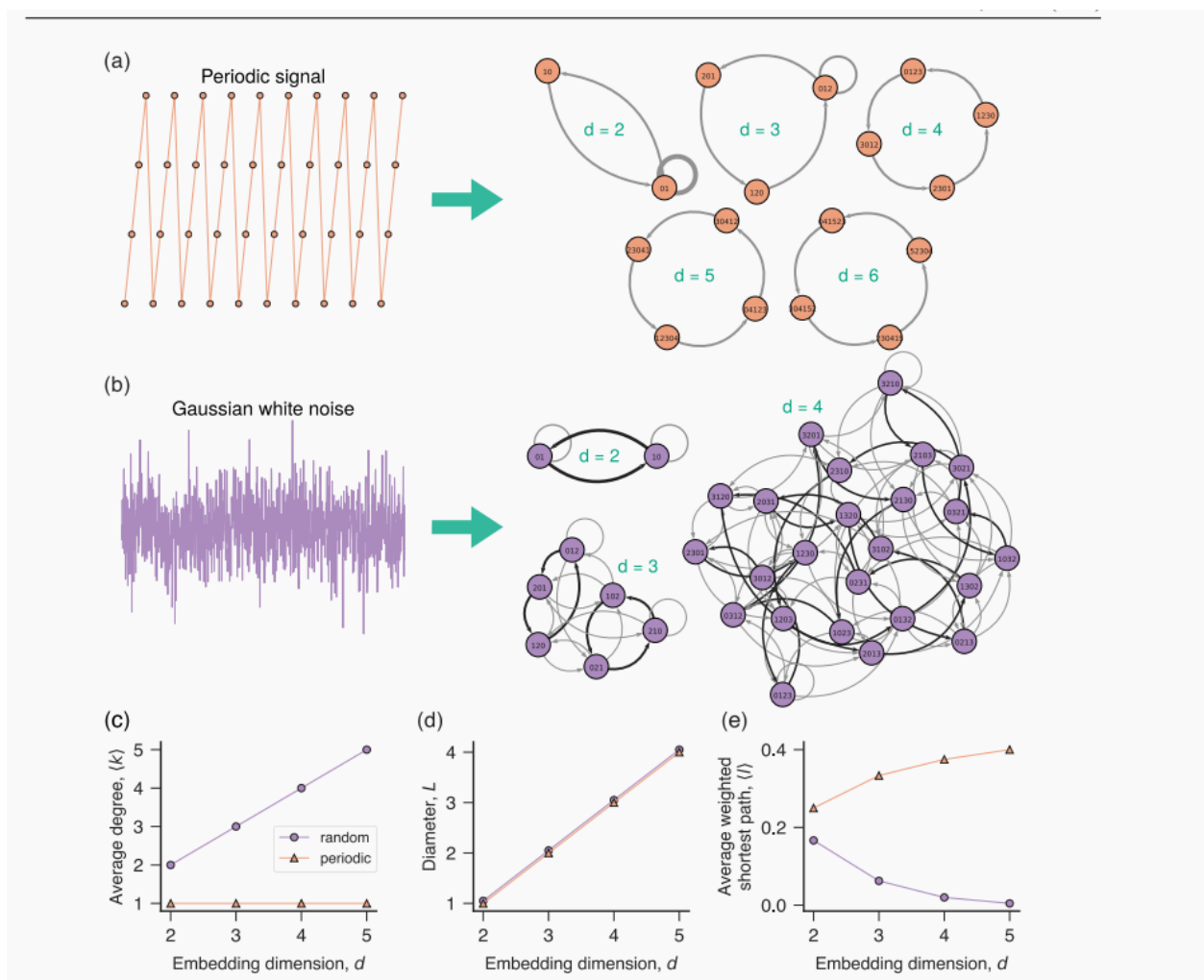


图 2.周期性和随机时间序列的序数网络。(a) 说明将周期信号映射到具有不同嵌入尺寸 d (在面板内表示) 的有序网络中的情况。(b) 将高斯白噪声映射到具有不同嵌入尺寸 d 的有序网络中 (在面板内显示)。黑色边缘表示发生过渡的可能性是两倍。(c) 对于从随机 (圆形) 和周期性 (三角形) 信号中出现的有序网络, 平均度数, (d) 直径 L 和 (e) 平均加权最短路径作为嵌入尺寸 d 的函数。周期信号的周期 T 与嵌入尺寸 ($d = T$) 相匹配。前三个属性的结果在足够长的时间序列内保持不变, 也就是说, 当可以获得所有排列转换的可靠估计时

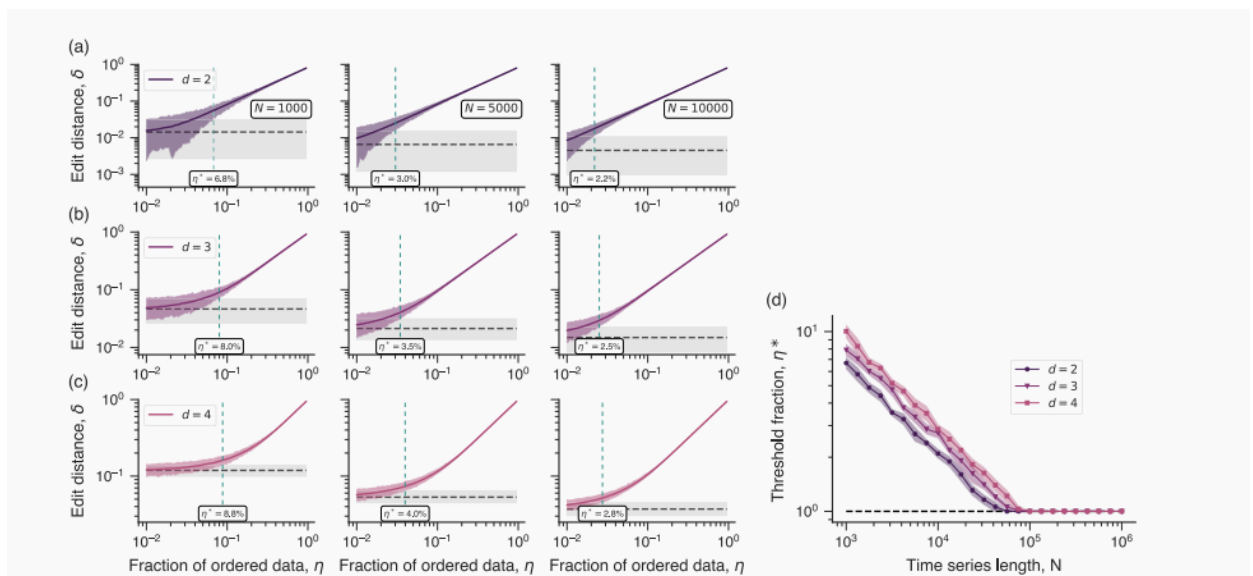


图 3.使用随机序数网络检测非随机行为。(a) - (c) 面板显示部分有序白噪声的有序网络 (η 是有序元素的分 数) 与理论上嵌入尺寸 $d=2$ 的随机有序网络之间的编辑距离 δ (彩色线) 的平均值, 3和4。在每个面板中, 第一列的结果是从1000个时间序列的集合中获得的, 每个集合的 $N = 1000$ 个数据点, 而第二列和第三列的结果是从1000个时间序列的集合中获得的, 其中 $N = 5000$ 和 10000 个元素, 分别。彩色阴影区域代表95%的置信带, 而灰色阴影区

域则是由有限随机序列计算得出的 δ 的95%随机置信带。垂直虚线表示有序元素的阈值分数 η^* ，在该阈值分数下，部分有序的随机序列变得与相同长度的完全随机序列有区别。(d) 对于不同的嵌入，阈值分数 η^* 是时间序列长度 N 的函数 $d \in \{2, 3, 4\}$ 。标记表示从十个副本中估计的 η^* 平均值，阴影区域是一个标准偏差带，虚线表示在我们的数值实验中使用的有序元素的最小比例 (1%)。

$p'_{\{i, j\}} = 2p_{\{i, j\}}$ 表示重新归一化的过渡概率，发生的可能性是其两倍。通过将此结果插入到公式 (5) $p'_{i=1}/d!$ (可置换)，我们发现不相关的随机数序列的全局节点熵为

$$H_{GN}^{(rand)} = \log_2(d+1) - (\log_2 4)/(d+1). \quad (8)$$

我们注意到， $H_{GN}^{(rand)}$ 的值始终小于所有跃迁都相等的情况下获得的值 (在这种情况下， $H_{GN}^{(equi)} = \log_2 d$)，只有当 $d \rightarrow \infty$ 时，我们才具有 $H_{GN}^{(rand)} \rightarrow H_{GN}^{(equi)}$ (等于)。因此，在存在 $H_{GN} > H_{GN}^{(rand)}$ 的时间序列的意义上， $H_{GN}^{(rand)}$ 的值不代表给定 d 的最大熵。

这些理论随机网络的一个有趣应用是检测经验时间序列中的非随机行为。一种选择是通过适当的图距测量直接将经验序数网络与其随机的理论对应网络进行比较[28,29]。为了说明这种可能性，我们研究了通过估计序数网络之间的编辑距离 δ [29]，是否可以部分有序的高斯白噪声与纯噪声区分开。对于加权有向图，此简单度量表示需要重新分配的边缘权重 (强度) 的数量，以使两个图变得相等。

我们生成了白噪声级数的集合，其中连续元素的给定分数 η 以升序放置。接下来，我们将这些时间序列映射到有序网络中，并根据有序数据 η 的比例估算到理论随机网络的编辑距离 δ 的平均值。图3 (a) –3 (c) 显示了嵌入尺寸 $d \in \{2, 3, 4\}$ 以及 $N \in$ 序列时， δ 平均值随 η 的变化以及95%置信区的行为。{1000, 5000, 10000}个元素。

正如预期的那样，我们观察到在所有情况下 δ 的值都随 η 的增加而增加。我们还估计了白噪声序列的经验序数网络 (具有相同的长度) 与用于构建随机置信带的理论随机网络之间的编辑距离的平均值。超出此随机置信带的 δ 值表示一个可靠的指标，表明经验时间序列显示出与纯随机行为的偏差。此外，这些置信区域还使我们能够识别阶数的阈值分数 η^* ，从这些阈值分数中，编辑距离 δ 能够准确地检测到此异常。图3 (a) –3 (c) 表明，这种方法能够检测到具有1000个元素的时间序列的排序，当 $d=2$ 时 $\eta \approx 6.8\%$ ，当 $d=3$ 时 $\eta \approx 8.0\%$ ，对于当 $d=4$ 时， $\eta \approx 8.8\%$ 。这些数字还表明 η^* 的值随时间序列长度 N 的增加而减小。但是，我们观察到 $d=2$ 显示最小的 η^* 值，而与 N 无关。

我们进一步估计了检测过程的十次实现中 η^* 的平均值与 N 的关系。图3 (d) 表明，对于 $d \in \{2, 3, 4\}$ ， η^* 随着 N 呈指数下降，并且对于很长的序列 ($N > 105$)，所有 d 值在检测所用有序元素的最小分数方面都同样有效。在我们的数值实验中 ($\eta = 1\%$)。我们观察到 $d=2$ 产生最小的阈值分数，但是 $d=2$ 和 $d=3$ 之间的一个标准偏差置信带的接近表明这些情况之间的差异很小。发生这种情况的原因是，当我们对一系列序列进行部分排序时，我们会创建不成比例的更多升序模式，并且由于 $d=2$ 的序数网络只有四个链接 (和两个节点)，所以与考虑使用具有更高的序数网络时相比，更容易发现这种不平衡 d 。如果异常模式变得更加复杂，则这种情况可能会改变，因此在这种情况下，具有更高嵌入维数的序数网络在检测异常方面可能会表现更好。但是，此示例说明，在嵌入维数和时间序列长度之间总会有一个权衡。

通过了解随机有序网络的确切形式，我们还可以估算经验时间序列中缺失过渡的比例。这种分析与Amigó和同事[30-32]引入并由几本著作[33-35]探索的缺失排列或禁止模式类似。这些工作已经观察到，在混沌系统中不会出现某些序数模式 (例如，在完全发展的混沌情况下，排列 (2,1,0) 永远不会出现在逻辑映射中)，甚至随机过程也可能呈现缺失的序数模式，具体取决于时间序列长度和嵌入维数的选择。但是，随时间序列变长，随机过程中丢失模式的数量会减少，因此，通常将其称为“错误禁止模式”。图4 (a) 说明了在不同嵌入尺寸 $d \in \{3, 4, 5, 6, 7, 8\}$ 的情况下，高斯白噪声中缺失置换的分数如何随时间序列长度的增加而减少。为了扩展这些结果，我们测试了嘈杂的级数是否也存在“错误的禁止转换”。为此，我们从长度为 N 的高斯白色序列中评估了序数网络，并将其转换链接的数量与随机序数网络的确切形式的转换链接数量进行了比较。图4 (b) 显示了不同嵌入尺寸的这些值，在该图中我们观察到白噪声序列也根据“ N ”和“ d ”的值而呈现“假缺失过渡”。我们注意到，这些曲线的形状类似于缺失排列的一部分所观察到的形状。但是，使用 N 时，丢失过渡的数量下降的速度要比丢失排列的数量下降得慢。还值得一提的是，在这些经验序数网络和随机序数网络的确切形式之间计算出的编辑距离不仅可以量化缺失过渡的存在，还可以量化其在时间序列中的出现频率差异。

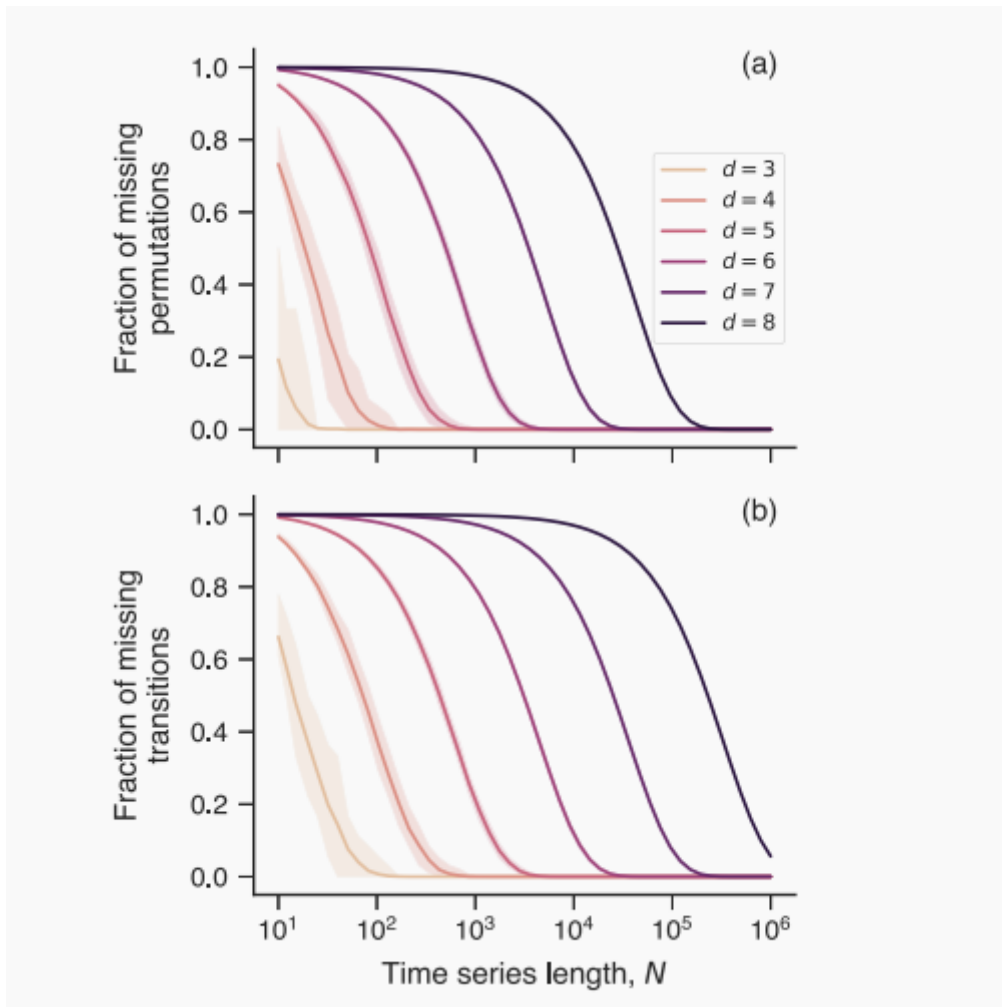


图 4.随机序列中缺少序数模式和排列变换。(a) 对于不同尺寸的嵌入维数 $d \in \{3, 4, 5, 6, 7, 8\}$ ，一元方差的高斯白噪声序列中的缺失置换的分数和 (b) 缺失置换转变的分数作为时间序列长度 N 的函数。在两个图中，彩色曲线表示每个嵌入维的100个实现的白噪声序列的平均值，阴影区域为95%置信区间。

C. 嘈杂的周期性时间序列的序数网络

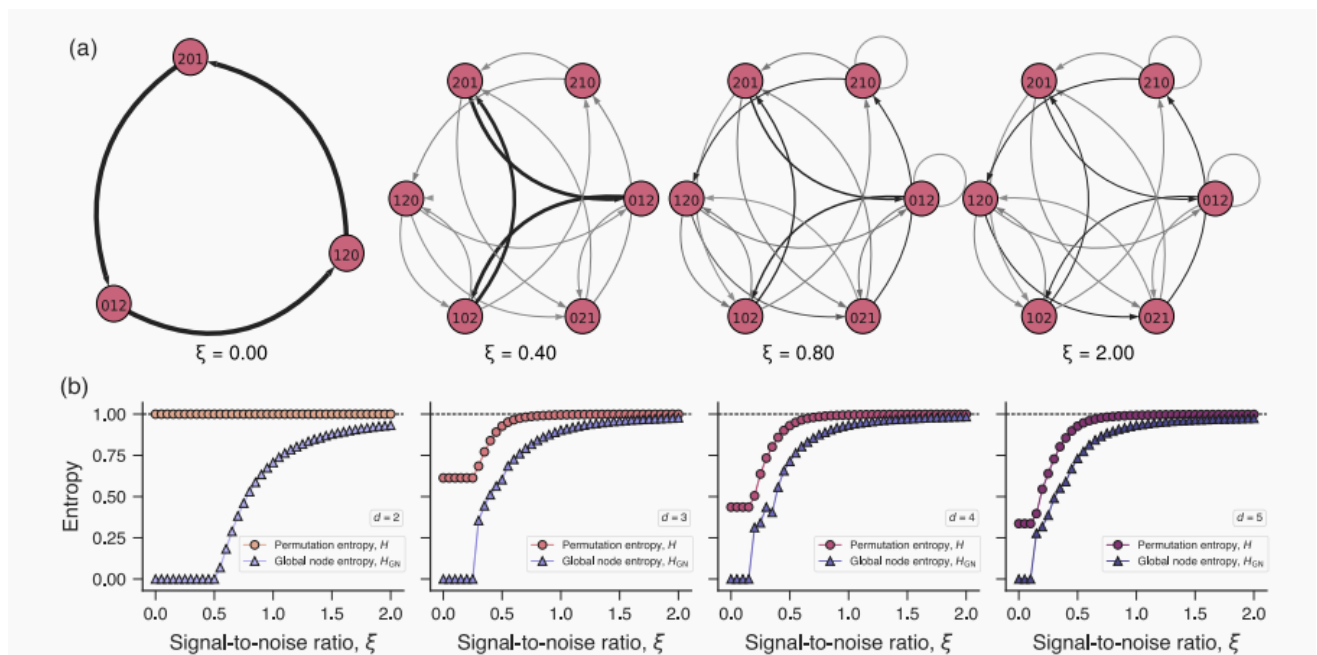


图5. 置换噪声和全局节点熵对噪声相加的鲁棒性。(a) 与 $d = 3$ 和不同信噪比 ξ (在面板内显示) 相关的有周期噪声信号相关的有序网络的可视化。(b) 从 $d = 2, 3, 4$ 和5的100个有噪声周期信号实现的集合中获得的置换熵 H (圆形) 和全局节点熵 H_{GN} (三角形) 的平均值。时间序列包含104个元素, 周期 T 设置为等于嵌入维数 ($T = d$)。置换熵通过其最大值归一化, 全局节点熵通过随机有序网络的值归一化[公式 (8)]。细阴影区域对应一个标准偏差带, 虚线是随机有序网络的值。

研究当我们向这样的时间序列添加噪声时, 周期性信号的序数网络如何转换为随机序数网络也很有趣。为此, 我们生成长度为104且周期 T 的锯齿状信号, 其形式为 $x_t = \{0, 1, 0, 1, \dots\}$ ($T = 2$), $x_t = \{0, 1/2, 1, 0, 1/2, 1, \dots\}$ ($T = 3$), 依此类推。接下来, 在时间间隔 $[-\xi, \xi]$ 中, 这些时间序列的所有元素以均匀分布的噪声递增, 其中 ξ 是控制信噪比的参数。图5 (a) 显示了从这些有噪声的周期信号获得的有序网络的示例, 其中 $\xi \in \{0, 0.4, 0.8, 2\}$ 并且 $d = T = 3$ 。正如所料, 周期有序网络的环状结构 ($\xi = 0$) 随着我们增加 ξ 的值而接近随机序数网络。我们注意到, 这种渐进过程始于所有可能的网络节点的出现 (对于 ξ 的较小值会发生), 然后对于 $\xi = 0$ 的网络的三个初始链路逐渐衰落, 而其他链路则得到增强。我们还证实, 随着信噪比的增加, 在嘈杂的周期信号的有序网络与随机有序网络的精确形式之间计算出的编辑距离接近于零。

这些嘈杂的周期性时间序列还使我们能够测试全局节点熵 H_{GN} 的噪声添加的鲁棒性。[公式 (5)]与标准排列熵 H 相比。[公式 (2)]。为此, 对于每个 $\xi \in \{0, 0.05, 0.1, \dots, 2\}$, 我们生成100个长度为104的锯齿状信号的集合。接下来, 我们估计每个时间序列的 H_{GN} 和 H 值, 以及所有 ξ 值的这些量的平均值。我们进一步标准化了这些数量, 将 H_{GN} 值除以 H_{GN} (兰特) [公式 (8)]和 H 由 $\log_2 d!$ 。图5 (b) 显示了 $d = 2, 3, 4$ 和5时 H_{GN} 和 H 的归一化值的平均值与 ξ 的关系。我们注意到, H_{GN} 的值在抗噪声添加方面比 H 更为健壮, 并且因此, 在区分具有不同 ξ 值的时间序列时, 效率更高。

对于 $d = 2$ (和 $T = 2$), 此特征更为明显, 因为在这种情况下 $x_t = \{0, 1, 0, 1, \dots\}$, 因此即使 $\xi < 0.5$, 两个序数模式也是等概率的, 这使得 H 的归一化值始终等于1。相反, 对于 $\xi < 0.5$, 全局节点熵为零, 对于 ξ 大于0.5, 则开始增大。对于更大的嵌入尺寸, 我们观察到置换熵 H 的逼近速度比 H_{GN} 快1。例如, 如果 $\xi > 1$, 则 H 的值无法区分这些嘈杂的周期序列, 而 H_{GN} 的值仍能够区分它们。

D. 分数高斯噪声和分数布朗运动的序数网络

我们还研究了分数高斯噪声和分数布朗运动的时间序列中出现的序数网络[36,37]。分数布朗运动是一个随机过程, 其特征参数 $h \in (0, 1)$ (Hurst指数) 表现出长期相关性。分数布朗运动的增量为高斯分布, 均值为零, 平稳且通常称为分数高斯噪声。赫斯特指数 h 控制着这些随机过程的粗糙度。对于 $h < 1/2$, 分数高斯噪声是反持久性的, 这意味着 (大致而言) 正值后跟负值的频率更高 (反之亦然), 而不是偶然。另一方面, 如果 $h > 1/2$, 则分数高斯噪声将持续存在, 这意味着正值后跟正值, 而负值后跟负值的频率要高得多。不相关的 (白色) 噪声对应于 $h \rightarrow 1/2$ 。对于分数布朗运动, h 值越大, 生成的时间序列越平滑。

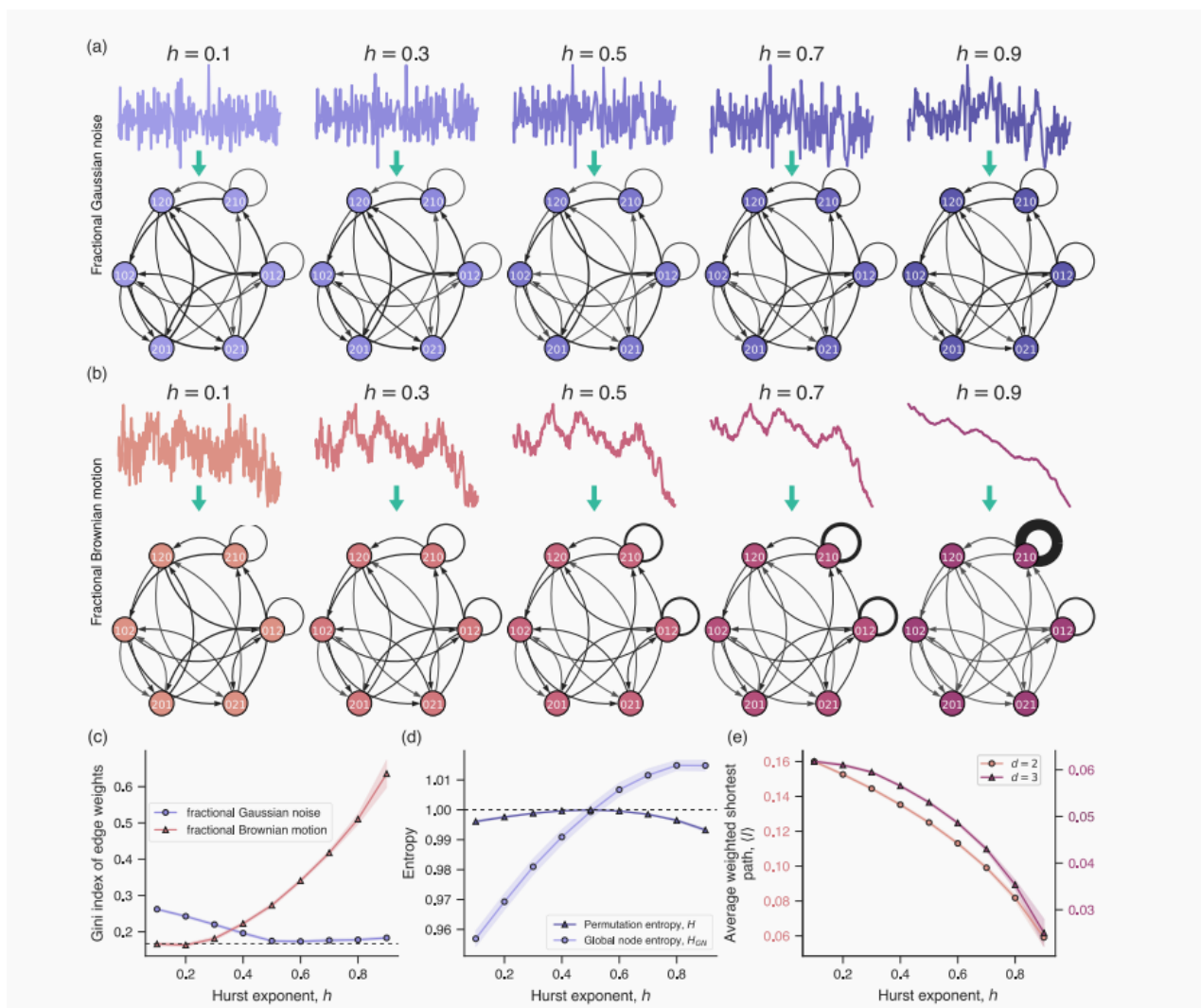


图6.分数高斯噪声和分数布朗运动的序数网络。(a) 将具有不同Hurst指数 (h , 显示在面板中) 的分数高斯噪声样本映射到有序网络中的图示。(b) 说明具有不同Hurst指数 (h , 显示在面板中) 的分数布朗运动样本到有序网络的映射。在前面的面板中, 链接的厚度与边缘权重成比例, 并且所有时间序列都是使用相同的起始种子生成的。(c) 对于分数高斯噪声 (圆形) 和分数布朗运动 (三角形), 与边缘权重分布相关的基尼指数与赫斯特指数 h 的相关性。(d) $d = 3$ 的置换熵 H (三角形) 和全局节点熵 H_{GN} (圆形) 作为分数高斯噪声的Hurst指数 h 的函数。置换熵通过其最大值归一化, 全局节点熵通过随机有序网络的值归一化。[公式 (8)]。(e) 对于分数布朗运动, 平均加权最短路径作为赫斯特指数 h 的函数, 其中 $d = 2$ (圆形) 和 $d = 3$ (三角形)。在图 (c) - (e) 中, 所有曲线代表每个Hurst指数从1000个系列的集合估计的平均值, 阴影区域代表一个标准偏差带。(c) 和 (d) 面板中的虚线表示从随机有序网络获得的度量值。

图6 (a) 和6 (b) 显示了用Hosking [38]程序生成的不同 h 值的分数高斯噪声和分数布朗运动的例子。这些图还描绘了获得的 $d = 3$ 时对应的有序网络的可视化。对于这两个过程的每个特定实现。我们观察到这些网络中存在所有可能的节点, 并且每个节点都建立了所有允许的连接。因此, 这些网络在没有边缘权重的情况下将是相同的。确实, 对这些顺序网络的权重模式进行目视检查已经使我们了解了每个时间序列的特殊性。对于分数高斯噪声, 对于较小的Hurst指数 h , 我们观察到权重分布不均匀, 而分数布朗运动对于 h 的低值显示权重更加平衡。在分数布朗运动的情况下, 值得注意的是, 对于较大的 h 值, 与置换 (2,1,0) 相关的自动循环权重变得非常强烈, 反映了随机过程的这些特定实现的下降趋势。

量化边缘权重不平等的一种可能性是计算基尼系数[39]。该系数已在多个学科中广泛使用 (尤其是在经济学中), 代表了概率分布的统计离散度的一种度量。基尼系数接近零的值表明权重是均匀分布的, 而接近1的值表明权重分布中的尖锐不等式。我们通过分数高斯噪声和分数布朗运动 (具有不同的Hurst指数) 相关联的有序网络的整体估计了基尼指数 (每个 $h \in \{0.1, 0.2, \dots, 0.9\}$ 的100个实现)。基尼系数的平均值如图6 (c) 所示, 其结果证实了我们先前的视觉分析。对于分数高斯噪声, 我们观察到, 随着 h 的值上升到 $h \approx 0.5$, 基尼系数会降低, 并从该处显示出一个平稳的平台, 其值非常接近随机序数网络的基尼系数。相反, 分数布朗运动的基尼系数随着 h 的增加而系统地增加, 反映了持续行为的增加。

我们还评估了归一化全局节点熵HGN与归一化置换熵H的平均值。图6 (d) 显示了分数高斯噪声的结果。我们注意到，H与h的行为比HGN与h的行为更加凹入，并且HGN的值显示出比H的变化范围大得多的行为。这种行为类似于我们针对嘈杂的周期性信号所报告的行为 [图5 (b)] 并为以下假设提供了进一步的支持：全局节点熵具有比普通置换熵更大的识别能力。我们进一步计算了分数布朗运动的平均加权最短路径作为Hurst指数h的函数。图6 (e) 表明，对于嵌入尺寸d = 2和d = 3，的值随h单调递减（分数高斯噪声表现出非常相似的行为）

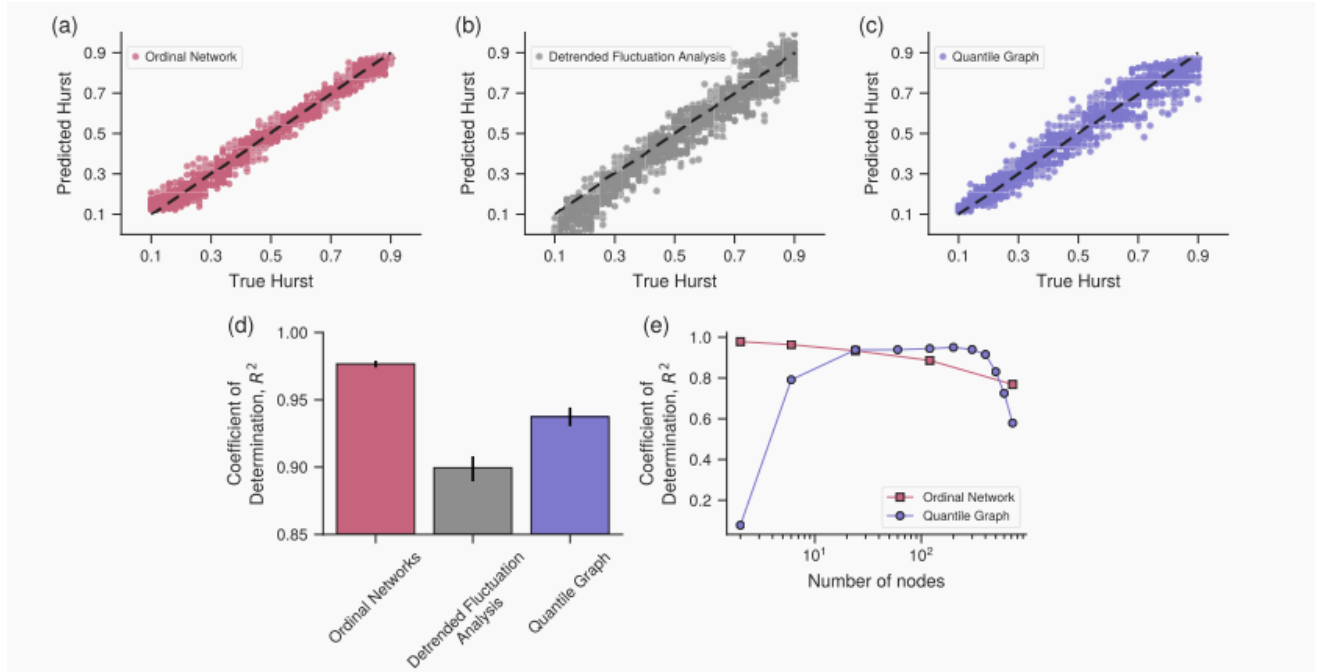


图 7.用序数网络估计赫斯特指数。面板 (a) – (c) 分别显示了通过有序网络 (d = 2)，去趋势波动分析 (DFA 具有线性趋势) 和分位数图 (q = 50) 获得的赫斯特指数真实值和预测值之间的关系。虚线表示1: 1关系。通过将训练后的K最近邻回归算法应用于与1024个数据点的分数布朗运动序列有关的平均加权最短路径 (对于有序网络) 和顶点度 (对于分位数图) 的值，即可获得预测结果。对于DFA，h的值是通过以对数对数标度 [即 $\log_2 F(s) \propto h \log_2 s$] 拟合波动函数 $F(s)$ 和标度参数s之间的关系的最小二乘获得的。(d) 条形图显示了由确定系数 R^2 测得的每种逼近方法的精度。误差线是95%的自举置信区间，其中包含1000次重采样。(e) R^2 对序数网络 (正方形) 和分位数图 (圆形) 中节点数量的依赖。

与h的单调且定义明确的行为表明，我们可以使用平均加权最短路径来预测Hurst指数。为了系统地检验这种可能性，我们建立了一个统计学习回归任务，其中使用d = 2的值作为唯一协变量来预测分数布朗运动的赫斯特指数。通过Hosking方法，对于每个 $h \in \{0.1, 0.12, 0.14, \dots, 0.9\}$ ，我们还生成了100个长度为1024的分数布朗运动样本的合奏。通过使用该数据集，我们用所有序列的75%训练了K近邻回归算法[40] (请参阅附录)，并使用五重交叉验证方法选择了最优的邻数k。系列的其余25% (从未接触过学习算法) 用于测试预测的准确性。图7 (a) 显示了Hurst指数的真实值与预测值之间的关系，其中， R^2 的准确度 (由确定系数 R^2 测量) = 97.7%。这代表了非凡的精度，尤其是考虑到该系列仅包含1024个元素时。为了提供基线精度，我们将序数网络的结果与去趋势波动分析 (DFA) 的结果进行了比较[14]，这是一种广泛使用的用于评估赫斯特指数的方法，被认为是一种尖端且可靠的方法[15]。图7 (b) 显示了通过将DFA应用于相同系列的25%获得的h的真实值与预测值之间的关系。我们立即注意到，可以通过 $R^2 = 89.9\%$ 的值来量化此关系的较大分散。

我们进一步将常规网络方法的准确性与另一种称为分位数图的时间序列到网络图进行了比较[41,42]。分位数图是加权有向网络，其中节点表示与时间序列相关的经验概率分布的给定数量q的分位数，并且基于与每个时间序列元素相关的分位数的时间顺序绘制链接。我们使用与序数网络相同的数据集构建了分位数网络，并测试了几种网络指标 (平均顶点度，平均加权最短路径，直径，聚类系数) 作为Hurst指数的预测特征。我们发现平均顶点度 (进，出或两者结合) 显示了此回归任务的最佳性能。平均加权最短路径对于分位数图是有问题的，因为出现了无限距离，该无限距离要么与一个以上的网络组件相关联，要么与通过定向路径的不可达节点相关联。图7 (c) 显示了具有50个分位数的分位数图的赫斯特指数真实值和预测值之间的关系，其中确定系数为 $R^2 = 93.5\%$ 。如图7 (d) 所示，该精度高于DFA，但低于使用有序网络获得的精度。

由于两个参数都定义了映射网络中节点的数量，因此分位数图中分位数的数量与嵌入维 d 的作用相似。因此，我们需要微调这些参数，以便在两种方法之间进行更公平的比较。为此，我们使用 d 的不同值和分位数 q 训练了学习算法，并针对每种组合估算了测试集中的 R^2 值。图7 (e) 显示了两种方法中 R^2 与网络节点数的关系。我们观察到， d 的增加会系统地降低回归任务的性能，并且对于这些具有1024个元素的时间序列， $d = 2$ 时会出现最佳值[在图7 (a)]中报告的结果相同]。另一方面，当分位数的数量太少 ($q \lesssim 10$) 或太大 ($q \lesssim 500$) 时，分位数图显示的精度非常低，在 $q \approx 50$ 左右具有最佳分位数。[图7 (c)]。因此，我们得出结论，在有序网络 ($d = 2$ 的 $R^2 = 97.7\%$) 下，对于 d 和 q 的优化值，回归任务的性能明显优于分位数图 (对于 $q = 50$ 的 $R^2 = 93.5\%$) 。

E.地震烈度级数网络

作为最后的应用，我们研究了地球地震活动时间序列中出现的序数网络。特别是，我们分析了1990年至2019年之间来自南加州地震台网[43]的地震震级时间序列。我们询问了序数网络是否能够检测由大地震的发生而引起的震级序列行为的变化。地震事件（主震）。为了测试这种可能性，我们选择了所有大于7.0的事件，并建立了两个时间序列，分别由主震前的 N 个事件和主震后的 N 个事件组成。图8 (a) 显示了与2010年4月4日在墨西哥下加利福尼亚州的一个小城市瓜达卢佩维多利亚发生的2010年下加利福尼亚州M 7.2地震有关的 $N = 200$ 的时间序列前后的示例。[44]。我们的数据中还存在另外两次大地震：1992年的7.3级Landers (1992年6月28日) [45]和1999年的Hector 7.1级地震 (1999年10月16日) [46]。

因此，我们已将这三个主震的震级序列的前后序列映射到 $d = 2$ 和 $N = 200$ 的有序网络中，如图8 (a) 所示。我们计算前后网络的平均加权最短路径。图8 (b) 的结果表明，的值在大地震发生后总是减小。-300，如图8 (c) 所示。大地震发生后的减小可能与大森定律[47]有关，大森定律是其中一项基本地震定律，该定律确定了每单位时间余震的数量会随大功率定律而衰减。自主震以来经过的时间。我们已经证实，Omori衰减增加了主震事件后时间序列的持续性，这可以通过主震后lag-1自相关系数从0.1增加到0.2 ($N = 200$ 的三个事件的平均值) 来量化。因此，该结果表明，的减少可能与主震事件之后时间序列的持久行为的增加有关。还值得一提的是，这些时间序列的长度太短 ($N \leq 300$) 使我们无法获得更高嵌入尺寸的排列转换 (以及有序网络的转换) 的可靠估。

IV.结论

我们提出了根据随机自然时间序列映射的有序网络的研究。特别是，我们分析了由随机序列，嘈杂的周期性信号，分数布朗运动和地震震级序列产生的序数网络。我们提供了对随机序数网络的详细描述，揭示了一些违反直觉的属性，例如边缘权重的不均匀分布以及仅在仅与增加或减少排列有关的节点中存在自动环。我们还提出了一种能够构建随机有序网络确切形式的方法，该网络依次用于检测时间序列中的非随机行为和缺失的排列转换。我们对有噪声周期信号的结果表明，从序数网络估计的全局节点熵比标准排列熵更能抵抗噪声。我们进一步证明了序数网络对于估计时间序列的赫斯特指数和检测大地震事件发生后地震震级序列突然变化的有用性。

因此，我们认为我们的工作有助于更好地理解序数网络的一般属性，从而阐明与随机性质的时间序列有关的结果和应用。

附录：统计学习算法

K最近邻 (KNN) 是一种用于分类或回归任务的监督统计学习算法[40]。监督一词表示该算法分析了数据集 (训练集) 的一部分以产生推断的函数，然后将其用于预测数据实例的行为。特别是，KNN算法通过对 K 个最近的数据点的值进行平均 (使用训练集中独立变量所跨越的空间) 来确定新观测值。因此，最近邻居的数量 K 是该算法的参数，通常是通过同时最小化预测的偏差和方差来确定的[40]。当学习模型过于简单以致于无法充分描述数据时，就会发生偏差错误。另一方面，当复杂模型可以很好地适应训练集但无法推广到看不见的实例时，就会出现方差误差。

如正文所述，为了使用KNN算法估计分数布朗运动样本的赫斯特指数 h ，我们使用了序数网络的平均加权最短和分位数图的平均顶点度 k 。为此，我们为每个 $h \in \{0.1, 0.12, 0.14, \dots, 0.9\}$ 生成100对值 (h_i) 和 (h_i) 的集合。然后，我们从这些数据中随机选择75%进行KNN算法的训练 (训练集)，其余25%用于测试预测的准确性。通过使用训练集，我们采用了五重交叉验证方法[40]来确定参数 K 的最佳值。此过程包括将训练集分为五个子集，使用其中一个子集来验证算法，其余的四个用于训练。重复此过程五次，以便将每个子集用于验证。在每个步骤中，我们从训练子集 (训练得分) 和验证子集 (验证得分) 估计准确性 (此处为确定系数 R^2)。经过五次重复，我们估计了训练和验证分数的平均值以及他们对时间间隔的信心。这些分数作为最近邻居 K 数的函数 (所谓的验证曲线) 的图表使我们能够确定此参数的最佳值，即与验证分数的最高值相对应的最优值。通过此过程，我们发现，对于序数网络，对于正文中报告的结果， K 的最佳值为 $K = 125$ [图2]。图7 (a)]，分位数图的 $K = 77$ [图7]。7 (c)]。

- [1] A. Ziletti, D. Kumar, M. Schefflfler, and L. M. Ghiringhelli, *Nat. Commun.* **9**, 2775 (2018).
- [2] H. Y. Sigaki, M. Perc, and H. V. Ribeiro, *Proc. Natl. Acad. Sci. USA* **115**, E8585 (2018).
- [3] C. A. Mattmann, *Nature (London)* **493**, 473 (2013).
- [4] G. Nicolis, A. G. Cantu, and C. Nicolis, *Int. J. Bifurcation Chaos* **15**, 3467 (2005).
- [5] J. Zhang and M. Small, *Phys. Rev. Lett.* **96**, 238701 (2006).
- 042304-12CHARACTERIZING STOCHASTIC TIME SERIES WITH ... *PHYSICAL REVIEW E* **100**, 042304 (2019)
- [6] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, *Proc. Natl. Acad. Sci. USA* **105**, 4972 (2008).
- [7] Y. Zou, R. V. Donner, N. Marwan, J. F. Donges, and J. Kurths, *Phys. Rep.* **787**, 1 (2019).
- [8] L. Lacasa and R. Toral, *Phys. Rev. E* **82**, 036120 (2010).
- [9] J. Schnakenberg, *Rev. Mod. Phys.* **48**, 571 (1976).
- [10] M. Small, in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)* (IEEE, Beijing, China, 2013), pp. 2509–2512.
- [11] C. Bandt and B. Pompe, *Phys. Rev. Lett.* **88**, 174102 (2002).
- [12] M. Small, M. McCullough, and K. Sakellariou, in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)* (IEEE, Florence, Italy, 2018), pp. 1–5.
- [13] J. Zhang, J. Zhou, M. Tang, H. Guo, M. Small, and Y. Zou, *Sci. Rep.* **7**, 7795 (2017).
- [14] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, *Phys. Rev. E* **49**, 1685 (1994).
- [15] Y.-H. Shao, G.-F. Gu, Z.-Q. Jiang, W.-X. Zhou, and D. Sornette, *Sci. Rep.* **2**, 835 (2012).
- [16] C. Bandt and F. Shiha, *J. Time Ser. Anal.* **28**, 646 (2007).
- [17] Y. Cao, W.-w. Tung, J. B. Gao, V. A. Protopopescu, and L. M. Hively, *Phys. Rev. E* **70**, 046217 (2004).
- [18] O. A. Rosso, H. A. Larrondo, M. T. Martin, A. Plastino, and M. A. Fuentes, *Phys. Rev. Lett.* **99**, 154102 (2007).

- [19] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
- [20] M. McCullough, M. Small, H. H. C. Iu, and T. Stemler, Philos. Trans. R. Soc. A **375**, 20160292 (2017).
- [21] A. M. Unakafov and K. Keller, Phys. D (Amsterdam, Neth.) **269**, 94 (2014).
- [22] J. Jonsson, *Simplicial Complexes of Graphs* (Springer, New York, 2008).
- [23] P. Baudot, M. Tapia, D. Bennequin, and J.-M. Goillard, Entropy **21**, 869 (2019).
- [24] M. Andjelković, N. Gupte, and B. Tadić, Phys. Rev. E **91**, 052817 (2015).
- [25] M. McCullough, M. Small, T. Stemler, and H. H.-C. Iu, Chaos **25**, 053101 (2015).
- [26] X. Sun, M. Small, Y. Zhao, and X. Xue, Chaos **24**, 024402 (2014).
- [27] C. Masoller, Y. Hong, S. Ayad, F. Gustave, S. Barland, A. J. Pons, S. Gómez, and A. Arenas, New J. Phys. **17**, 023068 (2015).
- [28] C. Donnat and S. Holmes, Ann. Appl. Stat. **12**, 971 (2018).
- [29] P. Wills and F. G. Meyer, arXiv:1904.07414.
- [30] J. M. Amigó, L. Kocarev, and J. Szczepanski, Phys. Lett. A **355**, 27 (2006).
- [31] J. M. Amigó, S. Zambrano, and M. A. F. Sanjuán, Europhys. Lett. **79**, 50001 (2007).
- [32] J. Amigó, S. Zambrano, and M. A. Sanjuán, Europhys. Lett. **83**, 60005 (2008).
- [33] M. McCullough, K. Sakellariou, T. Stemler, and M. Small, Chaos **26**, 123103 (2016).

- [34] K. Sakellariou, M. McCullough, T. Stemler, and M. Small, *Chaos* **26**, 123104 (2016).
- [35] F. Olivares, L. Zunino, and D. G. Pérez, *Phys. A (Amsterdam, Neth.)* **534**, 122100 (2019).
- [36] B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
- [37] B. B. Mandelbrot and J. W. Van Ness, *SIAM Rev.* **10**, 422 (1968).
- [38] J. R. M. Hosking, *Water Resour. Res.* **20**, 1898 (1984).
- [39] C. Gini, *Econ. J.* **31**, 124 (1921).
- [40] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R* (Springer, New York, 2014).
- [41] A. S. L. O. Campanharo, M. I. Sirer, R. D. Malmgren, F. M. Ramos, and L. A. N. Amaral, *PLoS One* **6**, e23378 (2011).
- [42] A. S. Campanharo and F. M. Ramos, *Phys. A (Amsterdam, Neth.)* **444**, 43 (2016).
- [43] California Institute Of Technology And United States Geological Survey Pasadena, Southern California Seismic Network, Available at <http://www.fdsn.org/doi/10.7914/SN/CI>, accessed: 28 Jun 2019.
- [44] Tectonic Summary—M 7.2 - 12km SW of Delta, B.C., MX, Available at <https://earthquake.usgs.gov/earthquakes/eventpage/usp000habu/executive>, accessed: 28 Jun 2019 .
- [45] Tectonic Summary—M 7.3 - Landers, California Earthquake, Available at <https://earthquake.usgs.gov/earthquakes/eventpage/usp00059sn/executive>, accessed: 28 Jun 2019 .
- [46] Tectonic Summary—M 7.1 - 16km SW of Ludlow, CA, Available at <https://earthquake.usgs.gov/earthquakes/eventpage/usp0009fwb/executive>, accessed: 28 Jun 2019.
- [47] T. Utsu, Y. Ogata, R. S. and Matsu'ura, J. *Phys. Earth* **43**, 1 (1995).

